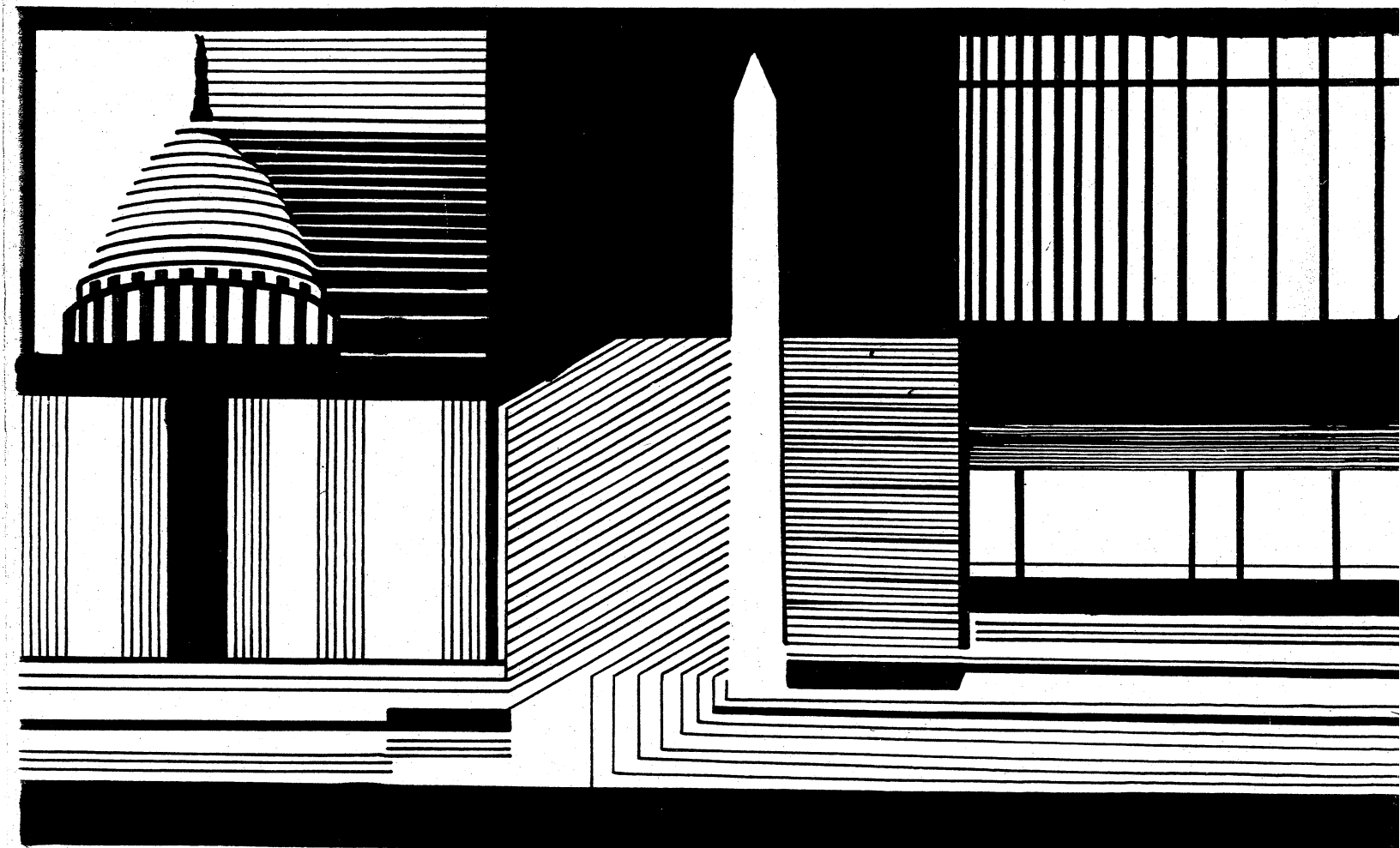


ADVANCES IN CONSUMER RESEARCH

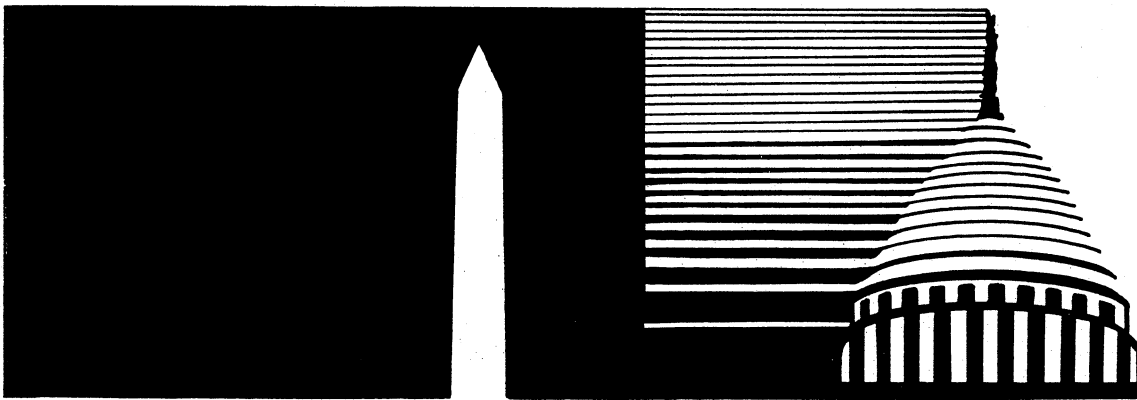
VOLUME VIII



KENT B. MONROE, EDITOR

ADVANCES IN CONSUMER RESEARCH

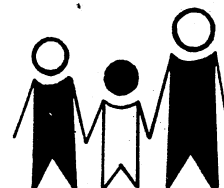
VOLUME VIII



PROCEEDINGS OF THE
ASSOCIATION FOR
CONSUMER RESEARCH
ELEVENTH ANNUAL
CONFERENCE

ARLINGTON, VIRGINIA
OCTOBER 1980

KENT B. MONROE, EDITOR



1981 copyright

© ASSOCIATION FOR CONSUMER RESEARCH

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

International Standard Book Number: 0-915552-08-6

International Standard Serial Number: 0098-9258

Monroe, Kent B., Editor. *Advances in Consumer Research*, Volume VIII. (Ann Arbor: Association for Consumer Research, 1981).

ORDER FORMS ARE PROVIDED AT THE BACK OF THIS VOLUME

1979-80 Officers

Association for Consumer Research

BOARD OF DIRECTORS

President, William L. Wilkie, University of Florida
President Elect, Jerry C. Olson, Pennsylvania State University
Past President, H. Keith Hunt, Brigham Young University
Executive Secretary, Thomas C. Kinnear, University of Michigan
Treasurer, Donald R. Lehman, Columbia University
Director, Michael Mazis, American University
Director, Stanley D. Shores, Jr., Procter & Gamble
Director, Peter L. Wright, Stanford University

COMMITTEES

Conference Chairman, Kent B. Monroe, Virginia Polytechnic Institute and State University
Newsletter, Richard C. Reizenstein, University of Tennessee

ADVISORY COUNCIL

Kenneth L. Bernhardt, Georgia State University
John A. Dimling, Jr., National Association of Broadcasters
John L. Evans, Consumer & Corporate Affairs, Canada
Robert Ferber, University of Illinois
Stephen A. Greyser, Harvard University and Marketing Science Institute
Harold H. Kassarian, U.C.L.A.
David Schoenfeld, J. C. Penney
William D. Wells, Needham, Harper & Steers, Inc.
E. Scott Maynes, Cornell University

REVIEWERS FOR PAPER COMPETITION - 1980 ANNUAL CONFERENCE

Albaum, Gerald - University of Oregon
 Allen, Chris - University of Massachusetts
 Anderson, Beverlee - Wright State University
 Arbeit, Stanley - Canisius College, Buffalo
 Barnes, James - University of Georgia
 Bearden, William - University of South Carolina
 Bechtel, Gordon - University of Florida
 Beckwith, Neil - University of Pennsylvania
 Belch, Mickey - San Diego State University
 Belk, Russell - University of Utah
 Bell, Russell - Holiday Inns Inc.
 Bennett, Peter - Pennsylvania State University
 Berkowitz, Eric - University of Minnesota
 Best, Roger - University of Arizona
 Bettman, James - Univ. of California, Los Angeles
 Block, Carl - University of Missouri
 Bonfield, E. H. - Temple University
 Burnkrant, Robert - Ohio State University
 Calder, Bobby - Northwestern University
 Cattin, Philippe - University of Connecticut
 Choudhury, Pravat - Howard University
 Cohen, Joel - University of Florida
 Cohen, Dorothy - Hofstra University
 Cosmas, Stephen - Virginia Polytechnic Inst. & State Univ.
 Danes, Jeffrey - Virginia Polytechnic Inst. & State Univ.
 Day, Ralph - Indiana University
 Della Bitta, Albert - University of Rhode Island
 Dolich, Ira - University of Nebraska
 Douglas, Susan - New York University
 Downes, Phillip - Florida State University
 Dyer, Robert - George Washington University
 Engel, James - Wheaton College
 Gardner, David - Univ. of Illinois, Urbana
 Gelb, Betsy - University of Houston
 Gibson, Lawrence - General Mills Inc.
 Ginter, James - Ohio State University
 Goldberg, Marvin - McGill University
 Golden, Linda - Univ. of Texas, Austin
 Gorn, Gerald - McGill University
 Granbois, Donald - Indiana University
 Green, Paul - University of Pennsylvania
 Greyser, Stephen - Harvard University
 Grossbart, Sandy - University of Nebraska
 Gutman, Jon - University of Southern California
 Haines, George - University of Toronto
 Heeler, Roger - York University
 Hempel, Donald - University of Connecticut
 Holbrook, Morris - Columbia University
 Holman, Rebecca - Young and Rubicam Inc.
 Houston, Michael - Univ. of Wisconsin, Madison
 Huber, Joel - Duke University
 Hunt, Keith - Brigham Young University
 Jain, Subhash - University of Connecticut
 Johnston, Wesley - Ohio State University
 Kamen, Joseph - Indiana University, Gary
 Kassarian, Harold - Univ. of California, Los Angeles
 Kaufmann, Maureen - Survey Research Center, Ann Arbor
 Kernan, Jerome - University of Cincinnati
 King, Robert - Virginia Polytechnic Inst. & State Univ.
 Kinnear, Thomas - University of Michigan
 Kuehl, Philip - University of Maryland
 Lambert, Zarrel - Auburn University
 Lautman, Martin - Associates for Research in Behavior, Inc.
 Lehmann, Donald - Columbia University
 Lutz, Richard - Univ. of California, Los Angeles
 MacLachlan, Douglas - Univ. of Washington, Seattle
 Mason, Barry - University of Alabama
 Mathews, Lee - Ohio State University
 Mazis, Michael - American University
 McNeill, Dennis - University of Denver
 Mentzer, John T. - Virginia Polytechnic Inst. & State Univ.
 Miller, John - Univ. of Colorado, Colorado Springs
 Miller, Steve - Oklahoma State University
 Mittelstaedt, Robert - University of Nebraska
 Moschis, George - Georgia State University
 Mowen, John - Oklahoma State University
 Munson, Micael - University of Santa Clara
 Newman, Joseph - University of Arizona
 O'Brien, Terrence - Kansas State University
 O'Connor, P. J. - University of Kentucky
 Olshavsky, Richard - Indiana University
 Ostlund, Lyman - University of Arizona
 Permut, Steven - Yale University
 Perreault, William - University of North Carolina
 Peter, J. Paul - Ohio State University
 Peterson, Robert - Univ. of Texas, Austin
 Popper, Edward - University of Florida
 Prasad, Kanti - Univ. of Wisconsin, Milwaukee
 Raju, P. S. - Univ. of Illinois, Chicago Circle
 Rao, Vithala - Cornell University
 Rao, T. R. - Univ. of Wisconsin, Milwaukee
 Ring, Lawrence - University of Virginia
 Roberts, Lou Mary - Boston University
 Robertson, Thomas - University of Pennsylvania
 Roering, Kenneth - University of Missouri
 Ross, Ivan - University of Minnesota
 Rossiter, John - Columbia University
 Rothchild, Michael - University of Wisconsin
 Rubin, Ronald - Central Florida University
 Russ, F. A. - University of North Carolina
 Russo, J. Edward - University of Chicago
 Ryan, Michael - Columbia University
 Satow, Kay - Needham, Harper, & Steers
 Scammon, Debra - University of Utah
 Schaninger, Charles - University of Massachusetts
 Schewe, Charles - University of Massachusetts
 Schiffman, Leon - City University, New York
 Schlinger, Mary - Univ. of Illinois, Chicago Circle
 Scott, Carol - Univ. of California, Los Angeles
 Sewall, Murphy - University of Connecticut
 Sheffet, Mary - Indiana University
 Sheth, Jagdish - Univ. of Illinois, Urbana
 Shocker, Allan - University of Pittsburgh
 Shoemaker, Robert - Dartmouth College
 Sirgy, M. Joseph - Virginia Polytechnic Inst. & State Univ.
 Smith, Theodore - Old Dominion University
 Smith, Scott - University of Oregon
 Upah, Gregory - Virginia Polytechnic Inst. & State Univ.
 Venkatesan, M. - University of Oregon
 Warshaw, Paul - McGill University
 Weinberger, Marc - University of Massachusetts
 Weitz, Barton - Univ. of California, Los Angeles
 Wells, William - Needham, Harper, & Steers
 Wildt, Albert - University of Georgia
 Wiley, James - Temple University
 Wilkes, Robert - Texas Tech University
 Wilkie, William - University of Florida
 Wilson, Dale - Batten, Barton, Durstine & Osborn, Inc.
 Wilson, David - Pennsylvania State University
 Witt, Robert - University of Texas, Austin
 Zaltman, Gerald - University of Pittsburgh

PREFACE

In October, 1980, the Association for Consumer Research (ACR) held its Eleventh Annual Conference. This conference opened a new decade for consumer research marking a major step in maturity for ACR, and presenting new challenges and opportunities for the advancement of consumer research.

In recognition of these challenges and opportunities, the program committee sought to have sessions that highlighted current and future issues in consumer research. Moreover, the committee sought to utilize the conference location, Washington D.C., and include public policy as a major programmatic focus.

Volume VIII of the Advances in Consumer Research series contains most of the presentations made at the 1980 conference. All of the competitive papers and most of the discussants' commentaries are included here. We were also able, sometimes with a little persuasion, to publish most of the presentations made at the special topics sessions and workshops. The Presidential Address by Dr. William L. Wilkie is included as are the special remarks made by the first two fellows of the Association for Consumer Research, Dr. James F. Engel and Dr. John A. Howard.

A number of people contributed to the success of the conference and to this volume. The key contributions were made by the authors and the organizers of the special sessions and workshops. The paper reviewers listed on p. iv made substantive contributions by judging the 113 paper submissions and providing authors of accepted papers with suggestions for improving their papers. The Program Committee evaluated all special session proposals, offered me encouragement when needed, and took a lead in developing sessions fitting the theme of the conference. Special thanks goes to Paul N. Bloom and to Gary T. Ford for helping with arrangements and registration. We also received excellent support from the staff of the Key Bridge Marriott Hotel.

At Virginia Tech, many people made outstanding contributions. John McGinnis, William Middleton and Micky Nguyen provided early assistance in organizing the tasks to be done, stuffing envelopes, and recording reviewers' judgments. R. Krishnan provided invaluable assistance in computerizing our operations and providing the day-to-day assistance throughout the registration period. Ramon Avila continued the post-registration computer work and Nelson Craige compiled, paged, and indexed the Proceedings. Wanda Belcher and Janice Blevins provided secretarial assistance, and my secretary Becky Glick handled the correspondence and telephone calls for me. Molly Hobson, University of Maryland, assisted Gary Ford and provided registration help during the conference. Artistic work was done by John Board.

Special thanks goes to Allyn Kosenko who proofed, edited and re-typed the manuscripts. She simply did a fantastic job! The support provided by Virginia Tech and the Marketing faculty is also gratefully acknowledged.

Kent B. Monroe
Editor

Department of Marketing
Virginia Tech
Blacksburg, VA 24061
January 1981

"Knowledge exists in minds, not in books."

Kenneth Boulding

"[W]here is the knowledge we have lost in information?"

T.S. Eliot, *The Rock*

"The thirteenth stroke of a clock is not only false of itself, but casts grave doubts on the credibility of the preceding twelve."

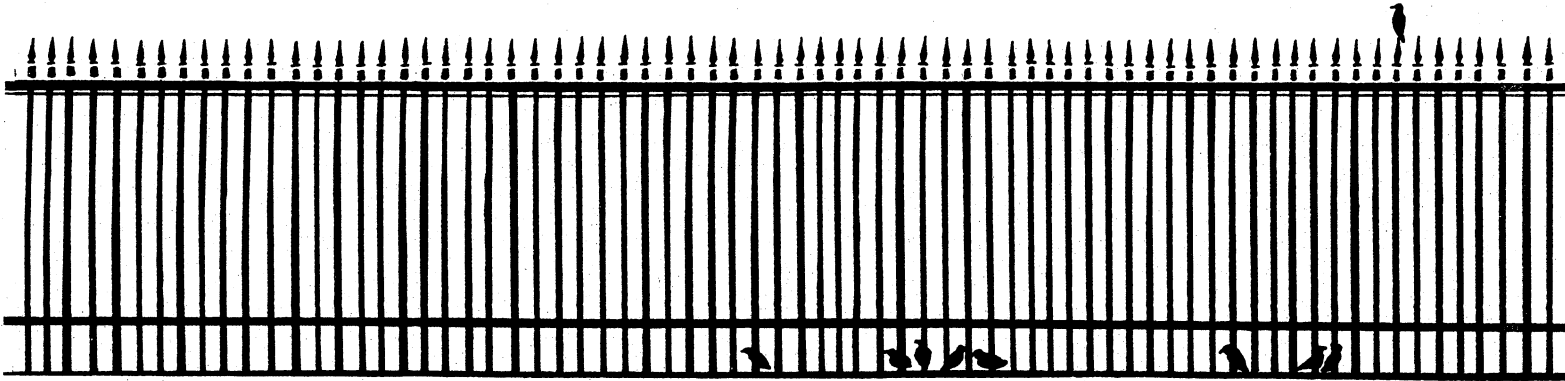
Mark Twain, *Autobiography*

"It's not what we don't know that gives us trouble. It's what we know that ain't so."

Will Rogers

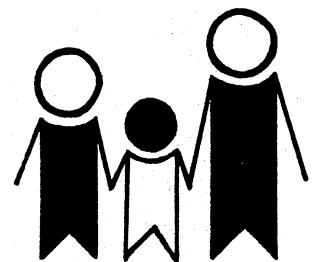
*ASSOCIATION
FOR CONSUMER RESEARCH
ELEVENTH ANNUAL CONFERENCE*

*KEY BRIDGE MARRIOTT
ARLINGTON, VIRGINIA
OCTOBER 16-19 1980*



KENT B. MONROE, PROGRAM CHAIRMAN
PAUL N. BLOOM
GARY T. FORD ARRANGEMENTS CHAIRMEN

PROGRAM COMMITTEE:
KENNETH L. BERNHARDT
WILLIAM R. DARDEN
ELIZABETH C. HIRSCHMAN
E. LAIRD LANDON, JR.
JERRY C. OLSON
ANDREW A. MITCHELL
ALAN G. SAWYER



1980 PROGRAM: ANNUAL CONFERENCE OF THE ASSOCIATION FOR CONSUMER RESEARCH

Thursday, October 16, 1980

Executive Committee Meeting: 10:00 a.m. - 3:00 p.m., *Monroe Room*
Registration opens at 2:00 p.m. and closes at 8:00 p.m., Lobby Level

Special Event: 3:30 p.m. - 5:15 p.m. (attendance limited)
PUBLIC POLICY UPDATE: PERSPECTIVES ON THE FEDERAL TRADE COMMISSION

Commissioners' Conference Room, Federal Trade Commission

The purpose of this special session is to provide ACR members with an update on current and prospective issues under consideration at the FTC.

CHAIR: Kenneth L. Bernhardt, Georgia State University

Participants:

1. Bob Reich, Director of the Office of Policy Planning, FTC. He will discuss consumer issues for the 1980's outlining his views on issues that ACR members should be tracking. He will comment on the increased orientation toward providing consumers with useful information, use of performance standards, sunset legislation, and increased use of cost/benefit analysis.
2. Ron Stiff, Regulatory Analysis/Impact Evaluation Office, Bureau of Consumer Protection. He will discuss current and future consumer research activities at the agency.
3. Michael Pertschuk, Chairman, Federal Trade Commission. He will provide his perspective on current and future policy issues being studied by the Commission. He will also comment on the use of consumer research in FTC policy.

EARLY BIRD RECEPTION: 6:00 p.m. - 8:00 p.m. (cash bar), Potomac Ballroom

FRIDAY, OCTOBER 17, 1980

Registration opens at 8:00 a.m. and closes at 6:00 p.m.

MORNING SESSIONS - 1

8:30 a.m. - 10:00 a.m.

Special Topic Session (3 hrs): EMERGING ISSUES IN LOW INVOLVEMENT THEORY

Francis Scott Key Room, Salon A

CHAIR: Richard J. Lutz, Univ. of California, Los Angeles
John T. Cacioppo, University of Iowa

This session seeks to clarify the conceptual properties of the low involvement construct and to provide direction for future research. The papers address the implications of involvement for understanding consumer behavior.

What is Low Involvement Low In?

Clark Leavitt, Ohio State University
Anthony G. Greenwald, Ohio State University
Carl Obermiller, Ohio State University

Issue Involvement as a Moderator of the Effects on Attitude of Advertising Content and Context

Richard E. Petty, University of Missouri
John T. Cacioppo, University of Iowa

The Dimensions of Advertising Involvement

Andrew A. Mitchell, Carnegie-Mellon University

Reconceptualizing Consumer Involvement

Richard J. Lutz, Univ. of California, Los Angeles

Discussant: Harold H. Kassarjian, Univ. of California, Los Angeles

Special Topic Session: THE NEGATIVE SIDE OF CONSUMER INFORMATION

Francis Scott Key Room, Salon B

CHAIR: Marc G. Weinberger, Univ. of Massachusetts, Amherst
Chris T. Allen, Univ. of Massachusetts, Amherst
William R. Dillon, Univ. of Massachusetts, Amherst

The full impact of negative product information on consumers is unknown, but preliminary research suggests that negative information can have a powerful influence on buyers. This session summarizes prior research on negative information, discusses current research, and offers some theoretical explanations for the empirical results.

Negative Information: Perspectives and Research Directions

Marc G. Weinberger, Univ. of Massachusetts, Amherst
Chris T. Allen, Univ. of Massachusetts, Amherst
William R. Dillon, Univ. of Massachusetts, Amherst

Factors Influencing Consumer Responses to Product Recalls: A Regression Analysis Approach

John C. Mowen, Oklahoma State University
David W. Jolly, Oklahoma State University
Gary S. Nickell, Oklahoma State University

Theoretical Perspectives on the Impact of Negative Information

Carol A. Scott, Univ. of California, Los Angeles
Alice M. Tybout, Northwestern University

Negative Information: A Commentary

Stephen A. Greyser, Harvard University/Marketing Science Institute

BREAK: Coffee and Soft Drinks

Lobby Level and Lower Level

10:00 a.m. - 10:30 a.m.

Special Topic Session: PUBLIC POLICY AND CONSUMER
RESEARCH IN THE 1980's

Georgetown Room, Salons A,B

CHAIR: Paul N. Bloom, Marketing Science Institute

This session identifies probable public policy issues of the 1980's that might be resolvable with appropriate consumer research. The purpose of the session is to provide an "early warning" on issues so that relevant consumer research could be completed in time to influence formulated policies

The Future of Consumer Protection Regulation
Michael B. Mazis, American University

Consumer Policy Issues: Global Trends for the 1980's
Graham T. T. Molitor, Public Policy Forecasting,
Inc.

Judging Marketing in the 1980's
Alan R. Andreasen, University of Illinois, Urbana

*The Swedish Experience: Its Transferability and Related
Research Implications*
Hans B. Thorelli, Indiana University

Paper Session: CONSUMER RESEARCH AND QUALITY OF LIFE

Georgetown Room, Salon C

CHAIR: Pravat K. Choudhury, Howard University

*Leisure and the QOL Construct: A Review and Some Modest
Proposals*

Lynette S. Unger, Miami University
Jerome B. Kernan, University of Cincinnati

The Consumer and the Health Care Process
Lynn Langmeyer, Wright State University
George Miaoulis, Wright State University

*Attitudes Toward Public Policy Alternatives to Reduce Air
Pollution*

David A. Aaker, Univ. of California, Berkeley
Richard P. Bagozzi, Massachusetts Institute of
Technology

Discussant: William G. Zikmund, Oklahoma State Univ.

Paper Session: COGNITIVE PROCESSES

Potomac Ballroom, Salon D

CHAIR: Albert R. Wildt, University of Georgia

Toward Conceptualizing and Measuring Cognitive Structures
Rajesh Kanwar, The Pennsylvania State University
Jerry C. Olson, The Pennsylvania State University
Laura S. Sims, The Pennsylvania State University

Cognitive Maps and Shopping Convenience
Sanford Grossbart, University of Nebraska
Balusu Rammohan, University of Nebraska

*A Method for Determining the Sequencing of Cognitive Pro-
cesses in Judgment: Order Effects on Reaction Times*
John G. Lynch, Jr., University of Florida

Discussant: Michael J. Ryan, Columbia University

BREAK: Coffee and Soft Drinks Lobby Level and Lower Level

10:00 a.m. - 10:30 a.m.

Special Topic Session: EMERGING ISSUES IN LOW INVOLVEMENT THEORY

*Francis Scott Key Room, Salon A*CHAIR: Richard J. Lutz, Univ. of California, Los Angeles
John T. Cacioppo, University of Iowa

[Continued from 8:30]

Special Topic Session: USING THE BUREAU OF LABOR STATISTICS CONSUMER EXPENDITURE SURVEY IN CONSUMER RESEARCH

*Potomac Ballroom, Salon D*CHAIR: Seymour Sudman, Univ. of Illinois, Urbana
Gregory D. Upah, Virginia Tech

The Consumer Expenditure Survey consists of: (1) a quarterly interview panel, and (2) a diary completed by each respondent for two weeks. The data are used in the formulation of public policy. The purposes of this session are to describe the CES and its methodology and to illustrate how this longitudinal data may be used in consumer research.

Overview of the Consumer Expenditure Survey
Gail Hoff, Bureau of the Census*Uses of the Consumer Expenditure Survey Data*
Eva Jacobs, Bureau of Labor Statistics*An Evaluation of the 1972-73 Consumer Expenditure Survey Data*

Robert Pearl, Univ. of Illinois, Urbana

*The Consumer Expenditure Survey: Prospects for Consumer Research*Gregory D. Upah, Virginia Tech
Seymour Sudman, Univ. of Illinois, Urbana

Paper Session: CONSUMER RESPONSE TO ADVERTISING

Georgetown Room, Salon C

CHAIR: Mary Jane Schlinger, Univ. of Illinois, Chicago Circle

*Linear Effects of Cognitive Response to Advertising*Larry Percy, Creamer, Inc.
Martin R. Lautman, Associates for Research in Behavior, Inc.*On Explaining and Predicting the Effectiveness of Celebrity Endorsers*John C. Mowen, Oklahoma State University
Stephen W. Brown, Arizona State University*Guilt Arousing Marketing Communications: An Unexplored Variable*

Morry Ghingold, Pennsylvania State University

Discussant: Robert L. King, Virginia Tech

Special Topic Session: UNIVERSITY PANELS FOR CONSUMER RESEARCH

*Georgetown Room, Salons A,B*CHAIR: Kenneth L. Bernhardt, Georgia State Univ.
William R. Darden, Univ. of Arkansas

This session introduces consumer researchers to academic-sponsored consumer research panels. The purposes of the session are to discuss the design of these panels and to illustrate their value for consumer research. The discussion will also focus on the operation and maintenance of these consumer research panels.

*Research Panels in Consumer Behavior*Robert Ferber, Univ. of Illinois, Urbana
Linda B. Lannom, Univ. of Illinois, Urbana*Perspectives on Academically-Oriented Panel Research*Arch Woodside, University of South Carolina
Jesse Teal, University of South Carolina*Surveys of Consumer Attitudes: Survey Research Center, University of Michigan*

Richard Curtin, Consumer Sentiment Panel

Discussants: Richard B. Ross, Market Facts, Inc.
C. Whan Park, University of Pittsburgh

Paper Session: DEVELOPING NEW RESEARCH TRADITIONS IN CONSUMER RESEARCH

Francis Scott Key Room, Salon B

CHAIR: Debra L. Scammon, University of Utah

*The Impact of Program Evaluation Needs on Research Methodology*R. Bruce Hutton, University of Denver
Dennis L. McNeill, University of Denver*Risk-Benefit Analysis and the Determination of Acceptable Risk*Rachel Dardis, University of Maryland
Julie Stremel, University of Maryland*Scientific Progress and Research Traditions in Consumer Research*Liisa Uusitalo, International Institute for Environment and Society, West Germany
Jyrki Uusitalo, University of Helsinki

Discussant: Peter D. Bennett, Pennsylvania State Univ.

LUNCHEON Potomac Ballroom
Guest Speakers - James F. Engel - John A. Howard

12:00 p.m. - 1:30 p.m.

Special Topic Session (3 hrs.): THE ROLE OF CONSUMER RESEARCH IN THE FORMATION OF PUBLIC POLICY: AN EXAMPLE OF AN INTEGRATED RESEARCH AND DEVELOPMENT PLAN

Francis Scott Key Room, Salon A

CHAIR: James T. Heimbach, Food & Drug Administration

Since Spring 1978, the Food and Drug Administration (FDA) has been executing a large, integrated research plan for the purpose of revising the current information content of the food label. This research plan is a unique example of the application of consumer research to an issue of public policy. This session presents the research plan and highlights the current research results.

Defining the Problem: The Scope of Consumer Concern with Food Labeling

James T. Heimbach, Food & Drug Administration

Food Label Information: What Consumers Say They Want and What They Need

Edwin C. Hackleman, Food & Drug Administration

Food Label Information: What Consumers Say They Use and What They Actually Use

Michael H. Baumgardner, Food & Drug Administration

Commercial Market Tracking Systems: Applications to Policy Formation Regarding Labeling

Michael L. Stewart, Food & Drug Administration

Labels That Communicate: Design and Evaluation

Raymond C. Stokes, Food & Drug Administration

Implementation of a Revised Food Labeling Policy: Evaluation and Tracking

Raymond E. Schucker, Food & Drug Administration

Workshop (3 hrs.): ASSESSING CORRECTIVE ADVERTISING EFFECTS

Georgetown Room, Salons A,B

CHAIR: William L. Wilkie, Univ. of Florida

This session presents the empirical research results from several corrective advertising campaigns. Previously unreported data dealing with the impact of STP and Listerine advertising will be presented. Jack Jacoby will present his very controversial research study examining the efficacy of the FTC's corrective advertising strategy.

Viewer Miscomprehension of Televised Communications

Jacob Jacoby, Purdue University
Wayne D. Hoyer, Purdue University
David A. Sheluga, Purdue University

Impact of Publicity on Corrective Advertising Effects

Kenneth Bernhardt, Georgia State University
Thomas Kinnear, University of Michigan

Correcting Corrective Advertising

Jacob Jacoby, Purdue University
Margaret C. Nelson, Purdue University
Wayne D. Hoyer, Purdue University

Recall of Corrective Commercials

Dennis McNeill, University of Denver

Attitude and Belief Changes from Corrective Ads

Gary Armstrong, University of North Carolina
Fritz Russ, University of North Carolina

A Longitudinal Study of Corrective Advertising

Michael Mazis, American University

Discussants: Peter Rheinstein, Food & Drug Administration
Richard D. Mizerski, Federal Trade Commission

Special Topic Session: PATRONAGE THEORY

Potomac Ballroom, Salon D

CHAIR: William R. Darden, University of Arkansas

This session and the following session on patronage research focuses on patronage choice behavior. In recent years, store choice behavior research has been limited both in terms of theoretical development and the research methods used, despite the large amounts spent on location research by retail firms. The patronage theory session develops the linkage between the traditional gravitational models and the behavioral models of patronage choice behavior. The patronage research session that follows discusses the newer research methods that are being used to gather evidence on store choice behavior.

Integration of Economic Geography and Social Psychological Models of Patronage Behavior

Robert F. Lusch, University of Oklahoma

Women's Self-Ascribed Occupational Status and Retail Patronage

Elizabeth C. Hirschman, New York University

Consumer Socialisation Factors in a Patronage Model of Consumer Behavior

William R. Darden, University of Arkansas

Donna K. Darden, University of Arkansas

Roy Howell, University of Illinois

Shirley Jo Miller, University of Arkansas

Discussant: Robert Peterson, Univ. of Texas, Austin

Paper Session: CONSUMER PERCEPTIONS

Francis Scott Key Room, Salon B

CHAIR: E. Laird Landon, University of Houston

A Multivariate Analysis of the Perception of Value from Retail Price Advertisements

Albert J. Della Bitta, University of Rhode Island

Kent B. Monroe, Virginia Tech

The Effect of Generic Products on Consumer Perceptions and Brand Choice

John J. Wheatley, University of Washington

Determinants of Consumption Cue Utilization in Impression Formation: An Associational Derivation and Experimental Verification

Russell W. Belk, University of Utah

Discussant: Eric N. Berkowitz, University of Minnesota

Paper Session: PERCEPTION OF TIME AND CONSUMER BEHAVIOR

Georgetown Room, Salon C

CHAIR: J. Paul Peter, Ohio State University

Temporal Links Between Preference and Perception

James McCullough, The University of Arizona

Douglas MacLachlan, University of Washington

Reza Moïnpour, University of Washington

Temporal Incongruency in Consumer Behavior

Philip E. Hendrix, University of Michigan

Claude R. Martin, Jr., University of Michigan

The Imagination of the Future: A Hidden Concept in the Study of Consumer Decision Making

Rebecca H. Holman, Young and Rubicam, Inc.

Discussant: Wesley T. Johnston, Ohio State Univ.

Special Topic Session: THE ROLE OF CONSUMER RESEARCH IN THE FORMATION OF PUBLIC POLICY: AN EXAMPLE OF AN INTEGRATED RESEARCH AND DEVELOPMENT PLAN

Francis Scott Key Room, Salon A

CHAIR: James T. Heimbach, Food & Drug Administration

[Continued from 1:30]

Workshop: ASSESSING CORRECTIVE ADVERTISING EFFECTS

Georgetown Room, Salons A,B

CHAIR: William Wilkie, University of Florida

[Continued from 1:30]

Paper Session: ATTITUDE RESEARCH AND BEHAVIORAL INTENTIONS

Francis Scott Key Room, Salon B

CHAIR: David T. Wilson, Pennsylvania State Univ.

Attitude Structure and Search: An Integrative Model of Importance Directed Information Processing

Morris B. Holbrook, Columbia University
David A. Velez, Columbia University
Gerard R. Tabouret, Columbia University

Examining the Diagnostic Validity of the Fishbein Behavioral Intention Model

Paul W. Miniard, Ohio State University

A Comparison of Two Behavioral Intention Models

David Brinberg, University of Maryland

Discussant: Paul R. Warshaw, McGill University
Jeffrey E. Danes, Virginia Polytechnic and State University

Special Topic Session: PATRONAGE RESEARCH APPLICATIONS AND METHODS

Potomac Ballroom, Salon D

CHAIR: William R. Darden, University of Arkansas

An Exploratory Investigation of Mediating Factors in Retail Store Image Responses

Robert Peterson, Univ. of Texas, Austin

A Comparison Between Discriminant Analysis and Multinomial Logit in the Prediction of Store Choice

Douglas Tigert, University of Toronto
Victor Roth, University of Toronto

Research into Shopping Mall Choice Behavior

Roy Howell, Univ. of Illinois, Urbana
Jerry Rogers, Southwest Missouri State University

Discussant: Thomas Stanley, Georgia State Univ.

Paper Session: SOCIAL CLASS AND SOCIAL INFLUENCE

Georgetown Room, Salon C

CHAIR: Ivan Ross, University of Minnesota

Perceived Differences in Product and Brand User Stereotypes Across Upper, Middle and Lower Social Classes

J. Michael Munson, University of Santa Clara
W. Austin Spivey, University of Santa Clara

Proposed Extensions of the Basic Model of Social Class Employed in Consumer Research

Terence A. Shimp, University of South Carolina
J. Thomas Yokum, University of South Carolina

An Analysis of Alcohol Advertisements Using French and Raven's Theory of Social Influence

Scott B. Mackenzie, Univ. of California, Los Angeles
Judy L. Zaichkowsky, Univ. of California, Los Angeles

Discussant: Barton Weitz, Univ. of California, Los Angeles

BUSINESS MEETING	Georgetown Room, Salons A,B	5:00 p.m. - 5:30 p.m.
	EVERYONE IS WELCOME TO ATTEND	
RECEPTION (Cash bar)	Potomac Ball Room	6:00 p.m. - 8:00 p.m.
DINNER ON YOUR OWN		

SATURDAY, OCTOBER 18, 1980

Registration opens 8:00 a.m., and closes at 3:00 p.m.

Jogging Along the Potomac, Shirts Sponsored by Prentice-Hall, Inc.
Meet Gary Ford in the Lobby 6:15 a.m.

MORNING SESSIONS - 1

8:30 a.m. - 10:00 a.m.

Special Topic Session: CONSUMER INFORMATION: IMPROVING
THE INTERACTION BETWEEN SUPPLY
AND DEMAND

Georgetown Room, Salons A,B

CHAIR: Hans B. Thorelli, Indiana University

Past research on the need for and design of consumer information has primarily focused on the consumer as a buyer, but has largely neglected the fact that early research and public policy focused on the idea that consumer information was necessary for the smooth functioning of the market system. The purpose of this session is to focus on research that examines the impact of consumer information on the interdependencies among sellers behaviors, market structure, and market performance.

Consumer Information and Workability of Competition - A Theoretical Framework

Guenter Schoeppe, University of Frankfurt
Christine Czerwonka, University of Frankfurt

The Influence of Product Test Reports on Suppliers' Behavior

Hans Raffee, University of Mannheim
Guenter Silberer, University of Mannheim
Wolfgang Fritz, University of Mannheim
Harald Hilger, University of Mannheim
Bernd Kierdorf, University of Mannheim

Sources for Product Ideas: A Proactive View on the Consumer

Gerd Fleischmann, University of Frankfurt

Perceived Risk as a Hint for Better Consumer Information and Better Products: Some New Applications of an Old Concept

Klaus G. Grunert, University of Hohenheim
Ingrid Gottschalk, University of Hohenheim
Konrad Dedler, University of Hohenheim

Discussant: Edward J. Heiden, Kirschner Associates

Special Topic Session (3 hrs.): WOMEN'S ROLES - AT WORK,
AT HOME, AND IN THE
MARKET PLACE

Francis Scott Key Room, Salon A

CHAIR: Mary Lou Roberts, Boston University

This session brings together leading scholars from several areas of study about women. The speakers present areas of basic research in the behavioral sciences that form the foundation for research into various aspects of women's consumer behavior. The purpose of the session is to stimulate more applied research in women's consumer behavior.

Behavioral Science Research on Women - Status, Problems and Prospects

Virginia E. O'Leary, American Psychological Association

An Evaluation of Sex Role Theories: The Clash Between Idealism and Reality

Susan Hesselbart, Florida State University

Housewives, Breadwinners, Mothers and Family Heads: The Changing Family Roles of Women

Janet Kohen, San Diego State University and the Institute for Social Research, University of Michigan

Psychological and Sociological Perspectives on Women's Paid and Unpaid Work Choices

Ruth B. Ekstrom, Educational Testing Service

Women in Paid Work: Some Consequences and Questions for Family Income and Expenditures

Hilda Kahne, Wheaton College

Women's Changing Roles: A Consumer Behavior Perspective

Mary Lou Roberts, Boston University

REAK: Coffee and Soft Drinks Lobby Level and Lower Level

10:00 a.m. - 10:30 a.m.

Panel Discussion: THE TRAINING AND CARING OF A CONSUMER RESEARCHER

Francis Scott Key Room, Salon B

CHAIR: Alan Sawyer, Ohio State University

The members of this panel will discuss ways to train a skilled consumer researcher. Panel participants and the audience will discuss issues such as theory vs application, how-to-do research courses vs practicum courses, business or government research career objectives vs an academic research career. The panel will also discuss priorities among different types of courses or seminars.

Panelists: Joel Cohen, University of Florida
Fred Reynolds, University of Georgia
Gerald Zaltman, University of Pittsburgh
Lawrence D. Gibson, General Mills, Inc.

Paper Session: NEW INSIGHTS FOR CONSUMER INFORMATION PROCESSING

Potomac Ballroom, Salon D

CHAIR: Murphy Sewall, University of Connecticut

A Consumer Based Approach for Establishing Priorities in Consumer Information Programs: Implications for Public Policy

Rohit Deshpande, University of Texas, Austin
S. Krishnan, Pennsylvania State University

Inferential Belief Formation Through the Use of Non-Information: An Example

David W. Finn, Texas Christian University

An Examination of Information Processing Traits: General Social Confidence and Information Processing Confidence

James M. Munch, Pennsylvania State Univ.
John L. Swasy, Pennsylvania State Univ.

Discussant: Jagdish Sheth, University of Illinois, Urbana

Paper Session: DIFFERENT CULTURES AND SEGMENTATION ANALYSIS

Georgetown Room, Salon C

CHAIR: P. J. O'Connor, University of Kentucky

The Influence of Geographic Subcultures in the United States

Kenneth A. Coney, Arizona State University
Del I. Hawkins, University of Oregon
Donald Roupe, University of Oregon

Consumerism in Developing Countries: The Brazilian Experience

J. Stanton, Temple University
R. Chandran, Temple University
J. Lowenhar, Temple University

Testing the Stability of Market Segmentation Analysis

D. A. Schellinck, Dalhousie University
Ian Fenwick, Northeastern University

Discussant: Roger M. Heeler, York University

Special Topic Session: WOMEN'S ROLES - AT WORK, AT HOME,
AND IN THE MARKET PLACE*Francis Scott Key Room, Salon A*

CHAIR: Mary Lou Roberts, Boston University

[Continued from 8:30]

Special Topic Session: BROADENING THE DIMENSIONS OF THE
AGE VARIABLE IN CONSUMER BEHAVIOR*Potomac Ballroom, Salon D*CHAIR: Leon G. Schiffman, City University of New York
(Baruch College)

The three papers in this session discuss the measurement and/or analysis of age in ways not usually considered in consumer research. These papers offer some new thinking on age as a variable including cohort analysis, age stereotypes, and self-perceived age. Implications for future consumer research are also discussed.

*Separating Age Cohort and Period Effects in Consumer Behavior*Fred D. Reynolds, University of Georgia
Joseph O. Rentz, University of Virginia*Cognitive Age: A Nonchronological Age Variable*Leon G. Schiffman, City University of New York
(Baruch College)
Benny Barak, Rutgers University*Issues in the Measurement and Use of Chronological and Non-chronological Age*Gerald Zaltman, University of Pittsburgh
Mark I. Alpert, University of Texas, Austin
Michael Heffring, University of Calgary

Discussant: Nora Ganim Barnes, Boston College

Special Topic Session: THE CONSUMER IN THE PUBLIC SECTOR:
APPROACHES TO ACCEPTABLE RISK*Georgetown Room, Salons A,B*

CHAIR: M. Venkatesan, University of Oregon

The purpose of this session is to discuss research approaches to determine the "safe" levels relative to products and public goods where there may be a number of hazards involved. Such issues range from the containment problems around nuclear plants to carcinogenicity of saccharine. The issues discussed are at the frontier of consumer research.

Approaches to Acceptable Risk: A Brief Summary
Paul Slovic, Decision Research*Approaches to Acceptable Risk: Their Relevance to Research in Consumer Behavior in the Public Sector*
M. Venkatesan, University of Oregon*Towards Determinants of Acceptable Risk: The Case of Product Risk*Arno Rethans, Pennsylvania State University
Gerald S. Albaum, University of Oregon

Discussant: Jacob Jacoby, Purdue University

Workshop: SURVEY RESPONSE RATES: DO WE KNOW WHERE THESE
PROBLEMS CAN LEAD?*Francis Scott Key Room, Salon B*

CHAIR: Joy Williams-Jones, Marketing Science Institute

Survey research is widely used for market analysis and is becoming accepted as evidence in legal proceedings. However, the application of many statistical tests to survey data is suspect unless a relatively high response rate is obtained. Moreover, some governmental agencies now require high response rates for a research-based testimony to be credible. The difficulty is there seems to be no mutually agreed upon definition for "response rate." The papers in this session highlight the nature of this research problem in terms of research replication, manuscript evaluation, and expert testimony. Research into the issue of an increased tendency of people not to respond will also be reported.

Non-Response in Consumer Surveys

Fred Wiseman, Northeastern University

Reporting of Response Rates in Professional Literature and the Ability to Replicate

Robert A. Mittelstaedt, Univ. of Nebraska, Lincoln

The Role of Response Rates in Evaluating Manuscripts for Publications

Robert Ferber, University of Illinois

Controversy in Washington: An FTC Case

Irving Roshwalb, Audits & Surveys Inc.

Lack of Agreement on the Standardization of Response rate Terminology in the Survey Research Industry

Joy Williams-Jones, Marketing Science Institute

Discussant: Keith Hunt, Brigham Young University

Paper Session: MULTI ATTRIBUTE SCALING MODELS

Georgetown Room, Salon C

CHAIR: Naresh Malhotra, Georgia Institute of Technology

*Two Models for Representing Unrestricted Choice Data*Paul E. Green, University of Pennsylvania
Wayne S. DeSarbo, Bell Laboratories*Limits to Accuracy in Conjoint Analysis*Franklin Acito, Indiana University
Richard W. Olshavsky, Indiana University*On Alternative Methodological Procedures for Conjoint Measurement*Thomas W. Leigh, Indiana University
David B. MacKay, Indiana University
John O. Summers, Indiana UniversityDiscussants: Jeffrey E. Danes, Virginia Tech
Philippe Cattin, Univ. of Connecticut

LUNCHEON

Awards Presentation,

Potomac Ballroom

Presidential Address: William L. Wilkie, University of Florida

12:00 p.m. - 1:30 p.m.

Special Topic Session: APPROACHES FOR DEVELOPING CAUSAL MODELS IN CONSUMER RESEARCH

Georgetown Room, Salons A,B

CHAIR: Kent B. Monroe, Virginia Tech

It has become increasingly apparent that an appropriate method to test empirically the complex models of consumer behavior is through the use of structural equation models. Indeed, the issue of model identification made explicit by the structural equation approach raises the question of whether most consumer research efforts are valid. Each speaker, in a tutorial approach, shows the steps from theory to model building, to estimation and hypothesis testing, to issues of validity. Step-by-step empirical examples are used.

Causal Modeling: A General Method for Developing and Testing Theories in Consumer Research

Richard P. Bagozzi, Massachusetts Institute of Technology

A Comparison of Causal Path and Econometric Modeling Approaches

Leonard Parsons, Georgia Institute of Technology

Investigating Causal Systems with Qualitative Variables: Goodman's Wonderful World of Logits

William R. Dillon, Univ. of Massachusetts, Amherst

Special Topic Session (3 hrs.): THE CONSUMER IN THE PUBLIC SECTOR

Francis Scott Key Room, Salon A

CHAIR: Kristian S. Palda, Queens University

When individuals vote for a political candidate or vote for a bond issue for public goods, behaviorally they are acting similarly as the consumer who "votes" for a brand of cereal at the supermarket. However, researchers recently have raised the question of whether "consumers" indeed have sufficient information about these public goods they are "buying." Viewing the citizen-taxpayer as a consumer raises many economic and behavioral research questions needing further attention. The papers in this session present current research on the consumer as a taxpayer who acquires public goods and the consumer as a voter.

Individual Preferences and Public Policy

Robert T. Deacon, Univ. of California, Santa Barbara

Substitution Between Public and Private Goods: An Overview of The Market Meeting Consumer Preferences

Charles L. Vehorn, General Accounting Office

*Consumer Information About Public Goods and The Workability of the Welfare State*Martin Pfaff, International Institute for Empirical Social Economics
Ernst Kistler, International Institute for Empirical Social Economics*Voting Participation in a Public Consumption Perspective*Randall G. Chapman, Queens University
Kristian S. Palda, Queens University*Measuring Subjective Valuation and Demand for Government Services*

Eli M. Noam, Columbia University

Fiscal Illusion and Consumer Sovereignty: An Exploratory Study

W. W. Pommerehne, Universitaet Zurich

Discussant: Robert Ferber, Univ. of Illinois, Urbana

Special Topic Session: HIGH TECHNOLOGY RESEARCH IN CONSUMER BEHAVIOR

Francis Scott Key Room, Salon B

CHAIR: Martin R. Lautman, Associates for Research in Behavior, Inc.
Richard C. Reizenstein, Univ. of Tennessee

The purpose of this session is to present applications of the "newest" research techniques in applied settings. Each panelist presents several examples of how state-of-the-art research has been used in a study within their company. Special emphasis is placed on "new methodological wrinkles" necessitated by doing large scale consumer research.

Forecasting Military Enlistment Decision Making
Al Martin, Office of the Secretary of Defense, Pentagon

Quantitative Modeling and Advertising Accountability
Lew Pringle, Director of Research, BBDO

From the Consumer Decision Process to Marketing: the Case for Micro Behavioral Simulation
Harry Sunenshine, Director of Research, Heublein
Richard Westwood, Beaumont Organization

Paper Session: THE EFFECT OF PRODUCT FAMILIARITY ON CONSUMER BEHAVIOR

Potomac Ballroom, Salon D

CHAIR: Melanie R. Wallendorf, Univ. of Michigan

The Moderating Effects of Cognitive Complexity and Prior Product Familiarity on the Predictive Ability of Selected Multi-Attribute Choice Models for Three Consumer Products

Chin Tiong Tan, University of Singapore, Singapore
Ira J. Dolich, Univ. of Nebraska, Lincoln

Toward a Cognitive Structure Conceptualization of Product Familiarity

Lawrence J. Marks, Pennsylvania State Univ.
Jerry C. Olson, Pennsylvania State Univ.

Product Familiarity and Learning New Information

Eric J. Johnson, University of Chicago
J. Edward Russo, University of Chicago

Discussant: M. Joseph Sirgy, Virginia Tech

Paper Session: THE ENVIRONMENTALLY CONCERNED CONSUMER

Georgetown Room, Salon C

CHAIR: Zarrel Lambert, Auburn University

Trade-Offs in Attribute Levels Made by Ecologically Concerned and Unconcerned Consumers When Buying Detergents

Karl E. Henion II, The Univ. of Texas, Austin
Russell Gregory, The Univ. of Texas, Austin
Mona A. Clee, The Univ. of Texas, Austin

A Causal Path Analysis of Ecological Behavior Relating to Marketing

Lawrence A. Crosby, University of Nebraska, Lincoln
James D. Gill, University of Nebraska, Lincoln

The Evolution of Distribution Channels for Solar Products: Consumer Decision Making in Perspective

Jerald W. Blakely, University of Oregon
Scott M. Smith, University of Oregon

Discussant: Michael Rothschild, Univ. of Wisconsin, Madison

Special Topic Session: THE CONSUMER IN THE PUBLIC SECTOR

Francis Scott Key Room, Salon A

CHAIR: Kristian S. Palda, Queens University

[Continued from 1:30]

Special Topic Session: METHODOLOGICAL ISSUES IN USING
PANEL DATA FOR MODEL BUILDING*Francis Scott Key Room, Salon B*

CHAIR: Donald R. Lehmann, Columbia University

Panel data are important for the development and testing of models of consumer behavior. Yet, panel data have a number of characteristics that affect the procedures used to draw inferences from such data. This session explores the problems involved in developing models using panel data by focusing on how these special characteristics of panel data must be recognized during analysis.

Attrition Bias in the Estimation of Econometric Models From Panel Data

Russell S. Winer, Columbia University

Scientific Advancement in Consumer Research: Some Problems Encountered When Using Consumer Panel Data

R. Dale Wilson, Batten, Barton, Durstine & Osborn, Inc.

*Decomposing the Correlation Matrix in Panel Data*John U. Farley, Columbia University
Donald R. Lehmann, Columbia University

Paper Session: INVOLVEMENT, LEARNING, AND ATTITUDES

Georgetown Room, Salon C

CHAIR: Donald J. Hempel, Univ. of Connecticut

An Exploration Into the Scaling of Consumers' Involvement with a Product Class

Peter H. Bloch, Portland State University

*Consumer Learning Through Experience: A Study and Experimental Paradigm*Joel Huber, Duke University
Terry Elrod, Columbia University*Measuring Consumer Attitudes Toward Alternative Check Verification Systems*David J. Barnaby, University of Tennessee
Richard C. Reizenstein, University of Tennessee
John W. Philpot, University of Tennessee

Discussant: David M. Gardner, Univ. of Illinois, Urbana

Special Topic Session: COGNITIVE APPROACHES TO CONSUMER
CATEGORIZATION OF PRODUCT/BRAND
CATEGORIES*Georgetown Room, Salons A,B*

CHAIR: Elizabeth C. Hirschman, New York University

A research area of increasing theoretical and strategic importance is how consumers develop perceptions of product/brand categories. The papers in this session present four theoretical frameworks from cognitive psychology that provide an explanation of consumer choice. Each theoretical framework is documented by a review of the relevant literature, and an attempt at empirical validation.

*Hierarchical Cognitive Content in Four Domains of Consumption*Elizabeth C. Hirschman, New York University
Susan Douglas, New York University*Usage-Situational Influences on Perceptions of Product Markets: Theoretical and Empirical Issues*

Rajendra K. Srivastava, Univ. of Texas, Austin

Product Perception as a Feature Matching Process

Michael Johnson, University of Chicago

A Means-End Model for Facilitating Analysis of Product Markets Based on Consumer Judgment

Jonathan Gutman, Univ. of Southern California

Paper Session: DIFFERENT INFLUENCES ON CONSUMER DECISION
MAKING*Potomac Ballroom, Salon D*

CHAIR: Betsy Gelb, University of Houston

*Mother's Attitudes and Perceptions of Children's Influence and Their Effect on Family Consumption*Mary Lou Roberts, Boston University
Lawrence H. Wortzel, Boston University
Robert L. Berkeley, Boston University*Four Situations and Their Perceived Effect on Husband and Wife Purchase Decision Making*Alvin C. Burns, Louisiana State University
Stephen P. DeVere, Louisiana State University*The Elderly Consumer: Past, Present, and Future*H. Lee Meadow, Virginia Tech
Stephen C. Cosmas, Virginia Tech
Andy Plotkin, Bridgewater State College

Discussant: E. H. Bonfield, Temple University

SUNDAY, OCTOBER 19, 1980

MORNING SESSIONS - 1

8:30 a.m. - 10:00 a.m.

Workshop: BEHAVIOR RESEARCH IN CONSUMER DECISION MAKING
AT A CROSSROADS

Georgetown Room, Salons A,B

CHAIR: Andrew A. Mitchell, Carnegie-Mellon University

The purpose of this workshop is to delineate the conceptual definitions and measurement procedures necessary for developing a theoretical understanding of consumer decision making. Each speaker addresses a particular research issue involved for developing such a theory of consumer decision making. The discussion will include the speakers and the audience.

Decision Strategies: Issues Surrounding the Dependent Variable in Decision Studies

James R. Bettman, Univ. of California, Los Angeles

Problem Framing Research

Peter L. Wright, Stanford University

Is Consumer Decision Making a Legitimate Offspring of Decision Making

J. Edward Russo, University of Chicago

Knowledge Structures, Production Systems and Decision Strategies

Merrie Brucks, Carnegie-Mellon University

Andrew A. Mitchell, Carnegie-Mellon University

Special Topic Session: CONSUMER DECISION MAKING: SOME
NEW DIRECTIONS FROM MANAGEMENT
SCIENCE

Georgetown Room, Salon C

CHAIR: R. Dale Wilson, Batten, Barton, Durstine &
Osborn, Inc.

Papers in this session concern research that interfaces the methodological orientation of the management or decision scientist with the behavioral orientation of the consumer researchers.

A Normative Model of Consumer Information Processing

David A. Aaker, Univ. of California, Berkeley

Michael R. Hagerty, Univ. of California, Berkeley

The Aggregate Effects of Induced Changes in Consumer Decision Structures

David J. Curry, University of Iowa

Jordan J. Louviere, University of Iowa

Michael J. Augustine, University of Iowa

The Cost of Thinking: Its Implications

Steven M. Shugan, University of Chicago

Discussant: Jagdish N. Sheth, University of Illinois,
Urbana

Paper Session: MEASURING THE EFFECT OF ADVERTISING

Francis Scott Key Room, Salon A

CHAIR: Dorothy Cohen, Hofstra University

The Effect of Recall on Belief Change: The Corrective Advertising Case

Neil K. Allison, University of South Carolina

Richard D. Mizerski, Federal Trade Commission

An Experimental Study of the Effects of Commercial TV Advertising and Pro-Consumer Product Test Results on TV

Emmanuel J. Cheron, Universite du Quebec a Rimouski

Jean Perrien, Universite du Quebec a Rimouski

Inquiry Response Rates, Cost and Revenue per Inquiry of Repetitive Print Advertising

Arch G. Woodside, University of South Carolina

Discussant: Robert F. Dyer, George Washington University

Paper Session: INFORMATION PROCESSING

Francis Scott Key Room, Salon B

CHAIR: Allan D. Schocker, University of Pittsburgh

Evoked Set Formation and Composition: An Empirical Investigation Under A Routinized Response Behavior Situation

Jacques E. Brisoux, Univ. of Quebec, Trois-Rivieres

Michel Laroche, Concordia University, Montreal

The Compensatory Dimension in Subjective Evaluation Processes: A Multimethod Validation

Ruby Roy Dholakia, Kansas State University

Developing an Instrument to Identify Individual Differences in the Processing of Pictorial and Other Non-Verbal Information

Flemming Hansen, University of Copenhagen

Neils Erik Lundsgaard, Frederiksberg Hospital

Discussant: George H. Haines, Jr., University of Toronto

BREAK: Coffee and Soft Drinks Lobby and Lower Levels

10:00 a.m. - 10:30 a.m.

Panel Discussion: REVITALIZING ATTITUDE THEORY AND RESEARCH: INCORPORATING IDEAS FROM COGNITIVE PROCESSING THEORY

Georgetown Room, Salons A,B

CHAIR: Jerry C. Olson, Pennsylvania State University

The purpose of this session is to present new theoretical ideas that may revitalize consumer attitude research. Drawing on cognitive psychology and information processing theory, the papers present perspectives that may be useful in explaining difficult conceptual issues in consumer attitude research, or in accounting for previously puzzling aberrations in the extant attitude data.

The Cognitive Structure Approach to Attitudes: A Review of Theoretical and Empirical Issues

Richard J. Lutz, Univ. of California, Los Angeles

A Memory Schema Account of Situational Influences on Attitude

Bobby Calder, Northwestern University

Belief/Knowledge Structures and Attitudes from a Cognitive Processing Perspective

Jerry C. Olson, Pennsylvania State University

Cognitive Processes in Attitude Formation and Change: Alternatives to Strict Cognitive Structure Explanation

Andrew Mitchell, Carnegie-Mellon University

Paper Session: CONSUMER SURVEY RESEARCH

Georgetown Room, Salon C

CHAIR: Philip G. Kuehl, University of Maryland

Investigating 'Income Refusals' in a Telephone Survey by Means of Logit Analysis

Robert A. Peterson, Univ. of Texas, Austin

Robert P. Leone, Univ. of Texas, Austin

Mohammad H. Sabertehrani, Univ. of Texas, Austin

The Effects of Salutation, Monetary Incentive and Degree of Urbanization on Mail Questionnaire Response Rate, Speed, and Quality

Linda L. Golden, Univ. of Texas, Austin

W. Thomas Anderson, Jr., Univ. of Texas, Austin

Louis K. Sharpe IV, Sharpe's Departmental Stores, Checotah, Oklahoma

What If Opinion Leaders Didn't Know More? A Question of Nomological Validity

Jacob Jacoby, Purdue University

Wayne D. Hoyer, Purdue University

Discussant: James H. Barnes, Jr., University of Georgia

Paper Session: THE CONSUMER IN THE RETAIL MARKET PLACE

Francis Scott Key Room, Salon B

CHAIR: Robert E. Wilkes, Texas Tech University

Retail Shopping Area Image: Structure and Congruency Between Downtown Areas and Shopping Centers

Michael J. Houston, Univ. of Wisconsin, Madison

John R. Nevin, Univ. of Wisconsin, Madison

Deviance and Dissatisfaction: An Exploratory Study

Michael K. Mills, Univ. of Southern California

Comprehensive Identification of Consumers' Marketplace Problems and What They Do About Them

F. Kelly Shuptrine, University of South Carolina

Gerhard Winglorz, University of South Carolina

Discussant: Donald Granbois, Indiana University

Paper Session: CONSUMER SATISFACTION AND BEHAVIOR

Francis Scott Key Room, Salon A

CHAIR: John Miller, University of Colorado

Testing Comparison Level and Predictive Expectations Models of Satisfaction

John E. Swan, Univ. of Alabama, Birmingham

Warren S. Martin, Univ. of Alabama, Birmingham

Complainers and Noncomplainers Revisited: Another Look at the Data

Kjell Grønhaug, Norwegian School of Economics and Business Administration

Gerald Zaltman, University of Pittsburgh

Effect of Satisfaction and Its Antecedents on Consumer Preference and Intention

Richard L. Oliver, Washington University (St. Louis)

Gerald Linds, Marsteller, Inc., Chicago

Developing Better Measures of Consumer Satisfaction: Some Preliminary Results

Robert A. Westbrook, University of Arizona

Richard L. Oliver, Washington University (St. Louis)

Discussant: Ralph L. Day, Indiana University

12:00 p.m.

CONFERENCE ENDS

Have a Very Successful Year

SEE YOU IN ST. LOUIS IN 1981!

TABLE OF CONTENTS

1979-80 Officers, Association for Consumer Research	iii
Reviewers for Paper Competition - 1980 Annual Conference	iv
Preface	v
1980 Conference Program	viii
Presidential Address	1
<i>William L. Wilkie, University of Florida</i>	
Presentation of the ACR Award	
<i>Harold H. Kassarian</i>	
"Fellow in Consumer Behavior" to John A. Howard and James F. Engel	6
<i>John Howard</i>	
Promotion in a Static Market	9
<i>James F. Engel</i>	
The Discipline of Consumer Research: Permanent Adolescence or Maturity?	12
PART I: THEORETICAL ISSUES	
A. EMERGING ISSUES IN LOW INVOLVEMENT THEORY	
<i>Clark Leavitt, Anthony G. Greenwald and Carl Obermiller</i>	
What is Low Involvement Low In?	15
<i>Richard E. Petty and John T. Cacioppo</i>	
Issue Involvement as a Moderator of the Effects on Attitude of Advertising Content and Context	20
<i>Andrew A. Mitchell</i>	
The Dimensions of Advertising Involvement	25
DISCUSSION PAPER:	
<i>Harold H. Kassarian</i>	
Low Involvement: A Second Look	31
B. ATTITUDE RESEARCH AND BEHAVIORAL INTENTIONS	
<i>Morris B. Holbrook, David A. Velez and Gerard J. Tabouret</i>	
Attitude Structure and Search: An Integrative Model of Importance-Directed Information Processing	35
<i>Paul W. Miniard</i>	
Examining the Diagnostic Utility of the Fishbein Behavioral Intentions Model	42
<i>David Brinberg</i>	
A Comparison of Two Behavioral Intentions Models	48
DISCUSSION PAPERS:	
<i>Paul R. Warshaw</i>	
A Discussion of Attitude Research and Behavioral Intentions	53
<i>Jeffrey E. Dames</i>	
Attitude Research and Behavioral Intentions: A Critical Review	57
C. INVOLVEMENT, LEARNING AND ATTITUDES	
<i>Peter H. Bloch</i>	
An Exploration into the Scaling of Consumers' Involvement with a Product Class	61
<i>Joel Huber and Terry Elrod</i>	
Consumer Learning Through Experience: A Study and Experimental Paradigm	66
<i>David J. Barnaby, John W. Philpot and Richard C. Reizenstein</i>	
Measuring Consumer Attitudes Toward Alternative Check Verifications	71

DISCUSSION PAPER:	
<i>David M. Gardner</i>	
Discussion: Involvement, Learning and Attitudes	76
D. CONSUMER SATISFACTION AND BEHAVIOR	
<i>John E. Swan and Warren S. Martin</i>	
Testing Comparison Level and Predictive Expectations Models of Satisfaction	77
<i>Kjell Grønhaug and Gerald Zaltman</i>	
Complainers and Noncomplainers Revisited: Another Look at the Data	83
<i>Richard L. Oliver and Gerald Linda</i>	
Effect of Satisfaction and Its Antecedents on Consumer Preference and Intention	88
<i>Robert A. Westbrook and Richard L. Oliver</i>	
Developing Better Measures of Consumer Satisfaction: Some Preliminary Results	94
E. COGNITIVE APPROACHES TO CONSUMER CATEGORIZATION OF PRODUCT/BRAND CATEGORIES	
<i>Elizabeth C. Hirschman and Susan P. Douglas</i>	
Hierarchical Cognitive Content: Towards a Measurement Methodology	100
<i>Rajendra K. Srivastava</i>	
Usage-Situational Influences on Perceptions of Product-Markets: Theoretical and Empirical Issues	106
<i>Michael E. Johnson</i>	
Context Effects in Product Perception	112
<i>Jonathan Gutman</i>	
A Means-End Model for Facilitating Analyses of Product Markets Based on Consumer Judgement	116
F. COGNITIVE PROCESSES	
<i>Rajesh Kanwar, Jerry C. Olson and Laura S. Sims</i>	
Toward Conceptualizing and Measuring Cognitive Structures	122
<i>Sanford L. Grossbart and Balusu Rammohan</i>	
Cognitive Maps and Shopping Convenience	128
<i>John G. Lynch</i>	
A Method for Determining the Sequencing of Cognitive Processes in Judgment: Order Effects on Reaction Times	134
G. THE EFFECT OF PRODUCT FAMILIARITY ON CONSUMER BEHAVIOR	
<i>Chin Tiong Tan and Ira J. Dolich</i>	
The Moderating Effects of Cognitive Complexity and Prior Product Familiarity on the Predictive Ability of Selected Multi-Attribute Choice Models for Three Consumer Products	140
<i>Larry J. Marks and Jerry C. Olson</i>	
Toward a Cognitive Structure Conceptualization of Product Familiarity	145
<i>Eric J. Johnson and J. Edward Russo</i>	
Product Familiarity and Learning New Information	151
DISCUSSION PAPER:	
<i>M. Joseph Sirgy</i>	
Product Familiarity: Critical Comments on Selected Studies and Theoretical Extensions	156
H. CONSUMER PERCEPTIONS	
<i>Albert J. Della Bitta and Kent B. Monroe</i>	
A Multivariate Analysis of the Perception of Value from Retail Price Advertisements	161
<i>John J. Wheatley</i>	
The Effect of Generic Products on Consumer Perceptions and Brand Choice	166

<i>Russell W. Belk</i>	
Determinants of Consumption Cue Utilization in Impression Formation: An Association Derivation and Experimental Verification	170
DISCUSSION PAPER:	
<i>Eric N. Berkowitz</i>	
Consumer Perceptions: A Discussion	176
I. PERCEPTION OF TIME AND CONSUMER BEHAVIOR	
<i>James McCullough, Douglas MacLachlan and Reza Moïnpour</i>	
Temporal Links Between Preference and Perception	178
<i>Philip E. Hendrix and Claude R. Martin, Jr.</i>	
Temporal Incongruity in Consumer Behavior	182
<i>Rebecca H. Holman</i>	
The Imagination of the Future: A Hidden Concept in the Study of Consumer Decision Making	187
DISCUSSION PAPER:	
<i>Wesley J. Johnston</i>	
Time and Consumer Behavior	192
PART II: METHODOLOGICAL ISSUES	
A. APPROACHES FOR DEVELOPING CAUSAL MODELS IN CONSUMER RESEARCH	
<i>Richard P. Bajocchi</i>	
Causal Modeling: A General Method for Developing and Testing Theories in Consumer Research	195
<i>Leonard J. Parsons</i>	
A Comparison of Causal Path and Econometric Modeling Approaches	203
<i>William R. Dillon</i>	
Investigating Causal Systems with Qualitative Variables: Goodman's Wonderful World of Logits	209
B. METHODOLOGICAL ISSUES IN USING PANEL DATA FOR MODEL BUILDING	
<i>Russell S. Winer</i>	
Attrition Bias in the Estimation of Econometric Models from Panel Data	220
<i>A. Dale Wilson</i>	
Scientific Advancement in Consumer Research: Some Problems Encountered When Using Consumer Panel Data	227
<i>Donald R. Lehman and John U. Farley</i>	
Decomposing the Correlation Matrix in Panel Data	233
C. UNIVERSITY PANELS FOR CONSUMER RESEARCH	
<i>Robert Farber</i>	
<i>Linda R. Lammom</i>	
Research Panels in Consumer Behavior	238
D. USING THE BUREAU OF LABOR STATISTICS CONSUMER EXPENDITURE SURVEY IN CONSUMER RESEARCH	
<i>N. Gail Hoff</i>	
Overview of the Consumer Expenditure Surveys	245
<i>Eva N. Jacobs</i>	
Consumer Expenditure Survey Data: Uses by BLS	251
<i>Robert B. Pearl</i>	
An Evaluation of the Methodology of the 1972-73 U.S. Consumer Expenditure Survey	254
<i>Gregory D. Uppah and Seymour Sudman</i>	
The Consumer Expenditure Survey: Prospects for Consumer Research	262

E. SURVEY RESPONSE RATES: DO WE KNOW WHERE THESE PROBLEMS CAN LEAD?

Fred Wiseman
 Nonresponse in Consumer Surveys 267

Robert Mittelstaedt
 Response Rates and Internal Validity: Some Implications for Our Literature 270

Robert Ferber
 The Role of Response Rates in Evaluating Manuscripts for Publication 274

Irving Roshwalb
 Recent Controversy in Washington: An FTC Case 276

Joy Williams-Jones
 Lack of Agreement on the Standardization of Response Rate Terminology in the Survey Research Industry 281

F. CONSUMER SURVEY RESEARCH

Robert A. Peterson, Robert P. Leone and Mohammad H. Sabertehrani
 Investigating "Income Refusals" in a Telephone Survey by Means of Logit Analysis 287

Linda L. Golden, W. Thomas Anderson and Louis K. Sharpe
 The Effects of Salutation, Monetary Incentive, and Degree of Urbanization on Mail
 Questionnaire Response Rate, Speed and Quality 292

Jacob Jacoby and Wayne D. Hoyer
 What if Opinion Leaders Didn't Know More? A Question of Nomological Validity 299

DISCUSSION PAPER:
James H. Barnes, Jr.
 A Discussion of Survey Research 304

G. MULTI ATTRIBUTE SCALING MODELS

Naresh K. Malhotra
 Multiattribute Scaling Models: Some Observations 306

Paul E. Green and Wayne S. DeSarbo
 Two Models for Representing Unrestricted Choice Data 309

Franklin Acito and Richard W. Olshavsky
 Limits to Accuracy in Conjoint Analysis 313

Thomas W. Leigh, David B. MacKay and John O. Summers
 On Alternative Experimental Methods for Conjoint Analysis 317

DISCUSSION PAPER:
Jeffrey E. Danes and Philippe Cattin
 Multiattribute Choice Models: A Critical Review 323

H. HIGH TECHNOLOGY RESEARCH IN CONSUMER BEHAVIOR

Shel Feldman and A.J. Martin
 Initial Market Research Steps Toward A Model of the Military Enlistment Decision 329

Richard A. Westwood and Harry S. Sunenshine
 Marketing Action Based on Consumer Decision Processes: The Case for Micro-behavioral Simulation 332

PART III: THE ROLE OF INFORMATION

A. NEW INSIGHTS FOR CONSUMER INFORMATION PROCESSING

Rohit Deshpande and S. Krishnan
 A Consumer Based Approach for Establishing Priorities in Consumer Information Program:
 Implications for Public Policy 338

David W. Finn
 Inferential Belief Formation Through the Use of Non-Information: An Example 344

James M. Munch and John L. Swasy
 An Examination of Information Processing Traits: General Social Confidence and Information
 Processing Confidence 349

DISCUSSION PAPER:	
<i>Jagdish N. Sheth</i>	
Discussion	355
B. INFORMATION PROCESSING	
<i>Jacques E. Brisoux and Michel Laroche</i>	
Evoked Set Formation and Composition: An Empirical Investigation Under a Routinized Response Behavior Situation	357
<i>Ruby Roy Dholakia</i>	
The Compensatory Dimension in Subjective Evaluation Processes: A Multimethod Validation	362
<i>Flemming Hansen and Niels Erik Lundsgaard</i>	
Developing an Instrument to Identify Individual Differences in the Processing of Pictorial and Other Non-Verbal Information	367
C. CONSUMER INFORMATION: IMPROVING THE INTERACTION BETWEEN SUPPLY AND DEMAND	
<i>Gunter Schoppe and Christine Czerwonka</i>	
Consumer Information and Workability of Competition: A Theoretical Framework	374
<i>Wolfgang Fritz, Harald Hilger, Günter Silberer and Hans Raffée</i>	
The Impact of Non-Commercial Product Test Information on Commerce and Industry: Outline of the Study and Preliminary Findings	381
<i>Gerd Fleischmann</i>	
Sources for Product Ideas: A Proactive View on the Consumer	386
<i>Konrad Dedler, Ingrid Gottschalk and Klaus G. Grunert</i>	
Perceived Risk as a Hint for Better Consumer Information and Better Products: Some New Applications of an Old Concept	391
D. THE NEGATIVE SIDE OF CONSUMER INFORMATION	
<i>Marc Weinberger, Chris T. Allen and William R. Dillon</i>	
Negative Information: Perspectives and Research Directions	398
<i>John C. Mowen, David Jolly and Gary S. Nickell</i>	
Factors Influencing Consumer Responses to Product Recalls: A Regression Analysis Approach	405
<i>Carol A. Scott and Alice M. Tybout</i>	
Theoretical Perspectives on the Impact of Negative Information: Does Valence Matter?	408
E. ASSESSING CORRECTIVE ADVERTISING EFFECTS	
<i>Jacob Jacoby, Wayne D. Hoyer and David A. Sheluga</i>	
Viewer Miscomprehension of Televised Communication: A Brief Report of Findings	410
<i>Kenneth L. Bernhardt, Thomas C. Kinnear, Michael B. Mazis and Bonnie B. Reece</i>	
Impact of Publicity on Corrective Advertising Effects	414
<i>Jacob Jacoby, Margaret C. Nelson and Wayne D. Hoyer</i>	
Correcting Corrective Advertising	416
F. MEASURING THE EFFECT OF ADVERTISING	
<i>Neil K. Allison and Richard W. Mizerski</i>	
The Effects of Recall on Belief Change: The Corrective Advertising Case	419
<i>Emmanuel J. Cheron and Jean Perrien</i>	
An Experimental Study of the Effects of Commercial TV Advertising and Pro-Consumer Product Test Results on TV	423
<i>Arch G. Woodside</i>	
Inquiry Response Rates, Cost and Revenue per Inquiry of Repetitive Print Advertising	428
G. CONSUMER RESPONSE TO ADVERTISING	
<i>Mary Jane Schlinger</i>	
Consumer Response to Advertising: Implications for Copy Testing and Copy	432

<i>Larry Percy and Martin R. Lautman</i> Linear Effects of Cognitive Response to Advertising	434
<i>John C. Mowen and Stephen W. Brown</i> On Explaining and Predicting the Effectiveness of Celebrity Endorsers	437
<i>Morry Ghingold</i> Guilt Arousing Marketing Communications: An Unexplored Variable	442
DISCUSSION PAPER:	
<i>Robert L. King</i> Consumer Response to Advertising: A Discussion of Three Studies	449

PART IV: PUBLIC POLICY ISSUES

A. PUBLIC POLICY AND CONSUMER RESEARCH IN THE 1980'S

<i>Kenneth L. Bernhardt and Ronald Stiff</i> Public Policy Update: Perspectives On the Federal Trade Commission	452
<i>Michael B. Mazis</i> The Future of Consumer Protection Regulation	455
<i>Graham T.T. Molitor</i> Consumer Policy Issues: Global Trends for the 1980's	458
<i>Hans B. Thorelli</i> Swedish Consumer Policy, Its Transferability and Related Research Implications	467

B. THE ROLE OF CONSUMER RESEARCH IN THE FORMATION OF PUBLIC POLICY: AN EXAMPLE OF AN INTEGRATED RESEARCH AND DEVELOPMENT PLAN

<i>James T. Heimbach</i> Defining the Problem: The Scope of Consumer Concern with Food Labeling	474
<i>Edwin C. Hackleman</i> Food Label Information: What Consumers Say They Want and What They Need	477
<i>Robert J. Vandenberg</i> Food Label Information: What Consumers Say They Use and What They Actually Use	484
<i>Michael L. Stewart</i> Commercial Market Tracking Systems: Applications to Policy Formation Regarding Nutrition Labeling	488
<i>Raymond E. Schucker</i> Implementation of a Revised Food Labeling Policy: Evaluation and Tracking	493

C. THE CONSUMER IN THE PUBLIC SECTOR: APPROACHES TO ACCEPTABLE RISK

<i>Paul Slovic, Daruch Fischhoff and Sarah Lichtenstein</i> Facts and Fears: Societal Perception of Risk	497
<i>M. Venkatesan</i> Consumer Behavior and Acceptable Risk: Some Research Issues	503
<i>Arno J. Rethans and Gerald S. Alba</i> Towards Determinants of Acceptable Risk: The Case of Product Risks	506
DISCUSSION PAPER:	
<i>Jacob Jacoby</i> Some Perspectives on Risk Acceptance	511

PART V: BROADENING CONSUMER RESEARCH

A. THE CONSUMER IN THE PUBLIC SECTOR

<i>Robert T. Deacon</i> Individual Preferences and Public Policy	517
---	-----

<i>Charles L. Vehorn</i> Substitution Between Public and Private Goods: An Overview of the Market Meeting Consumer Preferences . . .	523
<i>Martin Pfaff and Ernst Kistler</i> Consumer Information About Public Goods and the Workability of the Welfare State	527
<i>Randall G. Charman and Kristian S. Palda</i> Voting Participation in a Public Consumption Perspective	530
<i>Eli M. Noam</i> Measuring Subjective Valuation and Demand for Government Services	534
<i>Werner W. Pommerehne</i> Fiscal Illusion and Consumer Sovereignty: An Exploratory Study	539
DISCUSSION PAPER:	
<i>Robert Ferber</i> Comments	545
B. DEVELOPING NEW RESEARCH TRADITIONS IN CONSUMER RESEARCH	
<i>R. Bruce Hutton and Dennis L. McNeill</i> The Impact of Program Evaluation on Research Methodology	547
<i>Rachel Dardis and Julie Stremel</i> Risk-Benefit Analysis and the Determination of Acceptable Risk	553
<i>Liisa Uusitalo and Jyrki Uusitalo</i> Scientific Progress and Research Traditions in Consumer Research	559
<i>Johan Arndt, Kjell Grønhaug, Richard E. Homans, R. Neil Maddox and Frederick E. May</i> Toward a Replication Tradition in Consumer Behavior: Cross-Cultural Replication of Bennett and Mandell's Study of the Learning-Information Seeking Hypothesis	564
DISCUSSION PAPER:	
<i>Peter D. Bennett</i> Methodology and Meta-Methodology in Consumer Research: A Commentary	568
C. WOMEN'S ROLES - AT WORK, AT HOME, AND IN THE MARKET PLACE	
<i>Susan Hesselbart</i> An Evaluation of Sex Role Theories: The Clash Between Idealism and Reality	570
<i>Janet A. Kohon</i> Housewives, Breadwinners, Mothers, and Family Heads: The Changing Family Roles of Women	576
<i>Ruth B. Ekstrom</i> Psychological and Sociological Perspectives on Women's Paid and Unpaid Work Choices	580
<i>Hilda Kahne</i> Women in Paid Work: Some Consequences and Questions for Family Income and Expenditures	585
<i>Mary Lou Roberts</i> Women's Changing Roles -- A Consumer Behavior Perspective	590
D. BROADENING THE DIMENSIONS OF THE AGE VARIABLE IN CONSUMER BEHAVIOR	
<i>Joseph O. Rents and Fred D. Reynolds</i> Separating Age, Cohort and Period Effects in Consumer Behavior	596
<i>Benny Barak and Leon G. Schiffman</i> Cognitive Age: A Nonchronological Age Variable	602
E. CONSUMER RESEARCH AND QUALITY OF LIFE	
<i>Lynette S. Unger and Jerome B. Kerman</i> Leisure and the QOL Construct: A Review and Some Modest Proposals	607
<i>Lynn Langmeyer and George Miaoulis</i> The Consumer and the Health Care Process	612

<i>David A. Aaker and Richard P. Bagozzi</i> Attitudes Toward Public Policy Alternatives to Reduce Air Pollution	616
DISCUSSION PAPER:	
<i>William K. Zikmund</i> Conceptual and Methodological Issues in Research on the Quality of Life	622
F. THE ENVIRONMENTALLY CONCERNED CONSUMER	
<i>Karl E. Henion, Russell Gregory and Mona A. Clee</i> Trade-offs in Attribute Levels Made by Ecologically Concerned and Unconcerned Consumers When Buying Detergents	624
<i>Lawrence A. Crosby and James D. Gill</i> A Causal Path Analysis of Ecological Behavior Relating to Marketing	630
<i>Jerald W. Blakely and Scott M. Smith</i> The Evolution of Distribution Channels for Solar Products: Consumer Decision Making in Perspective	637
DISCUSSION PAPER:	
<i>Michael L. Rothschild</i> Providing Reinforcers for Environmentally Unconcerned Consumers	642
PART VI: PATRONAGE THEORY AND RESEARCH	
A. PATRONAGE THEORY	
<i>Robert F. Lusch</i> Integration of Economic Geography and Social Psychological Models of Patronage Behavior	644
<i>Elizabeth C. Hirschman</i> Women's Self-Ascribed Occupational Status and Retail Patronage	648
<i>William R. Darden, Donna K. Darden, Roy Howell and Shirley J. Miller</i> Consumer Socialization Factors in a Patronage Model of Consumer Behavior	655
B. PATRONAGE RESEARCH APPLICATIONS AND METHODS	
<i>Robert A. Peterson</i> An Exploratory Investigation of Mediating Factors in Retail Store Image Responses	662
<i>Stephen J. Arnold, Victor Roth and Douglas J. Tigert</i> Conditional Logit Versus MDA in the Prediction of Store Choice	665
<i>Roy D. Howell and Jerry D. Rogers</i> Research into Shopping Mall Choice Behavior	671
C. THE CONSUMER IN THE RETAIL MARKET PLACE	
<i>Michael J. Houston and John R. Nevin</i> Retail Shopping Area Image: Structure and Congruency Between Downtown Areas and Shopping Centers	677
<i>Michael K. Mills</i> Deviance and Dissatisfaction: An Exploratory Study	682
<i>F. Kelly Shuptrine and Gerhard Wenglorz</i> Comprehensive Identification of Consumers' Marketplace Problems and What They Do About Them	
DISCUSSION PAPER:	
<i>Donald Granbois</i> An Integrated View of the Store Choice/Patronage Process	693
PART VII: INFLUENCES ON CONSUMER BEHAVIOR	
A. SOCIAL CLASS AND SOCIAL INFLUENCE	
<i>J. Michael Munson and W. Austin Spivey</i> Product and Brand User Stereotypes Among Social Classes	696

<i>Terence A. Shimp and J. Thomas Yokum</i>	
Extensions of the Basic Social Class Model Employed in Consumer Research	702
<i>Scott B. MacKenzie and Judy L. Zaichkowsky</i>	
An Analysis of Alcohol Advertising Using French and Raven's Theory of Social Influence	708
B. DIFFERENT CULTURES AND SEGMENTATION ANALYSIS	
<i>Del I. Hawkins, Don Roupe and Kenneth A. Coney</i>	
The Influence of Geographic Subcultures in the United States	713
<i>John Stanton, Rajan Chandran and Jeffrey Lowenhar</i>	
Consumerism in Developing Countries - The Brazilian Experience	718
<i>Tony Schellinek and Ian Fenwick</i>	
Testing the Stability of Market Segmentation Analysis	723
DISCUSSION PAPER:	
<i>Roger M. Heeler</i>	
Vive Les Differences	728
C. DIFFERENT INFLUENCES ON CONSUMER DECISION MAKING	
<i>Mary Lou Roberts, Lawrence H. Wortzel and Robert L. Berkeley</i>	
Mothers' Attitudes and Perceptions of Children's Influence and Their Effect on Family Consumption	730
<i>Alvin C. Burns and Stephen P. DeVere</i>	
Four Situations and Their Perceived Effects on Husband and Wife Purchase Decision Making	736
<i>H. Lee Meadow, Stephen C. Cosmas and Andy Plotkin</i>	
The Elderly Consumer: Past, Present, and Future	742
DISCUSSION PAPER:	
<i>E. H. Bonfield</i>	
Different Influences on Consumer Decision Making	748
D. BEHAVIOR RESEARCH IN CONSUMER DECISION MAKING	
<i>Merrie Brucks and Andrew Mitchell</i>	
Knowledge Structures, Production Systems and Decision Strategies	750

William L. Wilkie, University of Florida

Introductory Comments

[Before beginning my formal address, I would like to thank Kent Monroe, who acted as Conference Chairman for this event. Having had the pleasure of doing this several years ago, I can appreciate the substantial effort that he, his Program Committee, and his Arrangements Co-Chairmen, Gary Ford and Paul Bloom, have put into this conference.

I would also like to note that it gives me particular pleasure to be able to deliver a Presidential Address in Washington, D.C. ...especially in this election year. While thinking about what I might say, I began to appreciate the power inherent in the Presidency. As I stand before this distinguished assemblage, therefore, I am tempted to announce that I have ordered -- and there will in fact be -- a massive cut in government spending this coming year..... This will be coupled with a major tax cut for both businesses and individuals.

Inflation will end, and employment will rise..... Our savings rate will increase..... Productivity gains will be phenomenal.....

I am today placing our entire military forces on maneuvers in the Mid-East; this will result in immediate cessation of all hostilities in that region, and will shortly bring world peace and tranquility.....

Overall, I am pleased to announce that we shall shortly reach all of our goals in life, both individually and as a society.....

That is all I have to announce for today. Unfortunately, due to the press of world affairs, I will be unable to take questions at this time.....

Beyond the pleasure I get from being able to make such Presidential pronouncements, it is also a personal pleasure to return to Washington, D.C. for this occasion. I once lived here and came to feel great affection for the town. Some of you may not know, by the way, that Washington is considered to be the point at which the North and South of the United States come together. This special characteristic once led President John F. Kennedy to describe Washington as a city that managed to uniquely blend "old-time Southern speed and efficiency, with all the graciousness of urban Northern charm."

At any rate, unlike some other Presidential addresses this year, my speech today will not be mean..... Unlike some others, it won't be taken from note cards, and won't be memorized. Hopefully, though, it will be brief.]

ACR's Mission

My topic today deals with the central mission of the Association for Consumer Research, and how well we seem to be accomplishing this mission. After reviewing why ACR was founded, and how it proceeded at its start (Pratt 1974), I think that ACR's two key goals are quite clear:

- The pursuit and development of knowledge.
- A sharing in this pursuit.

The pursuit of knowledge is a basic goal for this type of organization; I intend to discuss it in more detail in the body of my speech. The explicit "sharing" goal is less typical, however. ACR was intended to be interdisciplinary in nature, and the pursuit of knowledge was intended to be accomplished across institutional structures. It was intended that persons in business, in government, and in academia should come together with this common goal,

each contributing significant perspectives and findings to the body of knowledge. As we all have seen, the pairing of these objectives presents a challenging task.

It turns out that our field's progress and problems have provided by far the most popular topics for past Presidential Addresses to this Association. They have been approached from several directions, including:

- The amount of work.
- The quality of work.
- Problems in theory and methods.
- Relevance of the work.
- Proper domain of our work.
- Some ethical issues in dissemination of our work.

Today, I would like to take yet another approach to this topic, mainly to ask in a broader sense, "When we, as an Association, assert that we want to pursue knowledge, what do we really mean?"

Overview of the Address

There seem to be three special issues involved here, and I have organized my talk according to these. First, there is the question of each member's personal choice whether or not to attempt to contribute to the development of knowledge. I would like to briefly address this in terms of "Individual Choice: Contribution and/or Participation."

Choosing to contribute to knowledge is no light matter. This is a point that I'd like to amplify upon in a second section, which discusses the literature on "The Meaning of Knowledge." In requiring that we possess knowledge before presuming to add to it, the choice of attempting to contribute suggests that past achievement is important.

In addition, of course, contributions will require future performance. The quality of such future performance -- or lack of same -- depends on one's willingness to work toward the goal. I would like to raise some basic issues here in the third section, entitled "Personal Acceptance of the Responsibilities."

On a personal note, because of the nature of this topic, there is a risk that this topic might seem pretentious and/or presumptuous. I want to begin by saying that I didn't choose this topic without reservations, and am not raising it as if I personally feel especially knowledgeable about it. I think, furthermore, that the real success or failure of this speech will actually depend on each of you, and how willing you are to deal with these issues at a personal level. They are important issues for our field, our Association, and for many of us as individuals.

Individual Choice: Contribution and/or Participation

As its base, the development of knowledge requires that each individual choose whether or not to undertake this effort. With respect to ACR, we should recognize that membership does not necessarily mean contribution to knowledge. For example, I expect that many of you have come to the conference solely for what you might take away from it, either to use in decision making, or because you simply find it to be of personal interest. There is certainly nothing wrong with this; one of our Associa-

tion's chief purposes is the dissemination of knowledge. On the other hand, we should not equate the roles of receiving knowledge and contributing to knowledge.

There are two key issues that I'd like to raise in this regard. The first issue concerns the motivation to contribute, while the second involves institutional barriers to such contributions. I expect that all of us can appreciate the personal motivations to contribute to knowledge, so will not attempt to delve into this area in any depth (whether this reflects insight or cowardice is not exactly clear to me).

At any rate, there is a book entitled The Cultivated Mind (Hodnett 1963) which bears on the motivation to contribute. I mention this book especially to this audience because Dr. Hodnett served both in academe and in business, gaining a background which should allow him to communicate well with almost all of us. His description of the "cultivated mind" seems an especially appropriate summary of why an individual might choose to attempt to contribute to knowledge. In brief, he asserts that the cultivated mind is distinguished by three qualities. It is a mind that is:

- Conceptual
- Discriminating
- Humane

The "conceptual" property refers to a mind that seeks understanding, desires to know, to know why, and is willing to conceptualize and speculate. The "discrimination" characteristic suggests that the cultivated mind is aware of and sensitive to value. The cultivated mind is "humane" in the sense that it is able to move beyond an obsession with self and a pervasive focus on its personal day-to-day affairs. The cultivated mind is thus capable of adopting a serious concern with mankind and the nature of the human existence.

Barriers and Facilitators

While individual motivation is basic to the personal choice to try to contribute to knowledge, it would be inappropriate not to recognize that some substantial barriers presently exist for certain of our members. The most obvious of these are institutional in nature. Most ACR members who work in business or government find themselves in roles which neither provide for, nor reward, attempts to contribute to general knowledge. This is a chronic problem, and well recognized by all of us. I certainly don't have a clear resolution for it, but would urge you to consider ways in which such contributions can be facilitated within the Association's rubric.

For example, I feel that our field is in need of more descriptive research on consumer markets and aggregate behavior. This form of contribution is likely to be readily available within business and government institutions, and would not require much additional effort to transmit to the field. Similarly, persons in business have closer and more specific contact with the actual consumer marketplace and purchasing behavior. This experience can provide a fertile ground for the development of insights into consumer behavior. These insights, while not achieving the rigorous standards of science, could provoke significant research investigations and refinements by others attempting to follow up on them. The Association's role would seem to provide vehicles for the presentation of these types of contributions, and to provide a receptive atmosphere for them. The stimulus, however, still needs to occur at the individual level.

With respect to other forms of barriers, I believe that the academic sector possesses some substantial constraints as well. These, however, are more informal in nature, and harder to pin down. I would like to raise them in a later portion of my talk.

What does it Mean to Add to Knowledge?

A Shared Heritage and Value Legacy

I mentioned earlier that a personal choice to attempt to contribute to knowledge is not an empty or minor gesture. The first perspective which I encountered in beginning to address the question was a realization that, by opting to assert a goal of "adding to knowledge," we have elected to join the thousands of men and women who have also sought this goal, and who have gone before us. In pursuing the goal, they have provided us with the knowledge which we now have available. In addition, they have passed on insights, philosophies, and warnings about the nature of the quest. It is sobering to realize the nature of the goal, and humbling to appreciate our significance in the larger picture.

The recognition of the others who have gone before us, and of those in other fields who currently share our broader goal, provides a useful perspective, an appropriate humility, and an important source of guidance for approaches to our development of knowledge. Before turning to consumer behavior matters, I would like to briefly share with you some things I've discovered in reading about the development of knowledge.

The Great Disciplines of the Mind: The Keys to Knowledge

In reading about the development of knowledge, it appears that there is a general agreement that certain disciplines are fundamental; additions to knowledge in any area must necessarily depend upon prior knowledge in these basic disciplines. In a sense, this suggests that before we can presume to add to knowledge, we must possess knowledge.

While there is not complete agreement on the exact number of basic disciplines of the mind, the following five receive great support:

- Mathematics
- Language
- History
- Logic
- Philosophy

Mathematics is the key to the natural and material sciences; it allows us to comprehend space and time, and to understand the basis of our existence. Language allows us to add human reason, and to communicate and transmit knowledge. History, which in some sense is our repository of knowledge, requires language for its basic material. Logic introduces the principles of reasoning, and employs the elements of language to arrive at new facts and new knowledge.

Philosophy, which can include logic, also transcends the narrow principles to provide us with consideration of norms and ideals for mankind. Within this area, metaphysics, the study of the first principles of being and reality, is perhaps the highest science. Within the melding of logic and philosophy, then, we are able to address the true (logic) and the good (ethics), and thus to approach some significant ideals of mankind. In addition (and beyond this framework) are the aesthetic areas, which aim to enhance our appreciation of the beautiful elements of life. Here are included art, music, and literature.

I think there are two distinct conclusions which derive from even an elementary consideration of these thoughts. First, the store of knowledge which humanity already possesses is awesome. One way to become more familiar with the scope of knowledge, if not with its depths and substance, is to examine the Propaedia of the Encyclopaedia Britannica. This volume wrestles with the question of how to conceptualize and organize "knowledge" in order to be able to summarize and present it in a complete and coherent fashion. In this regard, Mortimer Adler's remarks in that volume are especially illuminating.

The second conclusion that I reached pertains more directly to ACR and our topic today. The consideration of all knowledge places consumer behavior in an appropriate perspective; as a very small part of the puzzle, we should be open to appreciating that more developed areas might provide important insights for us. At the same time, though, it is important to see that we do belong in the puzzle, and this is reassuring. In joining the puzzle, however, we must accept that our additions to knowledge will come from, and depend upon, our mastery of the underlying disciplines or keys to knowledge -- the five areas listed above. It thus becomes appropriate for us to ask how well we have mastered these areas.

This is an uncomfortable question to raise. At the individual level, the most troublesome question might well be, "Can I contribute to knowledge if I do not possess mastery of any of the great disciplines?" Subsequent to this, one might inquire, "Can I contribute to knowledge if I am ignorant of several of the great disciplines?"

As a field of study, it would be worthwhile to assess which of these disciplines are relatively more significant for knowledge of consumer behavior, and whether some of the disciplines (philosophy leaps to mind) can be continually overlooked within our field. It is also important that we at least tentatively address the issues of whether sets of specialists, steeped in the application areas of one discipline, offer the most chance for advancing knowledge across the field of consumer behavior.

Trends

More broadly, recent trends are not especially reassuring on these grounds. For example, I recently noticed that only 12% of all college graduates are now matriculating in the traditional humanities areas of history, philosophy, language, literature, and the fine arts. This suggests a dwindling pool of persons available to offer these special forms of perspectives on our work in the future. Conversely, 90% of all scientists and engineers in the history of the world are alive and working today.

On a somewhat different note, the quality level at which the disciplines are being transmitted may be falling -- it is interesting to consider that there is no necessary reason that knowledge must increase over generations. For example, it is disturbing that the College Boards SAT scores, have just been reported to have declined for the 18th consecutive year. Whether we might wish to quarrel with the validity of these scores as predictive devices or not, they do make an effort to tap the level of understanding of three of the five disciplines listed above -- mathematics, language, and logic -- and these measures are falling.

In summary, I have found that reading about knowledge can provide a rich perspective on our field and our efforts. Before rushing to do research, it is reasonable to contemplate what most needs to be done. In addition, it is also reasonable to question our capability in certain areas. We should be willing to examine our training, our education, and our interests, and should be honest in appraising areas of weakness. When we find areas of weakness, moreover, it is appropriate to consider that attempts to build strength in those areas will pay longer term dividends than attempting to proceed as if these deficiencies weren't a problem.

Personal Acceptance of the Responsibilities

Not only does the choice to attempt to contribute to knowledge require a strong knowledge base, but it also brings with it a substantial set of responsibilities. Without a willingness to shoulder these, a decision to attempt to contribute to knowledge can be virtually meaningless.

In this regard, I would like to especially recommend The House of Intellect, a book by Professor Jacques Barzun (1959). In it, Professor Barzun delves into the requirements for -- and threats to -- high standards of intellect and contributions to knowledge. It is a powerfully written volume, and certainly does not attempt to please everyone. Because of this, however, it can serve as both an inspirational volume for some of us, and as a hard-line test of our personal willingness to take on the inherent responsibilities.

As to the responsibilities, there are two key types which stand out: effort and standards. With respect to effort, we should recognize that contributing to knowledge is a difficult and demanding task. It is work. At the same time, of course it brings benefits.

It is impossible to discuss effort, however, without also referring to standards, as these will directly affect the way in which effort is expended. As I will discuss shortly, a reallocation of effort is probably more needed than an increase in effort per se. Standards will provide a valuable resource in this reallocation of effort. One way to view this relationship is to recall that the choice to attempt to contribute to knowledge brings with it an implicit agreement to have one's work subjected to external judgments, which are expected to employ rigorous criteria. A most significant accompaniment to the development of scholarship -- and in some sense a measure of such development -- is the internalization of these standards or criteria, so that the contributor practices self-criticism at a high level before inflicting his or her output on the external world.

Standards of Quality

One type of standard relates to the inherent quality level of the work; how accurate, precise, valid, and/or insightful should a contribution have to be? Many persons (myself included) believe that our field suffers from the general acceptance of too low a quality standard. Knowledge is not advanced when a reader or listener must approach an offering with a legitimate scepticism as to its underlying veracity.

I believe that ACR, and our field more generally, would benefit from a consensus judgment that higher quality standards are in order. We should not be willing to adopt the minimal acceptable standards of our sister disciplines, or sister associations, but instead should strive for an absolute level of quality in our contributions.

I noted earlier that the issue of standards interacts closely with the question of effort. Here I feel encouraged that higher quality standards would not necessarily require more effort, although they would likely lead to reallocation of the effort. (This of course depends upon the underlying training and knowledge available to the individual; if this is insufficient to allow quality contributions, then remedial work seems clearly to be needed. Assuming that a reasonable background is available, however, it does seem likely that a reallocation of our effort would be fruitful.)

At present, it seems obvious that much effort is currently being expended fruitlessly, at least in terms of having impact on the body of knowledge in our field. For example, we have again this year failed to make a JCR/ACR Award in recognition of a doctoral dissertation article's contribution to knowledge in consumer behavior. None of us doubts that substantial work and worry was expended on dissertations in this past year, as in every year. Also, I don't doubt that some contributions did occur in each of these efforts. The degree of contribution, however, was viewed again to be insufficient to merit an award.

This is beginning to look like a chronic condition, and no one is pleased about it. While some minor adjustments in the process may be called for, it does seem inappropriate for us to lower our criteria in order to make an award; on the contrary, it would seem more appropriate

for doctoral candidates and their chairpersons to raise the criteria for doctoral dissertations.

Similarly, the acceptance rate at our major journals runs about fifteen percent each year. This means that six out of seven papers fail to meet minimal requirements for publication. Even allowing for adjustments to this figure to account for multiple submissions, lack of topic match-ups, etc., I am sorry to see such considerable effort involved in writing a manuscript go to such negative results. Again, wouldn't our field, and we as individuals, be better off if we simply raised the average quality level of our papers? Note that this could be done by lowering the number of papers, and expending more effort on the remaining projects.

As mentioned earlier, this would have to be accompanied by an acceptance and internalization of higher standards for our attempts to contribute to knowledge. The internalization process seems especially important, because often the external standards for publication are not high enough. Acceptance for publication shouldn't be seen as a vindication of our personal standards, nor as the primary purpose of our work. A preferable approach, I think, is for each of us to seriously examine whether we are adding to knowledge, broadly viewed.

Standards of Relevance

This brings me to my second issue under standards, which pertains to the "relevance" of our work. It is interesting to discover that "relevance" has always played an important role as a criterion for directing the pursuit of knowledge. In one sense, in fact, relevance can be viewed as the ultimate purpose of this pursuit.

However, the great thinkers have not chosen to define relevance in as narrow a sense as many of us apparently do. That is, they view relevance as a general measure of the fit or essential appropriateness of a project of the body of knowledge with which it is affiliated. Impact on the body of knowledge is one submeasure, but not the only one. Usefulness for immediate purposes, such as managerial or governmental decision making in our case, is another appropriate measure. Stress on this measure alone, however, can lead to a narrow conception of both the purposes of knowledge and of the most reasonable representations of phenomena within our field of inquiry. This can have deleterious impacts on basic forms of research, and might retard or even eliminate significant breakthroughs and contributions.

My particular concern in this area involves the feeling on the part of many persons in our field that consumer behavior is essentially an engineering discipline. If this position is advanced as a description of the current interests and motivations of most persons in the field, then I accept it as being accurate and appropriate. On the other hand, if this engineering orientation is advanced as a normative requirement or definition of the field of consumer research, I must emphatically disagree. Consumer behavior is an increasingly significant phenomenon in our world; it deserves description and understanding, independent of interests in improving, influencing, or otherwise benefiting from it.

Whatever the resolution to this issue, the more general problem of relevance is an appropriate standard for us to consider as individuals, and as a field. As individuals, this suggests that we should attend carefully to our substantive goals and motivations for each undertaking: What exactly am I attempting to add to knowledge with this undertaking? What questions need to be answered and will this potential project be able to answer them sufficiently well? If not, can it be revised or refined to do so? If not, should a different question be pursued, or should a new undertaking be substituted?

Although it sounds trite, I think that we typically don't spend nearly enough mental effort on exactly these questions. I would point to this deficiency as one of the major causes of rejection rates at our journals and conferences, and as a major reason that persons looking for

application help in decision making tend to openly complain about the "irrelevant" work done by academics in our field.

In concluding this discussion of responsibilities, I think that the major implication is that each of us would benefit from a hard, objective assessment of both our willingness to accept the inherent responsibilities, and of the extent to which our preliminary efforts are likely to meet the required levels of quality and meaning of the problem at hand. This is inherently a personal, individual process. It can be greatly assisted, however, if external feedback and/or yardsticks are integrated into that process. By raising our standards as a field, I believe that the individual's decision making in this area would be both enhanced and rewarded.

Conclusions and Closing Remarks

As I indicated at the start, I hope that the most significant conclusions from this talk will occur at the individual level, in the form of personal contemplation and decisions about our individual roles in the field. I also expect that these conclusions will differ substantially within this group, and I think that this is quite appropriate. Therefore, I am left with only the possibility of forwarding some of my conclusions from a more aggregate, association point of view. I hope that you will accept these in that light.

Really a Small Group

First, I think that if we were to honestly apply the three criteria in this speech -- possessing knowledge, choosing to contribute, and accepting the responsibilities -- to our field as a whole, we would have to conclude that the number of people actively contributing to knowledge is considerably smaller than the number of people affiliated with this area. I would not expect consensus on the underlying knowledge requirements in regard to the five great disciplines of the mind, but certainly this must impair our ability to develop knowledge. Further, it is obvious that a number of us have chosen to participate, but not especially to contribute. This leaves a smaller group who may have tentatively made the choice to attempt contributions.

Insufficient recognition and/or acceptance of the concurrent responsibilities, however, remains a serious barrier to the actual realization of contributions. It is my impression that a number of persons find this barrier inconsistent with their personal motivations, and decide to avoid it. This takes another major cut into the ranks of contributors, leaving us with a fairly small group, especially when contrasted to the wide range of issues in the field of consumer behavior.

ACR needs More Contributions

I think this point is clear, when we consider how much work remains to be done, and the pace at which we are overcoming areas of ignorance. My comments earlier on the desirability of more contributions from business and government sectors are pertinent here as well.

In addition, there are some closing perspectives I wish to raise regarding the academic sector. I would divide these into two categories.

The first pertains to the "older, established" contributors, most of whom are affiliated with academic institutions. To this group I wish to offer no recommendations. In the spirit of my topic, I wish to pay my respects to you, as you are the people who have brought the field this far along. You know what high standards mean, and you know the effort required to produce contributions to knowledge. I wish only to transmit the message that the field will benefit from your decisions to stay active,

to continue to contribute, and to act as role models for others who wish to make contributions.

There are several points I would like to offer for the "younger, not-quite-yet-established" members. First, in evaluating your decisions to decide to contribute to knowledge, please take the discussions of accepting responsibilities in a most serious way. Many -- and probably most -- persons have not really done this, and most institutional entities in our field do not really require it. As a result, you may find little overt guidance towards internalization of the standards and effort-levels which really ought to be your goals.

Unfortunately, if you adopt what appears to be "accepted" standards, together with diffuse and unexplicated goals, the chances for "success" in academia are sharply curtailed. In addition, but perhaps less obvious, the chances of your making contributions to knowledge drop significantly. For many persons, it appears that this combination of factors leads to a later conception that one is merely "playing a game," and that the deck is stacked against the player. Substantial time, efforts, and emotional costs may then be expended in attempting to outfox the opposition and obtain "another pub" for the resume. If we would step back a bit and abstractly view the needs of our field, we could easily see that this system is misdirected.

In fact, however, the more straightforward system is the better and assuredly more rewarding to pursue. The leaders of our field recognize its need for contributions. The established contributors in the field are able to recognize new contributions, and generally await the appearance of same. Unlike the perceptions that I know that many of you either hold or have heard, this is not an "in-group" association, nor is it a "private club" running the field. In fact, this association was born, as I understand it, in large part to avoid these very characteristics. In short, the field wants to look positively on your work. It needs your work, and it knows it.

As I listen to myself saying this, I realize that it may not be very convincing to you, especially if you already hold the opposite view. Perhaps an example might be useful in this regard, because it strikingly demonstrated to me the field's real receptivity to quality work. One such situation occurred about five years ago at the ACR Conference, in a particularly negative setting. In brief, we had been involved in an all-day special topics session, with a reasonably small group -- maybe 30 or 40 people -- with a sincere interest in the subject. By the end of the day, we were hot, tired, and generally in a bad mood. As the papers went on, criticism became more severe. By the time the final paper was scheduled, most of us, I think, were quite ready to get to the football game where we could take out our frustrations. The last thing in the world anyone wanted to hear was a mediocre or bad paper.

Unfortunately, the last speaker had to walk into this situation. He was not a "big name" in the field; to the contrary, he was a young academic, trained at not one of the major prestige universities, and he was just beginning his career. As he started to speak, some of the testier members of the audience began to sound him out on various detailed issues of the topic. If he had failed to deal with these questions adequately, they would have felt comfortable in dismissing the remainder of his effort, and probably appropriately so. I don't know if I have vividly portrayed the whole context, but it surely was a severe test.

However, he stepped up and did a superb job. He handled the questions, sometimes adding additional issues which hadn't been included. He demonstrated an awareness of the key work on the topic, and an interest in improving the conceptualization and research in the area. In short, he gave a fine talk, and made a contribution with it.

My point in relating this episode involves what happened next. As soon as the discussion ended, I watched as leading scholars from three or four of the major in-

stitutions in the country walked up and introduced themselves. In the space of a half-hour, he had moved dramatically upward in their evaluations. He later moved to one of the top schools in the nation, where he is now a veteran on the faculty.

The moral of the story is two-fold. First, this didn't happen easily. This man had to be able to meet a relatively rigorous set of standards to be viewed as making a contribution. This had required long hours of work on his part, and a sincere interest in the topic. He had already put in those hours and made that commitment long before arriving at the ACR session. Second, when others recognized the contribution, they were pleasantly surprised, interested in discussing the topic further with him, and wanted to encourage him to continue his work. This is what our field is really about, and I hope that the younger members will recognize it and appreciate it.

Summary: The Goal

In this speech I have ranged over a variety of issues, and have attempted to discuss both pragmatic and abstract questions. I hope that these comments may stimulate some further consideration of "what it means to add to knowledge" by all of us. In closing the substantive portion of this talk, I would like to quote Dr. Barzun, as he summarizes his discussion of the House of Intellect:

"These considerations make only more imperative the safeguard of the master virtues of Intellect. They are, once again: concentration, continuity, articulate precision, and self-awareness. Intellect needs the congregation of talents spurring one another to high achievements by the right degree of proximity and intercourse; it needs the language and the conversation that maintain its unity like a beneficent air; it needs precision to dispel the blinding fogs of folly or stupidity; it needs self-awareness to enjoy its own sport and keep itself from vainglory."

Closing Comment

Given the nature of the topic I chose, it is almost necessary that we be dissatisfied with the level of our work, and with the progress we have made. It would be inappropriate, however, for me to close this talk on that note. Yesterday it was mentioned that ACR grew out of an initial conference at Purdue, then took shape at a subsequent meeting at Ohio State. The shocking fact is that this started only 14 years ago! In this very short period of time, we have come a long way, and can be proud of our progress.

On a personal note, I would like to close by saying that I am truly appreciative that you were willing and saw fit to elect me as a President of ACR. I have enjoyed the chance to serve during this past year, and to have seen firsthand many of the finer elements of the Association and its membership. Thank you very much.

References

- Barzun, Jacques (1959), The House of Intellect, New York: Harper & Brothers.
- Hodnett, Edward (1963), The Cultivated Mind, New York: Harper & Row.
- Pratt, Jr., Robert W. (1974), "ACR: A Perspective," in S. Ward and P.L. Wright, eds., Advances in Consumer Research, vol. I, 1-8.

PRESENTATION OF THE ACR AWARD:
"FELLOW IN CONSUMER BEHAVIOR" TO JOHN A. HOWARD AND JAMES F. ENGLE

Harold H. Kassanjian, University of California, Los Angeles

I consider it a great honor to have been selected to make these first awards for "Fellow in Consumer Behavior", and to share this podium with two unique scholars. For each, in his own way, has had a long term significant impact on the field of consumer behavior. And this award was created to honor just that type of contribution, a long term, major, scholarly impact.

The idea of such an award was not a rapid decision by the Board of Directors. Discussions concerning an ACR award were initiated as early as under the presidency of Joel Cohen, now almost a decade ago. Over the years, numerous individuals were involved and committees appointed to consider the suggestion. It was felt that the fellow award should not be based simply on longevity in the field, or only because a given colleague was well known. It was not intended to be given for a single work, article, book, or monograph. Unlike awards given for "the Best Article", a single major scholarly work would not necessarily qualify for this award. Also a stream of research carried on for several years that may have had an impact on the field might well not qualify for the award.

The status of "Fellow in Consumer Behavior" was to be reserved as recognition to the individual for scholarly contributions spanning not just a few years but rather a distinguished long term commitment.

After several years of deliberation on goals and procedures, the final decision was made that there should be a selection committee composed of a chairman appointed by the president of ACR, and as committee members, the three immediate past presidents of the association. This, it was hoped, would represent a broad spectrum of interests with a committee regarded highly enough to have been elected by the membership as president over a three year span.

The decision of the committee was to be unanimous, for if consensus could not be achieved, the award should not be given that year. The decision was then to be presented to the Board of Directors for confirmation by a 75% affirmative vote. The award will not become a routine event given each year to whomever is still around. There will be no award next year, for example. Also a three year "sunset" provision was enacted to review the program.

The first year, the award committee consisted of James Myers, chairman, appointed by the then president of the association, Keith Hunt. The other members, the three immediate past presidents, were Jerome Kernan, David Gardner and myself. This committee solicited nominations from the membership but under no conditions felt obligated by a vote count or obligated to pay attention to the fact that some potential candidates were or were not nominated. For popularity was not to be one of the criteria. Although enthusiasm by colleagues might be considered, it was clearly not sufficient.

The final decisions were unanimous and enthusiastic. The Board of Directors enthusiastically and unanimously approved the nominations. From zero to three awards may be made each year. The first year, two individuals were thus selected.

¹Written while visiting professor at the Copenhagen School of Business Administration and Economics, Denmark.

John A. Howard

The first is Professor John A. Howard, Columbia University. Professor Howard did not start his career in consumer behavior. As you well know, consumer behavior as a field of scientific endeavor did not exist at that time.

His bachelor's degree in economics was from the University of Illinois in 1939 and his Master of Science in Economics in 1941. He received his doctorate from Harvard University in 1952 with a dissertation on British Monopoly Policy, prepared while at Oxford University in England.

Professor Howard's first teaching assignment was at the University of Illinois. From 1950 to 1958 he was professor at the University of Chicago, then for five years at the University of Pittsburgh. Since then he is professor at Columbia University.

In 1959 he was commissioned by the chairman of the Scott Paper Company to investigate the need for basic research in marketing. That study led to the establishment of the Marketing Science Institute. He has been a consultant to the Department of Commerce and to the Federal Trade Commission. His work at the FTC led to the acclaimed monograph, Advertising and the Public Interest, co-authored with James Hulbert. In 1975 he was awarded the Paul D. Converse award for "contributions to theory of marketing and buyer behavior" for four of his publications:

Marketing Management, 1963
Marketing: Executive and Buyer Behavior, 1963
Marketing Theory, 1965
The Theory of Buyer Behavior, 1969,
with Jagdish N. Sheth.

As with many other early pioneers in this field, Ford Foundation money was to take on a significant role in Professor Howard's career. Today many of us are unaware of the pivotal role the Ford Foundation had played. In their effort to up-grade business schools, marketing scholars were encouraged to incorporate the behavioral sciences, mathematics, and other basic scientific disciplines into their research, teaching, and thinking.

In 1960, Howard was commissioned by the Foundation to do a two-year study on the state of knowledge in Marketing. The results of that study led to the monograph, Marketing: Executive and Buyer Behavior. This was the embryo from which Howard and Sheth's, The Theory of Consumer Behavior was to emerge, and in fact that monograph was quite influential on the development of the Engle, Kollat and Blackwell model as presented in their text.

In turn, the roots of Executive and Buyer Behavior appeared earlier yet in Howard's intellectual career. For in the first edition of his Marketing Management textbook, Howard discussed buyer behavior from a behavioral science perspective. His very first model of the consumer was an adaptation of Clark Hull's learning theory to behavior found in the marketplace. By the time of the second edition in 1963, this view had been expanded to almost a chapter length discussion of behavioral science in marketing. And tangentially, another new topic was given an embryonic introduction - mathematical models and stochastic learning theory. This work was expanded in the

monograph, Marketing Theory, in 1965 and culminating, as already mentioned, in The Theory of Buyer Behavior in 1969. Part of a one-half million dollar program at Columbia to build, test, and apply a theory of consumer research, this stream of activity led to several doctoral dissertations as well as to follow-up research and conceptualization by Howard and his colleagues. This work led to numerous papers and the monograph, Consumer Behavior: Applications of Theory in 1977. Unlike so many builders of grand theory, Howard has been most interested and involved in testing his models. And this may have been his major accomplishment - the one he can be quite proud of. But, Howard played another role in Consumer Behavior - a role, that now, few may recognize. Howard was not a renegade, he was not a rebel with a bag full of new gimmicks called behavioral science and learning theory with which to dazzle the staid establishment in marketing. For Howard was part of the marketing establishment - in a time when marketing consisted of a bit of economic theory, wholesaling, transportation, etc. New fangled ideas were not easily being accepted into the establishment and marketing people simply tended to be hostile to behavioral science and mathematics. Programs at the American Marketing Association were typically not available to those interested in the behavioral sciences. Papers on psychological topics or even landmark work on stochastic models were generally not acceptable to the journals and programs of the day.

John Howard was a member of that establishment - he had authored a widely used text in marketing management, and yet he was willing to integrate some of these new ideas into the field and to encourage unorthodox research. If for nothing else, it is for that role of creative innovator, open-mindedness, scholarly rigor and respectability, and a leadership role of which he may not even have been aware, that we are in debt to Professor Howard. The rest of us could point to the work of Howard to support the introduction of basic behavioral research into the business schools. For here was a respected member of the establishment using the term buyer behavior.

Fittingly, and with humility, Professor Howard was selected as one of the first two recipients of the "Fellow in Consumer Behavior" award, (and I quote), "For focusing attention on the importance of consumer behavior and for pioneering work in conceptualizing consumer decision making". Professor Howard, we thank you.

James F. Engle

The second person to receive the first awarding of the "Fellow in Consumer Behavior" recognition is Professor James F. Engle. Professor Engle is from another academic generation than his co-recipient and revolved in a different orbit. Not trained in economics and not a member of the establishment at the time, Engle was the beginning of a new breed. He entered Drake University in 1952 with an interest in music and a drama scholarship only to switch his major to marketing because several of his friends were majoring in marketing. And so, a career was launched.

Under the influence of Professor Martin Zober, one of the early people who felt that marketing should take the consumer seriously, Engle took electives in psychology, social psychology and cultural anthropology - something mighty unusual at the time.

Turning down attractive fellowships for graduate work from other Big 10 Universities, Engle matriculated at the University of Illinois at the end of an era of ferment and excitement - the time of P. D. Converse, Huey, Hugh Wales and the young new kid on the block - Bob Ferber. Under the influence of Wales and the work Wales and Ferber had been doing on motivation research, again Engle turned to social psychology for his minor area and to a

methodological study of motivation research for his dissertation.

His first professional position was to be at the University of Michigan where he began the integration of the behavioral sciences into an advertising course - the same course, by the way, where other behavioral types throughout the land were being relegated. And it was here that he conceived his Promotional Strategy book with Wales and published his laudatory work on cognitive dissonance. For its time, this paper was a fine piece of empirical research integrating social science concepts and methodology into advertising and marketing. That, along with the promotion book and a paper on perception were among the early attempts at borrowing concepts, middle-range theories, and methodologies from psychology and applying them to consumer behavior in the marketplace.

But, at his university, as in so many other business schools, this type of work and this type of research thrust was neither appreciated nor understood. It did not mesh with the established norms of academic activity. The time had come to move to The Ohio State University. Here, within a short time, a critical mass, a symbiotic relationship, emerged as a consumer behavior group was formed by Engel, Kollat and Blackwell and the contributions of a graduate student, Larry Light. The culmination of that effort was the first true textbook in the field, Engel, Kollat and Blackwell's consumer behavior trilogy. Now in its third edition, and the fourth being conceptualized, that text has unquestionably become the standard work in the field by which all other texts, course outlines and course content are compared and categorized. Throughout the land, courses in Consumer Behavior were introduced with the EKB Table of Contents as the outline. What was left for other authors and publishers of texts was to attempt to carve out a small market share below or around the EKB positioning, for it is still the standard work in the field.

But the Engel contribution is far more than an article on dissonance, a paper on longitudinal analysis, a study on the use of self-medication drugs, or the co-authoring of a text, or any of his 60 other papers, books and articles. For the name James Engel, in many ways is synonymous with the Association for Consumer Research and the field of consumer behavior.

Perhaps the key turning point of the academic discipline of consumer behavior can be traced back to a meeting at Purdue University in 1966. King, Bass, Pessimeir and others at Purdue invited the promising young men in the field of marketing to attend a week-long conference. Engel and Light presented a paper on cognitive dissonance and a model of consumer behavior - the nucleus of the EKB model. It was here that various people interested in consumer behavior began to realize that throughout the land, there were others with similar interests and experiences and on-going research. For this was the debut, the debutante ball, not only for Engle and others of similar vision but for the field itself. Consumer Behavior need not be the ugly duckling of marketing but rather might be an interdisciplinary field of scientific endeavor in its own right.

In the end, it was Engle who did not let the fledging swan wilt but with his colleagues nursed and nurtured it to its outgrowth - the 1969 Ohio State Workshop on Consumer Behavior (three dozen people meeting with a thrill of intellectual companionship that few had experienced before - under the guiding hands of James Engle).

When this association was formed that Saturday afternoon, with everyone tossing in a ten dollar bill to get it launched, I think I shall never forget glancing at Jim Engle. He had just smiled at a colleague - I believe

either Kollat or Blackwell - a smile I perceived as a self satisfied grin of accomplishment - proclaiming with his eyes, "we pulled it off, we launched the organization".

Yes, professor, you pulled it off. You fathered, you wet nursed an interdisciplinary field of scientific endeavor. And now, for his intellectual stimulation, for his writings and research, for his leadership, for his "significant impact on scholarly work in consumer behavior", the Association for Consumer Research presents him with the award, "Fellow in Consumer Behavior", (and I quote) "in recognition of his singular contribution to the systematic study of consumer behavior". Professor Engle, we thank you.

PROMOTION IN A STATIC MARKET*

John Howard, Columbia University

Introduction

I am deeply honored to receive this symbol of the esteem of my colleagues. It is one of the finest achievements that I could hope for. As my response in accepting it, I would first like to make a brief assessment of the state of knowledge in the field of consumer behavior.

State of Field

Remarkable progress has been made since about 1955 in developing theory that has relevance for practice. Let me apply this pragmatic definition of progress--creation of theory relevant for practice--and briefly evaluate research on each of the stages of choice that we, economists and psychologists have found useful (Howard, 1977) and especially now that Jim Bettman and Michael Zinns (1977) have given some empirical legitimization to that typology.

In the most common case--a buyer confronted with an unfamiliar brand in a familiar product class (limited problem solving)--research has gone exceedingly well by my pragmatic definition of progress. I need only cite the recent splendid technologically-oriented review by Shocker and Srinivasan (1979) which was possible only because of theory development.

Research on the case of a buyer confronted with an unfamiliar product class (extensive problem solving)--innovation--is going well. We have broken out of the confines of the rural sociologist's work in diffusion, which emphasizes the social processes and individual differences among buyers. Instead, we are focusing upon information processing (Bettman 1979) and upon values as the source of choice criteria (Howard 1979). Also, especially encouraging is Peter Wright's (1980) fascinating recent experiment on product class advertising.

Research on the third case--a buyer facing a familiar brand in a familiar product class (routinized response behavior)--however, has been disappointing, even though economists under the stimulus of Stigler's seminal article have also been lending a hand (Rosen 1976). I would like to address it at greater length.

Obvious Contradiction of Theory

Our RRB theory seems to me to be beset with a serious contradiction. Stated most baldly, our behavioral theory predicts that in a stable market consumers will learn about the array of brands and zero in upon the most preferred brand. This loyal behavior, then, will continue ad infinitum unless a change in the market occurs. Our prediction implies there should be no advertising because there is nothing to communicate.

*Paper given before the Annual Meeting of the Association for Consumer Research in reply to the honor of being appointed a Fellow in Consumer Behavior, Washington, D.C. October 17, 1980. I am grateful to Professor Wilfried Vanhonacker, my colleague, for his splendid comments on the paper.

But if we look around at the real world, it seems to me that we see far more contradictions of this prediction than we see consistencies, e.g. cigarettes before health warnings triggered a spate of minor product innovations. This conclusion, of course, assumes that managers spend their advertising budgets wisely. From the decision net studies of a number of executive decision processes in the early 1960's, I believe that for short-term decision-making this assumption is reasonably valid (Clarkson 1962); Howard and Morgenroth 1968). Also, I am ignoring the possibility of "reminder promotion" which tells the consumer, in effect, that the product class is available, which can cause shifts among product classes such as in food where there is flexibility across product class boundaries.

This contradiction of our theory has troubled me for several years. Not only is it blatantly obvious, but it involves a considerable proportion of our economy. Hence, whether our theory fits the facts in this case makes substantial difference. Further, we must remember that the success of our theory in the most common case--LPS--came as really quite a surprise to most of us. I was not nearly as confident that consumers used information in their buying as my words indicated. Because of my general skepticism, our failure in the simplest case--RRB--has had a profound effect in creating serious doubts in my mind about our entire venture into understanding the consumer.

In fact, in my concern about the contradiction, I have felt somewhat akin to Professor Weyl, the great mathematician who, in 1946, wrote, "...we are less certain than ever about the ultimate foundations of (logic and) mathematics... Like everybody and everything in the world today we have our 'crisis'... Outwardly it does not seem to hamper our daily work, and yet I for one confess that it has had a considerable practical influence on my mathematical life; it directed my interests to fields I considered relatively 'safe' and has been a constant drain on the enthusiasm and determination with which I pursued my research work."

Now, having recited my tale of woe, let me take a more positive view and conclude that the field even in the area of routinized response behavior is in better condition that I had thought.

Contradiction of Theory Removed

In the 1960's, I became interested in Daniel Berlyne's work on the effect of novelty in esthetics and other areas of human behavior. Out of it I developed the ambiguity-arousal hypothesis which some of you might remember from the book with Professor Sheth (1969, pp. 202-220).

This hypothesis predicted that once consumers have zeroed in on their preferred brand in a stable market, they would, after a period of brand loyalty, become bored and shop around trying some alternative brands that are more novel, more interesting, and then again zero in on a preferred brand. A cyclic pattern of behavior would result: stability of brand loyalty, instability, stability, etc.

Not being able to think of an effective way to test it, I had forgotten all about it. Then, three years ago, I received in the mail a monograph on testing stochastic choice models done as a dissertation at Rotterdam University in Holland.

The author, B. Wierenga (1974), creatively develops the concept of "pool size." In this way his work went beyond that characterizing the stochastic choice research prior to the late 1970's. By defining "pool size" as the number of different brands bought in the last 10 purchases, he makes pool size the critical measure of behavior.

He uses three product classes of low-priced, frequently-purchased items to determine if the cyclic patterns predicted by the ambiguity-arousal hypothesis could be said to characterize the behavior of consumers. The products were beer, margarine, and a third unspecified item called "fopro." At least three of the brands for each product were national brands distributed over Holland as a whole and promoted by newspapers, television and radio on a national scale. The information was collected from a 2000-member commercial purchase panel for the two year period 1967 and 1968.

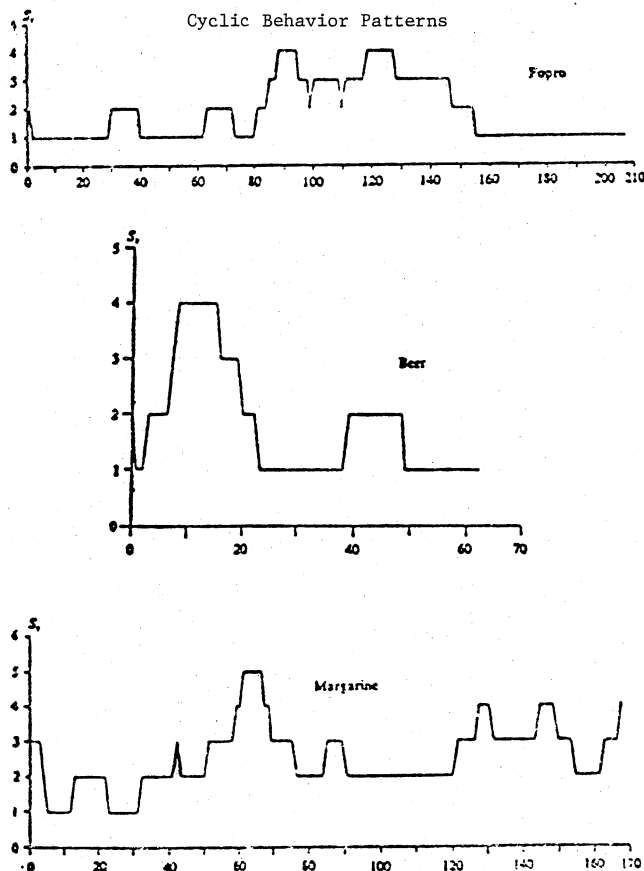
To be more specific, I will describe his method first as applied to the purchasing of a single family, but a different family for each product, as shown in Figure 1. S is the poolsize measured on the vertical axis and it is the number of different brands bought by the family in its last 10 purchases. Of course, t is time. At purchase S_t for fopro, for example, the poolsize is 2, as you will note. In the purchase sequence t-9, t-8, t-7, t-6, t-5, t-4, t-3, t-2, t-1, t, two different brands were purchased by this family. S was immediately reduced, however, to one particular brand because at purchase S_{t+1} , over the purchase sequence t-8, t-7, t-6, t-5, t-4, t-3, t-2, t-1, t and t+1, the family bought the same brand. This single brand purchasing continues with fopro for 28 purchases, at which time another brand, other than that originally bought, was purchased, which brings the total to t=29. Figure 1 implies that during the next 9 purchases, the pool will contain at least two brands and it is possible that a consumer in one time period completely switches to the new brand. If he did, the pool size of 2 lasts for exactly 9 purchases. After 10 purchases (t=39), the pool size returns to 1, which indicates that no further purchase of the new brand was made while no other brands were tried in the meantime. After that, there is a period of no brand-switching which lasts for 23 purchases (t=62). Again, an incidental purchase of a different brand occurs followed by a short rest (t=80). Then a period of intensive brand-switching begins where the pool size increases to as much as 4. Following this turbulent period, the family seems to have made its choice and continues for a long time buying the same brand.

As you see in Figure 1, fopro and beer show each one period of extreme search. "Search" is defined as when, during a certain period, the buyer tries a number of different brands before settling on a particular brand as contrasted with a straightforward switch from one favorite brand to another favorite brand. Margarine exhibits two such periods. Also, fopro was much more inclined to single brand periods than the other two products.

In Figure 1 we have seen in single households evidence of cycles where, after periods of intensive search, the family alternates with periods of continuous buying of the same brand within a period but not necessarily across periods. A similar picture emerges when all 2000 households are thrown together.

Unfortunately purchase patterns cannot be shown diagrammatically for more than a single family. Instead, they can be described by using the range of difference between the maximum and minimum of pool size over all families. The results for families buying two or more brands are shown in Table 1. Families that bought only one brand during the two-year period of the data are excluded because we want to show the variation in switching required to conform to the search-loyalty cycle. These continuous repeat

Figure 1



purchasers constitute roughly a fourth (fopro = 28%, beer = 27%, margarine = 16%) for all three brands.

As you can see in Table 1, where RANGPLZ means range of variation in pool size, a substantial difference in levels of switching activity occurred. In fopro, for example, 2.69% exhibited no variation, while 46.69% showed a variation of 1. This is consistent with the search-disloyalty cycle concept. Specifically, the pool size is not constant but has periods of low values and periods of high values.

Table 1

Relative Frequency of Range of Values for Households with More than One Brand

RANGPLZ	Fopro	Beer	Margarine
0	.0269	.1204	.0190
1	.4669	.5055	.3814
2	.3223	.2582	.2975
3	.1157	.0875	.1745
4	.0475	.0241	.0738
≥ 5	.0207	.0044	.0537
	1.0000	1.0000	1.0000
Number of Households with ≥ 2 brands	484	457	894

But we should ask ourselves whether these patterns of cyclic behavior could have been the result of promotion or other marketing activity instead of the ambiguity-arousal hypothesis? If we examine the pool sizes of different consumers at a given point in time, we find they

are different. They are not low and high together. If the source of the cyclic pattern was a common marketing activity, the lows and highs of different consumers would tend to occur together in time.

But once they have decided to shift, marketing effort does appear to make a difference in which way they shifted, to which brands. Specifically, the start of search is often associated with changes in price, deals, shop where bought and unit-size bought, which supports the notion that promotion is justified as implied by the ambiguity arousal hypothesis. The absolute number and proportion of each of these promotional events associated with brand switches are seen in Table 2.

Table 2

Relative Occurrence of Different Types of Concomitant Events of Search-Initiating Brand Switches

	Fopro		Beer		Margarine	
	abs*	perc**	abs	perc	abs	perc
Number of brand-switches observed	76	100	34	100	211	100
With concomitant event:						
Price change	29	36.7	5	14.7	141	66.8
Deal	6	7.6	0	0	18	8.5
Shop change	50	63.3	18	52.9	115	54.5
Unit-size change	33	41.8	22	64.7	-	-
None of these	4	5.1	5	14.7	26	12.3

* Absolute

** Percent

Additional support was also given the cyclic hypothesis in this study in that the linear learning model, which implies cyclic behavior, gave the best fit to the data among the various stochastic models (Wierenga p. 184).

We must be careful, however, in generalizing the results of the Dutch study to the United States. My colleague, Professor Wilfried Vanhonacker, informs me that most American studies have typically found zero order probability and with a pool size of about 4, although with some exceptions.

What I conclude is that hidden within this high shifting in the American market is a substantial number of customers exhibiting cyclic behavior as predicted by the ambiguity-arousal hypothesis. Not as high as the roughly 80% in Holland, but enough to justify the marketing effort we see in stable markets in the United States. This will, of course, have to be verified and I throw it out as a challenge to fellow researchers.

Conclusion

This finding renews my faith in progress in our work now that I have been provided an answer for the one gnawing question that has plagued me for years. It has implications for both private and public policy. I have emphasized the private policy, but its public policy significance can be still greater. It can be at least a partial answer to charges of pseudo-differentiability and shared monopoly, for example.

References

Bettman, J. R., (1976). An Information Processing Theory of Consumer Choice. Reading, Mass: Addison-Wesley.

_____ and Zinns, M. A.,

Bettman, J. R. and Zinns, M. A., (1977). "Constructive Processes in Consumer Choice," Journal of Consumer Research 4, 75-85.

Clarkson, G. P. E., (1962). Portfolio Selection: A Simulation of Trust Investment. Englewood Cliffs, New Jersey, Prentice-Hall, Inc.

Howard, J. A., (1977). Consumer Behavior: Application of Theory. New York: McGraw-Hill Co.

_____ and Morgenrath, W. M., (1968). "Information Processing Model of Executive Decision," Management Science 14, March, 416-428.

_____ and Sheth, J. M., (1969). The Theory of Buyer Behavior. New York: John Wiley & Sons, 157-165.

Rosen, S., "Advertising Information and Product Differentiation," University of Rochester, mimeo, 1976.

Weyl, H., (1946). "Mathematics and Logic," American Mathematics Monthly 53, 1-13.

Wierenga, B., (1974). An Investigation of Brand Choice Processes, University of Rotterdam.

THE DISCIPLINE OF CONSUMER RESEARCH: PERMANENT
ADOLESCENCE OR MATURITY?

James F. Engel, Wheaton Graduate School

Abstract

It goes without saying that it is a high honor to be named as one of the first two Fellows of the Association for Consumer Research. It is gratifying to hear that voice of colleagues saying, "we think you're OK," and all I can do is to express a heartfelt thank you. This makes all of those years of digging in the trenches worthwhile.

I am grateful also for the opportunity to reflect on the state of the art in this polyglot discipline. This will not be the somewhat tedious methodological and conceptual critique of the type previously voiced by Jacoby, Kassarian, Cohen, my colleagues and me, and others. While these have their role, it is my intent first of all to look backward focusing on some signs of progress along the way. Indeed, it is fair to say that our field has grown from crawling infancy to a fairly healthy state of adolescence in a short period of time. My fear, however, is that we are destined to permanent adolescence unless we do some serious stocktaking. Therefore, the primary purpose of this paper is to stimulate careful analytical assessment in the hopes that we can finally emerge as a discipline showing some much-needed signs of maturity.

The Road to Respectability

I have always dated the birth of consumer research as a serious discipline in and of itself back to the late 1950's. Obviously the roots go back much further than that, but this period marked the appearance of a handful of "marketing types" who possessed at least some training in the behavioral sciences. These names are well known to most, and I happened to be one of them pretty largely by luck and by accident.

Our early work, by and large, consisted of the "theory of the month club," to borrow one of Jack Jacoby's apt phrases. We worked independently at first and frankly were on a fishing expedition throughout the behavioral sciences. This soon began to change a bit, however, and the American Marketing Association was the catalyst for the beginnings of some focused inquiry. Conceptual and empirical thrusts then began to appear under such rubrics as risk and diffusion of innovations. The early models of consumer behavior were published, and the field began to receive the shape and direction which continues to this day.

The most meaningful change along the way has been the gradual shift from wholesale borrowing from related behavioral sciences to something much broader. Consumer research has emerged as a legitimate discipline. In the 1960's, few in the parent fields would dirty their hands in our arena. Those active here were all too often typed as the great contaminators of pristine truths as we began to see whether or not the stuff published in lofty journals proved out in real life. But that was to change pretty quickly. My Ohio State colleagues and I noted this most graphically when one of our noted psychologists did a real about face. Rather than the frequent volleys across campus and occasional harassment of courageous graduate students, he began to ask us to cut him in on the action, especially the consulting. I cannot condemn his earlier reactions; to some extent they were entirely deserved. The later response simply

underscored a growing recognition that consumer research was approaching legitimacy. Now it is a valid area of human life for scholarly study, and research from here has percolated its way back into these parent fields in a notable way.

Yes, we have moved beyond infancy. Let me point out some things that I feel are real signs of progress.

A Decline in the "Theory of the Month Club"

This club referred to above received some of its early impetus from me as I dabbled around with the theory of cognitive dissonance. Quite rightly I reasoned that this was a sure way to break into print, and off my career went. Those early publications will not go down in history as permanent landmarks by any means; nor will a majority of the other stuff which followed. Yet, looking back, this was more a sign of an infant discipline than anything else. After all, we had to start somewhere.

Embracing the theory of the month club today, however, is a sad vestigial remain of a bygone era. It is declining, because most of us don't rise up and uncritically salute the latest thing anymore. I am glad, for example, to see the advocates of attribution theory challenged to show what, if anything, it has contributed to existing knowledge.

Gains in Methodological Sophistication

In sharp contrast to earlier times, a large percentage of those who publish today evidence signs of good methodological grounding. While I feel this has led to undue preoccupation with "number crunching," it is a definite sign of maturation. The same can be said of the growing focus on validity.

A Tradition of Replication

Quite a few who read this paper will recoil at the very mention of multi-attribute models. To be sure, some have ridden this horse into the wall. Nevertheless, our preoccupation with this model has reflected a very good thing--the development of a tradition of replication. Because of the evidence that has accumulated over the past decade, we now are able to reflect meaningfully on the contribution this model makes. Frankly, progress can occur in no other way as I have long argued, because nothing firm can be concluded from the results of an isolated study here and there on a given issue. We are now seeing significant replication in such areas as information processing and choice heuristics, and I applaud it.

Some Consensus on Research Priorities

The first meetings of ACR were little more than everyone doing our own thing to get on the program and have our way paid. This also was reflected in the broader literature of the field. We are not far beyond that point today, but there is at least the beginning of some consensus on those areas which most need investigation. This too is an encouraging sign. While this is not the place to state my own priorities (that will appear in the fourth edition of Consumer Behavior), some welcome developments are the recent focuses on the dynamics of information processing,

theories of involvement, satisfaction, and non-product consumer behavior, just to mention a few examples.

The Crossroads: Maturity or Not?

Progress is being made, and let's be grateful for it. It certainly makes the revision of Consumer Behavior a less traumatic experience. The upheaval in the literature between the first couple of editions forced us into some really tough reconceptualization. While the volume of research has not abated recently, the field has settled a good bit into a more definitive pattern. This allows us to focus in the fourth edition on new insights into the old questions and the implications which emerge. This is a liberating experience.

I cannot escape the conclusion, however, that this discipline still is far from maturity. It still is possible to write something roughly categorized as "behavioral" and get it published with precious little inquiry into its methodological validity or substantive contribution. In a sense we are now at a crossroads if we want to maintain progress. It is time to take stock of what we are doing and where we are going. Here are some of the major issues to be faced.

The Criterion of Relevance

In the first years of ACR some of us used to debate the issue of basic versus applied research with some acrimony. Nothing has changed my opinion that this issue is a straw man based on a misconception of the very nature of our field.

Two decades ago, Ithiel de sola Pool described marketing as an engineering discipline drawing heavily upon related sciences and fields of inquiry to bring insight into the solution of practical problems. While consumer research reaches far beyond the boundaries of marketing, in what sense is it anything other than an engineering discipline as well? Is not the ultimate concern to shed light on the dynamics of consumer behavior so that those involved in the practical arena can do a better job? Contrast it with my parent field of social psychology which is not engineering and is a distinct field unto itself. If its professional contributors want to confine their research to such issues as prisoner dilemma games and other concerns pretty far removed from practice, it is perfectly fair game. We do not have that same right, however,

Some will disagree, but I do not think our discipline will go far beyond adolescence until relevance becomes a more central criterion. What I mean is that research priorities must be based on substantive issues reflecting the needs of decision makers. If research does not ultimately have some practical impact, no matter how small, how can we justify it? In this sense, we have a mighty long way to go. To paraphrase a little doggerel attributed I think to Kenneth Boulding:

Research is like a blunderbuss.
For all our muss and fuss,
We fire a monstrous charge of shot,
And sometimes hit but mostly not.

Frankly, there was a time in my own professional career where the sole concern was publication by any means and building a reputation among my academic peers. Then, I began to ask this disturbing question: "Of all the things I have written, what impact has it had on business practice?" My conclusion was that the impact was pretty nearly zero. Should I have cared? After all, most of you active at that time were using my books and giving the usual professional accolades. But I could not escape

de sola Pool's insights, and the whole focus of my life changed.

Increasingly I found myself trying to sort out the issues being faced by practitioners and the ways behavioral research could contribute. Obviously it became clear that this would at times require basic research which, at that point in time, would be pretty far removed from the firing line so to speak. But that focus has remained, and I make a plea that it be the ultimate criterion underlying all we do. I am tired of reading that obligatory section in professional papers entitled "practical implications." The straining we do approaches absurdity. Pragmatic issues should be upfront at the very outset rather than a forced and frequently trivial conclusion. When this criterion is applied rigorously, a large part of our published erudition quickly dissolves into obscurity.

Let me now put the same question to all of us. What concepts and methods have diffused from our research into the various fields of practice? Some will quickly identify psychographics and MDS/conjoint analysis as prime examples. But where did these contributions originate and have their greatest development? I think you will find that the diffusion process was just the reverse--from practice into the academe. So who is leading whom? At times it seems as if the practitioner and the academician live in different worlds.

Perhaps it will help if I describe briefly how the diffusion process has occurred in one area of non-product marketing. This is the field of Christian communication in which I have been actively engaged since coming to Wheaton. Our focus from the outset has been on the needs of decision makers in the local church, publishers, broadcasters, mission boards, and so on. They continue to ask for serious help, and many of the answers can be found in research. The Wheaton group alone has completed over 200 research studies on such bread and butter issues as readership, listenership, felt needs, and so on. By and large, these are methodologically unsophisticated, but they have found their way into practice. In the process we have derived a simple hierarchy of effects type model of spiritual decision making which now is the basis of evangelistic strategy around the world. This diffusion has occurred over a mighty short time period. We have no choice but to be relevant, and the outcome is an absence of dichotomy between academic and practical. Even more revealing is that an academic field of inquiry has arisen in the process. This is how it should be in my opinion.

Isn't it time for us to quit playing games? How many practitioners are at this conference? There are some, but ACR has always had a thin mix in spite of the contrary vision we had at the outset. Whose fault is it that an advertising researcher and I finally gave up on the sessions at a recent ACR meeting and decided we would benefit much more by informal conversation with those we could find who were doing some relevant things.

This perhaps is a lengthy and wordy way of saying something very simple. It is time to go beyond mere number crunching and insist that researchers have a firm and clear answer to the question, "so what?" If this is not our central focus, we are destined to permanent adolescence.

Moving Beyond Ethocentrism

This issue will require fewer words, but it is a serious challenge for all of us. I am appalled at the western domination in the consumer literature. Here and there is it possible to find a cross-cultural application, but these almost invariably reflect a profound western bias. Must we cast the rest of the world into our mold? Naive application of western models and methods can lead to decidedly erroneous conclusions and set back the hopes

and dreams of non-western leaders seeking solutions in their unique environment.

Serious cross cultural consumer research is a must in a world characterized by rising materialistic aspirations and imbalanced distribution of resources. Such research must be grounded in cultural anthropology, however, if we are to take account of profound behavioral variation. I have learned this lesson the hard way through working in research and training in more than 50 countries since 1972. Here is a vast frontier where we can join hands meaningfully with properly trained research partners throughout the world. In this context, the training of researchers in non-western countries is an absolute necessity. I see this need unmet wherever I go, and I often wonder why so many of you who first came to the United States as students remain here when there are such desperate needs at home. You above all have an obligation to meet these international responsibilities.

A Recognition of Moral Issues

I cannot imagine a discipline which is more amoral than ours. Only rarely do we address the social implications of what we are doing. What are the longrun implications, for example, of our continued focus on a high and rising standard of living in a era of diminishing natural resources? Is our growing knowledge of the mechanics of information processing gradually giving us the means of persuasion against a person's will? These are just two examples of issues which must come to the forefront rather than being relegated to the sidelines.

Moving From Adolescence to Maturity

I have presented a mixed picture so far--encouragement on the one hand and serious concerns on the other. Hopefully, the concerns expressed have struck a responsive note with some, and it is these colleagues to which these closing comments are addressed.

The starting point in the move from adolescence to maturity is for all serious contributors in this field to rethink their own conception of professional success. As I mentioned earlier, much of my own career was based on success as defined by publication in a "refereed" journal and recognition by peers. Relevance of what I had to say was secondary. Success, in short, is to be found in the length and nature of a publication list.

I have only given part of the picture on the changes which occurred in my life years ago, and I would be remiss by not filling in the remainder. I also had to face the deeper question of whether or not recognition and achievement really is a sufficient foundation for life itself, because it became increasingly non-satisfying. Fortunately my wife and I began a great adventure 15 years ago as we began for the first time to take our Christian faith seriously and place this eternal dimension at the center of our lives. That adventure continues to this day. It has led to a profound change in personal motivation as I have been freed to focus on basic issues from a perspective of some well-formed and functioning values which have their roots beyond myself.

While my own story is unique to me in many ways, I have found that many in this field are walking to the same dummer I have described here. The tragedy is that the reward system within our institutions and within this field itself only encourages it. I would again plead for the establishment of practical relevance as the guiding focus in all that we do. This would mean, of course, that our reward system must take account of the extent to which our research makes a genuine contribution in the context

of a substantive issue that ultimately will have some influence on practice.

Since this type of criterion is not likely to replace the narrowly focused publish or perish criterion used in our academic institutions, ACR will have to take the lead. We must become a true forum of researchers and decision makers, of academicians and practitioners. We have reached the point of sufficient maturity that we can begin to come to some real closure on research priorities. These priorities, in turn, must be enforced rigorously by those who serve as research reviewers and referees.

These priorities also must be based in serious reflection on the moral and social dilemmas which confront us. Yes, this is an area which goes beyond numbers into values, but we must not lose sight of the fact that we are dealing with human beings and not just rats in laboratories.

Can we rise to these challenges? Only time will tell, but I will do my best to be a part of the solution. While I am not so visible in "mainline" circles as I once was, this is only because of the arena in which I work. I hope to remain as a functioning and responsible contributor to this field for years to come.

WHAT IS LOW INVOLVEMENT LOW IN?

Clark Leavitt, The Ohio State University¹
Anthony G. Greenwald, The Ohio State University
Carl Obermiller, The Ohio State University

Abstract

Consumer psychologists have disagreed in their theories about what constitutes high involvement, necessarily leaving the important phenomenon of low involvement ambiguous. After reviewing the relevant consumer and cognitive psychology concepts, we offer a 3-stage information processing model in which high involvement is associated with the third stage--elaborated encoding of (that is, extensive cognitive responding to) advertising messages. Our model associates different types of low involvement with each of its first two stages--preattentive processing and focal attention. Both of these types of low involvement--inattentive and attentive--can lead to advertising impact.

* * * * *

One of us (Leavitt and Walton 1978) has suggested elsewhere that the paradigm that has been applied to persuasive communication is an adversary model of interaction which pictures persuasion as a kind of contest. The persuader's role is to move the audience to action by rhetoric; by promise of reward, or any other means. The target--the audience member--resists, is critical, and uses reasoning and skepticism to emphasize the flaws in the appeals.

An alternative to the adversary paradigm is unclear. Perhaps it is a kind of informal learning in which the self plays an instrumental role rather than acting as a goal. In any case, the adversary model is congenial to the notion of ego involvement which is a topic of some interest recently to one of us (Greenwald, in press). We began this paper with the notion that the adversary paradigm and its embodiment in ego involvement in particular, would provide a key to understanding the concept of involvement in consumer psychology. Consumer psychologists have been interested particularly in low involvement, and ego involvement is high involvement. By using little more than reverse English, we thought, we should arrive at the meaning of low involvement.

Perhaps fortunately, the conceptual analysis of consumer involvement turned out to provide a more complex challenge than we anticipated. We shall retrace our path by looking first at prominent attempts to define involvement in consumer research, next reviewing the relevant theoretical concepts in social and cognitive psychology, and lastly presenting our conclusions about the nature of consumer involvement.

1. The Involvement Concept in Consumer Research

Krugman

Most of the paths followed by consumer psychologists in their recent investigations of involvement emanate from an article that appeared in Public Opinion Quarterly in 1965, by Herbert Krugman. Because of the importance of that article to subsequent treatments of involvement, it is worth indulging in a complete quotation of the definition of involvement Krugman offered at that time:

¹Draft of August 11, 1980. Address reprint requests to Clark Leavitt, The Ohio State University, 1775 College Road, Columbus, Ohio 43210.

(There are) two entirely different ways of experiencing and being influenced by mass media. One way is characterized by lack of personal involvement, which, while perhaps more common in response to commercial subject matter, is by no means limited to it. The second is characterized by a high degree of personal involvement. By this we do not mean attention, interest, or excitement but the number of conscious "bridging experiences," connections, or personal references per minute that the viewer makes between his own life and the stimulus. This may vary from none to many.

The significance of conditions of low or high involvement is not that one is better than the other, but that the processes of communication impact are different. That is, there is a difference in the change processes that are at work. Thus, with low involvement one might look for gradual shifts in perceptual structure, aided by repetition, activated by behavioral-choice situations, and followed at some time by attitude change. With high involvement one would look for the classic, more dramatic, and more familiar conflict of ideas at the level of conscious opinion and attitude that precedes changes in overt behavior.

For our purposes, the most significant aspects of Krugman's analysis were (a) his specific disavowal of the interpretation that involvement might mean level of attention, interest, or excitement, (b) his reference to lack of personal connections in the case of low involvement, and (c) his suggestion that information processing followed a different route when involvement was low rather than high. In an article appearing two years later, Krugman (1967) added the important proposal that television was a medium that characteristically elicited low involvement, while print elicited high involvement.²

Ray et al.

In a 1973 article, Michael Ray along with several others incorporated a modification of Krugman's low-involvement processing concept into a broader analysis of variations in the sequence of changes in affect, behavior, and cognition that might occur following receipt of a communication. Krugman's high-involvement processing corresponds to Ray et al.'s learning hierarchy. It is not immediately clear whether Ray et al.'s low involvement hierarchy or their dissonance-attribution hierarchy should be taken as more nearly equivalent of Krugman's low-involvement processing. In a recent discussion of just this point, Calder (1979) suggested that low involvement might be best conceived in terms of a direct impact of communication on behavior

²In a 1971 article, Krugman added the speculation that brain wave patterns might be used to assess low vs. high involvement, but it may be best to regard that suggestion as a digression, since it seems to lead the concept of involvement away from his disavowal of an interpretation in terms of attention and excitement.

rather than an impact that is mediated by cognitive change.³

Houston & Rothschild

Houston and Rothschild (1977) have distinguished three types of involvement in the context of their decision-making analysis of consumer behavior. Their concept of involvement corresponds approximately to personal importance, which is a common feature with Krugman's analysis. The significance of their contribution is in recognizing that importance or involvement may be a consequence not only of the circumstances in which a message is received (they refer to this as situational involvement), but also of unique characteristics of the recipient (which they identify as enduring involvement). Their third type of involvement, response involvement, is defined as "the complexity or extensiveness of cognitive and behavioral processes characterizing the overall consumer decision process." (p. 5). Also, they conceive response involvement as reflecting the combined influence of situational and enduring involvement. Thus, the first two types of involvement represent sources of involvement (in the situation and person, respectively) while the third type captures the role of involvement in causing detailed cognitive processing. The association of involvement with an elaborated form of cognitive processing is an assumption shared with both Krugman and Ray et al., but the decision-making framework of the Houston and Rothschild analysis of cognitive processing is quite different from the cognitive models used by Krugman and Ray et al.

Mitchell

Houston and Rothschild's analysis of involvement in terms of a detailed set of cognitive processes provides a contrasting background for Andrew Mitchell's (1979) proposal that involvement be defined as a state, rather than a process (p. 194). Mitchell starts from a conception of involvement as meaning relevance to an important or high priority personal goal--a concept that fits well enough with the meaning of personal importance that underlies the other conceptions we have reviewed. Mitchell, however, proceeds in a different direction when he defines involvement as indicating "the amount of arousal, interest, or drive evoked by a particular stimulus or situation" (1979, p. 194). It is appropriate to recall Krugman's exclusion of "attention, interest, or excitement" from his (1965) conception of high involvement.

In summary, although there has been consensus that high involvement means personal importance, consumer behavior theorists have shown little agreement regarding the psychological analysis of importance--alternatively interpreting it in terms of personal connections, sequence of information processing, complexity of processing, and degree of arousal. This uncertainty about high involvement necessarily leaves the theoretical analysis of low involvement unsettled. We turn for help to several involvement-related concepts from social and cognitive psychology.

2. Involvement-Related Concepts in Psychology

First we consider three psychological approaches to the nature of information processing associated with high involvement--cognitive response analysis, encoding elaboration theory, and ego task analysis.

³Krugman's original (1965) statement, "with low involvement one might look for gradual shifts in perceptual structure, aided by repetition, activated by behavioral-choice situations, and followed at some time by attitude change," (p.) is ambiguous. Its interpretation in terms of the Ray et al. hierarchies depends on whether we care to interpret "gradual perceptual shifts" as a full-fledged cognitive stage or not.

Cognitive response analysis (Greenwald 1968; Petty, Ostrom & Brock 1980) holds that the cognitive reactions of the recipient of a persuasive communication are more important than is the actual communication content in determining the impact of a persuasive message. The effect of a message thus depends on whether the recipient's cognitive responses support or oppose the message's conclusion. Petty and Cacioppo (1979) recently showed that an involvement manipulation resulted in increased cognitive response activity--a finding that provides a nice link between Krugman's analysis of involvement in terms of personal connections and the cognitive response theory.

However, personal connections or bridging experiences are not the only kinds of cognitive responses that may occur in response to a persuasive message. Recent cognitive psychological theory about the role of degree of elaboration in encoding of information provides further connections among the concept of consumer involvement, cognitive response theory, and a variety of memory phenomena.⁴ The more elaborate, or detailed, is the encoding of an item of information (i.e., the more different ways it is categorized), the more likely it is to be recalled later. Some recent work by T. B. Rogers and his colleagues (Rogers et al. 1977; Kuiper & Rogers 1979) has shown that the task of judging the relevance of trait words to oneself has the properties of elaborated encoding, producing even better memory for the trait words than do other tasks thought to involve high degrees of elaboration (such as making a judgment about the meaning of the trait). In other words, relating information to the self entails substantial elaboration. Extending this theory to the advertising communication domain, we have a strong basis for predicting that the greater the self relevance of the topic of a message, the more elaborate the processing of the content and therefore the better the memory for the message.

The cognitive response and encoding elaboration theories offer useful accounts of effects of involvement on cognitive responses to, and memory for, an advertising message. Important questions still needing answers are: (a) How do people vary in what they find important? and (b) Do such variations between people have significance for the processing or impact of advertising communications? Greenwald's (in press) recent review of the ego-involvement concept in social psychology led to an approach called ego task analysis, which provides a basis for addressing these questions. An ego task is a long-persisting task, which means that it is directed to a distant--perhaps unattainable--goal, and subsumes tasks that are directed at immediately available goals--goals that may be only temporarily important.

Two ego tasks that seem pervasive in daily social existence are impression management (achieving a satisfactory image in the eyes of others) and self-image management (achieving a satisfactory self-evaluation). The analysis of ego tasks can be applied to consumer involvement by considering the relation of advertising content to the various short-term and long-term tasks of the consumer, and to the impression and self-image management tasks in particular. It is interesting to observe that, although most heavily advertised products are instrumental in achieving some immediate and concrete goals, advertisers often choose to appeal to the more generalized goals of ego tasks. Thus, cigarettes are advertised not in terms of their effects on level of arousal or on reduction of social tension, but in terms of the image of the person who uses the advertised brand.

⁴The encoding elaboration view previously was identified as a theory of levels of processing (Craik & Lockhart 1972), but the "levels" metaphor has seemed inappropriate on the basis of more recent findings (e.g., Craik & Tulving 1975).

Automobiles are advertised partly in terms of their transportation virtues, but even more in terms of the image of the person who drives the advertised model. Advertisers are imaginative in manufacturing connections of products to ego tasks--one of our favorites is the dishwasher detergent advertised by showing a housewife whose clean, shiny dishes cast "a nice reflection" on her. To the extent that the audience member shares such ego-task goals of the advertisement's protagonist--who achieves a favorable self-image or a favorable impression by using the recommended brand--the ad helps the involvement process; that is, it helps to establish personal connections to the brand. Further, to the extent that various audience segments are characterized by different ego tasks, the same type of appeal will not work best for all. Thus, it is not surprising that some cigarette brands appeal to health by advertising low tar and filters, while others appeal to an image of social or sexual success.

Ego tasks are not the only form of activity that might give us some clues about the nature of involvement. Ego-involvement is a subset of task-involvement, the highest level of that broader category. Lower level task involvement may also be intrinsically motivating and may elicit elaborated processing of advertising messages. Unlike ego involvement, task involvement occurs where there is intrinsic motivation but no special self-relevance (de Charms 1968). In fact, Csikszentmihalyi (1975) found that his more task-involved subjects were less self-conscious.

Although the cognitive response, encoding elaboration, and ego task analyses provide a coherent conception of high consumer involvement, they still only characterize low involvement indirectly. (This is where our reverse English fails!) To get psychologically closer to low involvement we consider a few concepts that have become familiar in research on learning and on selective attention.

As a characterization of low-involvement processes, the study of animal learning gives us the concept of latent learning (learning in the absence of relevant motivation), and research on human verbal learning gives us the somewhat related concept of incidental learning (learning without intention to learn). These principles lead us to expect that learning will occur under low-involvement conditions, but it may be weaker learning than when motivation or intention are present.

Cognitive psychology has provided us with the contrast between focal attention and preattentive processing. Information is assumed to be processed via a limited-capacity channel when it achieves focal attention, and this channel (sometimes called short-term memory, active memory, or primary memory) is assumed to be the entry route to long-term memory. Information that is outside of focal attention is expected to receive some immediate processing, but to leave no long-term memory residue. This contrast is suggestive, but the problem arises--where to locate low-involvement processing. Is it characterized by focal attention or does it occur instead outside of focal attention?

Figure 1 provides a basis for interrelating the several psychological principles that have been discussed in this section. The basic scheme of the figure employs the sequential-stages conception of information processing, which has been the dominant paradigm of cognitive psychology in recent years. (Although we find this paradigm heuristically useful, we will question one aspect of it in the third part of this paper). In Figure 1, the interpretations of high involvement as cognitive responding or as establishing personal connections are grouped together as variations on the theme of encoding elaboration. With high involvement, the elaboration provided by the audience member should vary from person to person; ego task analysis can be used as one way to analyze these individual differences. People who are highly involved may

nonetheless be engaged in very different tasks.

Figure 1 reveals our dilemma in the conceptual location of low involvement. If it is identified with failure of the message to reach a stage of focal attention, then we should expect no memory for the message, and presumably no impact. Since the major analyses of low consumer involvement lead us to expect impact, we are inclined to identify low involvement instead with achievement of focal attention, accompanied by minimal encoding. These two options are identified by different subscripts in Figure 1. It is low involvement₂ that corresponds to the principles of latent and incidental learning in the experimental psychology literature.

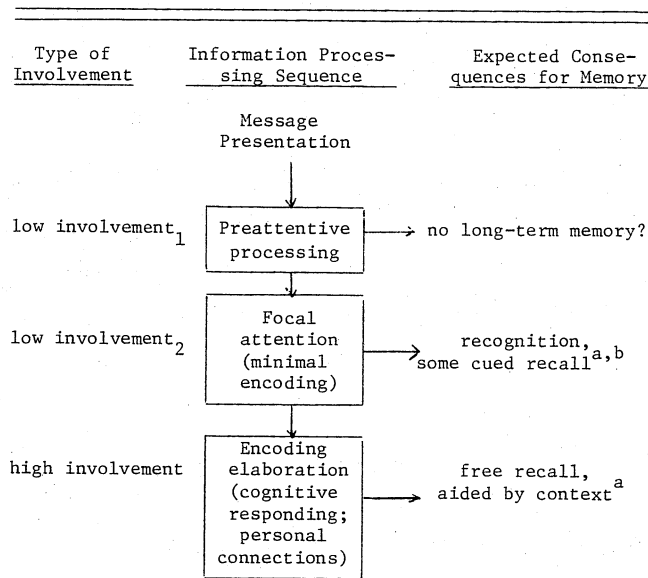
3. The Effectiveness of Low Involvement

Two Types of Low Involvement

Our consideration of a form of low consumer involvement that operates outside of focal attention is encouraged by Zajonc's (1980) recent criticism of the prevailing view that affective judgments are cognitively mediated--that is, his criticism of the view that we use our beliefs or cognitions about an object as the basis for evaluating the object. Zajonc's critique rested partly on the empirical observation that affective reactions often occur too rapidly to be cognitively mediated. If, as Zajonc suggests, affect and cognition are independent systems, then affective reactions could be influenced by information that impinges on the nervous system but does not penetrate the cognitive system of focal attention.

FIGURE 1

Message Processing, Involvement, and Memory



^aThe principle of encoding specificity (Tulving & Thomson 1973) is used to predict that recall will be aided by the cues or categories used in original encoding. To the extent that these cues can be supplied by the audience member unaided, "free" recall should be likely.

^bWith minimal encoding, situational cues present during original exposure to the message may effectively elicit recall.

A reformulation of Figure 1 that incorporates this possibility is given in Figure 2, which describes a segment of a multisystem analysis of human information processing (Greenwald in press b). In Figure 2, the system of focal attention and long-term memory (that is, the cognitive system) is not the only route by which input information can influence behavior. A second (noncognitive) path is proposed. In addition to Zajonc's evidence for affective-cognitive independence, some other support for the existence of the noncognitive path comes from several staple phenomena of contemporary social psychology: (a) discrepancies between verbal and nonverbal communications that emanate from the same person (Ekman & Friesan 1969), (b) inconsistencies between stated attitude and overt behavior (Wicker 1969), and (c) the pervasive occurrence of influences on human behavior that cannot be reported accurately by the cognitive system (Nisbett & Wilson, 1977).

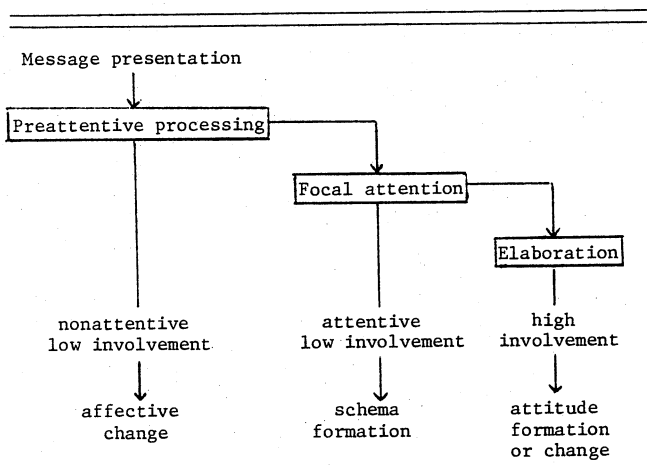
We shall refer to the two types of low involvement shown in Figure 2 as nonattentive and attentive low involvement. In proposing the nonattentive variety of low involvement, we do mean to suggest that an unattended advertising message may influence behavior without the audience member having any verbally reportable knowledge of this influence.

Three Bases for Effectiveness of Low Involvement

In most cases, high-involving advertising messages should be more effective than low-involving ones because the former should be better remembered. Nonetheless, there are at least three bases for expecting that low-involving messages can be effective with repeated exposures. In discussing the effects of low involvement we shall assume that information can be processed simultaneously by the cognitive and noncognitive paths shown in Figure 2.

FIGURE 2

Varieties of Involvement



Note: As processing proceeds from preattentive processing to focal attention and to encoding elaboration, the later processes should add to (rather than replace) the earlier ones. However, since the later processes may also be likely to have more powerful consequences for behavior, their effects may predominate. The effect of the high-involvement path is referred to as formation or change of an attitude because it represents an integrated effect on affect and cognition.

1. Reduced counterarguing. First we consider a circumstance under which low involvement may enhance the

effectiveness of a focally attended message. If the cognitive response elicited by a message would tend to oppose its effect--as in the case of a political advertisement for a candidate one intends to vote against--then the audience's lack of elaborative encoding may be more desirable than the presence of such elaboration from the advertiser's perspective.

2. Learning by repetition. Certain effects do not require elaborative encoding, and can therefore occur quite successfully with low involvement. In particular the aim of establishing a simple schema in the audience's memory can be achieved by means of repetition when there is only minimal encoding. You don't have to have been highly involved with the relevant ads to find yourself tending to respond to a bartender's "What'll you have?" by thinking (if not saying) "Pabst Blue Ribbon." The low-involving repetition of this message, with the accompaniment of a musical theme that aids in organizing the response together with the question, unavoidably establishes the desired scheme in listeners who are exposed to it often enough. One of us (Leavitt 1961) has previously analyzed the role of stimulus qualities of advertising messages in facilitating recall that may be dependent only on minimal encoding. The attention-holding characteristic that Leavitt designated intrigue may be all that is needed to produce this sort of learning.⁵

3. Learning without attention. Zajonc (1980), in his paper on the separateness of cognitive and affective systems, reported several studies in which repeated exposure to a stimulus increased the liking for that stimulus without apparent cognitive mediation. In one of these studies (Wilson 1979) the repeated stimuli were presented in the unattended ear in a dichotic listening task, to assure that their effect was not mediated by focal attention. It is such results that prompt us not to dismiss the importance to consumer psychology of the nonattentive variety of low involvement.

Conclusion

The first two of our suggested mechanisms of low-involvement persuasion--reduction of counterarguing and learning via repetition--fit well with existing theory in social and cognitive psychology. The third--learning without attention--does not. Nonetheless, and even though we count ourselves among those who look skeptically at claims of "hidden persuasion" and "subliminal influence," we believe that the proposal of noncognitively mediated influence should receive serious consideration. Needless to say, it will be a challenge to researchers to develop designs and procedures that can convincingly demonstrate influences on consumer behavior that occur without self-conscious cognitive mediation.

⁵This perceptual learning perhaps resembles the process by which very young children develop object schemas--before they have much of a cognitive structure that is capable of providing elaborative encoding. This similarity to perceptual learning may allow insight into a phenomenon that must have been noticed by every parent who has ever observed a child in front of a television set--a degree of rapt attention to advertisements that often exceeds attention to other programming. Although this attention may be very focused, it may often not meet our conceptual criterion of high involvement because there may be only minimal elaborative encoding (or cognitive responding). The resulting schema may be easily recognized or elicited by appropriate cues (as when the child accompanies the parent to market), even though it is not readily producible in free recall.

References

- Calder, B. S. (1979), "When Attitudes Follow Behavior--A Self-Perception/Dissonance Interpretation of Low Involvement," in J.C. Maloney and B. Silverman (eds.), Attitudes Play for High Stakes, Chicago: American Marketing Association.
- Craik, F. I. M., and Lockhart, R. S. (1972), "Levels of Processing: A Framework for Memory Research," Journal of Verbal Learning and Verbal Behavior, 11, 671-684.
- _____, and Tulving, E. (1975), "Depth of Processing and the Retention of Words," Journal of Experimental Psychology: General, 104, 268-294.
- Csikszentmihalyi, C. (1975), Beyond Boredom and Anxiety, San Francisco: Sassey-Bass.
- de Charms, R. (1968), Personal Conversation: The Internal Affective Determinants of Behavior, New York: Academic Press.
- Ekman, P., and Friesen, W. V. (1969), "Nonverbal Leakage and Clue to Deception," Psychiatry, 72, 118-137.
- Greenwald, A. G. (1968), "Cognitive Learning, Cognitive Response to Persuasion, and Attitude Change," in A. G. Greenwald, T. C. Brock, and T. M. Ostrom (eds.), Psychological Foundations of Attitudes, New York: Academic Press.
- _____. (1981 a), "Ego Task Analysis: An Integration of Research on Ego-Involvement and Self-Awareness," in A. Hastorf and A. Isen (eds.), Cognitive Social Psychology, New York: Elsevier-North Holland, in press.
- _____. (1981 b), "Social Psychology from the Perspective of Self," in J. Suls (ed.), Social Psychological Perspectives on Self, Hillsdale, NJ: Erlbaum, in press.
- Houston, M. J. and Rothschild, M. L. (1977), "A Paradigm for Research on Consumer Involvement," unpublished paper, Graduate School of Business, University of Wisconsin, Madison.
- Krugman, H. E. (1965), "The Impact of Television Advertising: Learning Without Involvement," Public Opinion Quarterly, 29, 349-356.
- _____. (1967), "The Measurement of Advertising Involvement," Public Opinion Quarterly, 30, 583-596.
- _____. (1971), "Brain Wave Measures of Media Involvement," Journal of Advertising Research, 11, 3-9.
- Kuiper, N. A., and Rogers, T. B. (1979), "Encoding of Personal Information: Self and Other Differences," Journal of Personality and Social Psychology, 37, 499-512.
- Leavitt, C. (1961), "Intrigue in Advertising: The Motivating Effects of Visual Organization," Proceedings of the 7th Annual Conference of the Advertising Foundation, New York.
- _____, and John Walton (1978), "Personality and Adoption Behavior," in Clark Leavitt (ed.), Proceedings of the Division 23 Program, American Psychology Association, Columbus, Ohio.
- Mitchell, A. A. (1979), "Involvement: A Potentially Important Mediator of Consumer Behavior," in W. G. Wilkie (ed.), Advances in Consumer Research, Vol. 6, Ann Arbor: Association for Consumer Research.
- Nisbett, R. E. and Wilson, T. D. (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes," Psychological Review, 84, 231-259.
- Petty, R. E., and Cacioppo, J.T. (1979), "Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Responses," Journal of Personality and Social Psychology, 37, 1915-1926.
- _____, Ostrom, T. M., and Brock, T. C. (eds.) (1980), Cognitive Responses in Persuasive Communication, Hillsdale, NJ: Erlbaum, in press.
- Ray, M. (1973), "Marketing Communication and the Hierarchy of Effects," in P. Clarke (ed.), New Models for Mass Communication Research, Beverly Hills, CA: Sage Publications.
- Rogers, T. B., Kuiper, N. A., and Kirker, W. S. (1977), "Self Reference and the Encoding of Personal Information," Journal of Personality and Social Psychology, 35, 677-688.
- Tulving, E., and Thomson, D. M. (1973), "Encoding Specificity and Retrieval Processes in Episodic Memory," Journal of Experimental Psychology, 80, 352-373.
- Wicker, A. W. (1969), "Attitudes Versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects," Journal of Social Issues, 25, 41-78.
- Wilson, W. R. (1979), "Feeling More Than We Can Know: Exposure Effects Without Learning," Journal of Personality and Social Psychology, 37, 811-821.
- Zajonc, R. B. (1980), "Feeling and Thinking: Preferences Need No Inferences," American Psychologist, 35, 151-175.

ISSUE INVOLVEMENT AS A MODERATOR OF THE EFFECTS ON ATTITUDE
OF ADVERTISING CONTENT AND CONTEXT

Richard E. Petty, University of Missouri-Columbia
John T. Cacioppo, University of Iowa

Abstract

Two persuasion experiments were conducted manipulating the personal relevance of the message, the quality of the arguments employed, and the characteristics of the message source. The results suggested that message content factors are more influential than source characteristics under high involvement conditions. The reverse tends to occur under low involvement.

The Involvement Concept

Persuasion researchers within both social and consumer psychology have recently emphasized the need to distinguish between high and low involvement situations (cf., Petty, Cacioppo, & Heesacker 1981, Ray 1976). Although there are many specific definitions of "involvement" within both disciplines, there is considerable agreement that in high involvement situations, the persuasive message under consideration has a high degree of personal relevance to the recipient, whereas in low involvement situations, the personal relevance of the message is rather trivial. In social psychology, this view is best represented by the work of the Sherifs who have argued that high involvement occurs when the message has "intrinsic importance" (Sherif & Hovland 1961, p. 197) or "personal meaning" (Sherif et al. 1973, p. 311), and in consumer psychology this view is well-represented by Krugman (1965) who defines involvement as the number of "personal references" or connections that recipients make between the message and their own lives (p. 355).

Despite the widespread agreement that involvement has something to do with the personal relevance of a message, there is little agreement on the research operations employed in studying involvement. Some of the research on involvement conducted by both social (e.g., Hovland, Harvey, & Sherif 1957) and consumer (e.g., Newman & Dolich 1979) psychologists has investigated existing groups that differed in the extent to which an issue or product was personally important, or has employed designs allowing subjects to assign themselves to high and low involvement groups. These methods, which are correlational in nature, confound involvement with all other existing differences between the high and low involvement groups. Other social (e.g., Rhine & Severance 1970) and consumer psychologists (e.g., Lastovicka & Gardner 1979) have defined involvement in terms of the specific issue or product under consideration. This procedure, of course, confounds involvement with aspects of the issue or product that are immaterial to their personal relevance. Finally, some researchers have studied involvement by varying the medium of message presentation. Interestingly, however, some investigators have argued that television is a more involving medium than is print (Worchel et al. 1975), whereas others have argued just the opposite (Krugman 1967). A preferable procedure that keeps recipient, message, and medium characteristics constant for high and low involvement conditions was introduced by Apsler & Sears (1968) and is the method employed in the studies to be reported here. In this procedure subjects are randomly assigned to high and low involvement conditions and receive the same message via the same medium, but high involvement subjects are led to believe that the issue or product has some personal relevance whereas low involvement subjects are not.

In addition to the methodological differences that have plagued the involvement concept, another area of disagreement concerns the effects on persuasion that involvement is expected to have. The Sherifs have argued that increased involvement is associated with increased resistance to persuasion (cf., Sherif, Sherif, & Nebergall 1965). The notion is that on any given issue, highly involved persons exhibit more negative evaluations of a communication because high involvement is associated with an extended "latitude of rejection." Thus, incoming messages on high involvement issues are thought to have an enhanced probability of being rejected because they are more likely to fall within the unacceptable range of a person's implicit attitude continuum. This view has received considerable acceptance within social psychology (e.g., Eagly & Manis 1966, Greenwald 1980). Krugman (1965) has proposed an alternative view that has achieved considerable acceptance within consumer psychology (e.g., Ray 1974, Rothchild 1979). Under this second view, increasing involvement does not invariably increase resistance to persuasion, but instead shifts the sequence of communication impact. Krugman argues that under high involvement, a communication affects cognitions, then attitudes, then behavior, whereas under low involvement a communication affects cognitions, then behavior, and then attitudes. The focal goal of the present paper is to present and test a third view of how involvement affects persuasion.

Involvement as a Determinant of
Content-based Persuasion

Elsewhere we have proposed that the level of involvement directs the focus of a subject's thoughts about a persuasive communication (Petty & Cacioppo 1979). Specifically, we have suggested that under high involvement conditions, the focus of thought is on the content of the persuasive message, whereas under low involvement conditions, the focus of thought is on non-content cues. Thus, under high involvement, if the communication presents arguments that are subjectively cogent and compelling, the recipient's thoughts will be primarily favorable and persuasion will result. If the communication presents arguments that are subjectively specious and subject to counterargumentation, resistance to persuasion (and perhaps boomerang) will occur. Thus, contrary to the Sherifs' view, increasing involvement can lead to either enhanced or reduced persuasion depending upon the quality of the arguments presented in the message.

In contrast to this focus on the content of a message under high involvement conditions, we have suggested that subjects who are not involved are more likely to focus on such non-content cues as the rewards available for adopting a certain attitude, the attractiveness, credibility, or power of the communication's source, and the number of others who advocate a certain position. Focusing on each of the latter aspects of a communication allows a person to evaluate a message or decide what attitudinal position to adopt without engaging in any extensive cognitive work relevant to the issue or product under consideration. As Miller et al. (1976) noted: "It may be irrational to scrutinize the plethora of counter-attitudinal messages received daily. To the extent that one possesses only a limited amount of information processing time and capacity, such scrutiny would disengage the thought processes from the exigencies of daily life" (p. 623). Thus, when a

person is not highly involved with a persuasive message (i.e., when the message has no personal consequences), we propose that the person relies on a short-cut means of evaluation. Although, like Krugman, we are proposing that there are separate processes governing persuasion under high and low involvement, unlike Krugman we believe that the sequence of communication impact is the same--cognitions, attitudes, then behaviors. The difference between the two processes lies in what cognitions are affected--cognitions dealing with issue-relevant argumentation (high involvement), or cognitions dealing with non-content features of the influence situation (low involvement).

Empirical Tests of the Involvement Concept

In an initial test of our two-process model of involvement (Petty & Cacioppo 1979), subjects heard a counterattitudinal message containing either strong or weak arguments under conditions of either high or low involvement. As expected, increasing involvement enhanced the production of counterarguments to the weak arguments and increased the production of favorable thoughts to the strong arguments. Consistent with this finding, increasing involvement increased the persuasiveness of the strong arguments, but decreased the persuasiveness of the weak arguments. Although the results of this initial study did support the view that subjects do more thinking about the content of a message under high involvement than under low, it did not directly address whether subjects are more attentive to content-irrelevant cues under low involvement than under high.

Next, we report two experiments designed to test the full two-process model of involvement. In each study, subjects were exposed to a persuasive communication. In Experiment 1 (conducted in collaboration with Rachel Goldman) the message was presented on audio tape and concerned a change in a campus regulation. In Experiment 2, the message was presented in print form and concerned a new consumer product. The following variables were manipulated in each study: (a) the personal relevance of the message (high and low involvement), (b) the quality of the arguments which subjects heard or read in support of the advocated conclusion (strong or weak arguments), and (c) a characteristic of the source presenting the message (high or low expert source in Experiment 1, high or low attractive source in Experiment 2). The two-process model of involvement would expect that under high involvement conditions, persuasion would be affected most by the quality of the message arguments employed, but that under low involvement conditions, persuasion would be tied most strongly to the credibility or attractiveness of the message source.

Experiment 1

One hundred forty-five male and female undergraduates at the University of Missouri participated in order to earn extra credit in an introductory psychology course. The design was a 2 (Issue involvement: high or low) \times 2 (Argument quality: strong or weak) \times 2 (Source credibility: high or low) factorial. Subjects were run in groups of 3 to 16 in cubicles designed so that no subject could have visual or verbal contact with any other subject. Upon arrival at the laboratory, subjects read that they would be rating tapes for their sound quality. After reading these instructions, subjects heard one of four tapes over headphones. After listening to the appropriate communication, subjects completed the dependent variable booklets, and were then debriefed, thanked, and dismissed.

Independent Variables

Argument quality. All subjects heard a communication that advocated that seniors be required to pass a comprehensive exam in their major area as a requirement for graduation. The strong version of the message provided persuasive evidence (statistics, data, etc.) in support of the

exam (e.g. institution of the exams has led to a reversal in the declining scores on standardized achievement tests at other universities). In contrast, the weak version of the message relied more on quotations and personal opinion and examples to support its position (e.g. the author's major advisor took a comprehensive exam and now had a prestigious academic position). The strong arguments were taken from a pool that had elicited primarily favorable thoughts in a pretest, and the weak arguments were taken from a pool that had elicited primarily counterarguments.

Issue involvement. Subjects in the high involvement conditions heard the speaker advocate that the comprehensive exams should begin in the 1979-1980 academic year at their university (in which case they would all be affected personally by the proposal). In the low involvement conditions, the speaker advocated that the exams be initiated in the 1989-1990 academic term.

Source credibility. The high credible source was described as a professor of education from Princeton University who had conducted a study of comprehensive exams nationally. The low credible source was a junior at a local high school who had prepared a term paper on the topic.

Dependent Variables

After hearing the tape, subjects completed two measures of opinion about the topic. First, subjects rated the concept "senior comprehensive exams" on four 9-point semantic differential scales (harmful-beneficial, wise-foolish, good-bad, favorable-unfavorable) that were summed to form a general measure of evaluation. Next, subjects responded to an 11-point Likert-type scale concerning their extent of agreement with the speaker's proposal. The responses to these two attitude measures were converted to standard scores and averaged to form an index of communication acceptance.

Following the key attitude measures, subjects completed some manipulation check measures and other ancillary questions. Finally they were given 4 minutes to list as many of the arguments provided in the communication as they could remember. Two judges, blind to the experimental conditions rated each argument listed for accuracy ($r = .92$). Similar statements of the same argument were only counted once. Disagreements between judges were resolved by consulting a third judge.

Results

Analyses on the manipulation check measures indicated that the three independent variables were successfully varied. Subjects hearing the high credibility speaker rated him as more "qualified" ($M = 6.4$) than subjects who were exposed to the low credibility induction ($M = 5.8$), $F(1,137) = 4.86$, $p < .03$. Also, subjects hearing the strong arguments rated their quality as being significantly higher ($M = 8.9$) than subjects exposed to the weak arguments ($M = 4.5$), $F(1,137) = 51.02$, $p < .001$. Finally, subjects in the high involvement condition rated the likelihood that the University of Missouri would institute comprehensive exams during their stay as higher ($M = 5.5$) than subjects in the low involvement conditions ($M = 2.7$), $F(1,137) = 5.12$, $p < .02$.

The means for each cell on the measure of communication acceptance are presented in Table 1, A 2 \times 2 \times 2 analysis of variance on this measure yielded two main effects and two qualifying interactions. A main effect for credibility, $F(1,137) = 6.06$, $p < .02$, indicated that the high credibility communicator induced more acceptance ($M = .21$) than the low credibility communicator ($M = -.21$). A main effect for argument quality revealed that the strong arguments produced more agreement with the position advocated ($M = .36$) than did the weak arguments ($M = -.36$), $F(1,137) = 20.35$, $p < .001$.

Of most interest are the two interactions, however. An Involvement X Arguments interaction, $F(1,137) = 6.05, p < .02$, demonstrated that the strong arguments produced significantly more persuasion than the weak only under conditions of high personal involvement. Also, an Involvement X Credibility interaction, $F(1,137) = 3.92, p < .05$, revealed that the high credibility speaker produced significantly more persuasion than the low only under conditions of low personal involvement.

TABLE 1

EFFECT OF INVOLVEMENT, ARGUMENT QUALITY, AND SOURCE CREDIBILITY ON ATTITUDES TOWARD COMPREHENSIVE EXAMS

	High Involvement		Low Involvement	
	high cred.	low cred.	high cred.	low cred.
Strong Arguments	.62 _{bd}	.58 _{bd}	.38 _{bc}	-.14 _{acd}
Weak Arguments	-.40 _{ac}	-.61 _a	.24 _{bc}	-.65 _a

Note.--Means without a common subscript are significantly different at the .05 level by the Newman-Keuls procedure.

Finally, an analysis of the argument recall scores indicated that subjects were able to recall more of the strong ($M = 4.2$) than the weak ($M = 3.2$) arguments, $F(1,137) = 14.93, p < .001$, but involvement did not affect argument recall.

Discussion

The present study provided evidence for the view that the level of issue involvement is one determinant of whether content or noncontent factors dominate in producing persuasion. When the issue was of high personal relevance to the subjects, the manipulation of argument quality had a significant impact on attitudes, whereas the effect of source credibility was small and not significant. Under low involvement conditions, however, the reverse occurred--source credibility had a large effect on persuasion, but argument quality had little impact. This pattern of results provides strong support for the view that high issue relevance motivates diligent processing of the content of a message. When the issue is relatively uninvolved, however, subjects appear to be unwilling to engage in the effortful cognitive work necessary to evaluate the quality of the arguments.

Under low involvement conditions, subjects were content to form their attitudes on the basis of who said it, rather than on the merits of what was said. In the second experiment, a conceptual replication of Experiment 1 was conducted employing advertising stimuli.

Experiment 2

Two hundred and forty undergraduates at the University of Missouri participated in a study entitled "Evaluations of Advertisements," and received extra credit in an introductory psychology course. The design of the study was a 2 (Product involvement: high or low) X 2 (Message content quality: strong or weak) X 2 (Source attractiveness: high or low) X 2 (Sex of subject). Fifteen subjects were run in each cell of the design in groups of 4 to 8.

Upon arrival at the laboratory, the subjects were told that they would be examining a variety of advertisements and providing their impressions of them. Each subject was handed a booklet containing six magazine ads. Five of the ads were for real but relatively unfamiliar products (e.g. Lux cigarettes, Rolex watches), and one of the ads was for

a fictitious product (Vilance shampoo). The fictitious ad was fifth in the booklet of six and was prepared specifically for the present experiment. The subjects were told to turn through the booklet at their own pace and give a signal to the experimenter when they had finished examining the ads. Following perusal of the ad booklet, subjects completed a dependent variable booklet, and were then debriefed, thanked, and dismissed.

Independent Variables

Product involvement. Preceding each ad in the booklet was a brief description of the purpose of the ad. All subjects read the same descriptions for the real ads, but the description for the bogus Vilance shampoo ad differed for subjects in the high and low involvement conditions. Subjects in the low involvement conditions read:

The product you are about to see is being put out by the Europa Collection based in Vienna, Austria. Their main interest lies in introducing the product to the rest of Europe. Before launching a new European campaign, they have distributed their advertising materials to Journalism schools in the U.S. This is in order to determine if the ad has enough basic appeal to make it worth pursuing. The psychology department is assisting the journalism school in this evaluation.

Subjects in the high involvement conditions read a similar description except that they were told that the company was interested in introducing the product to the United States.

The University of Missouri has been chosen for research purposes. It is for this reason that the product will soon be introduced in the Columbia area. Since you will soon be able to purchase this product in Columbia, the company is asking University students to evaluate their proposed advertisement.

Source attractiveness. Four different advertisements for Vilance shampoo were created to vary the source and message variables. Each ad looked similar in that it presented a male and female in their early 20s giving the reasons why they liked Vilance shampoo. In the high attractive ads a photograph of a couple that previously had been rated as "extremely attractive" was used, and in the low attractive ads a photograph of a couple that previously had been rated as "somewhat unattractive" was used. An "extremely unattractive" stimulus was not used because it did not prove plausible in pilot testing.

Message quality. In the strong argument conditions, the text that accompanied the photograph presented arguments for the shampoo that previously had been rated as compelling and persuasive (e.g. Vilance contains minerals that strengthen each hair shaft so it helps prevent split-ends). In the weak argument conditions, the text presented arguments that previously had been rated as unpersuasive (e.g. Vilance has a down-to-earth brown color that makes us feel natural).

Dependent Variables

Subjects were asked to answer a number of questions about each product for which they had seen an ad. Different questions about the advertisements were also posed to maintain the cover story. The crucial measure of attitude toward the product consisted of the sum of subjects' ratings of Vilance shampoo on four 9-point semantic differential-type scales (good/bad, satisfactory/unsatisfactory, favorable/unfavorable, high quality/low quality).

Results

The means for each cell on the measure of attitude toward

Vilance shampoo are presented in Table 2. A 2 X 2 X 2 X 2 analysis of variance revealed no main effects nor interactions involving the sex of subject variable so this will not be discussed further. A main effect for the argument quality manipulation, $F(1,223) = 35.41, p < .0001$, indicated that the strong arguments produced more favorable attitudes toward the product ($M = 6.6$) than did the weak arguments ($M = 1.3$). A main effect for the attractiveness variable, $F(1,223) = 8.46, p < .004$, showed that the high attractive models induced more acceptance of the product ($M = 5.3$) than did the low attractive models ($M = 2.7$).

Of most interest in the analysis, however, was a significant Involvement X Arguments interaction, $F(1,223) = 10.84, p < .001$. This interaction revealed that increasing the personal involvement of the advertisement enhanced the importance of the message content in affecting attitudes. When the relevance of the ad was increased, subjects responded more favorably to the ad with strong arguments, but less favorably to the ad with weak arguments. The Involvement X Source interaction was not significant this time though the means were directionally consistent with the findings of Experiment 1.

TABLE 2

EFFECT OF INVOLVEMENT, ARGUMENT QUALITY, AND SOURCE ATTRACTIVENESS, ON ATTITUDES TOWARD AN ADVERTISED PRODUCT

	High Involvement		Low Involvement	
	high	low	high	low
	Attr.	Attr.	Attr.	Attr.
Strong Arguments	9.1	7.0	6.9	3.4
Weak Arguments	.9	-1.2	4.1	1.4

Discussion

Experiment 2 replicated the finding from Experiment 1 that involvement (personal relevance) is an important determinant of the extent to which content-based persuasion will occur. When the advertisement was high in personal relevance to the subjects, the quality or cogency of the arguments presented in the ad had a much greater impact on attitudes toward the advertised product than when the ad was of low relevance. The study did not find strong support for the view that source factors were more important when the ad was low rather than high in personal relevance, however. In retrospect, this effect may not have been strong in this study because how the models looked may have been viewed as a relevant persuasive argument for some subjects! In other words, for the specific product employed (shampoo), the attractiveness of the models (especially their hair) may have served as persuasive testimony for the effectiveness of the product.

The Role of Involvement in Persuasion

Taken together, the results of the two studies strongly indicate that under high involvement, message content is the prepotent determinant of the amount of persuasion that occurs. Less strongly, the studies suggest that under low involvement, non-content factors such as the credibility or attractiveness of the message source are more important. Thus, the present studies provide some evidence that attitude change is determined by different factors under high and low involvement conditions.

In another paper (Petty & Cacioppo 1981) we have argued that a persuasive message can induce attitude change via one of two routes. Under the first, or central route, thinking about issue-relevant information is the most

direct determinant of the direction and amount of persuasion produced. Attitude changes induced via this route tend to be relatively permanent and predictive of subsequent behavior. Under the second, or peripheral route, attitude change is the result of non-content cues in the situation. Changes induced via this route tend to be relatively temporary and are not highly predictive of subsequent behavior (see review by Cialdini et al., 1981).

According to this framework, involvement is a prime determinant of whether or not an attitude change is induced via the central or the peripheral route. The experiments reported here as well as the results of other recent studies are consistent with the view that under low involvement conditions, persuasion may typically be governed by such peripheral cues as source characteristics, concerns about desirable self-presentation, and/or one's social role, whereas under high involvement, persuasion may be governed more by message content factors such as the number, quality, and/or accuracy of the message arguments presented (cf., Cacioppo & Petty 1980, Chaiken 1980, Cialdini et al. 1976).

The level of involvement is not the only determinant of the route to persuasion, however. In addition to having the necessary motivation to think about issue-relevant argumentation, the message recipient must also have the ability to process the message if change via the central route is to occur. Thus, if involvement is high (and the person is motivated to think about the message content), but the arguments are too complex for the person to understand, or if too many distractions prevent issue-relevant thought, then the central route cannot be followed.

Finally, we note that attitude change via the central route is a very difficult way to change a person's attitudes. First, the message must show some personal relevance to the recipient. Second, the person must have the ability to process the message content. Third, the message must present arguments that elicit primarily favorable thoughts. If the recipient is able to counterargue the message, then increasing involvement will not facilitate persuasion. If a change can be produced via the central route however, the benefits are clear--the attitude change will tend to persist and be predictive of subsequent behavior. An alternative strategy is to induce attitude change via the peripheral route. Since the peripheral route induces only a temporary change, it will be necessary to constantly remind the recipient of the cue (e.g., attractive source) upon which the new attitude is based. Lutz (1979) provides the example of a person who drives Hertz Rent-a-Cars, not because the person has thought about the attributes of the company (central route), but only because he has been constantly reminded that O. J. Simpson endorses the company (peripheral route). If the favorable attitude about Hertz had been based on a full consideration of the positive features of the company, the favorable attitude would likely persist on its own. Since the favorable attitude is based on a positive peripheral cue however, the favorable attitude persists only so long as the cue remains salient (accomplished through advertising repetition). Such continually repeated positive cues may be sufficient to get a person to try the advertised product. Interestingly, once the person has tried the product, it may become more personally involving and may make the person more likely to think about the content of future advertisements about the product. In this manner a peripheral change can lead to a central one.

References

Apsler, R., and Sears, D.O. (1968), "Warning, personal involvement, and attitude change," Journal of Personality and Social Psychology, 9, 162-166.

Cacioppo, J.T., and Petty, R.E. (1980), "Sex differences in influenceability: Toward specifying the underlying process," Personality and Social Psychology Bulletin, in press.

Chaiken, S. (1980), "Heuristic versus systematic information processing and the use of source and message cues in persuasion," Journal of Personality and Social Psychology, in press.

Cialdini, R.B., Levy, A., Herman, P., Kozlowski, L., & Petty, R. (1976), "Elastic shifts of opinion: Determinants of direction and durability," Journal of Personality and Social Psychology, 34, 663-672.

Cialdini, R.B., Petty, R.E., and Cacioppo, J.T. (1981), "Attitude and attitude change," Annual Review of Psychology in press.

Eagly, A.H., and Manis, M. (1966), "Evaluation of message and communication as a function of involvement," Journal of Personality and Social Psychology, 3, 483-485.

Greenwald, A.G. (1980), "The totalitarian ego: Fabrication and revision of personal history," American Psychologist, 35, 603-618.

Hovland, C.I., Harvey, O.J., and Sherif, M. (1957), "Assimilation and contrast effects in reactions to communication and attitude change," Journal of Abnormal and Social Psychology, 55, 244-252.

Krugman, H.E. (1965), "The impact of television advertising: Learning without involvement," Public Opinion Quarterly, 29, 349-356.

Krugman, H.E. (1967), "The measurement of advertising involvement," Public Opinion Quarterly, 30, 583-596.

Lastovicka, J.L., and Gardner, D.M. (1979), "Components of involvement," in Attitude research plays for high stakes, eds., J. Maloney and B. Silverman, Chicago: American Marketing Association, 53-73.

Lutz, R.J. (1979), "A functional theory framework for designing and pretesting advertising themes," in Attitude research plays for high stakes, eds., J. Maloney and B. Silverman, Chicago: American Marketing Association, 37-49.

Miller, N., Maruyama, G., Beaver, R., and Valone, K. (1976) "Speed of speech and persuasion," Journal of Personality and Social Psychology, 34, 615-624.

Newman, L.M., and Dolich, I.J. (1979), "An examination of ego-involvement as a modifier of attitude changes caused from product testing," Advances in Consumer Research, 6, 180-183.

Petty, R.E., and Cacioppo, J.T. (1979), "Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses," Journal of Personality and Social Psychology, 37, 1915-1926.

Petty, R.E., and Cacioppo, J.T. (1981), Attitudes and persuasion: Classic and contemporary approaches, Dubuque, Iowa: William C. Brown Co.

Petty, R.E., Cacioppo, J.T., and Heesacker, M. (1981), "The use of rhetorical questions in persuasion: A cognitive response analysis," Journal of Personality and Social Psychology, in press.

Ray, M.L. (1974), "Consumer initial processing: Definitions, issues, and applications," in Buyer/Consumer information processing, eds., G. Hughes and M. Ray, Chapel Hill: University of North Carolina Press, 145-156.

Ray, M.L. (1976), "Attitude as a communication response," in Attitude research at bay, eds., D. Johnson and W. Wells, Chicago: American Marketing Association, 87-111.

Rhine, R., and Severance, L. (1970), "Ego-involvement, discrepancy, source credibility, and attitude change," Journal of Personality and Social Psychology, 16, 175-190.

Rothschild, M. L. (1979), "Advertising strategies for high and low involvement situations," in Attitude research plays for high stakes, eds., J. Maloney and B. Silverman, Chicago: American Marketing Association, 74-93.

Sherif, M., and Hovland, C. (1961), Social judgment, New Haven: Yale University Press.

Sherif, C. W. et al. (1973), "Personal involvement, social judgment, and action," Journal of Personality and Social Psychology, 27, 311-328.

Sherif, C. W., Sherif, M., and Nebergall, R.E. (1965), Attitude and attitude change, Philadelphia: Saunders.

Worchel, S., Andreoli, V., and Eason, J. (1975), "Is the medium the message? A study of the effects of media, communicator, and message characteristics on attitude change," Journal of Applied Social Psychology, 5, 157-172.

THE DIMENSIONS OF ADVERTISING INVOLVEMENT

Andrew A. Mitchell, Carnegie-Mellon University

Abstract

A conceptual definition of involvement and a conceptual model of how involvement affects the information acquisition process is presented. This definition views involvement as an internal state of an individual with both intensity and directional properties. As such, it is one of many different variables that may affect the information acquisition process. This definition and model is then used to discuss alternative definitions of involvement and previous research examining the effect of involvement on communication processes.

Introduction

Interest in understanding the effect of involvement on communication processes has been increasing in recent years. There seems to be a general understanding of communication effects under conditions of "high involvement" and a general agreement that communication effects under conditions of "low involvement" are different from those under "high involvement". However, these effects under "low involvement" conditions are not well understood. Recently, there have been a number of attempts to examine these differences experimentally (e.g., Petty and Cacioppo 1979, Mitchell 1980, Mitchell, Russo and Gardner 1980), however, the "involvement" manipulations differ in these studies so it is difficult to understand the results.

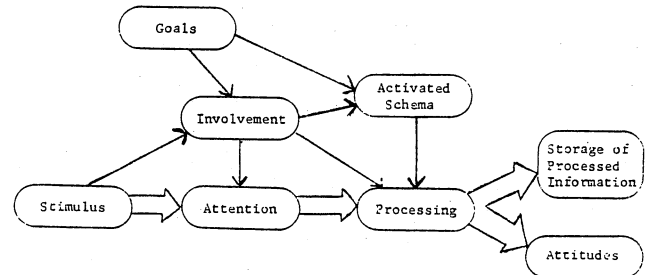
Much of the confusion surrounding involvement and its effect on communication processes seems to center on our failure to understand conceptually what involvement is and how it may affect the information acquisition process. In this paper, I will attempt to clarify this misunderstanding. In the next section I will present a definition of involvement and a conceptual model of how involvement affects the information acquisition process. Next, I will discuss our research which has been based on this model and then use the model to discuss previous research examining the effect of involvement on communication processes and alternative conceptual definitions of involvement.

Conceptual Model

In a previous paper, I defined involvement as an individual level, internal state variable whose motivational properties are evoked by a particular stimulus or situation (Mitchell 1979). As such, it has two dimensions, intensity and direction. Consequently, in discussing involvement in a particular context, we must define both the level of involvement (e.g., high vs. low) and also the direction of involvement (e.g., directed at the advertised brand or at an element of the advertisement).

A conceptual model of the effect of involvement on the information acquisition process is presented in Figure 1. The double arrows indicate the critical stages of the information acquisition process that affect attitude formation and the retrieval of processed information from long term memory. These critical stages are attention and processing. Attention is conceptualized as having capacity limitations (e.g., Kahneman 1973) which causes individuals to focus their cognitive resources on a limited number of stimuli in the environment. In processing information, then, one decision individuals must make is which stimuli to attend to and the amount of attention to devote to each stimuli.

FIGURE 1
CONCEPTUAL MODEL



The second stage is processing. Here individuals interpret the information from the environment, make inferences and evaluate the information. Numerous theories from cognitive psychology indicate that how individuals process information affects their ability to retrieve the information at a later point in time (Mitchell 1980). For instance, the depth of processing approach suggests that the use of semantic processing and the amount of elaboration that occurs during processing enhances the ability to retrieve the processed information (e.g., Craik and Tulving 1975, Craik 1979). Other theories hypothesize that other dimensions of processing affect recall (e.g., Paivio 1971, Nelson 1979). Although there is general agreement that the type of processing that occurs during exposure to a stimulus affects the recall of that stimulus, there is currently disagreement as to exactly what dimensions of processing affect recall.

There is also considerable research that indicates that the type of evaluative processing or the type of cognitive responses generated during exposure affects attitude formation and change (e.g., Greenwald 1968, Wright 1973, 1980). For instance, counterarguments reduce the favorableness of the attitudes that are formed while support arguments enhance their favorableness. In summary, then, the two critical stages in the information acquisition process are attention and processing. These two stages will affect what information can be retrieved from long term memory and the formation and change of attitudes.

The factors which affect attention and processing are the stimulus (e.g., advertisement) and the goals of the individual while exposed to the stimulus. These factors, in turn, determine the level and direction of involvement and the particular memory schema that is used to process the information. Here we define memory schema as an organized set of information about a concept in long term memory (e.g., Rumelhart and Ortony 1978).

To illustrate the model, assume that a number of individuals are exposed to an advertisement for a new fuel efficient automobile. In the advertisement, the automobile is shown on a dock surrounded by a number of sailboats. If an individual, for instance, was planning to purchase a small fuel efficient automobile in the near future and had not decided which brand to purchase, he or she would probably be very interested in this product and may be actively searching for information about the various alternatives on the market. Consequently, this individual would probably devote full attention to the advertisement and activate a memory schema that contains information about

fuel efficient automobiles so that he or she can critically evaluate the information in the advertisement. As a consequence of this processing, the information about the new fuel efficient automobile would be organized into a tightly integrated knowledge structure and, in most cases, an evaluation of the new automobile would be formed.

Another individual, however, may not be interested in purchasing an automobile or even obtaining information about new brands of fuel efficient automobiles. When this individual is exposed to the same advertisement, he or she may be attracted to the picture of the sailboat in the advertisement due to an interest in sailing. This individual might also devote full attention to the advertisement and may skim part of the copy, however, he or she would probably activate their memory schema for sailboats in processing information from the advertisement. This would result in a different type of processing than in the first case. Since the individual did not activate the memory schema that contains information about fuel efficient automobiles, he or she would not have comprehended the message (Rumelhart and Ortony 1978) and would generate few, if any, counterarguments or support arguments. As a result, he or she may have acquired some information about the advertised car, but it would be much less information than in the previous example and it probably would not be tightly organized into a knowledge structure about the advertised automobile. In addition, the individual probably would not have formed an evaluation of the automobile during exposure to the advertisement.

Finally, a third individual may have little interest in automobiles or sailboats and, therefore, this individual would probably devote little attention to the advertisement and would probably do little active processing of the information in the advertisement. In this case, the individual may be unable to retrieve any information about the new automobile and may not be aware of its existence. In summary, then, the content of the stimulus and the goals of the individual determine the amount and direction of involvement during exposure to the advertisement. The intensity of the involvement determines how much attention is devoted to the advertisement. The direction of the involvement determines which memory schema is activated which, in turn, determines the type of processing that occurs during exposure.

Our Research

In our research (Mitchell 1980, Mitchell Russo and Gardner 1980), we have directly manipulated processing strategy and attention levels to examine the resulting effects on attitudes, intentions and the retrieval of brand related information from long term memory. In this research, we differentiate between a brand and a nonbrand processing strategy. A brand processing strategy involves the active processing of brand information from an advertisement to either form an evaluation of the brand or acquire information about the brand. A nonbrand processing strategy involves the processing of information from the advertisement to achieve some other goal. This may be caused, for instance, by the use of an attention getting device such as an attractive model in the advertisement.

To manipulate processing strategy, we instruct subjects to either examine the advertisement to form an evaluation of the advertised brand (brand processing strategy) or to evaluate the effectiveness of the advertisement (nonbrand processing strategy). For the latter manipulation, we provide subjects with a set of criteria to use in evaluating the effectiveness of the advertisement. These two different instructions presumably result in the activation of different memory schema for use in processing the information.

In our initial study, (Gardner, Mitchell and Russo 1978, Mitchell, Russo and Gardner 1980), we examined the effect of the two different processing strategies with high attention levels on attitudes and the retrieval of information from long term memory. The subjects in each condition were shown advertisements for four new brands in four different product categories. The results indicated that subjects executing a brand processing strategy were able to retrieve more product related information faster and at a greater level of accuracy than subjects executing a nonbrand processing strategy (Mitchell, Russo and Gardner 1980). In addition, subjects using a brand processing strategy formed less favorable attitudes toward the advertised brands. Although the subjects using a brand processing strategy generated almost twenty times as many counterarguments, support argument and support derogation statements as subjects using a nonbrand processing strategy, these differences only partially explained the differences in attitudes. More specifically, the use of the number of counterarguments, support arguments and source derogations as covariates did not eliminate all the reliable difference in attitudes between the two groups. These results suggest the possibility that other mediators may affect attitude formation in these conditions.

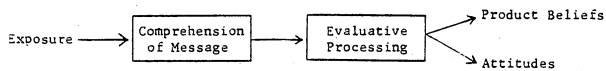
In a second study, these two different processing manipulations were used along with two different manipulations of attention levels (Mitchell 1980). For one attention level manipulation, subjects were shown advertisements for new products and were instructed to execute a brand processing strategy while watching a red and a blue light that randomly flashed at the rate of approximately 25 flashes per second. Subjects were instructed to raise their left hand whenever one of the two lights flashed. The second attention level manipulation required the subjects to count backward by three's from a three digit number while examining the advertisements. The subjects in each condition saw the same eight advertisements for new brands in eight different product categories. The results indicated differences in attitudes and the amount of product related information retrieved from long term memory between all conditions except the nonbrand processing condition and the second attention level manipulation where subjects had to count backward by three's.

These results suggest that differences in involvement may result in three different types of information acquisition processes. One process is caused by a "high involvement" condition and two are caused by "low involvement" conditions (Figure 2). Here "high involvement" refers to high interest levels in the advertised brand. Under the "high involvement" condition, individuals devote all their attention to the advertisement and execute a brand processing strategy. Consequently, they will critically evaluate the brand information in the advertisement and will, generally, form an overall evaluation of the advertised brand during exposure to the advertisement. This means that their verbal thought processes during exposure to the advertisement will contain a large number of counterarguments and support arguments.

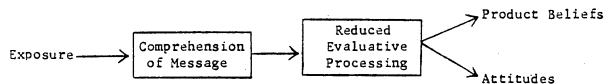
Under the first "low involvement" condition, individuals will also execute a brand processing strategy, however, this strategy will be executed with reduced attention levels. This will generally result in a reduction in the number of counterarguments and support arguments generated which, in turn, will frequently result in a difference between this condition and the "high involvement" condition in the attitudes that are formed. Under this condition, then, individuals will comprehend the message since the proper schema is activated, but they will not be as critical of the information as they were under conditions of "high involvement".

FIGURE 2
ALTERNATIVE MODELS OF INFORMATION ACQUISITION
UNDER DIFFERENT INVOLVEMENT CONDITIONS

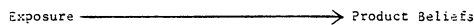
Model I: Information Acquisition Under Conditions of "High Involvement"



Model II: Information Acquisition Under Conditions of "Low Involvement"



Model III: Information Acquisition Under Conditions of "Low Involvement"



Finally, the second "low involvement" condition occurs when individuals either execute a nonbrand evaluation strategy or there are severe attention deficits. There will be few, if any, counterarguments or support arguments generated and the individuals will not totally comprehend the message. They will, however, acquire some information about the advertised brand.

These three different information acquisition processes should cause differences in the amount, content and organization of information about the advertised brand in memory. In the previously mentioned study (Mitchell 1980) subjects in the nonbrand brand processing condition and the second attention level condition could recall significantly less information about the advertised brands than subjects in the brand processing condition and the first attention level condition. It also took them longer to verify statements about the attributes of the advertised brands and their evaluation of them. No differences were found on these two measures between the brand strategy condition and the first attention level condition or the nonbrand strategy condition and the second attention level condition. Since attitudinal differences were found between the first two conditions, however, we might expect differences in the content of the information that would be recalled.

In summary, our conceptualization of involvement and our research suggest that differences in involvement may result in three different types of information acquisition processes -- one caused by a "high involvement" condition and two by "low involvement" conditions. In addition, we believe that the latter two processes may be caused by factors other than involvement. One of these may be a lack of knowledge about the topic of the communication.

Alternative Approaches To Involvement

There have been a number of articles over the last ten years either discussing the effect of involvement on the reception of communications or experimental studies which manipulate involvement and examine the resulting effects. As mentioned previously, these conceptualizations and manipulations have generally differed. In this section, I will discuss these alternative manipulations and conceptualizations using the model presented in the previous section.

Conceptual Approaches

Krugman: In his original article, Krugman (1965) suggested that the cognitive processes that occur during exposure to television advertisements were similar to those that occur during the learning nonsense syllables. As evidence of this similarity, Krugman mentioned that the recall of three consecutive television commercials displays the same U shaped relationship as the recall of a series of nonsense syllables -- strong primacy and recency effects. He further hypothesized that television advertising produces subtle shifts in our perceptions of brands which result in changes in the saliency of the different attributes of the product. In this article, and in a second article (Krugman 1967), he defines involvement as the number of "bridging experiences, connections or personnel references per minute that the viewer makes between his own life and stimulus", not as the "amount of attention, interest or excitement".

This definition of involvement differs from the one presented here in that Krugman defines involvement as one of the dimensions of the type of processing that occurs during exposure to the advertisement. In contrast, I define involvement as a particular state of the individual at a point in time. Our state variable conceptualization of involvement affects the type of processing that occurs during exposure, however, it is not the only variable that may affect it. The amount and type of knowledge that an individual has and the memory schema that is activated during exposure also affects the resulting processes. Therefore, it is entirely possible that according to our definition an individual may be highly involved, however, if this person has little knowledge about the advertised product, the resulting process might be a low involvement one according to Krugman's definition.

In general, we believe that information acquisition under the second low involvement condition is similar to the one Krugman describes in his original paper. Under our second low involvement condition individuals acquire information about the advertised brand but do not actively organize this information into a conceptual understanding of the product. At some later point in time, something may trigger the individual to organize this information about the advertised brand, and at this point, an attitude may be formed.

In an experiment which I am currently conducting, we find that subjects who executed a nonbrand processing strategy during exposure to a set of advertisements lose information more rapidly during a two week interval after exposure than subjects that executed a brand processing strategy during exposure to the same advertisements. This result also seems to be consistent with Krugman's conceptualization of what should occur under low involvement conditions.

Greenwald, Leavitt and Obermiller: Greenwald, et. al (1980) also define involvement as a process and suggest that there may be a "high involvement" process and two "low involvement" processes. Their "high involvement" process agrees with my conceptualization of information acquisition under conditions of "high involvement", however, we differ on the other two processes. The first "low involvement" process proposed by Greenwald, et. al (1980) seems to contain both of our information acquisition processes under conditions of "low involvement". We, however, believe that the distinction we make is an important one since, as discussed previously, these two processes result in differences in the amount, content and organization of the processed information in long term memory and may result in attitudinal differences.

The second "low involvement" process suggested by Greenwald, et. al (1980) is based on the "mere exposure"

literature (Zajonc 1980). This literature indicates that attitudes may be formed toward a stimulus even though the individual cannot determine, at better than chance levels, whether he or she has seen the stimuli (Wilson 1979). Greenwald, et. al (1980) refer to this as a pre-attentive process.

Although these ideas are interesting, they have yet to be demonstrated with complex as opposed to simple stimuli. Part of the problem with examining information acquisition under these conditions is defining exactly what non-attention means in terms of experimental manipulations. Kellogg (1980), for instance, has recently suggested that individuals can recognize unattended stimuli. This conclusion is based on the results from a series of experiments where subjects looked at pictures of faces while performing a complex multiplication task. Kellogg found that under these conditions his subjects could recognize the pictures at greater than chance levels, however, there was a decrement in recognition performance compared to when the subjects looked at the faces under conditions of full attention.

Problems arise in these types of studies, however, since individuals can acquire information through different channels. Research, for instance, has indicated that individuals have both a verbal and visual channel for acquiring information (e.g., Brooks 1968) and a single distractor task may not use up the capacity of both channels. Kellogg (1980), presents evidence that his complex multiplication task used up the capacity of the verbal channel, however, it is not clear whether his instructions to produce a mental image of the task used up the capacity of the visual channel.

In the experiment discussed previously, for instance, we found that when the advertisements contained visual information, our subjects formed different attitudes under the second attention level condition than a control group that did not see the advertisements (Mitchell 1980). This did not occur for advertisements that contained primarily verbal information. Consequently, in order to examine whether learning can occur under these conditions, we need to define more precisely what we mean by nonattention during exposure.

Issue Involvement

Concern with the effects of issue involvement on persuasion goes back at least as far as the research by Hovland and Janis on communication effects (Hovland, Janis and Kelley 1953). This early research led to the development of social judgment theory which provides an explanation of the effect of issue involvement on persuasion (Sherif, Sherif and Nebergall 1965). According to this theory, individuals that are highly involved with an issue will have large "latitudes of rejection" and will, therefore, be more likely to reject a persuasive message on this issue.

Recently, Petty and Cacioppo (1979) examined the effect of issue involvement on the number of counterarguments and support arguments generated during exposure to a persuasive communication. In two different experiments, Petty and Cacioppo (1979) manipulated issue involvement by presenting subjects with a persuasive communication advocating a change in coed visitation hours. In the high involvement condition, these changes were advocated for the student subjects' own university, while in the low involvement condition the change was advocated for another university. The other manipulations included the direction of change advocated and the strength of the arguments supporting the change. As might be expected, increased involvement enhanced persuasion for the proattitudinal message (i.e., increase visitation hours) and reduced persuasion for the counterattitudinal message (i.e.,

decrease visitation hours). Increasing involvement also enhanced persuasion for the message with the strong arguments even though the message was counterattitudinal, but reduced persuasion for the message with weak arguments. Perhaps one of the most interesting findings was that under low issue involvement conditions, the correlations between cognitive response measures (e.g., counterarguments) and attitudes were considerably lower than under conditions of high involvement (average absolute values of .29 vs. .69).

In terms of the model that was presented here, the topic of the message affected the subjects' involvement with the message. This, in turn, resulted in differences in attention levels and processing strategies. In each case, the authors found that low issue involvement resulted in a general reduction in the amount of evaluative processing (i.e., number of counterarguments and source arguments) that occurred. This indicates that the involvement manipulation used in these experiments resulted in information acquisition under our first low involvement condition.

As mentioned previously, they also found that the magnitude of the relationships between the number of counterarguments and support arguments and the resulting attitudes were greater under the high involvement manipulation than under the low involvement manipulation. It is interesting to note that in our study (Mitchell, Russo and Gardner 1980), we found that the magnitude of the relationship between a measure of attitude and attitudes predicted from the Fishbein model was greater with the brand processing strategy than with the nonbrand processing strategy. In other studies, we have generally found that the number of counterarguments and support arguments and predicted attitudes from the Fishbein model explained the same variance in independent measures of attitudes (Mitchell 1981). Therefore, it would seem that the type of semantic processing that occurs during exposure to a communication and the semantic information in memory after exposure explains more of the variance in attitudes when information is acquired under conditions of high involvement as opposed to low involvement. This suggests that either other mediators of attitude formation and change may be operating under these latter conditions or that the resulting attitudes may be less reliable.

Media Involvement

Krugman (1965, 1966) has suggested that, based on his definition of involvement, the media used in advertising a product determines the resulting level of involvement during exposure. According to Krugman, television advertising results in low involvement conditions while print results in high involvement conditions.

Wright (1974) tested this hypothesis by examining the effect of providing the same information both visually (print) and auditorily (radio) and found that his subjects generated fewer counterarguments and support arguments with the audio message as opposed to the visual message. Wright argued that these differences occurred because individuals do not have the opportunity to verbally respond to audio messages since they cannot control the rate at which they receive information. In other words, capacity constraints limit the amount of verbal processing that occurs during an audio message. It should be noted that if Krugman is correct in his hypothesis concerning the similarity in the type of processing that occurs during exposure to a television commercial and the processing that occurs during the learning of nonsense syllables, then Wright's explanation is consistent with the currently accepted explanation for the finding of primacy and recency effects in verbal learning experiments.

Much of this argument is based on how involvement is conceptualized. As mentioned previously, Krugman views involvement as a process -- essentially the type of processing that occurs during exposure to a communication. If, however, involvement is viewed as a state variable and only one of many variables that may affect the type of processing that occurs during exposure to an advertisement, it would be possible for someone to be highly involved with a television advertisement. However, because less time is available to process the information, he or she may execute a different type of processing than would occur if the same information is presented in print.

Brain Wave Activity and Involvement

In a provocative article, Krugman (1977) suggested that the processing of information from television advertisements may be primarily a right brain activity while the processing of information from print advertisements may be primarily a left brain activity. Based on this hypothesis, he also suggested that the primary effect of most television advertising may be the storing of visual images in the right hemisphere of the brain. Under these conditions, individuals would not necessarily be able to recall specific television commercials, however, individuals would be able to recognize whether or not they had seen a television commercial. Supposedly, according to Krugman, their recognition of a specific commercial may have an effect at the point of purchase.

To test this hypothesis, Appel, Weinstein and Weinstein (1979) measured the brain waves of subjects exposed to high recall and low recall television advertisements that were presented three times to each subject. The results indicated that advertisements with higher recall scores generated higher levels of brain wave activity (e.g., more seconds of alpha waves), however, there was no difference in the amount of brain wave activity between the left and right hemispheres. Krugman (1980), however, after re-examining the data, found brain wave activity in the left hemisphere declined with repetition, but brain wave activity in the right hemisphere remained constant with repetition. In a later study, Weinstein, Appel and Weinstein (1980) reported no difference in the amount of brain wave activity in the left and right hemispheres for subjects exposed to print and television advertisements. Magazine advertisements, however, generated a higher level of brain wave activity in both hemispheres as opposed to television advertisements.

These results seem to indicate that the amount of brain wave activity in each hemisphere is generally the same during exposure to advertisements in both television and print. There may, however, be differences after a number of repetitions. In addition, magazine advertisements seem to generate more brain wave activity. If there is a strong relationship between brain wave activity and the amount of cognitive response generated, and I suspect that there is, these latter results would tend to confirm the findings of Wright (1974). Cacioppo and Petty (1979), for instance, found a positive relationship between a number of physiological measures (e.g., oral muscle, cardiac and respiratory activity) and the number of counterarguments and support arguments generated.

Discussion

In this paper, I have presented a definition of involvement and a conceptual model of how involvement affects the information acquisition process. This definition, which defines involvement as an internal state variable whose motivational properties are evoked by a particular stimulus, differs from other definitions which define involvement in terms of the type of processing that occurs during exposure to an advertisement. Under the

conceptualization of involvement presented here, involvement is only one of many possible causes of these processes. Others are the amount of knowledge that an individual has about the topic of the communication and the opportunity to cognitively respond to the communication. In general, much of the confusion in the literature concerning involvement seems to center on different conceptualizations of involvement.

We have also suggested that different levels of involvement may cause three different types of information acquisition processes. One is caused by high involvement and two by low involvement. The issue involvement manipulation used by Petty and Cacioppo (1979) seems to have resulted in information acquisition under the first low involvement condition, while we suggest that Krugman (1965) is essentially discussing information acquisition under our second low involvement condition.

One interesting effect of involvement on the information acquisition process that has been found in two different experiments is a reduced relationship between semantic information and attitudes. Petty and Cacioppo (1979), found smaller correlations between counterarguments and attitudes and support arguments and attitudes under conditions of low issue involvement as opposed to high issue involvement. Mitchell, Russo and Gardner (1980) found smaller correlations between predictions from the Fishbein model and attitudes under a nonbrand processing strategy as opposed to a brand processing strategy. These lower correlations may occur for two different reasons. First, attitudes formed under conditions of low involvement may have lower reliability. This would result in lower observed correlations. Second, other mediators of attitude formation may have a greater effect under these conditions. Clearly, additional research is required to both verify this effect and, if valid, to determine its cause.

Finally, it should be noted, that the examination of factors affecting the information acquisition process can occur at different levels. Petty and Cacioppo (1979) and Wright (1974), for instance, examined the effects of message involvement and modality while Mitchell (1980) and Mitchell, Russo and Gardner (1980) examined the effects of different attention levels and processing strategies. Mitchell (1981) has referred to the latter manipulations as Level One variables affecting the information acquisition process while the former are called Level Two variables. Obviously, research at both levels is required to understand communication effects, however, it may be very difficult to obtain Model III information acquisition processes with Level Two variables in the laboratory. This may occur because knowledge that they are participating in a laboratory experiment may automatically raise the involvement level of subjects so that only Model I or Model II processes may occur.

References

- Appel, V., S. Weinstein and C. Weinstein (1979), "Brain Activity and Recall of TV Advertising," Journal of Advertising Research, 19, 7-15.
- Brooks, L. R. (1968), "Spacial and Verbal Components of the Act of Recall," Canadian Journal of Psychology, 22, 349-350.
- Cacioppo, J. T. and R. E. Petty (1979), "Attitudes and Cognitive Response: An Electrophysiological Approach," Journal of Personality and Social Psychology, 37, 2181-2199.

- Craik, F. I. M. (1979), "Levels of Processing: Overview and Closing Comments," in L. S. Cermak and F. I. M. Craik, eds., Levels of Processing in Human Memory, Hillsdale, N.J.: Lawrence Erlbaum Associates, 447-462.
- Craik, F. I. M. and R. S. Lockhart (1972), "Levels of Processing: A Framework for Memory Research," Journal of Verbal Learning and Verbal Behavior, 11, 671-684.
- Craik, F. I. M. and E. Tulving (1975), "Depth of Processing and the Retention of Words in Episodic Memory," Journal of Experimental Psychology: General, 104, 268-294.
- Gardner, M., A. A. Mitchell and J. E. Russo (1978), "Chronometric Analysis: An Introduction and an Application to Low Involvement Perception of Advertisements," in H. K. Hunt, ed., Advances in Consumer Research, Vol. 5, Ann Arbor: Association for Consumer Research, 581-589.
- Greenwald, A. G. (1968), "Cognitive Learning, Cognitive Response to Persuasion and Attitude Change," in A. G. A. G. Greenwald, et. al. eds., Psychological Foundations of Attitudes, New York: Academic Press, 147-170.
- Greenwald, A. G., C. Leavitt and C. Obermiller (1980), "What is Low Consumer Involvement?" Paper presented at the American Psychological Association Meeting in Montreal.
- Hovland, C., I. L. Janis and H. H. Kelley (1953), Communication and Persuasion, New Haven, Conn.: Yale University Press.
- Kahneman, D. (1973), Attention and Effort, Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Kellogg, R. T. (1980), "Is Conscious Attention Necessary for Long Term Storage?" Journal of Experimental Psychology: Human Learning and Memory, 6, 379-390.
- Krugman, H. E. (1980), "Sustained Viewing of Television," Journal of Advertising Research, 20, 65-68.
- Krugman, H. E. (1977), "Memory Without Recall, Exposure Without Perception," Journal of Advertising Research, 17, 7-12.
- Krugman, H. E. (1971), "Brain Wave Measures of Media Involvement," Journal of Advertising Research, 11, 3-10.
- Krugman, H. E. (1967), "The Measurement of Advertising Involvement," Public Opinion Quarterly, 30, 583-596.
- Krugman, H. E. (1965), "The Impact of Television Advertising: Learning Without Involvement," Public Opinion Quarterly, 29, 349-356.
- Mitchell, A. A. (1981), "Cognitive Processes Initiated by Exposure to Advertising," in R. Harris, ed., Information Processing Research in Advertising, Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Mitchell, A. A. (1980), "Using an Information Processing Approach to Understand Advertising Effects," in J. Olson, ed., Advances in Consumer Research, Vol. 7, Ann Arbor: Association for Consumer Research, 171-177.
- Mitchell, A. A. (1980), "The Relationship Between Cognitive Processes and Cognitive Structures," in I. Ross, ed., Proceedings, Division 23 Program, 87th Annual Convention, American Psychological Association, Vol. IX,
- Mitchell, A. A. (1979), "Involvement: A Potentially Important Mediator of Consumer Behavior," in W. Wilkie ed., Advances in Consumer Research, Vol. 6, Ann Arbor: Association for Consumer Research, 191-196.
- Mitchell, A. A., J. E. Russo and M. Gardner (1980), "Strategy Induced Low Involvement Processing of Advertising Messages," Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA. 15213.
- Nelson, D. L. (1979), "Remembering Pictures and Words: Appearance, Significance and Name," in L. S. Cermak and F. I. M. Craik, eds., Levels of Processing in Human Memory, Hillsdale, N.J.: Lawrence Erlbaum Associates, 45-76.
- Paivio, A. (1971), Imagery and Verbal Processes, New York: Holt, Rinehart and Winston.
- Petty, R. E. and J. T. Cacioppo (1979), "Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message - Relevant Cognitive Responses," Journal of Personality and Social Psychology, 37, 1915-1926.
- Rumelhart, D. E. and A. Ortony (1978), "The Representation of Knowledge in Memory," in R. C. Anderson, R. J. Spiro and W. E. Montague, eds., Schooling and the Acquisition of Knowledge, Hillsdale, N.J.: Lawrence Erlbaum Associates, 37-54.
- Sherif, C. W., M. Sherif and R. E. Nebergall (1965), Attitude and Attitude Change, Philadelphia, PA., Sanders.
- Weinstein, S., V. Appel and C. Weinstein (1980), "Brain-Activity Responses to Magazine and Television Advertising," Journal of Advertising Research, 20, 57-68.
- Wilson, W. R. (1979), "Feeling More Than We Know: Exposure Effects Without Learning," Journal of Personality and Social Psychology, 37, 811-821.
- Wright, P. L. (1980), "Message-Evoked Thoughts: Persuasion Research Using Thought Verbalizations," Journal of Consumer Research, 7, 151-175.
- Wright, P. L. (1974), "Analyzing Media Effects on Advertising Responses," Public Opinion Quarterly, 38, 192-205.
- Wright, P. L. (1973), "The Cognitive Processes Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-62.
- Zajonc, R. B. (1980), "Feeling and Thinking: Preferences Need No Inferences," American Psychologist, 35, 151-175.

LOW INVOLVEMENT: A SECOND LOOK

Harold H. Kassarian, University of California, Los Angeles¹

Abstract

Low involvement decision-making seriously challenges the cognitive orientation of present-day consumer research. However, product involvement may well be more complex than assumed thus far in that there may be an interaction effect with individual or personality characteristics. This paper proposes a six-fold classification of involvement including both high and low product involvement and also high and low involved personality types.

Introduction

In most fields of endeavor, new ideas are occasionally introduced which profoundly alter the prevailing status quo. In the physical sciences these may actually alter conceptions of physical reality - for example, the strange ideas of Copernicus thoroughly discredited the earlier views of philosophers and theologians alike. The systematic relationship of the ideas of men like Maxwell, Hertz and Einstein severely limited, and perhaps crippled, Aristotelian and Newtonian thinking. And in turn, what was to become known as field theory in physics had a profound effect on conceptualizations in biological and social sciences as earlier viewpoints crumbled.

In art and architecture, literature, humanities and music, similar turning points can be pointed out. For example, Elvis Presley and the Beatles altered musical conceptions as New Orleans Jazz had done earlier. Similarly in psychology, the contributions of Freud and his followers succeeded in destroying earlier views on the behavior of man as the unconscious and the impact of society on the mind were emphasized.

In less far-reaching areas of inquiry such as consumer behavior, similar patterns are evident. The elegant, yet simple, utility theory of economics was to give way to the psychoanalytically oriented motivation researchers. In turn, the clinical complexity of Freud was to make room for Hullian conceptions of learning theory. Even the idea of determinism began to be challenged with the introduction of mathematical learning theory into the field, culminating in Bass's (1974) proclamation that "man is substantially stochastic." And through the influence of social psychology and Bauer's (1964) claim of the "active audience," cognitive theory began to reign imperial - the notion, "... of a highly rational goal-striving, problem solving, information seeking, extracting, processing consumer ..." (Markin and Narayana, 1975). Topics such as attitude decision rules, as well as virtually every flow chart model of consumer decision-making have unquestioningly joined the "cognitive revolution" (McGuire 1976).

Perhaps it is too early to make the claim, but it may well come to pass that the cognitive view may also have to be modified. The concept of low-involvement might profoundly alter our present conceptions of the consumer.

Two Fields of Decision-Making

There is certainly no argument that under some conditions, at least for the purchase of some categories of goods and services, consumers do behave as information processing, problem solving, cognitive individuals reaching for a reasoned decision. These, of course, are the high involvement situations involving important, expensive, high risk, ego-related or value-expressive products. Actions and decisions are often highly sophisticated. And these decisions we have studied. We have analyzed, dissected, mapped and flow-charted the cognitive processes assumed to exist.

The problem is that the world of the consumer consists of more than highly involved decisions. In fact, very few of the hundreds of decisions made daily by the consumer are of an involved nature. In others, the low involvement decisions, the consumer unconcernedly purchases and consumes the product, tries new products, switches brands, obviously ignores promotional activities and worries about the important events in his or her life - the automobile's need for repairs, the children's grades in school, irritants at work or what have you.

In the last few years, numerous empirical and conceptual papers and a splendid review by Olshavsky and Grandbois (1979) clearly point out that the behavior of the consumer toward the high involvement, high risk, important product is simply different than toward the unimportant, low commitment, trivial product. Research in the future is simply going to have to take into account these two different sets or fields of decision making. For most types of consumer research the distinction will have to be made as to whether the products under study are of a high involvement nature or low involvement for the sub-set of respondents that comprise the subject pool. For example, Ray and his colleagues (1973) have proposed two hierarchies of effects, one for low involvement and another for high involvement situations.

Research on the hierarchy of effects in marketing communications is an excellent example of the fact that cognitive activity for low involvement and high involvement are simply different and that one cannot generalize research results from one situation to the other. Thus, research must measure the concept of involvement first, then turn to the hypotheses at hand. At this point it is unfortunate that a simple instrument or tool has not yet been developed to measure the concept of involvement but if in fact, "necessity is the mother of invention," that will come in time - for the measure of involvement is unquestionably a necessity - one that can no longer be ignored (See Hupfer & Gardner 1971).

Other topics in consumer research that may be seriously affected by the concept of low involvement are presented in Table 1.

The Involved Consumer

Segmenting products by the level of involvement although seemingly essential, may not fully account for all of the variance associated with the concept of high and low involvement. It may come to pass that we discover or personality variable related to the concept.

¹Written while visiting professor, Copenhagen School of Business Administration and Economics, Denmark.

Table 1

TOPICS HYPOTHESIZED TO BE RELATED TO LEVELS OF INVOLVEMENT

- Advertising Effectiveness
- Attitudes and Beliefs
- Attitude Change, Formation and Immunization
- Attribution Theory and Self Perception
- Brand Loyalty, Preference, Choice
- Cognitive Dissonance
- Cognitive Response
- Consumer Satisfaction
- Deceptive Advertising
- Decision Rules
- Family Decision Making
- Halo Effects
- Hierarchy of Effects
- Information Acquisition
- Information Processing
- Innovation and Diffusion
- Models of Consumer Behavior
- Perceived Risk
- Personal Influence
- Personality Correlates
- Persuasion
- Physiological Measures
- Political Marketing
- Pre-Purchase Anxiety
- Price and Packaging Research
- Reference Groups
- Repetition in Advertising
- etc.

It is undeniable that independent of the product class, there are some persons that tend to be more involved in the consumer decision process. They may be the addicted reader of *Consumer Reports*, those who pay greater attention to advertising and personal influence, and to the business and consumer sections of the newspaper. Some individuals may well be more price conscious, more alert to brand differences, generally more capable of discriminating quality differences, the more alert, the more conscious, the more interested and involved consumer. The category of person might include the wise housewife, the consumer activist, marketing people and professors of consumer behavior, among others

The Non-Involved Consumer

Equally evident is the fact there are other individuals who do not wish to concern themselves with marketing and buying activity. These might be individuals holding strong theoretical or aesthetic values who are simply oblivious to the practical affairs of the day to day world including marketing, advertising, consumption of goods, or to the accumulation of wealth. They have other interests and other concerns and just do not care and are uninvolved about consumer activities. This group might be termed detached individuals, those who put an emotional distance between themselves and the marketplace.²

It might also include the "know-nothings" of days past, the uneducated, and those who simply care very little about anything outside of their self-centered orbit. This group might very well be the utterly "no opinion" group on political polls, those who do not respond to mail questionnaires no matter what the issue, and refuse to try

²The term "detached type" seems particularly descriptive but does necessarily imply that it is similar to Horney's Detached Orientation (1937, 1945) nor necessarily measurable on Cohen's CAD Scale (1968).

free samples, try new products, or be interested in any of the issues that are of interest to this audience.

Hence the low involvement consumer consists of two groups. Those that are concerned about more significant problems and issues than consumer topics and those that can be best described as Know Nothings.

Moreover, in consumer research, an interaction effect between product involvement and individual involvement may exist suggesting perhaps a six-fold classification schema as presented in Table 2.

Table 2

CLASSIFICATION OF INVOLVEMENT

		Situation Effect or Product Involvement	
		High	Low
Individual or Personality Factors	High Involvement	Much of Consumer knowledge as it exists today	Typical low Involvement Research
	Low Involvement "Detached Type"	Minimal interest but narrowly and intensely focused	Oblivious to Product Issues Other Interests
	Low Involvement "Know-Nothing"	Choice determined by: availability packaging affordability	Don't know, Don't care and No opinion

The upper left hand cell - the high involvement personality, high involvement product group - consists of those people and behavior patterns that have been heavily studied in the past. Research on information processing, attitudes and consumer behavior models of decision making, is discussing this sub-set of consumers. They are the ones that fill out questionnaires, allow researchers to examine their behavior, and sit still for the numerous inane tasks that are required of them in experimental and descriptive research.

The upper right hand corner - the high involved consumers with low involved products - refers to the field of low involvement research as it has evolved to this point. Interested, concerned, cooperative subjects that have been presented with products in which they simply are not involved. The low involvement products may include a variety of consumer goods, politicians, causes, or a host of other objects and issues about which the individual simply is not concerned.

The detached individual with a highly involved product perhaps causes the greatest conceptual difficulty. He or she generally is unconcerned about the practical affairs of marketing and yet from time to time a product or issue may emerge which is of great importance. In this case, it is hypothesized that the embryonic interest, although perhaps temporarily intense, would be extremely narrowly focused. Once the politician is elected (or defeated), the issue is resolved or the product purchase

decision consummated, he returns to his basic state of apathy and detachment.

The lower left hand cell consists of the "know-nothings" who from time to time may be placed in a position where they simply must become involved in a product decision. Under these conditions the decision process probably is not the analytic, cognitive approach of the high involvement or detached type, but choice is determined by what is most easily available or whether or not one simply has enough money to pay for the object. The influence of attractive packaging or a glib salesman may be far more significant than a cognitive analysis of the product characteristics. "But it looked so pretty, and the man was no nice," may better describe the decision process than compensatory or lexographic decision rules.

The final group in the lower right hand corner can best be described by the terms, "don't know," "don't care," and "no opinion." This group is seldom, if ever, concerned about the affairs of the world - be it politics or canned spinach. Under a low involvement product condition, their contribution to consumer research primarily consists of filling the "no opinion" cells of a research design and contributing to the error term in any statistic.

Perhaps if research on low involvement is to be meaningful, the personality characteristic of involvement should be accounted for in research designs. Typically the "know-nothings" particularly in a low involvement product condition are naturally eliminated from research designs by their unavailability, but the differences between detached individuals with high involvement products and high involved persons with low involvement goods may be confounded in data analysis at present.

Space and time constraints do not allow for further elaboration at this point but such a personality - product involvement interaction effect seems quite conceivable and researchable, once tools or instruments are available for the measurement of product involvement and personality types.

Conclusion

If in fact involvement is an important topic, we should see its effects permeating research other than in the usual cognitive processing type of study. For example, the work of Kroeber-Riel and his colleagues on psychobiology seems particularly amenable and relevant (Kroeber-Riel 1979, 1980; see also Ryan 1980). Furthermore, the fascinating research being carried out by Flemming Hansen and his colleagues seems to be taking these types of distinctions into account (Hansen and Lundsgaard 1981, Hansen (In Press)). Hansen reports in his review of right-brain, left-brain lateralization research that in low involvement situations right-brain processes dominate whereas higher degrees of involvement give rise to left-brain processes. Thus, under high involvement conditions, information processing and deliberate choice-making occur. Under low involvement conditions information is received holistically and choices made without any high degree of awareness. He continues that under low involvement the number of exposures needed before a sufficient amount of learning has occurred may be greater when right-brain processes dominate.

In addition, there is some evidence to indicate that there may be individuals who are right-brain dominated and others who are left-brain dominated. The Hansen and Lundsgaard paper to be presented at this conference on Sunday morning presents the introductory report of their attempt to measure this variable with a paper-and-pencil test. If their stream of research is successful and a reliable-valid instrument emerges for measuring right or left brain domination, that approach would probably be superior to the

high-low involvement personality characteristic proposed here. In either case, it is becoming more and more evident that high and low involvement is not merely a product or situational characteristic but in some way involves the individual either on the molar level of overall disposition or on a more molecular or physiological level such as brain hemisphere research.

In summary, the simple concept of involvement off-handedly introduced by Krugman some years ago, may well qualify as one of the more important scientific ideas to emerge in consumer research in recent years. If I am correct, the topic should have an impact that will alter many if not most of our conceptions of consumer behavior models and our middle range theories and seriously challenge the supreme role of cognitive theory in our thinking.

References

- Bass, Frank M. (1974), "The Theory of Stochastic Preferences and Brand Switching," Journal of Marketing Research, 9 (February), 1-20.
- Bauer, Raymond A. (1964), "The Obstinate Audience," American Psychologist, 19 (May), 319-328.
- Cohen, Joel B. (1967), "An Interpersonal Orientation to the Study of Consumer Behavior," Journal of Marketing Research, 4 (August), 270-278.
- Hansen, Flemming (In Press), "Hemispherical Lateralization: A Review and a Discussion of Its Implications for Consumer Behavior Research," Journal of Consumer Research.
- _____ and Lundsgaard, Niels Erik (1981), "Developing an Instrument to Identify Individual Differences in the Processing of Pictorial and Other Non-Verbal Information," in Advances in Consumer Research: Vol. 8, (ed.) Kent Monroe, Ann Arbor: Association for Consumer Research, In Press this volume.
- Horney, Karen (1937), The Neurotic Personality of Our Time, New York: Norton.
- _____ (1945), Our Inner Conflicts, New York: Norton.
- Hupfer, Nancy T. and Gardner, David M. (1971), "Differential Involvement with Products and Issues: An Exploratory Study," in Proceedings 2nd Annual Conference of the Association for Consumer Research, (ed.) David M. Gardner, College Park, MD: Association for Consumer Research, 262-270.
- Kroeber-Riel, Werner (1979), "Activation Research: Psychobiological Approaches in Consumer Research," Journal of Consumer Research, 5 (March), 240-250.
- _____ (1980), "Rejoinder," Journal of Consumer Research, 7 (June), 96-98.
- Krugman, Herbert E. (1965), "The Impact of Television Advertising: Learning Without Involvement," Public Opinion Quarterly, 29 (Fall), 349-356.
- _____ (1977), "Memory Without Recall, Exposure Without Recognition," Journal of Advertising Research, 17, 7-12.
- Markin, Rom J. and Narayana, Chem (1975), "Behavior Control: Are Consumers Beyond Freedom and Dignity?" in Advances in Consumer Research: Vol. 3, (ed.) Beverlee B. Anderson, Ann Arbor: Association for Consumer Research, 222-228.

McGuire, William J. (1976), "Some Internal Psychological Factors Influencing Consumer Choice," Journal of Consumer Research, 2, 302-319.

Olshavsky, Richard W. and Grandbois, Donald H. (1979), "Consumer Decision Making-Fact or Fiction," Journal of Consumer Research, 6 (September), 93-100.

Ray, Michael L., Sawyer, Alan G. Rothschild, Michael L., Heeler, Roger M. and Reed, J. B., "Marketing Communications and the Hierarchy of Effects," in New Models for Mass Communications Research: Vol. II. Sage Annual Review of Communication Research, (ed.) Peter Clarke, Beverly Hills, CA: Sage Publications, Inc.

Ryan, Maichel J. (1980), "Psychobiology and Consumer Research: A Problem of Construct Validity," Journal of Consumer Research, 7 (June), 92-96.

ATTITUDE STRUCTURE AND SEARCH:
AN INTEGRATIVE MODEL OF IMPORTANCE-DIRECTED INFORMATION PROCESSING

Morris B. Holbrook, Columbia University
David A. Velez, Columbia University
Gerard J. Tabouret, Columbia University

Abstract

In spite of extensive research on multiattribute attitude models and information-acquisition paradigms, the linkage between these two important facets of buyer behavior is not yet well understood. Theory and previous studies suggest an integrative model of importance-directed information processing. This model proposes (1) that attribute importance is positively related to the extent and order of search and (2) that the resulting set of acquired cues serves as the basis for attitude. The results suggest both the feasibility of a questionnaire-based search task and the desirability of a more integrative approach to the study of attitude formation.

Introduction

Extensive progress has recently been made by consumer researchers in understanding the structure of attitudes that shape brand choices and in tracing the process by which information is acquired as a basis for preference formation. Too often, however, studies of attitude structure and information acquisition have proceeded independently. As a result, the connection between these two crucial phenomena has not yet been fully investigated.

More specifically, multiattribute attitude research has tended to focus primarily on the static structure of attitude while neglecting the process by which new informational inputs are incorporated into the determination of affect (Holbrook 1978). By contrast, research on information acquisition has typically dealt in great detail with the search process without systematically examining its presumed impact on brand evaluation (Holbrook and Maier 1978). In short, each approach has, in its own way, been somewhat one-sided. As a result, the linkage between attitude structure and information acquisition remains a vitally important area of consumer research with neglected potential for integration. Accordingly, the main purpose of the present paper is to propose and test a simple integrative model of information processing.

A Model Of Importance-Directed Information Processing

Importance and Search

The Principle of Parsimony (Haines 1974) argues that limited cognitive capacity forces problem solvers to adopt heuristics restricting their attention to the minimum amount of data necessary for satisficing decisions (Bettman 1974; Payne 1976; Wright 1973, 1974). Plausibly, one such heuristic might be to select cues bearing on those aspects of the decision regarded as most important. Indeed, such a search strategy would be consistent with the lexicographic decision model's assumption that information on various brand attributes will be processed in order of their perceived importance (Payne 1976; Wright 1975).

The precise meaning of "attribute importance" as a hypothetical construct is an issue that has been hotly debated, but not fully resolved by consumer researchers (Cohen, Fishbein, and Ahtola 1972; Fishbein 1976; Myers and Alpert 1977; Olson, Kanwar, and Muderrisoglu 1979; Ryan and Holbrook 1979). One clear consensus, however, is that perceived

importance must be distinguished from such related constructs as salience, determinance, and value-satisfaction or evaluative aspect. For the present purposes, the term "attribute importance" will be defined, in the sense suggested by Myers and Alpert (1977), as an attribute's "consequence or significance in making choices among brands or in forming overall evaluations or rankings of products" (p. 106).

When attribute importance is thus defined, its hypothesized role in directing information acquisition receives indirect support from several theoretical perspectives. It is reasonable, for example, to assume that an attribute's perceived importance will depend upon the range in relative magnitudes of the payoffs associated with its different possible levels. When uncertainty characterizes such outcomes, Bayesian decision theory suggests that search in problem solving will depend in part upon this degree of discrepancy among conditional payoffs (Edwards 1965). Similarly, Berlyne's (1960) theory of cognitive motivation relates the receiver's degree of exploratory behavior to the strength or "importance" (p. 36) of the competing response tendencies aroused by a cue. Such competing response tendencies (e.g., approach versus avoidance) engender conflict, which is relieved by specific exploration of the stimulus array. The stronger or more important the response tendencies, the more extensive the resulting specific exploration. An analogous concept in consumer research is the conflict that occurs when the consumer feels low self-confidence in choosing between alternatives with potentially serious consequences. The theory of perceived risk (Cox 1967) argues that, where such uncertain choices involve "important" consequences, increased search for information is one possible coping strategy for reducing the resulting level of risk back down to a tolerable level.

Such propositions from Bayesian, Berlynean, and perceived-risk theory have typically been applied by consumer researchers at the brand or product-class level. Nevertheless, they may reasonably be expected to pertain with equal force to the acquisition of information about a brand's attributes. Such an hypothesized relation between attribute importance and search priority was first tested by Tigert (1966) and has since received increased empirical support in several more recent studies (Heeler, Okechuku, and Reid 1979; Holbrook and Maier 1978; Sheluga, Jaccard, and Jacoby 1979; Quelch 1978). Indeed, decision-net researchers have sometimes actually defined attribute importance as the adjusted average rank of an attribute in the sequence of information seeking (Bettman 1974; Jacoby, Chestnut, Weigel, and Fisher 1976; Nakanishi and Bettman 1974).

Importance and Attitude Structure

Numerous studies have documented the failure of attribute-importance weights to improve upon the predictive performance of summative multiattribute attitude models (for reviews, see Bass and Wilkie 1973; Cohen, Fishbein, and Ahtola 1972; Lutz and Bettman 1977). A second implication of the Principle of Parsimony, however, is that attribute importance might plausibly play another kind of role in multiattribute models by serving as the basis for a parsimonious information-processing heuristic in which cues are combined in a relatively simple, energy-conserving way

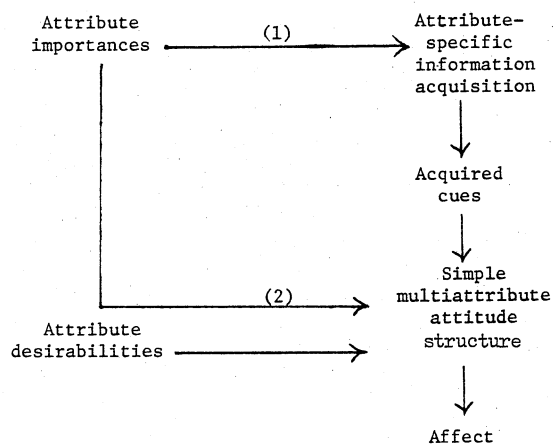
by restricting attention to a small set of most-important attributes. In this light, many studies have shown that a very simple model of attitude structure, summing only the scores on those attributes considered the few most important, predicts affect at least as well as more complex weighted forms of the multiattribute model (Bass and Wilkie 1973; Holbrook 1978; Holbrook and Hulbert 1975; Nakanishi and Bettman 1974; Wilkie, McCann, and Reibstein 1973). Attribute importance therefore appears to serve not so much as a weight in a complex multiattribute model, but rather as a factor directing the simplification of attitude structure.

An Integrative Model

The foregoing considerations may be summarized by the integrative model of importance-directed information processing shown in Figure 1.

FIGURE 1

A MODEL OF IMPORTANCE-DIRECTED INFORMATION PROCESSING



This model emphasizes the proposed role of attribute importance in linking information acquisition with attitude structure. According to Arrow (1), when confronted with the need to evaluate unfamiliar versions of some established product, the decision maker should seek information on those attributes viewed as most important. Arrow (2) then indicates that the perceived desirability values of the limited set of cues thus acquired should serve as the basis for appraising competing versions. Thus, the much-publicized debate concerning the role of attribute importance (e.g., Bass and Wilkie 1973) is resolved by a dynamic model hypothesizing that importance guides the acquisition of information that, in turn, determines the structure of attitude.

As indicated earlier, some empirical support already exists for the hypothesized relationships represented by Arrows (1) and (2). However, with few exceptions (e.g., Holbrook and Maier 1978; Sheluga et al. 1979), such support appears to be largely piecemeal in nature--focusing on one relationship or the other, but not on both at the same time. Thus, there is little firm evidence for the simultaneous operation of both arrows as part of a dynamic process in which importance guides the acquisition of cues that then compose the structure of attitude. Therefore, because this plausible, but inadequately tested integrative hypothesis appears to provide a useful way of linking research on information acquisition with that on attitude structure, the chief purpose of the present study was to provide a clearly-focused test of the importance-directed

model of information processing.

Methods for Studying Information Acquisition

Methods for studying the acquisition of information have recently been the topic for extended reviews, comparisons, and evaluations (Arch, Bettman and Kakkar 1978; Bettman 1977; Chestnut and Jacoby 1978; Jacoby, Chestnut, Hoyer, Sheluga and Donahue 1978; Payne and Ragsdale 1978; Russo 1978). Briefly, the most frequently advocated approaches for investigating search processes involve the use of: (1) eye-fixation tracking, in which mechanical instrumentation monitors the eye movements of subjects scanning some visual field; (2) verbal protocols, in which the subject provides a running account of his conscious thought processes while performing a decision task; or (3) information-display boards, in which the subject selects cards containing information from a multi-celled brands-by-attributes matrix.

One seldom-mentioned limitation of most such approaches is their implicit restriction to the confines of a laboratory or other closely monitored setting in which individual sessions are conducted with what, for cost reasons, usually turns out to be a small convenience sample. Given the need for marketing research on large, representative samples of consumers (Ferber 1977a), the standard repertoire of techniques for studying information acquisition may prove inadequate to the task of obtaining generalizable results. Toward this end, Holbrook and Maier (1978) introduced a modification of the information-display board intended to be usable in a mail questionnaire and, therefore, suitable for inclusion in studies employing large-scale survey techniques. Briefly, this procedure presents the respondent with a brands-by-attributes information-display sheet in which each cell is covered by a numbered sticker. The respondent acquires information by peeling stickers from the matrix and replaces them at the bottom of the page in a manner that provides a permanent record of the order in which they were removed. Holbrook and Maier (1978) found that this task was easily understood and performed by student subjects. A subsidiary purpose of the present study was to provide a further test of the usefulness of this questionnaire-based information-acquisition task.

Previous Findings Concerning the Importance-Directed Information-Processing Model

The Holbrook-Maier study. Using the sticker-pulling task in a study simulating the evaluation of phonograph records, Holbrook and Maier (1978) found evidence consistent with the importance-directed information-processing model. Across six attributes, there were highly significant mean intra-individual correlations between rated attribute importance and (1) number of cues selected on that attribute ($\bar{r} = .59, p < .0001$) and (2) average sequential order of selecting cues on that attribute ($\bar{r} = .63, p < .0001$). Using desirability ratings of the acquired cues in multi-attribute models to predict preference produced equally good intra-individual rank-order correlations whether additive ($\bar{r} = .57$ and $.56, p < .0001$) or averaging ($\bar{r} = .54$ and $.55, p < .0001$) models were employed. This finding was interpreted as exploratory evidence bearing on the debate between adding versus averaging formulations (Anderson 1967, 1976; Bettman, Capon and Lutz 1975; Fishbein 1976; Fishbein and Ajzen 1975; Lutz 1976; Troutman and Shanteau 1976). Apparently, when selective information acquisition is permitted rather than forcing exposure to factorially-designed stimuli, averaging models may not enjoy any predictive superiority over additive versions.

The Quelch study. Quelch (1978) produced support for the first Holbrook-Maier finding by asking housewives to choose among six brands of cold cereal after acquiring information on either four or five attributes. The mean intra-individual rank-order correlation between the order of information selection and a subsequent measure of attribute

The Høler study. Høler, Okechuku, and Reid (1979) contributed further support for the first Holbrook-Maier finding. Two groups of student subjects provided importance ratings or display-board search measures for 10 attributes of electric blenders. Rank-order correlation of aggregate scores across attributes found a moderately strong relationship between average rated importance and mean frequency of acquisition: $\bar{r} = .53$ ($p < .10$).

The Sheluga study. Finally, in a study performed concurrently with that reported in the present paper, Sheluga, Jaccard, and Jacoby (1979) extended Holbrook and Maier's (1978) "integrative approach" to include a comparison of several competing scaling methods. Though several of their findings concerning the evaluation of cameras are of considerable interest in their own right, the two most relevant to the issues addressed here are: (1) significant average Spearman correlations between retrospective "search importance" and extent ($\bar{r} = .80$, $p < .001$) and order ($\bar{r} = .68$, $p < .001$) of attribute-specific information acquisition; (2) relatively strong choice predictions (percentage correct) based upon acquired cues scored for utility by rating scales (64%), conjoint measurement (70%), and graded paired comparisons (100%).

Limitations in the Previous Findings

One limitation in Holbrook and Maier's (1978) design was the fact that questionnaires were distributed by hand, thus raising questions concerning the usefulness of the sticker-pulling technique in mail-survey research. A subsidiary objective of the present study, therefore, was to establish the feasibility of this method for use in surveys that are distributed through the postal system.

A second weakness in the Holbrook-Maier study raised a possible alternative hypothesis to explain their finding concerning the effect of attribute importance on search direction. Specifically, attributes were inadvertently arrayed on the information-display sheet from left to right roughly in order of their overall average perceived importance. In accord with the well-established effect of task format on performance in information-acquisition strategies (Bettman and Kakkar 1977), it might therefore be argued that the relationship between attribute importance and cue selection could have resulted from respondents' tendencies to move from left to right in searching for information (Quelch 1978). Accordingly, another subsidiary objective of the present study was to rule out this alternative explanation by randomly rotating the order of the attribute columns in the information-display sheet.

A further limitation in the Holbrook-Maier design raised doubts concerning the apparent equality of predictive performance by additive and averaging attitude models. Since respondents selected information by self-exposure, it is possible that they tended to acquire the same number of cues for each stimulus object. Such a naturally-occurring lack of variation in the extent of search among objects could account for the absence of predictive differences between additive and averaging models since the two kinds of formulation give identical results when the number of cues processed on each stimulus is constant (Anderson 1967; Fishbein and Ajzen 1975). The present study therefore attempted, as a third subsidiary objective, to encourage variation in the number of cues chosen on various stimulus objects by constraining the respondent to choose exactly 24 cues in evaluating 8 objects. Thus, if a number of cues other than three were chosen on any one stimulus, some variation in cue utilization among objects would be insured.

To maintain comparability with previous research using the sticker-pulling task, phonograph records were chosen as the product category to be tested. In accord with a suggestion by Ferber (1977b), the sample was composed of first year MBA students who demonstrated an interest in this product class by virtue of having purchased at least two pop or jazz recordings within the past six months. Questionnaires were mailed to 136 potential respondents, of whom 101 returned complete and usable sets of data for a response rate of 74%.

Evaluation Task

Respondents were asked to evaluate eight vocal recordings that differed on six alphabetically-listed attributes. Each attribute was defined by the following set of dichotomous characteristics: (1) Jacket--informative or visual; (2) Label--major or independent; (3) Singer's Style--traditional or contemporary; (4) Type of Music--pop or jazz; (5) Type of Production--studio or live; (6) Type of Songs--standards or originals and recent hits. These 6 attributes and 12 defining characteristics were described on the introductory page of the questionnaire by short paragraphs (omitted here for brevity) and were identical to those used by Holbrook and Maier (1978), except that Price in their study was replaced by Type of Music to avoid the somewhat ambiguous implications of price for preference formation.

Attribute-Specific Measures of Importance and Desirability

The six attributes were rated on 7-point check-mark scales for their degree of importance to the respondent's evaluation of a vocal recording from "not at all important" to "extremely important." In addition, the 12 defining characteristics were rated on comparable scales for their desirability in a vocal album from "extremely undesirable" to "extremely desirable." Importance ratings were scored from 1 to 7 while, in acknowledgement of the debate surrounding the proper coding of evaluative scores (Fishbein 1976; Holbrook 1977; Lutz 1976), two alternative codings of the desirability ratings were investigated: -3 to +3 and 1 to 7.

Test Objects

The dichotomous characteristics for each attribute were used to specify eight vocal albums according to the kind of fractional factorial design described by Green (1974) as suitable for conjoint measurement studies. The rows of the resulting 8 x 6 matrix were randomized and labelled from "Record A" to "Record H." To guard against the possibility of any position effects on the extent or order of cue acquisition (Bettman and Kakkar 1977; Quelch 1978), the columns of this basic information matrix were then arranged into 10 different random sequences. These 10 sequences were, in turn, assigned randomly to respondents.

The Information-Acquisition Task

As a basis for evaluating the vocal albums, the questionnaire provided an information-display sheet consisting of one of the 10 randomly-arrayed matrices with each cell covered by an appropriately labelled gummed sticker. The respondent was instructed to remove the 24 stickers that would best provide the information needed to rank the eight records in order of preference. Further instructions asked the respondent to replace each sticker, in its order of removal, into another matrix entitled "Order of Sticker Removal" with cells labelled from "1st" to "24th." Following this task, the respondent ranked the eight recordings in order of preference.

Operational Definitions

Extent and order of attribute-specific information acquisition. The extent of attribute-specific information acquisition was defined operationally as the number of cues acquired across all eight records on that attribute. The order of acquisition was defined as the average rank of selecting information on that attribute across the eight records (with the rank order for non-acquired cues scored arbitrarily as 36.5—i.e., half way between 25th and 48th.)

Models of attitude structure. The full set of competing attitude-structure models developed by Holbrook and Maier (1978) was tested in the present study. Specifically, the comparative predictive performance of additive and averaging models was explored, both with and without importance weights and with desirability scored both from -3 to +3 and from 1 to 7. In accord with the previous findings, however, no discernible differences arose from the inclusion of importance weights or from the alternative scorings of desirability. To save space, the present discussion will therefore be confined to the most conceptually appropriate attitude models—namely, those without importance weights and with desirability scored from -3 to +3 (Fishbein 1976; Fishbein and Ajzen 1975; cf. Rosenberg 1956). With this restriction, the competing models investigated may be represented as follows:

$$\text{Partial Additive Model} = \sum_{k=1}^6 B_k \cdot E_k \cdot ACQ_k \quad (1)$$

$$\text{Partial Averaging Model} = \left(\sum_{k=1}^6 B_k \cdot E_k \cdot ACQ_k \right) / \left(\sum_{k=1}^6 ACQ_k \right) \quad (2)$$

$$\text{Full Additive Model} = \sum_{k=1}^{12} B_k \cdot E_k \cdot ACQ_k \quad (3)$$

$$\text{Full Averaging Model} = \left(\sum_{k=1}^{12} B_k \cdot E_k \cdot ACQ_k \right) / \left(\sum_{k=1}^{12} ACQ_k \right) \quad (4)$$

where k refers to the 12 characteristics that a given record does ($k = 1, \dots, 6$) and does not ($k = 7, \dots, 12$) possess; $B_k = 1$ if the record possesses characteristic k or $B_k = -1$ if it does not; E_k is the respondent's evaluation of the desirability of characteristic k ; and ACQ_k is a zero-one dummy variable representing the acquisition of information on characteristic k .

More colloquially, the Partial Additive Model may be described as the summed desirability scores of those characteristics known, after search, to be present in a given record album (where it is assumed that the respondent believes the information he has received to be true). The Partial Averaging Model is the mean desirability of those known characteristics. The Full Additive Model further subtracts the desirabilities of those characteristics revealed to be absent from a given recording (on the assumption that the respondent makes inferences about what characteristics are missing from the record). And the Full Averaging Model computes the mean desirability of the present characteristics less that of those absent.

The full models assume that preference may be enhanced (reduced) by the absence of undesirable (desirable) characteristics, thereby incorporating logic analogous to that arguing for the bipolar coding of components in the Fishbein model (Fishbein 1976; Fishbein and Ajzen 1975; cf. Rosenberg 1956). Even these fuller versions of the attitude model do not, of course, account for the effects of additional beliefs that may be inferred on the basis of the information that has been acquired (Fishbein and Ajzen 1975; Lutz and Swasy 1977; Olson 1978). This, of

course, is a limitation that is shared by the vast preponderance of attitude research.

In comparing the additive and averaging models, it is important to recognize that any differences between them in the present design result from the fact that respondents are free to select different numbers of cues for different vocal albums. Where the variation among records in the extent of acquisition is relatively large, the two types of formulation may make substantially different preference-rank predictions. The degree of such differences is an empirical issue, however, and not part of the research design controlled by the investigators themselves. Thus, this part of the study is correlational rather than experimental in nature.

Operational Hypotheses

Given the arguments and findings reviewed earlier, four specific hypotheses were tested:

- H₁. Extent and order of attribute-specific information acquisition are positively related to attribute importance on an intra-individual basis;
- H₂. Attitude-structure models based on the cues acquired predict intra-individual preference ranks;
- H₃. The partial attitude models perform at least as well as the full models in predicting preference ranks;
- H₄. The predictive performance of additive attitude models is at least as good as that of averaging models.

Results

In accord with the findings of Holbrook and Maier (1978), respondents appeared to experience few problems with the self-administration of the mail questionnaire. Only three questionnaires had to be discarded because of failures to follow the written instructions. The likely effect of such difficulties on the results obtained in an information-acquisition task therefore appears to be minimal.

H₁. The first hypothesis was supported by highly significant mean intra-individual correlations of attribute importance with (1) extent of information acquisition ($\bar{r} = .665$, $Z_{\bar{r}} = 21.08$, $p < .0001$) and (2) order of information acquisition ($\bar{r} = .674$, $Z_{\bar{r}} = 22.81$, $p < .0001$).

Because of the presence of a small number of low-correlation outliers, the median intra-individual correlations were considerably stronger than the means: .775 and .782 respectively. These results compare favorably with the previously cited findings.

H₂. The results for all four versions of the attitude-structure model supported the second hypothesis by performing fairly well in predicting preference rank, with mean intra-individual rank-order correlations as follows: Partial Additive Model— $\bar{r} = .642$ ($Z = 19.63$, $p < .0001$); Partial Averaging Model— $\bar{r} = .602$ ($Z = 16.91$, $p < .0001$); Full Additive Model— $\bar{r} = .598$ ($Z = 7.73$, $p < .0001$); Full Averaging Model— $\bar{r} = .597$ ($Z = 16.86$, $p < .0001$). Because of skewed distributions, the median correlations were again somewhat stronger than the means, ranging from .673 to .756. These relationships were, if anything, somewhat better than those found by Holbrook and Maier.

H₃. In further accord with the Holbrook-Maier results, the third hypothesis was supported by the failure of either full attitude-structure model to perform better than its corresponding partial version. This finding contradicts the assertion that affect is determined, in part, by the

evaluative aspects of those characteristics that are believed to be missing from some object.

H_4 . Finally, in line with previous findings and the fourth hypothesis, neither averaging model improved upon the predictive performance of its additive counterpart. It appears, however, that constraining respondents to the selection of 24 pieces of information was only moderately successful in encouraging variation in the number of attribute cues acquired across stimulus objects. Thus, 49 out of 101 respondents showed no variation in the number of cues selected across recordings. As pointed out earlier, for such respondents, additive and averaging attitude models make identical preference predictions. The removal of these 49 respondents from the sample would only result, however, in an even greater relative superiority in the predictive performance of the Partial Additive Model versus the Partial Averaging Model ($\bar{r} = .622$ versus $.542$). Such a procedure, therefore, would not change the basic conclusion that averaging models do not improve upon the predictive performance of additive versions. Accordingly, the more conservative comparison, based on the full sample of 101 respondents, should be used in summarizing the results for H_3 and H_4 .

In this light, the findings concerning H_3 and H_4 may be summarized by comparing the predictive performance of the Partial Additive Model with that of the other three versions. In each case, the Partial Additive Model performed slightly, but significantly better in predicting preference: $\bar{r} = .642$ versus $.602$ ($t_{100} = 3.09$, $p < .01$); $\bar{r} = .642$ versus $.598$ ($t_{100} = 3.02$, $p < .01$); and $\bar{r} = .642$ versus $.597$ ($t_{100} = 2.89$, $p < .01$). It therefore appears that, in the kind of information-acquisition task investigated here, the Partial Additive Model is to be preferred over the full and/or averaging versions--partly because of its marginal predictive superiority, but primarily because of its greater parsimony.

Discussion

Limitations

Like most presently available research on attitude formation or information acquisition, the present study is characterized by certain limitations. The attitude model, for example, required the plausible, but untested assumption that the respondents believed the information to which they were exposed and, in effect, assigned each cue a subjective likelihood of 1.0. Also, the model could not account for the effects of any additional beliefs that might have been inferred on the basis of the cues acquired (Fishbein and Ajzen 1975; Lutz and Swasy 1977; Olson 1978). In these respects, however, the present approach shares assumptions that are fundamental to the vast preponderance of attitude research, including most of the work on conjoint analysis and information integration. Moreover, the importance of inferential beliefs in this study is called into question by the absence of any predictive improvement to be gained by including characteristics implicitly revealed to be absent from the recording (i.e., $E_k = -1$).

In common with other information-acquisition paradigms (eye-fixation tracking, display boards, protocols, etc.), the present search task was highly artificial in many respects. Restricting respondents to the use of 24 cues, for example, was one way of setting a reasonable limit on the amount of information acquired so as to permit some variance in cue utilization across attributes. Other methods for limiting cue selection--such as charging five cents per cue--would have been less practical in a mail questionnaire. Moreover, it appears doubtful whether any one way of restricting information intake is inherently more realistic than another.

In contrast to many applications of the information-display board, the cues uncovered by respondents in the present study remained visible and did not, therefore, place any demands on memory capacity. The authors would argue that this procedure has the advantage of removing the confounding effects of memory overload, but some might feel that this constitutes a departure from realism.

Further, the recordings employed in the present study were hypothetical rather than real, though respondents were not informed of the fictitious nature. The advantage of using such artificial products is that it permits an otherwise unattainable degree of experimental control. Accordingly, this kind of artificiality appears in numerous other studies of information processing--including many of those on cognitive algebra (e.g., Bettman, Capon, and Lutz 1975), conjoint measurement (e.g., Green 1974), and attitude change (e.g., Lutz 1975).

Finally, it should be noted that--though the approach used here is of some use in comparing the predictive efficacy of simple additive versus averaging models (Fishbein and Ajzen 1975)--it cannot discriminate between an additive model and the more elaborate weighted averaging model proposed by Anderson (1967) to take account of the "set size" effect. This latter formulation assumes that affect begins at an initial value that is then adjusted by the averaging-in of subsequent information. Such a model produces predictions that are quite similar to those generated by a simple additive representation. As pointed out by Fishbein and Ajzen (1975, p. 242), the inability to distinguish between a simple additive model and Anderson's (1967) differentially weighted averaging version is a limitation shared by most of the research in this area.

Conclusions

Subject to such limitations, the present study supports the proposed model of importance-directed information processing. Specifically, its results suggest (1) that attribute importance guides the acquisition of attribute-related cues and (2) that these cues are then incorporated into a relatively simple importance-shaped attitude structure so as to determine affect. Though both these effects have been supported independently by earlier studies, previous research has tended not to treat them together as parts of one dynamic importance-directed process. The present findings therefore suggest the usefulness of research attempting to integrate information-acquisition phenomena with the nature of multiattribute attitude structure.

In addition, the present study continues to support the adequacy of the relatively parsimonious Partial Additive Model when compared with more elaborate competing versions. To be sure, the research design left the respondent free to select among the six cues available for each object instead of being exposed to factorially-arranged combinations of cues so that the comparative tests of predictive performance are correlational rather than experimental in nature. Moreover, like most other research on consumer attitudes, the present study failed to account for the effects of additional beliefs that might be inferred from those explicitly manipulated and/or measured. Nevertheless, within this restricted framework, the findings may further reinforce the doubts of those questioning the often-claimed superiority of the averaging model (Fishbein 1976; Fishbein and Ajzen 1975; cf. Anderson 1967, 1976; Bettman, Capon, and Lutz 1975; Lutz 1976; Troutman and Shanteau 1976).

Finally, the study confirms the feasibility of the sticker-removal task as a device for incorporating the information-acquisition paradigm into mail-survey research designs. Though the respondents in the present study were students, the authors would anticipate no difficulties in applying the technique in mail questionnaires directed at larger scale, more representative samples. Obviously, the external

validity of such an approach remains to be tested. But, on the basis of the evidence presented here, it appears that the label-pulling task may be useful in future attempts to generalize findings concerning information acquisition to the overall consumer population.

References

- Anderson, Norman H. (1967), "Averaging Model Analysis of Set-Size Effect in Impression Formation," Journal of Experimental Psychology, 75, 158-65.
- _____ (1976), "How Functional Measurement Can Yield Validated Interval Scales of Mental Quantities," Journal of Applied Psychology, 61, 677-92.
- Arch, David C., Bettman, James R., and Kakkar, Pradeep (1978), "Subjects' Information Processing in Information Display Board Studies," Proceedings, Association for Consumer Research, Eighth Annual Conference, 555-60.
- Bass, Frank M., and Wilkie, William L. (1973), "A Comparative Analysis of Attitudinal Predictions of Brand Preference," Journal of Marketing Research, 10, 262-9.
- Berlyne, Daniel E. (1960), Conflict, Arousal, and Curiosity, New York: McGraw-Hill.
- Bettman, James R. (1974), "Toward a Statistics for Consumer Decision Net Models," Journal of Consumer Research, 1, 71-80.
- _____ (1977), "Data Collection and Analysis Approaches for Studying Consumer Information Processing," Proceedings, Association for Consumer Research, Seventh Annual Conference, 342-8.
- _____, Capon, Noel, and Lutz, Richard J. (1975), "Multiattribute Measurement Models and Multiattribute Attitude Theory: A Test of Construct Validity," Journal of Consumer Research, 1, 1-14.
- _____ and Kakkar, Pradeep (1977), "Effects of Information Presentation Format on Consumer Information Acquisition Strategies," Journal of Consumer Research, 3, 233-40.
- Chestnut, Robert W., and Jacoby, Jacob (1978) "Methods and Concepts in Consumer Information Processing: Toward an Integrated Framework," Research Paper No. 141a, Columbia University.
- Cohen, Joel B., Fishbein, Martin, and Ahola, Olli T. (1972), "The Nature and Uses of Expectancy-Value Models in Consumer Attitude Research," Journal of Marketing Research, 9, 456-60.
- Cox, Donald F., ed. (1967), Risk Taking and Information Handling in Consumer Behavior, Cambridge: Harvard University Graduate School of Business Administration.
- Edwards, Ward (1965), "Optimal Strategies for Seeking Information: Models for Statistics, Choice Reaction Times, and Human Information Processing," Journal of Mathematical Psychology, 2, 312-29.
- Ferber, Robert (1977a), "Research By Convenience," Journal of Consumer Research, 4, 57-8.
- _____ (1977b), "What is the JCR Editorial Policy on Samples and Sample Requirements?" Newsletter, Association for Consumer Research, 7, No. 4, 16.
- Fishbein, Martin (1976), "Extending the Extended Model," Proceedings, Association for Consumer Research, Sixth Annual Conference, 491-7.
- Fishbein, Martin and Ajzen, Icek (1975), Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading, Mass.: Addison-Wesley.
- Green, Paul E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," Journal of Consumer Research, 1, 61-8.
- Haines, George H., Jr. (1974), "Process Models of Consumer Decision Making," in Buyer/Consumer Information Processing, ed. G. David Hughes and Michael L. Ray, Chapel Hill: University of North Carolina Press.
- Heeler, Roger M., Okechuku, Chike, and Reid, Stan (1979), "Attribute Importance: Contrasting Measurements," Journal of Marketing Research, 16, 60-3.
- Holbrook, Morris B. (1977), "Comparing Multiattribute Attitude Models By Optimal Scaling," Journal of Consumer Research, 4, 165-71.
- _____ (1978), "Beyond Attitude Structure: Toward the Informational Determinants of Attitude," Journal of Marketing Research, 15, 545-56.
- _____ and Hulbert, James M. (1975), "Multi-Attribute Attitude Models: A Comparative Analysis," Proceedings, Association for Consumer Research, Fifth Annual Conference, 375-88.
- _____ and Maier, Karl A. (1978), "A Study of the Interface Between Attitude Structure and Information Acquisition Using a Questionnaire-Based Information-Display Sheet," Proceedings, Association for Consumer Research, Eighth Annual Conference, 93-8.
- Jacoby, Jacob, Chestnut, Robert W., Hoyer, Wayne D., Sheluga, David A., and Donahue, Michael J. (1978), "Psychometric Characteristics of Behavioral Process Data: Preliminary Findings on Validity and Reliability," Proceedings, Association for Consumer Research, Eighth Annual Conference, 546-54.
- _____, Chestnut, Robert W., Weigl, Karl C., and Fisher, William A. (1976), "Pre-Purchase Information Acquisition: Description of a Process Methodology, Research Paradigm, and Pilot Investigation," Proceedings, Association for Consumer Research, Sixth Annual Conference, 306-14.
- Lutz, Richard J. (1975), "Changing Brand Attitudes Through Modification of Cognitive Structure," Journal of Consumer Research, 1, 49-59.
- _____ (1976), "Conceptual and Operational Issues in the Extended Fishbein Model," Proceedings, Association for Consumer Research, Sixth Annual Conference, 469-76.
- _____, and Bettman, James R. (1977), "Multi-attribute Models in Marketing: A Bicentennial Review," in Consumer and Industrial Buyer Behavior, ed. Arch G. Woodside, Jagdish N. Sheth, and Peter D. Bennett, New York: North-Holland.
- _____ and Swasy, John L. (1977), "Integrating Cognitive Structure and Cognitive Response Approaches to Monitoring Communications Effects," Proceedings, Association for Consumer Research, Seventh Annual Conference, 363-71.
- Myers, James H., and Alpert, Mark I. (1977), "Semantic Confusion in Attitude Research: Salience Vs. Importance Vs. Determinance," Proceedings, Association for Consumer Research, Seventh Annual Conference, 106-10.

Nakanishi, Masao, and Bettman, James R. (1974), "Attitude Models Revisited: An Individual Level Analysis," Journal of Consumer Research 1, 16-21.

Olson, Jerry (1978), "Inferential Belief Formation in the Cue Utilization Process," Proceedings, Association for Consumer Research, Eighth Annual Conference, 706-13.

_____, Kanwar, Rajesh, and Muderrisoglu, Aydin (1979), "Clarifying the Confusion Regarding Salience, Importance, and Determinance Concepts in Multiattribute Attitude Research," Working Paper, The Pennsylvania State University.

Payne, John W. (1976), "Heuristic Search Processes in Decision Making," Proceedings, Association for Consumer Research, Sixth Annual Conference, 321-7.

_____ and Ragsdale, E.K. Easton (1978), "Verbal Protocols and Direct Observation of Supermarket Shopping Behavior: Some Findings and a Discussion of Methods," Proceedings, Association for Consumer Research, Eighth Annual Conference, 571-7.

Quelch, John A. (1978), "Behavioral Measurement of the Relative Importance of Product Attributes: Process Methodology and Pilot Application," Working Paper, University of Western Ontario.

Rosenberg, Milton J. (1956), "Cognitive Structure and Attitudinal Affect," Journal of Abnormal and Social Psychology, 53, 367-72.

Russo, J. Edward (1978), "Eye Fixations Can Save the World: A Critical Evaluation and a Comparison Between Eye Fixations and Other Information Processing Methodologies," Proceedings, Association for Consumer Research, Eighth Annual Conference, 561-70.

Ryan, Michael J., and Holbrook, Morris B. (1979), "On the Relationship of Importance, Elicitation Order, and Expectancy X Value," Working Paper, Columbia University.

Sheluga, David A., Jaccard, James, and Jacoby, Jacob (1979) "Preference, Search, and Choice: An Integrative Approach," Journal of Consumer Research, 6, 166-76.

Tigert, Douglas J. (1966), "An Investigation of Experimental Techniques to Determine the Relative Importance of Various Product Attributes in the Consumer Decision-Making Process," Working Paper, Krannert Graduate School of Industrial Administration, Purdue University.

Troutman, C. Michael, and Shanteau, James (1976), "Do Consumers Evaluate Products by Adding or Averaging Attribute Information?," Journal of Consumer Research, 3, 101-12.

Wilkie, William L., McCann, John M., and Reibstein, David J. (1973), "Halo Effects in Brand Belief Measurement: Implications for Attitude Model Development," Proceedings, Association for Consumer Research, Fourth Annual Conference, 280-90.

Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-62.

_____ (1974), "The Harassed Decision Maker: Time Pressures, Distractions, and the Use of Evidence," Journal of Applied Psychology, 59, 555-61.

_____ (1975), "Consumer Choice Strategies: Simplifying Vs. Optimizing," Journal of Marketing Research, 12, 60-7.

EXAMINING THE DIAGNOSTIC UTILITY OF THE
FISHBEIN BEHAVIORAL INTENTIONS MODEL¹

Paul W. Miniard, The Ohio State University

Abstract

Evidence and issues relevant to the diagnostic utility of the Fishbein Behavioral Intentions model are discussed. Two experiments are then presented that examine the degree to which the weight associated with the normative component of the model can accurately reflect the absence or presence of salient normative influences.

Introduction

One promising approach to understanding the important motivations underlying a given behavior is represented by the Fishbein Behavioral Intentions (FBI) model. On the basis of a predictive criterion, the model's performance has been substantiated for a variety of behavioral objects such as alcohol (Schlegel, Crawford, and Sanborn 1977), birth control pills (Davidson and Jaccard 1975), female occupations (Greenstein, Miller, and Weldon 1979), financial loans (Ryan and Bonfield 1980), marijuana (Bearden and Woodside 1978), swine flue vaccinations (Oliver and Berger 1979), and toothpaste (Wilson, Mathews, and Harvey 1975). However, the prediction of intention alone offers little pragmatic value. Rather, the promise of any such formulation lies in its potential for generating information that is useful in developing behavioral change strategies. But before the model can be confidently employed for diagnostic purposes, attention should be directed at examining the model's diagnostic utility; that is, the extent to which the model leads to correct inferences about the determinants of intention and behavior. It is to this issue that this paper addresses itself. After briefly reviewing the FBI formulation, discussion turns to the possible threats to the model's diagnostic utility. Two experiments are then presented that examine the model's diagnostic accuracy under conditions that differ in the salience of normative influence.

The Fishbein Behavioral Intentions Model

The FBI model postulates that intention, viewed as the immediate antecedent of behavior, is determined by an attitudinal or personal component and a normative or social component (Fishbein and Ajzen 1975):

$$B \sim BI = w_1 (A_B) + w_2 (SN), \quad (1)$$

where B is the behavior, BI is the behavioral intention to perform behavior B; A_B is the attitude toward performing behavior B; SN is the subjective norm; and w_1 and w_2 are empirically determined weights.

The first component, A_B , is the person's attitude toward performing a particular behavior under a given set of circumstances. This attitudinal component can be decomposed into the perceived consequences of performing the behavior weighted by the evaluation of these consequences:

$$A_B = \sum_{i=1}^n b_i e_i, \quad (2)$$

where b is the belief that performing behavior B leads to consequence i; e is the person's evaluation of consequence i; and n is the number of salient beliefs.

The second component, SN, represents the person's perception of what important others think s/he should do which is proposed to be a function of the person's beliefs about the expectations of important referent others, and her/his motivation to comply with these referents:

$$SN = \sum_{j=1}^n NB_j MC_j \quad (3)$$

where NB is the normative belief (i.e., the person's belief that reference group or individual j thinks s/he should or should not perform the behavior); MC is the person's motivation to comply with referent j; and n is the number of relevant referents.

The importance of these two components in determining intentions is expected to vary with the behavior, the situation, and individual differences between persons. Some individuals, for example, may be more "sensitive" to social demands and therefore may attach more weight to their normative considerations than other individuals. Similarly, behaviors that carry greater interpersonal significance or that are visible to important others may be more susceptible to social influences than behaviors that have less interpersonal implications or that are less observable. The component weights (i.e., w_1 and w_2), which are traditionally estimated by multiple regression procedures, presumably reflect both the existence and the relative importance of these two determinants.

Criteria for Establishing the Model's Diagnostic Utility

Before any model that attempts to identify the determinants of behavioral intention can be confidently employed for diagnostic purposes, there are several important criteria that the model should meet. First, a model that decomposes the determinants of intention is useful only to the extent that there exists a strong relationship between intentions and behavior. Second, the causal flow hypothesized by the model should be validated. Third, the model should be fully identified in the sense that exogeneous variables (i.e., variables other than those postulated by the model as the immediate antecedents of intention) will affect intentions only indirectly (i.e., the exogeneous variable's influence on intention is mediated by one or more of the postulated determinants). Fourth, the weights employed for representing the presence and relative importance of the hypothesized determinants of intention should accurately reflect "reality". Each of these criteria are in turn discussed below with a brief review of the research relevant to each.

The Intention-Behavior Relationship

One major assumption of the FBI model is that "behavioral intentions are the immediate determinants of the corresponding overt behaviors" (Fishbein and Ajzen 1975, p. 372). If behaviors are not predictable from intentions, there

¹Appreciation is extended to Joel B. Cohen for his assistance in the development of the experiments presented in this paper.

would be little value in specifying and validating a model that identifies the determinants of intentions. Thus, establishing the existence of an intention-behavior relationship as well as the factors that influence the magnitude of this relationship is of crucial importance.

Much of the work relevant to this issue is reviewed by Fishbein and Ajzen (1975, pp. 372-381) and thus need not be repeated here. In general, the assumption of an intention-behavior relationship has been supported, although Fishbein and Ajzen (1975, pp. 368-372) do point out that this relationship is dependent upon (1) the degree to which the measure of intention corresponds directly to the observed behavior, (2) the time interval between measurement of intention and the behavior's occurrence, and (3) the degree to which the person is able to act in accordance with his intention or without the assistance of others. A strong relationship between intention and behavior is expected when the intention measure is specific to the behavior of interest, performance of the behavior is temporally close to the measurement of intention, and the behavior is under volitional control.

Verifying the Hypothesized Causal Flow

As implied by Equations 1, 2, and 3, the FBI model postulates that (1) the influence of A_B and SN upon B are mediated by BI, (2) the influence of Σb_{ie_j} upon BI is mediated by A_B , and (3) the influence of $\Sigma NB_j MC_j$ upon BI is mediated by SN. For the most part, this hypothesized causal flow has been substantiated. In an examination of BI's mediating role, Ajzen and Fishbein (1970) held BI constant and tested the partials between B- A_B and B-NB. This procedure completely eliminated the previously significant B-NB relationship. The B- A_B partial correlation was significant, however, in only one of the two statistical tests. Similarly, Ryan and Bonfield (1980) found that initially significant B- Σb_{ie_j} and B- $\Sigma NB_j MC_j$ relationships became insignificant when BI was partialled out. Further support is provided by Ajzen and Fishbein (1974) where changes in BI were related to changes in A_B and $\Sigma NB_j MC_j$. In a product evaluation setting, Lutz (1977) tested the causal relationships assumed by the FBI model and concluded that the results provided "reasonably strong support for the postulated flow of effects" (p. 206), although several conceptual and methodological concerns regarding the study and the conclusions drawn have been raised (cf. Carnegie-Mellon University Marketing Seminar 1978; Dickson and Miniard 1978; Lutz 1978a,b).

Several investigations have failed to support the postulated causal flow. Miniard and Cohen (1981) found that, contrary to the causal flow involving the normative component, the addition of $\Sigma NB_j MC_j$ into a regression model containing SN provided a significant improvement in the prediction of BI. This finding casts doubt upon SN's ability to mediate the effects of $\Sigma NB_j MC_j$ upon BI. Further, Bentler and Speckart (1979) report that attitudes and past behavior accounted for a significant portion of behavior that was not mediated by intention.

The Model's Mediation Adequacy

One assumption underlying the model that has received considerable attention is that exogeneous variables will affect intentions only indirectly through either component of the model. Thus, any variable related to intentions should also be related to one of the model's components. Further, any relationship between such variables and intentions should be eliminated when A_B and SN are statistically held constant (e.g., the addition of an external variable should not provide a significant increment in the prediction of intentions).

Several approaches have been adopted for operationalizing these external variables. Survey-based investigations have examined the model's ability to mediate such

exogeneous variables as religion, age, occupational prestige, and assorted personality and social constructs. In an early study involving transplant donations, Schwartz and Tessler (1972) found a subset of these external variables increased the model's ability to predict intentions. Later investigations, however, have supported the model's mediational adequacy. Jaccard and Davidson (1975) concluded from their study of family planning that the addition of external variables did not provide substantive predictive improvements. Similarly, in the realm of adolescent alcohol use, Schlegel, Crawford, and Sanborn (1977) stated that "...present results could hardly be considered to provide a psychologically meaningful basis for any revision of the Fishbein theory" (p. 428). Attention has also been directed at the components' ability to mediate the effects of the traditionally employed attitude toward an object (A_0) upon intentions. Several investigations have shown that A_0 is mediated by the model's components (Ajzen 1971; Ajzen and Fishbein 1970, 1974) although others have found the components unable to completely mediate A_0 (Jaccard and Davidson 1975; Schwartz and Tessler 1972).

An alternative approach to this issue has been to examine the extent to which the components mediate the effects of assorted experimental manipulations upon intentions. Ajzen and Fishbein (1970) report that while holding A_B and NB constant strongly attenuated the effects of the two experimental manipulations upon intentions, one of the manipulations still remained significant. This latter finding was replicated by Songer-Nocks (1976b), although this and related findings have been debated (Fishbein and Ajzen 1976a,b; Songer-Nocks 1976a). Unconditional support for the components' ability to mediate the effects of experimental manipulations upon intentions is provided by Ryan (1977).

To summarize, evidence bearing on the model's mediational adequacy has been mixed, with some investigations supporting the model's ability to mediate exogeneous variables while other research has shed a less favorable light. However, even in those cases where the components have not completely mediated the effects of external variables, they have greatly attenuated the relationship between such variables and intentions.

Validity of the Component Weights

One key validity concern is the model's accuracy in identifying both the existence and relative importance of salient attitudinal and normative influences. If attitudinal and normative influences were equally important in determining intentions, one would hope that the component weights would correctly represent this condition. One should also expect a given determinant to receive a significant weight when the determinant is in fact important in guiding intentions and not to receive a significant weight when it does not influence intentions. If this were not the case, then the diagnostic usefulness of the model would be limited.

Several factors could threaten the validity of the component weights and thus the model's diagnosticity. Multicollinearity between the components, for instance, can lead to situations where salient influences receive an insignificant weight while a significant weight is assigned to nonsalient influences (this latter situation has been discussed by Cohen and Cohen 1975 under the label of suppressor variables). Reason to suspect the model on this basis has been provided by Miniard and Cohen (1979) as they found the attitudinal and normative measures currently used for model implementation to be inadequate in separating these two sources of influence. The weights' accuracy should also be dependent on the measures' ability to fully capture the causal influences that affect intentions. It has recently been shown that SN may not mediate the influence of MC upon intentions (Miniard and Cohen 1981). This implies

that the normative component may be less accurate when SN is used for measuring normative influences.

Evidence demonstrating that the measures are not distinct or that they fail to fully mediate influences which they should capture, however, is limited in that it only establishes the potential for inaccurate weights. It may be that, even in the presence of such problems, the weights correctly reflect the salient influences. What is needed, then, is a direct test of the weights' validity.

Two possible criteria for evaluating the validity of the component weights are (1) the weights' ability to represent the absence or presence of salient attitudinal and normative influences and (2) whether the weights differ in their relative sizes in expected ways. Under the first criterion, the normative weight, for example, should be significant when normative influences are important determinants of intentions. Similarly, the weight should not be significant when normative influences are unimportant. Failure to confirm these expectations would cast serious doubt upon the model's diagnostic usefulness. The second criterion is less concerned with the accuracy of a particular weight. Rather, the focus is on the relative magnitude of the weights. Drawing upon some conceptual framework, one might hypothesize that a given component should be more important (i.e., have a larger weight) in situation A with the remaining component dominating intentions in situation B. Note that support for one criterion does not automatically imply support for the other. For instance, one might be led to predict that both sources of influence are important (Criterion 1), although the attitudinal component should be more important than the normative component (Criterion 2). If the results yielded a large significant attitudinal weight but a small insignificant normative weight, only the latter prediction would be supported.

Research examining the weights' validity has, without exception, relied upon the second criterion. Using the Prisoner's Dilemma game, Ajzen (1971) and Ajzen and Fishbein (1970) told subjects to consider themselves partners (cooperation condition) or to do better than the other person (competitive condition). It was expected that normative considerations would carry greater weight in the cooperation condition while attitudinal considerations would be more important in the competitive condition. The observed patterns of regression weights supported these predictions. Similar results derived from this experimental paradigm are reported by Songer-Nocks (1976b).

Wilson, Mathews, and Monoky (1972) found partial support for hypothesized differences in the weights' relative magnitudes. As predicted, the attitudinal component received a larger weight when the other person in a buyer-seller dyad was described as being dissimilar to the subject. The attitudinal component was also dominant when the other was described as being similar, although it was expected that the normative component would be of greater importance.

Evidence failing to support expected differences in the components' relative weights within a marketing context is reported by Ryan (1978). When the model was applied to two brands of toothpaste, social influences were predicted to be more important for Ultra Brite than Crest since the former's promotional activities emphasized social interactions. The results contradicted this hypothesis as the normative weight was larger than the attitudinal weight for Crest while the reverse was true for Ultra Brite.

As stated above, evidence addressing the weights' accuracy in representing the presence or absence of salient normative and attitudinal influences is nonexistent. However, close inspection of the correlational and regression analyses reported in several investigations suggests that the weights may not be accurately reflecting the true importance of a given component. In particular, it is often the case that a component correlates significantly with

intentions yet receives an insignificant weight. Fishbein (1966), for example, reported that males' attitude toward premarital sexual intercourse had a significant correlation with their intention to engage in this behavior but received an insignificant weight in the regression equation. The same result with respect to $\Sigma NB_j MC_j$ is reported by Ajzen and Fishbein (1972) where the normative component weight was insignificant although $\Sigma NB_j MC_j$ was significantly correlated with intentions. An extreme instance of these inconsistencies where actual behavior was the criterion is provided by Ajzen and Fishbein (1970). As reported in Table 4 of their results, both A_j and NB were significantly related to behavior but neither measure received a significant regression weight. Further examples of significant correlations and insignificant weights can be found in Greenstein, Miller, and Weldon (1979), Schwartz and Tessler (1972), and Warshaw (1980).

Unfortunately, evidence that a given component receives a significant correlation but an insignificant weight can not be unambiguously interpreted as evidence against the model's diagnostic utility. It may be the case that the measure should not correlate significantly with intentions. If the measure used to operationalize the normative component was sensitive to attitudinal influences (i.e., the normative measure also tapped the attitudinal component), the normative component might correlate with intention even when the intention is solely under attitudinal control. In such situations, the lack of a significant normative weight is quite appropriate. However, if a significant correlation between the normative component and intentions is valid in the sense that normative influences do play an instrumental role in determining intentions, the model's failure to assign a significant weight to the normative component would provide a clear indication of the potential for misleading inference. Without knowing whether a given component is in fact important, one cannot easily reconcile inconsistencies between the correlational and regression analyses.

Therefore, the following two experiments will attempt to provide some initial evidence concerning the FBI model's ability to accurately reflect the absence or presence of salient attitudinal and normative influences. In particular, the validity of the normative component's weight will be examined under conditions where (1) the social environment does not influence intentions (Study I) and (2) social influences do affect intentions (Study II).

Study I

Method

Subjects and Procedures. A total of thirty male and female undergraduate marketing students participated in the experiment. Subjects were informed that they were participating in a study of "new product introductions" and that they would receive a booklet containing two role playing descriptions involving a new product. Both descriptions began by asking subjects to assume that they had participated in a nationwide marketing research study of consumers' reactions to a new product. In one of the descriptions, the subject was portrayed as being very favorable toward purchasing the product (Attitude-Positive). The remaining description stated that the subject was very unfavorable toward buying the product (Attitude-Negative). The order of presentation of these two descriptions was counterbalanced across all subjects.

For a third of the subjects, this attitudinal statement completed the description. This condition (Others' Expectations-Unknown) thus represented a situation that was devoid of information concerning the social environment (e.g., whether others thought the subject should purchase the

product). The remaining subjects, however, received descriptions that contained information concerning the opinions of others. Subjects in the Others' Expectations-Positive condition were informed that most others were very favorably toward the new product and that almost everyone had responded "yes" to the question "Do you think others should buy this product?". Conversely, others were described as being unfavorable toward the product in the Others' Expectations-Negative condition and as answering "no" to the question regarding whether others should purchase the product.

Subjects in these latter two conditions (where others' opinions are known) were then told that: "Of course, what others feel you personally should do here is a matter of no concern to you. Buying or not buying the new product is strictly your own decision." These statements were included in order to create a situation in which the social environment is not an important source of influence. If this is the case (which will be examined below), the question then becomes whether the FBI model can correctly reflect via the normative component weight the unimportance of the social environment.

Measures. Prior to completing the following measures, subjects were instructed how the response scales were to be used (e.g., the meaning of the scale categories). Subjects first responded to a measure of BI via three 7-point bipolar semantic differential scales (likely-unlikely, probable-improbable, possible-impossible). Subjects then indicated their attitude toward the behavior ("For me, buying the new product would be:") on four 7-point bipolar evaluative semantic differential scales (good-bad, foolish-wise, rewarding-punishing, harmful-beneficial). Subjects next responded to the SN measure (i.e., "Most people who are important to me think I should/should not buy the new product") via a 7-point scale ranging from "I should" (+3) to "I should not" (-3).

Results

The results of the 2 (Attitude) X 3 (Others' Expectations) ANOVA conducted on subjects' responses to the BI, AB, and SN measures are presented in Table 1. Verification that social influences were not an important determinant of BI required testing the effect of the Others' Expectations manipulation upon subjects' responses to the BI measure. If the provision of information concerning others' opinions did in fact play an influential role in the formation of subjects' intentions to purchase the product, then the Others' Expectations manipulation should have a significant impact on BI. This manipulation did not attain significance ($p < .22$), although the attitudinal manipulation was highly significant and accounted for 92% of the measure's variation, as calculated by use of Hays' (1963) omega-squared (ω^2) statistic. The lack of a significant normative manipulation indicates that beliefs regarding others' expectations were not an important determinant of subjects' intentions to purchase the product. Accordingly, the normative component as operationalized by SN should not receive a significant weight.

Examination of the normative weight's ability to reflect the nonsalient social influences existing in the Others' Expectations-Positive and Negative conditions involved conducting a regression analysis pooling (i.e., collapsing across the experimental cells after adjusting for differences in cell means) across these conditions. Consistent with prior research, the model provided a reasonable prediction of intentions ($R = .73$, $p < .001$). Support for the normative weight's validity was attained as only the attitudinal component received a significant ($p < .01$) weight while the normative weight was clearly insignificant. It should also be noted that, in this situation, the regression and correlational analyses were consistent as A_B achieved a significant ($p < .01$) correlation ($r = .72$) with intentions while SN did not ($r = .08$).

TABLE 1
ANOVA SUMMARY TABLE: STUDY I

Source	BI		A _B		SN	
	F	ω^2	F	ω^2	F	ω^2
Between Subjects						
Others (A)	1.4	.00	5.2*	.01	32.1*	.39
Within Subjects						
Attitude (B)	503.2*	.92	257.1*	.86	29.9*	.20
AB	0.2	.00	0.2	.00	1.9	.01

* $p < .01$

Study II²

Method

Subjects and Procedures. A total of 193 male and female undergraduate marketing students participated in the experiment. Each subject read one of four different scenarios in which a hypothetical woman was considering the purchase of a dress and that a female neighbor who was accompanying this woman thought she should purchase the dress. The four scenarios represented the product of a 2 X 2 manipulation of the scenario content. The first manipulation, No Manipulative Intent (NMI)--Manipulative Intent (MI), refers to the presence or absence of information that the woman's neighbor was exerting influence for self-serving reasons. The second manipulation, Agree-Disagree, altered the statement describing the woman's attitude toward the dress from positive to negative. Upon completion of the scenario, subjects were asked to play the role of this woman in responding to a questionnaire containing the measures necessary for operationalizing the FBI model. At this point, a third manipulation was introduced as subjects responded to one of three possible questionnaires that differed in the level of specificity at which MC was assessed (i.e., general, moderate, and situation specific). This factor was included since there presently exists some disagreement as to which level of measure specificity is most appropriate (cf. Ahtola 1976; Fishbein 1976; Glassman and Fitzhenry 1976; Miniard and Cohen 1981).

Measures. Subjects first responded to one of three alternative measures of MC via 7-point scales ranging from "I want to do" (+3) to "I want to do the opposite of" (-3). Depending upon whether level of specificity was general, moderate, or situation specific, the scale was prefaced by either "In general", "In the area of clothes shopping behavior", or "In this situation", respectively. Next, NB was assessed on 7-point scales with the endpoints "I should" (+3) and "I should not" (-3). NB and MC were obtained for three referents: the hypothetical woman's neighbor, husband, and parents. Subjects then reported their attitude toward the behavior on four 7-point bipolar evaluative semantic differential scales (good-bad, foolish-wise, rewarding-punishing, harmful-beneficial). Subjects

²This experiment has been previously reported (Miniard and Cohen 1979, 1981), although the data were not discussed in light of the present concerns. The reader is referred to these articles for a more detailed discussion of the experimental procedures.

next responded to the SN measure via a 7-point scale ranging from "I should" (+3) to "I should not" (-3). This was followed by an alternative attitudinal measure (PAB) not employed by Fishbein. This measure ("Forgetting what others think you should do and their reaction, you personally feel that buying the dress is:") was designed to minimize the potential for normative contamination by instructing subjects that the influence of others was not to be included in their responses to the measure. The major difference between PAB and AB was that the latter does not suggest that the influence of others is to be ignored. BI was next measured on three 7-point bipolar semantic scales (likely-unlikely, probable-improbable, possible-impossible).

Validation Condition. In addition to examining the impact of the NMI-MI manipulation on BI for inferring the importance of normative influences, further evidence was gathered in the form of subjects' self-reports. An additional 57 male and female undergraduate marketing students participated in this phase. Each subject read one of the four scenarios, responded to a measure of BI, and was then asked to estimate the impact of the social influences depicted in the scenario upon their intentions to purchase the dress via an 11-point scale with the endpoints "more likely to buy the dress" (11) and "less likely to buy the dress" (1). The scale midpoint of this "directional importance" measure was to be circled if such influences were unimportant. If social influences were an important determinant of intentions, subjects' responses to this measure should then significantly differ from the scale midpoint.

Results

Evidence supporting the importance of normative considerations in determining intentions was attained as the cell means for both the NMI ($\bar{X}=8.31$) and MI ($\bar{X}=4.14$) scenarios significantly ($p < .05$) differed from the scale midpoint (i.e., 6). Further support for the salience of social influences was also attained as the NMI-MI manipulation had a highly significant ($p < .001$) effect upon intentions and explained 16% of the variability. These findings suggest that the normative component should receive a significant regression weight.

TABLE 2

POOLED MULTIPLE REGRESSION ANALYSES FOR THE PREDICTION OF BEHAVIORAL INTENTIONS: STUDY II

Model	n	β_1^a	β_2^b	R	R ²	r ^c
(1) AB+SN	193	.55**	.08	.59**	.35	.43**
(2) AB+ Σ NBMCg	57	.54**	-.06	.53**	.28	.25
(3) AB+ Σ NBMCm	68	.53**	.22*	.64**	.41	.38**
(4) AB+ Σ NBMCs	68	.54**	.19*	.63**	.40	.32**
(5) PAB+SN	193	.49**	.19**	.56**	.31	.31**
(6) PAB+ Σ NBMCg	57	.70**	-.02	.69**	.48	.13
(7) PAB+ Σ NBMCm	68	.38**	.38**	.57**	.33	.10
(8) PAB+ Σ NBMCs	68	.48**	.23**	.59**	.35	.27*

Note: The symbols MCg, MCm, and MCs refer to whether motivation to comply was measured at a general, moderate, or situation specific level, respectively.

^aStandardized partial regression coefficient for attitudinal model component.

^bStandardized partial regression coefficient for normative model component.

^cPooled simple correlation (i.e., multicollinearity) between attitudinal and normative components.

* $p < .05$
** $p < .01$

Given the two alternative attitudinal measures (i.e., AB and PAB) and the four different normative representations (i.e., SN and the three Σ NBjMCj formulations that differ in the level of MC specificity), a total of eight regression models were tested. The results of these pooled regression analyses for the various attitudinal and normative operationalizations are summarized in Table 2. Focusing on those models employing Fishbein's standard AB measure (i.e., models 1-4 in the table), the FBI model fails to correctly identify the importance of normative influences when either SN (model 1) or a general level of MC specificity (model 2) is employed for operationalizing the normative component as an insignificant normative weight was attained for these models. Interestingly, replacement of AB by PAB (which is significantly less correlated with SN) improves the FBI model's diagnostic performance when SN is employed for representing normative influences. While the normative weight was insignificant for the AB + SN model, the weight attained significance for the PAB + SN model. With respect to the correlational results, all of the normative measures except for the Σ NBjMCj formulation where MC was assessed at a general level correlated significantly ($p < .05$) with intentions. Thus, in the one instance where the correlational and regression analyses conflicted (i.e., model 1), the correlational analysis was correct.

Conclusion

The ability of the FBI model to validly represent the importance of normative influences in two distinct situations was examined. Although the model correctly identified the lack of salient normative influences, inadequacies were found for certain normative operationalizations when normative influences were an important determinant of intention. Evidence relevant to the appropriate level of MC specificity was also uncovered as the general level failed to reflect the importance of social influences whereas the moderate and situation specific levels did adequately represent such influences. These results can be attributed to the fact that some of the experimental conditions (i.e., the MI conditions) under which MC was assessed involved situations where a generally positive referent became negative. The general level, which does not ask subjects to report the situational importance of a referent, was apparently unable to reflect these situational variations in the referent's influence potential. Since the normative component received an appropriate weight when either of the two remaining levels of MC specificity were employed, these findings suggest that such levels are perhaps most appropriate for measuring MC.

One interesting question that arises is why the normative component, when represented by SN, performs accurately under conditions lacking social influences but is inaccurate under settings involving salient normative influences. One possible explanation is the multicollinearity between the attitudinal and normative components. In Study I, AB and SN were unrelated ($r = -.90$). In Study II, however, there was a significant correlation between the two measures (as reported in Table 2) thus indicating that the likelihood for diagnostic error increases in situations where the components are not independent.

While the present findings do shed some light on an area that has thus far been ignored, it is clear that further work will be necessary before the FBI model can be confidently used as a diagnostic tool for identifying the importance of attitudinal and normative influences. Possible areas of improvement over the present studies could involve the use of more natural (i.e., non-role playing) settings that allow a test of the weights' ability to represent the absence or presence of salient attitudinal and normative influences within the same experimental framework rather than comparing across data sets. Research that identifies the conditions under which the model will

or will not accurately identify the importance of these influences would seem to hold the greatest promise.

References

- Ahtola, O. T. (1976), "Toward a Vector Model of Intentions," in Advances in Consumer Research, Vol. 3, (ed.) B. B. Anderson, Cincinnati, OH: Association for Consumer Research, pp. 481-4.
- Ajzen, I. (1971), "Attitudinal vs. Normative Messages: An Investigation of the Differential Effects of Persuasive Communications on Behavior," Sociometry, 34, 263-80.
- _____, and Fishbein, M. (1970), "The Prediction of Behavior from Attitudinal and Normative Behaviors," Journal of Experimental Social Psychology, 5, 400-16.
- _____. (1972), "Attitudes and Normative Beliefs as Factors Influencing Behavioral Intentions," Journal of Personality and Social Psychology, 21, 1-9.
- _____. (1974), "Factors Influencing Intentions and the Intention-Behavior Relationship," Human Relations, 27, 1-15.
- Bearden, W. O., and Woodside, A. G. (1978), "Situational and Extended Attitude Models as Predictors of Marijuana Intentions and Reported Behavior," Journal of Social Psychology, 106, 57-67.
- Bentler, P. M., and Speckart, G. (1979), "Models of Attitude-Behavior Relations," Psychological Review, 86, 452-64.
- Carnegie-Mellon University Marketing Seminar (1978), "Attitude Change or Attitude Formation? An Unanswered Question," Journal of Consumer Research, 4, 271-6.
- Cohen, J., and Cohen, P. (1975), Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davidson, A. R. and Jaccard, J. J. (1975), "Population Psychology: A New Look at an Old Problem," Journal of Personality and Social Psychology, 31, 1073-82.
- Dickson, P. R., and Miniard, P. W. (1978), "A Further Examination of Two Laboratory Tests of the Extended Fishbein Attitude Model," Journal of Consumer Research, 4, 261-6.
- Fishbein, M. (1966), "Sexual Behavior and Propositional Control," paper presented at the Psycholomic Society meetings, St. Louis.
- _____. (1976), "Extending the Extended Model: Some Comments," in Advances in Consumer Research, Vol. 3, (ed.) B. B. Anderson, Cincinnati, OH: Association for Consumer Research, pp. 491-7.
- _____, and Ajzen, I. (1975), Belief, Attitude, Attention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison Wesley.
- _____. (1976a), "Misconceptions About the Fishbein Model: Reflections on a Study by Songer-Nocks," Journal of Experimental Social Psychology, 12, 579-84.
- _____. (1976b), "Misconceptions Revisited: A Final Comment," Journal of Experimental Social Psychology, 12, 591-2.
- Glassman, M., and Fitzhenry, N. (1976), "Fishbein's Subjective Norm: Theoretical Considerations and Empirical Evidence," in Advances in Consumer Research, Vol. 3, (ed.) B. B. Anderson, Cincinnati, OH: Association for Consumer Research, pp. 477-80.
- Greenstein, M., Miller, R. H., and Weldon, D. E. (1979), "Attitudinal and Normative Beliefs as Antecedents of Female Occupational Choice," Personality and Social Psychology Bulletin, 5, 356-62.
- Hays, W. L. (1963), Statistics for Psychologists. New York: Holt, Rinehart and Winston.
- Jaccard, J., and Davidson, A. R. (1975), "A Comparison of Two Models of Social Behavior: Results of a Survey Sample," Sociometry, 38, 497-517.
- Lutz, R. J. (1977), "An Experimental Investigation of Causal Relations Among Cognitions, Affect, and Behavioral Intention," Journal of Consumer Research, 3, 197-208.
- _____. (1978a), "Rejoinder," Journal of Consumer Research, 4, 266-271.
- _____. (1978b), "Rejoinder," Journal of Consumer Research, 4, 276-8.
- Miniard, P. W., and Cohen, J. B. (1979), "Isolating Attitudinal and Normative Influences in Behavioral Intentions Models," Journal of Marketing Research, 16, 102-10.
- _____. (1981), "An Examination of the Fishbein Behavioral Intention Model's Concepts and Measures," Journal of Experimental Social Psychology, in press.
- Ryan, M. J. (1977), "Programmatic Research Based on Fishbein's Model," in Consumer and Industrial Buying Behavior, (eds.) A. G. Woodside, J. N. Sheth, and P. D. Bennett, New York: North Holland, pp. 151-66.
- _____. (1978), "An Examination of an Alternative Form of the Behavioral Intention Model's Normative Component," in Advances in Consumer Research, Vol. 5, (ed.) H.K. Hunt, Ann Arbor, MI: Association for Consumer Research, pp. 283-9.
- _____, and Bonfield, E. H. (1980), "Fishbein's Intentions Model: A Test of External and Pragmatic Validity," Journal of Marketing, 44, 82-95.
- Schlegel, R. P., Crawford, C. A. and Sanborn, M. D. (1977), "Correspondence and Medial Properties of the Fishbein Model: An Application to Adolescent Alcohol Use," Journal of Experimental Social Psychology, 13, 421-30.
- Schwartz, S. H., and Tessler, R. C. (1972), "A Test of a Model for Reducing Measured Attitude Behavior Discrepancies," Journal of Personality and Social Psychology, 24, 225-36.
- Songer-Nocks, E. (1976a), "Reply to Fishbein and Ajzen," Journal of Experimental Social Psychology, 12, 285-90.
- _____. (1976b), "Situational Factors Affecting the Weighting of Predictor Components in the Fishbein Model," Journal of Experimental Social Psychology, 12, 56-59.
- Warshaw, P. R. (1980), "A New Model for Predicting Behavioral Intentions: An Alternative to Fishbein," Journal of Marketing Research, 17, 153-72.
- Wilson, D. T., Mathews, H. L., and Harvey, J. W. (1975), "An Empirical Test of the Fishbein Behavioral Intention Model," Journal of Consumer Research, 1, 39-48.
- _____, _____, and Monoky, J. F. (1972), "Attitude as a Predictor of Behavior in a Buyer-Seller Bargaining Situation: An Experimental Approach," in American Marketing Association Combined Conference Proceedings, (eds.) B. W. Becker and H. Becker, Chicago: American Marketing Association, pp. 390-5.

A COMPARISON OF TWO BEHAVIORAL INTENTION MODELS

David Brinberg, University of Maryland

Abstract

The present investigation compared the predictor variables of Fishbein's and Triandis' behavioral intention models. The distinction Triandis makes concerning affect (gut feeling) was supported, that is, affect and attitude were found to measure two dimensions. In addition, the social influence constructs of both models seem to measure two dimensions: (1) what relevant others think and (2) what is appropriate to do.

Introduction

In recent years, there has been increased interest in determining the extent to which verbal reports, survey results, and attitude measures can predict behavior. Numerous models have been developed to predict behavior, and recently, some of these models have been compared (Zuckerman and Reis 1978; Jaccard and Davidson 1975; Brinberg 1979). The primary interest of these researchers comparing models was examining the relationship between the predictor variables and the criterion within each model. In other words, the researchers were interested in comparing the outcome of one model's predictions with the outcome of comparable predictions from the other models. For instance, in the studies cited above, the comparison of the different models involved comparing the multiple correlation (i.e., the percent of variance accounted for in the criterion) or the population estimate of the mean square error.

Often ignored are the relationships among two(or more) sets of predictor variables from the two(or more) models. It is possible that the outcome (e.g., multiple correlation) from a set of models are different(same) even though the predictor variables measure the same (different) underlying dimension(s) (Brewer, Campbell, and Crano 1970; Birnbaum and Mellers 1979). Thus, independent of the models outcomes, substantial similarities (or differences) may be found among the models. Since the dependent variables for these models are the same, differences between these models can exist: (1) among the predictor variables or (2) among the relationships of these predictor variables with the criterion. The major focus of this paper will be on the relationships among the predictor variables. The two models selected for comparison in this study are proposed by Fishbein (Fishbein and Ajzen 1975) and Triandis (1977). These models have been examined in a number of cultures (Davidson, Jaccard, Triandis, Morales, and Diaz-Guerrero 1976; Davidson and Jaccard 1975), with a number of behaviors (Songer-Nocks 1976; Zuckerman and Reis 1978; King 1975; Pomazal 1974) as well as compared using different behaviors (Jaccard and Davidson 1975; Seibold and Roper 1979; Brinberg 1979). However, in none of the studies comparing these two models was an attempt made to determine the interrelationships among the various predictor variables.

Fishbein's Model

Fishbein (Fishbein and Ajzen, 1975) states that a person's intention to perform a behavior is determined by: (1) one's attitude towards the performance of the behavior and (2) one's perception of whether significant others think s/he should or should not perform the behavior.

This model may be expressed by the following algebraic equation

$$\text{Intention} = \text{Attitude}(w_1) + \text{Subjective Norm}(w_2) \quad (1)$$

The attitude is viewed as a function of the beliefs concerning the act multiplied by the evaluative aspect of the beliefs. This may be represented as follows:

$$\text{Attitude} = \sum b_i e_i \quad (2)$$

where b_i = likelihood that performing the behavior will result in some outcome, e_i = evaluation of that outcome. The subjective norm consists of normative beliefs of significant others multiplied by the motivation to comply with these others. This may be represented as follows:

$$\text{Subjective Norm} = \sum N_b M_c \quad (3)$$

where N_b = belief that a particular referent thinks you should or should not perform the behavior, M_c = motivation to comply with that particular referent.

The w_1 and w_2 constructs are theoretical parameters reflecting the importance of the variables in determining intention. They are generally determined through regression techniques.

Triandis' Model

Triandis (1977) states that the determinants of a person's intention may be expressed by the following set of equations.

$$\text{Intention} = A(w_3) + C(w_4) + S(w_5) \quad (4)$$

where A = affect towards the behavior, that is, perceived enjoyment vs. disgust associated with performing the behavior, C = the perceived value of the consequences associated with the behavior, S = social determinants.

As before, w_3 , w_4 , and w_5 are theoretical parameters reflecting the importance of the constructs in determining intention. They, too, are generally determined through regression techniques.

The social determinants component consists of a number of constructs: norms, roles, the actor's self-concept, moral norm, ideals, and contractual agreements. In this study, norms, roles, the self-concept, and the moral norm were used as elements of the social component. The definitions of these constructs will be discussed later.

The second equation of the model specifies the "C" component of the first equation. Following instrumentality theory (Rosenberg 1956), this term is operationalized as follows:

$$\text{Consequences} = \sum p_c v_c \quad (5)$$

where p_c = likelihood that a behavior will result in some outcome, v_c = evaluation of that outcome.

Comparison of Fishbein's and Triandis' Models

One difference between Fishbein's and Triandis' models concerns a distinction between affect and attitude (evaluation). Triandis (1977) defines affect as the "gut

feeling" associated with the behavior and distinguishes this from attitude (evaluation). Evaluation, according to Triandis, is a cognitive judgment based on $\Sigma p_c v_c$. Triandis (1977) justifies this distinction by arguing that for some types of behaviors (e.g., sin and/or duty), both constructs will make separate contributions in predicting intention. Fishbein, on the other hand, has argued that both affect and attitude measure the same underlying evaluative dimension. Therefore, it is not necessary to postulate separate constructs since each is measuring the same underlying dimension. (It is important to note that $\Sigma p_c v_c$ is equivalent to $\Sigma b_i e_i$ and is used interchangeably throughout this paper.) Based on these hypothesized relationships, one aspect of this study will examine whether a single or multiple dimension can be used to describe affect, attitude, and consequences ($\Sigma b_i e_i$).

A second difference between these two models involves the normative component in Fishbein's model and the social determinants in Triandis' model. Fishbein's approach implies that the normative component is an indicant of what relevant others think. According to Fishbein (Fishbein and Ajzen 1975), the subjective norm measures "what most people who are important to me think I should do." This subjective norm is hypothesized to be determined by normative beliefs (what a particular referent thinks I should do) multiplied by the motivation to comply with that referent. Triandis (1977), on the other hand, defines a norm as the appropriateness of performing the behavior. In addition to norms, Triandis includes other constructs in the social determinants: roles, which is the appropriateness of performing the behavior for a person in a particular position in the social system; self-concept, which consists of beliefs the person perceives concerning the appropriateness of the act for himself(herself); and moral norm, which is the perceived moral obligation to perform the behavior.

All the constructs in both the social determinants of Triandis' model and the normative component in Fishbein's model may be seen to measure a social influence dimension. If this is the case, these constructs should measure the same underlying dimension. Another aspect of this study, then, will examine the relationships among the constructs that are proposed to measure social influence.

Numerous behaviors can be used to test these two models. However, when comparing models, it would be useful to select behaviors that have been examined by each model separately. One behavior that meets this criterion is blood donation (Zuckerman and Reis 1978; Pomazal 1974).

Method

Subjects

Subjects were obtained from the participant pool in partial fulfillment of a course requirement. The data were collected at three points in time, during the course of four months, with a total of 96 subjects completing all three phases.

Procedure

In order to obtain the salient beliefs, 100 subjects, who did not participate in the main phase of this study, were asked to indicate the beliefs they associated with donating blood. Twenty-five subjects indicated their beliefs concerning donating blood in each of three contexts (i.e., in the context of the blood bank, the hospital, and the blood mobile) as well as donating blood in general. The use of these different contexts will be made clear shortly. The subjects were asked to indicate the advantages, disadvantages and consequences of

donating blood. The modal set of beliefs were similar across all contexts. Thus, only one set of beliefs was used as the modal salient set.

Measuring Instrument

All the constructs in both models were measured on 7 point scales. The belief statements were measured on a subjective probability scale (e.g., Donating blood at the blood bank this semester would be painful: likely:unlikely). The evaluation of those beliefs were measured on an evaluative scale (e.g., Doing something painful is: good-bad). Motivation to comply with a referent was measured by asking the subject, "Generally speaking, I want to do what (Referent X) thinks I should do: want to do-want not to do." Normative beliefs were measured by asking the subject, "(Referent X)thinks I should donate blood at the blood bank this semester: likely-unlikely." A person's attitude toward the act (e.g., Donating blood at the blood bank this semester is:) was measured on three evaluative scales (i.e., good-bad, nice-awful, favorable-unfavorable). The subjective norm was measured by asking the subject what significant others think s/he should or should not do (e.g., Most people who are important to me think I should donate blood at the blood bank this semester: likely-unlikely). The subject's intention was measured on a subjective probability scale (e.g., I intend to donate blood at the blood bank this semester: likely-unlikely).

The affect construct in Triandis' model was measured using scales intended to measure the person's "gut" feeling (i.e., enjoyable-disgusting; pleasant-unpleasant; exciting-nauseating). Roles, as conceptualized by Triandis measure the appropriateness of the behavior for a person in a particular position in the social system (e.g., It is appropriate for students to donate blood at the blood bank this semester: appropriate-inappropriate). Moral norm was measured by asking the subject, "I have a moral obligation to donate blood at the blood bank this semester: have a moral obligation-have no moral obligation." The attributes used to measure the self-concept were obtained in the aforementioned elicitation phase (e.g., I consider myself the type of person who is concerned: likely-unlikely).

Results

In order to determine the generalizability of these models across situations, blood donation was examined in the context of the blood bank, the hospital, and the blood mobile. The predictive accuracy of both models was assessed and found not to be systematically different (R^2 for both models was approximately 30%). Since the main intent of this paper is the examination of the interrelationships among the different predictor variables of both models, the predictive accuracy of these models will not be discussed. The reader interested in a detailed presentation of the predictive accuracy of these models should see Brinberg (1979).

As mentioned earlier, Triandis draws a distinction between affect and attitude whereas Fishbein sees both constructs as measures of the same underlying evaluative dimension. In order to examine this theoretical distinction, two approaches may be used. The first would be to place the attitude and affect measures in a regression analysis and determine whether both make independent contributions in predicting intention. However, as pointed out by Brewer et al. (1970), Kenny (1976) and Birnbaum and Mellers (1979), interpreting the pattern (significance) of regression weights may lead the researcher to mistakenly interpret that more than one factor underlies a set of constructs. Since the regression technique does not allow for the effects of

measurement error, a researcher may conclude that two (or more) constructs exist, when, in fact, the constructs are imperfect measures of the same underlying dimension.

To circumvent some of the limitations of the regression technique, Brewer et al. (1970) suggest using a factor analytic approach to determine the number of factors underlying a set of constructs. Since a unique algebraic solution is always available for three variables (Brewer et al. 1970), it was decided to factor analyze the scales used to measure affect and attitude as well as $\Sigma b_i e_i$ in order to determine whether a single factor can be used to account for the relationships among these constructs. Table 1 contains a varimax rotation of a principal components analysis of attitude, affect, and consequences.

TABLE 1
VARIMAX ROTATION OF PRINCIPAL COMPONENTS ANALYSIS OF ATTITUDE, AFFECT AND CONSEQUENCES FOR THE INTENTION TO DONATE BLOOD

Time 1						
Scale	Blood Bank		Hospital		Blood Mobile	
	F.L.1	F.L.2	F.L.1	F.L.2	F.L.1	F.L.2
good-bad	.14	.84	.18	.80	.25	.75
pleasant-unpleasant	.87	.24	.91	.10	.82	.28
favorable-unfavorable	.22	.68	.65	.28	.44	.64
nice-awful	.68	.51	.64	.51	.78	.14
exciting-nauseating	.89	.11	.93	.15	.84	.20
enjoyable-disgusting	.85	.27	.85	.27	.85	.32
$\Sigma b_i e_i$.20	.70	.16	.77	.09	.80
Time 2						
good-bad	.21	.84	.33	.73	.18	.82
pleasant-unpleasant	.89	.20	.91	.22	.87	.33
favorable-unfavorable	.51	.51	.46	.64	.39	.68
nice-awful	.65	.50	.63	.57	.52	.67
exciting-nauseating	.88	.22	.91	.20	.91	.17
enjoyable-disgusting	.79	.15	.80	.31	.92	.27
$\Sigma b_i e_i$.13	.78	.07	.83	.08	.57
Time 3						
good-bad	.31	.79	.22	.84	.57	.57
pleasant-unpleasant	.90	.16	.86	.30	.86	.21
favorable-unfavorable	.68	.50	.56	.48	.65	.26
nice-awful	.78	.43	.74	.44	.73	.43
exciting-nauseating	.87	.10	.89	.12	.88	.06
enjoyable-disgusting	.87	.19	.82	.34	.88	.21
$\Sigma b_i e_i$.06	.83	.23	.72	.11	.94

A maximum likelihood factor analysis (Lawley and Maxwell 1971; Joreskog 1967; Humphreys and Montanelli 1975) was used to determine the number of factors underlying these concepts since it provides a χ^2 goodness of fit test for any K factor model. If the χ^2 value obtained from this analysis is significant for K factors, this indicates that at least K + 1 factors are needed to adequately fit the data, that is, a significant χ^2 indicates that a significant portion of the variance can still be systematically accounted for by extracting another factor. Thus, the number of factors used as an adequate fit of the data occur for the first non-significant χ^2 value.

For the factor analyses reported in Table 1, all yielded a significant χ^2 for the first factor ($p < .01$) and all but one yielded a non significant χ^2 for the second factor. The significant χ^2 for the second factor was in the context of the blood mobile at time 3 ($\chi^2 = 17.63$, $p < .05$). The χ^2 for the third factor was non-significant.

The findings from the maximum likelihood factor analysis indicate that two factors consistently provide a sufficient fit of the data. Based on the pattern of factor loadings found in Table 1, the first factor appears to measure an affect dimension, that is, the bi-polar adjectives specified a priori to measure affect (Triandis 1977), all loaded on the same dimension. The second factor appears to measure an evaluative dimension since the good-bad scale (which Fishbein uses as a measure of attitude) consistently has a high loading.

An alternative explanation to account for the pattern of factor loadings (i.e., the emergence of two factors) is that it is an artifact of the response language used by the subject. In other words, two dimensions may emerge since one set of adjectives (e.g., affect) have a greater subjective range than another set (e.g., attitude). However, the plausibility of this alternative explanation is reduced since an independent assessment of attitude (i.e., $\Sigma b_i e_i$) is consistently loaded with the attitude scales. This provides additional evidence for describing the second factor as evaluative.

Thus, based on the pattern of factor loadings as well as the maximum likelihood factor analysis, Triandis' distinction between affect and attitude is supported since two factors, not one, are needed to adequately describe the data.

The relationships among the constructs in Triandis' social determinants component as well as Fishbein's normative component also need to be examined in order to determine the number of dimensions underlying these concepts. Triandis has postulated that moral norm, roles, norms, and the self-concept can be summed to form an index of social determinants. This implies that these constructs measure the same underlying dimension. In order to determine the relationships among these constructs (along with Fishbein's normative measures) the same factor analytic techniques described earlier were used. Tables 2 contains the varimax rotation of a principal components analysis of the constructs that measure social influence.

A maximum likelihood factor analysis was performed in order to determine the number of factors that would adequately describe the data. In all cases, the χ^2 value for the first factor was highly significant ($p < .01$) and non-significant for the second factor. These patterns of results indicate that a two factor solution can be used to describe the constructs that measure social influence. This is contrary to Triandis' assumption that the moral norm, roles, norms, and the self-concept are measuring a single dimension. Based on the pattern of factor loadings found in Tables 2, the first factor appears to measure a dimension that deals with the appropriateness of the behavior, since norms and roles have a high loading on it. The second factor appears to measure what "relevant others" think since the subjective norm and ENbMc have a high loading on this dimension. Interestingly the moral norm generally has a high loading on this dimension. This suggests that people view a moral obligation as something society thinks they should do, in other words, measuring what relevant others think.

TABLE 2
VARIMAX ROTATION OF PRINCIPAL COMPONENTS ANALYSIS OF
SOCIAL INFLUENCE CONSTRUCTS FOR THE INTENTION TO
DONATE BLOOD

Time 1

scale	Blood Bank		Hospital		Blood Bank	
	F.L. 1	F.L. 2	F.L. 1	F.L. 2	F.L. 1	F.L. 2
Subjective Norm	-.02	.89	-.09	.84	-.04	.92
ENbMc	.16	.88	.20	.88	.07	.91
Moral Norm	.21	.56	.25	.60	.25	.42
Role Norm	.93	.12	.89	.08	.92	.12
Self-Concept	.49	.28	.61	.14	.42	.47

Time 2

Subjective Norm	.08	.87	-.01	.89	.13	.88
ENbMc	.11	.91	.14	.86	.11	.89
Moral Norm	.29	.69	.24	.73	.26	.71
Role Norm	.92	.10	.95	.08	.94	.11
Self-Concept	.48	.23	.46	.35	.51	.22

Time 3

Subjective Norm	-.11	.89	.11	.82	.14	.91
ENbMc	.15	.84	.04	.86	.17	.89
Moral Norm	.23	.69	.27	.66	.37	.63
Role Norm	.94	.13	.92	.18	.91	.21
Self-Concept	.32	.50	.23	.47	.42	.39

In summary, a single factor solution was not adequate to describe the relationships among the social influence constructs. Based on the maximum likelihood factor analysis, two factors were adequate to describe the data. These factors appear to measure: (1) what relevant others think and (2) what is appropriate to do.

Discussion

Based on the factor analyses of attitude, affect, and consequences two dimensions, not one, appear to be sufficient in describing the relationships among these constructs. This result is contrary to Fishbein's assumption that all three are measuring the same underlying dimension and supportive of Triandis' assumption that these constructs may be separated into two dimensions; affect and attitude.

A similar analysis of the social determinants in Triandis model and the normative components in Fishbein's model shows that two dimensions, not one, are needed to account for the relationships among these constructs. This is inconsistent with Triandis' assumption that the constructs of the social determinants component measure a single underlying dimension. The two dimensions appear to measure: (1) what relevant others think and (2) what is appropriate to do.

An implication of these findings is that neither model, alone, can account for the relationships among these constructs. This suggests that a composite model, incorporating features of both models may be used to predict intention. This composite model would contain: (1) an affect construct, (2) an attitude construct, (3) a "relevant others" construct, and (4) an "appropriateness"

construct. A preliminary analysis of this composite model indicated that it was an accurate predictor of intention and consistently was more effective in predicting intention than either Fishbein's or Triandis' model. However, the magnitude of this difference was small and it is possible that this composite model was "better" since it was tested in the same sample in which it was developed. Future research is necessary comparing this composite model with Fishbein's and Triandis' model in order to determine whether it is useful to integrate these two approaches in predicting intention.

Finally, past research (e.g., Jaccard and Davidson 1975, Seibold and Roper 1979) has focused on the comparison of the predictive accuracy of these two behavioral intention models and has found both models to be superior to the other in predicting intention. What the present study suggests is that substantial similarities and differences may be found among these models by examining the relationships among the predictor variables. Therefore, it is strongly recommended that researchers interested in comparing different models (which have the same criterion), examine the relationships among the independent variables in addition to examining the relative predictive efficiency of these models.

References

- Birnbaum, M. H., and Mellers, B. A. (1979), "Stimulus Recognition May Mediate Exposure Effects," Journal of Personality and Social Psychology, 37, 391-394.
- Brewer, M. B., Campbell, D. T., and Crano, W. D. (1970), "Testing a Single Factor Model as an Alternative to the Misuse of Partial Correlations in Hypothesis Testing," Sociometry, 33, 1-11.
- Brinberg, D. (1979). "The Comparison of Three Attitude Models for the Prediction of Blood Donation Behavior," Unpublished Doctoral Dissertation.
- Davidson, A. R. and Jaccard, J. J. (1975), "Population Psychology: A New Look at an Old Problem," Journal of Personality and Social Psychology, 31(6), 1073-1082.
- Davidson, A.R., Jaccard, J. J., Triandis, H. C., Morales, M. L. and Diaz-Guerrero, R. (1976), "Cross-Cultural Model Testing: Toward a Solution of the Etic-Emic Dilemma," International Journal of Psychology, 11, 1-13.
- Fishbein, M. and Ajzen, I. (1975), Beliefs, Attitudes, Intention, and Behavior: An Introduction to Theory and Research, Reading, Mass.: Addison-Wesley,
- Humphreys, L. G. and Montanelli, R. G. (1975), "An Investigation of the Parallel Analysis Criterion for Determining the Number of Common Factors," Multivariate Behavioral Research, 10, 193-206.
- Jaccard, J. J. and Davidson, A. R. (1975), "A Comparison of Two Models of Social Behavior: Results of a Survey Sample," Sociometry, 38,491-517.
- Joreskog, K. G. (1967), "Some Contribution to Maximum Likelihood Factor Analysis," Psychometrica, 32, 443-482.
- Kenny, D. A. (1976), "Data Analysis in the Social Psychological Experiment", Representative Research in Social Psychology, 7(2), 120-133.
- King, G. W. (1975), "An Analysis of Attitudinal and Normative Variables as Predictors of Intentions and Behaviors", Speech Monographs, 42, 237-244.
- Lawley, D. N. and Maxwell, A. G. (1971), "Factor Analysis as a Statistical Method," American Elsevier, New York.

Pomazal, R. S. (1974), "Attitudes, Normative Beliefs, and Altruism: Help for Helping Behavior," Unpublished Doctoral Dissertation, University of Illinois.

Rosenberg, M. J. (1956), "Cognitive Structure and Attitudinal Affect," Journal of Abnormal and Social Psychology, 53, 367-372.

Schwartz, S. H. and Tessler, R. C. (1972), "A Test of a Model for Reducing Attitude Behavior Discrepancies," Journal of Personality and Social Psychology, 24, 349-364.

Seibold, D. R., and Roper, L. C. (1979), "Psychosocial Determinants of Health Care Intentions: Test of the Triandis and Fishbein Models," In D. Nimmo (ed.), Communication Yearbook 3, New Brunswick, N.J. Transaction Books.

Songer-Nocks, E. (1976), "Situational Factors Affecting the Weighting of Predictor Components in the Fishbein Model," Journal of Experimental Social Psychology, 12(1), 56-70.

Triandis, H. C. (1977), Interpersonal Behavior, Monterey, CA. Brooks/Cole.

Zuckerman, M. and Reis, H. T. (1978), "Comparison of Three Models for Predicting Altruistic Behavior," Journal of Personality and Social Psychology, 36(5), 498-510.

A DISCUSSION OF ATTITUDE RESEARCH AND
BEHAVIORAL INTENTIONS

Paul R. Warshaw, McGill University

Abstract

Each of the three papers in this session focus upon issues related to the determinants of behavioral intentions. Holbrook et.al. focus upon attitude formation. Specifically, the role of attribute importance in attitude structure is addressed. Brinberg looks more at the dimensionality of attitudes and norms. Employing Fishbein's and Triandis' models he shows both attitude and normative constructs may be bi-dimensional. Miniard, on the other hand, questions whether measured attitudes and norms are really independent factors of intent.

Attitude Structure and Search:
An Integrative Model of Importance-Directed
Information Processing

The Holbrook, Velez and Tabouret paper examines how attribute importance is related to attitude structure. While the study essentially replicates (with some methodological improvements) an earlier piece by Holbrook and Maier (1978), the findings are quite interesting and highlight some theoretical and methodological issues that consumer researchers should attend closely. The primary thesis of Holbrook et.al. is that product attributes vary in their importance to consumers. That is, some attributes are perceptually more important than others in determining product-related attitudes. The principal is certainly not novel, since attitude researchers have long acknowledged this fact. The concept is an outgrowth of Fishbein's proposition that attitudes are formulated from a small set of salient beliefs. However, the Fishbein approach assumes that once an attribute is incorporated within the salient set, its only contribution to attitude comes from the subjective probability of it being associated with the target object or action and its goodness or badness. Other researchers, quite naturally, became intrigued with the notion that there must be some hierarchical importance structure among those attributes in the evoked set. After all, since the evoked set contains approximately 5-9 attributes, the remaining attributes are tautologically non-important. If so, doesn't it logically follow that the salient attributes are similarly disectable into most important and least important classes? In concert with this line of reasoning, attempts have been made to secure attribute importance ratings from subjects and to use these ratings as weights in expectancy value or other type attitude models. The problem is that importance weight inclusion does not seem to enhance predictions of affect. The Holbrook et.al. paper proposes and tests a conceptual framework for resolving this seemingly perplexing issue. They claim that attribute importance has a major impact, but not within the model structure per se. Rather, importance relates more directly to information search processes. That is, the more important an attribute, the more information is gathered about it, with attitudes formed thereafter on the basis of cues acquired about these important attributes. The study was fairly well done and seemingly verifies the authors' model. More significant, it implies an issue that consumer researchers have too often ignored. A common practice in marketing and related disciplines is for researchers to haphazardly decide that some variable (e.g., a personality trait like self-confidence) is related to some aspect of consumer behavior (e.g., purchase of name-brand goods). Typically, there is no systematic, careful development of theoretical

linkages between the variables studied (cf. Kassarian 1971). Rather, in knee jerk fashion, the variable is tossed into some equation model in an additive or multiplicative fashion (without theoretical rationale) and run through a regression to see if significant betas pop out. Alternatively, some type of discriminant, factor or canonical analysis is conducted to delineate 'underlying factors or relationships'. The results are typically meaningless and not at all generalizable. This scenario is typical of marketing studies pertaining to attitude theory issues. The so-called importance weight debate is no exception. Hence, in my opinion, the greatest contribution of the Holbrook et.al. manuscript is its orientation toward building theoretical linkages in systematic, logical fashion. This tight approach is certainly more difficult and time consuming, and requires broader literature reviews than the alternative approach which is based mainly on personal opinion, scant surveys of existent research and frantic applications of the 'latest', 'hottest' data analysis techniques.

The paper has some other features I'd also like to highlight:

1. An interesting methodological technique is presented that might usefully be employed by other researchers. Namely, use of a sticker-pulling procedure in a mail-survey questionnaire was attempted (apparently quite successfully). This raises the possibility of conducting studies by mail that heretofore required laboratory settings. A major feature of social science research is that most studies employ relatively small, student samples. This creates many problems (e.g., nongeneralizable results, sensitized subjects, lack of statistical validity, etc.). Everyone would like to use large, representative samples but the dollar costs are typically prohibitive (especially where laboratory type formats seem warranted). Experimentation by mail certainly allows for better and larger sampling. The current sticker-pulling technique and further developments along these lines could help bring the laboratory to the people in an inexpensive, simple manner.
2. Part strength and part weakness of Holbrook et.al. was that the authors were stately aware of certain key assumptions inherent in their methodology. For example, "it is assumed that the respondent believes the information he has received to be true...". It would have been better methodology to verify this assumption during debriefing. Many consumer researchers are unaware that this and other key assumptions are integral components of their research design. This is particularly true when treatment manipulations entail differential communications across subject groups. Without appropriate manipulation check analysis (which includes believability questions), one may not justifiably conclude that the findings resulted from the stated treatments. Rather, all sorts of demand characteristic and other explanations are equally possible, thereby threatening the experiment's internal validity. While Holbrook et.al. recognized the problem, they did not attempt to ameliorate it. Similarly, the authors correctly acknowledge that they are assuming subjects assigned a subjective probability of 1.0 to each cue, and that inferential beliefs may have been generated but were unassessed

by the research instrument. It is meritorious that Holbrook et.al. were aware of these assumptions, but it would have been much better to have dealt directly with these issues or, at the least, to explore them via manipulation checks.

A Comparison of Two Behavioral Intention Models

Brinberg's paper compares the Fishbein and Triandis intention models. Results purportedly support Triandis' notion that affective and evaluative components of attitude should be measured separately. Further, contradicting both models, Brinberg finds two dimensions of social influence (whereas both Fishbein and Triandis postulate alternate unidimensional constructs). However, when Brinberg attempts to predict intent from the revealed attitudinal (two) and social influence (two) dimensions, only small predictive improvements are realized.

This is certainly an interesting paper for consumer researchers, but not because the study itself evidences exceptional methodology or creative insight. Rather, its greatest contribution stems from the fact that it brings the Triandis model to the attention of the ACR audience. One major blind spot among marketing researchers concerns their acknowledgement and study of existing non-Fishbeinian type intention models. Unfortunately, most attitude researchers in management departments are totally unaware of the existence of these competing paradigms. Typically at best, only the Fishbein model is cited, and then frequently in erroneous fashion. Hence, Brinberg's conference paper performs the valuable function of publicizing one of the important alternate conceptualizations of intent. Let me take this opportunity to point out other models that consumer researchers should attend: Wickers paradigm incorporating judged influence of extraneous events (e.g., Wicker 1971); Schwartz's model, which discusses moral norms and ascription of responsibility, and Snyder's use of self-monitoring as a moderater variable (cf. Zuckerman and Reis 1978); Sheth's model (Raju, Bhagat and Sheth 1975); and Warshaw's behavioral intention paradigm (1980). While some of these approaches are on stronger conceptual footing than are others, each adds something to our understanding of attitudes and intentions. Hence, consumer research in this area would be enriched considerably by conscientious attention to all competing theoretical tracks.

Concerning Brinberg's research per se, a few limitations seem evident:

1. Disecting and separately measuring both affective and evaluative dimensions of attitude may be necessary only for a restricted class of behaviors. As defined by Triandis and herein applied, affect concerns strong emotional states associated with some criterion act (e.g., disgust, nausea). While these feelings may be germane and present within certain behavioral contexts (e.g., blood donation, hospital volunteer work, purchase of pornographic material), such strong affective responses are unlikely to be associated with most product-related settings. Hence, the author's findings may have limited applicability for the bulk of consumer research. Nevertheless, this is primarily an intuitive hypothesis on my part and is not intended to deprecate Brinberg's results. Certainly, further work is merited on the affect versus evaluation issue.
2. I have strong reservations concerning some of Brinberg's questionnaire items. For example, Fishbein's Aact was measured by the following item: "Donating blood at the blood bank this semester is good-bad, etc.". However, the criterion act concerns subjects' own intentions to donate blood. Clearly, a person might claim that 'donating blood' is good but ascribe non-intent to personally donate blood. The problem

here is that the target of Aact and BI differ. While Aact concerns the general concept of blood donation, BI focuses upon the target person per se. Ajzen and Fishbein (1977) clearly point out the necessity of employing the same target, object, time and context for both Aact and BI.

Similarly, Brinberg employs an imprecise intent measure. That is, he assesses intent by the following item: "I intend to donate blood at the blood bank this semester: likely-unlikely". This measures the likelihood that the subject will intend to donate blood, not the likelihood that he/she will donate blood. Clearly, these are not equivalent, and accurately measuring subjective probability of behavior requires the latter 'will donate' format. To clarify the distinction, think of a person who is self-professedly likely to intend to stop smoking, but admits that the probability of his stopping smoking is meager. Part of the problem here is that Triandis' definition of intent is not directly comparable to Fishbein's parallel measure. Brinberg uses the Triandis measure but then relates both Triandis' and Fishbein's independent variables to this singular construct. Fishbein's subjective probability of behavioral performance is more akin to the objective, derived probability of behavior (Pa) variable in Triandis model (where Pa equals the sum of habits (i.e., the number of times the behavior has been performed in the past) and intent multiplied by a facilitating conditions measure). Hence, to be fair, Brinberg should have employed two alternate items to assess 'intent'.

Finally, I am unconvinced that appropriate scales were used to measure affect (gut feeling). The adjective sets 'enjoyable-disgusting' and 'exciting-nauseating' do not constitute bipolar adjective pairs. 'Disgusting' is not the opposite of 'enjoyable' (in fact, some acts may be enjoyable because they are disgusting). Similarly, exciting and nauseating may be quite unrelated. A better pair, for example, would be exciting-dull. Hence, interpretation of revealed factors is subject to debate.

Hence, the present study has sufficient instrumentation problems to render suspect the thrust of its conclusions. Nevertheless, the issues raised are important and justify additional research employing multiple types of behavior.

Examining the Diagnostic Utility of the Fishbein Behavioral Intention Model

Miniard's piece contains a rather lengthy review of evidence concerning the validity of Fishbein's behavioral intention model. While far from exhaustive, this literature review is certainly the strongest aspect of the paper. Miniard's experiments, on the other hand, are very troublesome. Their stated purpose is to ascertain whether the social norm (SN) measure receives a significant regression weight when social influences 'do affect' intentions but is nonsignificant when social influences don't impact upon intent. However, methodological deficiencies, questionable logic and inattention to recent published research negate the usefulness of these studies.

Regarding Study I, the following problems seem evident:

1. In the 'others expectations unknown' treatment, Miniard claims the situation is devoid of information concerning the social environment. Without sufficient manipulation checks (which apparently weren't done), this is an unwarranted claim. The role playing scenario format makes this a projective technique. As in all such artificial, 'pretend' tasks, a wide range of inferences, projections from past experiences and other situation-structuring cognitions are triggered. Hence, it is unreasonable to assume that subjects perceived a blank, featureless social

environmental context. If one wants to make this rather counterintuitive claim, manipulation check verification is imperative.

2. Similarly, there were no manipulation checks for the "others expectations 'positive' and 'negative'" treatments. Given the nature of the research hypotheses, it is necessary to verify whether subjects believed and accepted these statements as factual truths. If subjects did, in fact, believe these statements, then the subsequent disclaimer, "of course, what others feel you personally should do here is of no concern to you" is highly unlikely to negate fully the normative manipulation. If so, this is not a 'no normative influence' experiment (as Miniard claims) and SN should relate to BI. Since this did not occur, I suspect that subjects were not properly manipulated into perceiving differential 'others expectations'. If so, the treatments are invalid. Nevertheless, this still may serve as a no-influence experiment. Lack of a significant beta weight on SN and n.s. cor (SN, BI) reinforce this conclusion. However, the logic behind the experimental format seems highly questionable. Why bother to manipulate normative influences here? If the sole purpose was to demonstrate that SN is nonsignificant in contexts devoid of social influence, a simple, more tightly controlled design would surely have sufficed.

Essentially, my primary criticism here is that the author makes a lot of claims about what people were thinking and how they approached the task without any sort of verification via debriefing procedures. Without this structure, numerous explanations for the findings are possible.

Study II constitutes a reanalysis of previously published data. Miniard claims that the experimental format implies the existence of social influence. Hence, SN should correlate with BI and carry a significant regression weight. Once again, there are methodological problems and strained logic:

1. Since the role-playing scenarios depicted some hypothetical female buying a dress, why were both male and female subjects employed? This task is so external to the experience of most males that their responses would clearly represent projections of stereotypic 'female' behavior rather than their personal cognitive processes. Females, on the other hand, are likely quite familiar and experienced with the scenario settings. This familiarity probably generated projections from similar experiences in memory. Hence, their responses might be based on factors external to the treatment manipulations. Given that the task was so different for males than females, either only female subjects should have been used (with extensive manipulation checks), or else data should be analyzed separately for males and females.
2. Miniard only manipulated situation-specific information concerning the neighbor. However, NE and MC were additionally measured for the hypothetical woman's husband and parents. This creates several problems. First, SN is supposed to concern subject's perceptions of his/her 'most important others'. Miniard is assuming that neighbor, husband and parents are the 'important set' for each subject. This claim is only justified if verified by manipulation checks, since there is good cause to believe that some or all of these 'referents' may be irrelevant for dress buying. Further, what about the woman's 'best friends'? They might be expected to be highly important others.

Further, Miniard had 57 additional subjects read the scenarios and estimate the impact of the social influences in the scenario on their BIs. Finding differences between NMI and MI, Miniard claims social influence was important in the scenario settings. However, social influence in the treatment paragraphs involved only the neighbor, while the measuring instrument assessed attributions concerning husband, parents and most important others as well as the neighbor. To say the neighbor exerted social influence is not equivalent to saying that a measure which incorporates the neighbor and several other referents is related significantly to BI. Hence, there is no compelling justification for Miniard's hypothesis that SN should receive a significant beta weight. If SN only concerned the neighbor, then this postulate might be warranted. But, SN includes attributions to many others besides the neighbor, and in a projection task, these other attributions are unlikely to remain constant across treatments. Miniard, unjustifiably, implicitly assumes this constancy condition.

Hence, Miniard's claim that versions 1 and 2 of the FBI model fail to correctly identify the importance of normative influences is highly suspect. Since SN includes important others, the neighbor may or may not be included in this set. Further, since no information concerning husband, parents and other friends was given, SN is probably largely a projection of one's own A_B , which would imply high $cor(A_B, SN)$, where A_B dominates and thus receives the significant beta weight. The nonsignificance of model 2 simply shows that general MC is unrelated to influence in this particular situation (which is not surprising given lack of general information about any member of the offered referent set). The significant results in models 3 and 4 support this analysis (in each case there was significant $cor(SN, BI)$ and SN had a significant beta). That is, treatment information was situation-specific, and when MC was similarly specific, having only data on the neighbor gave this referent a dominant influence on SN, which then behaved in line with expectations. The 57 subject influence check implied that this would occur, but did not imply that when general SN or MC were employed, significance would result. Only when measures were specific enough to capture the neighbor's stated influence in these specific treatment situations was SN likely to be significant - and it was.

The only point Miniard really makes is that SN and A_B are typically correlated, which questions the assumption that A_B and SN are independent components of intent. However, this issue has been covered extensively in Warshaw (1980) and elsewhere. The real issues, such as separating out identification, internalization and compliance aspects of normative influence were not even touched upon by Miniard.

References

- Ajzen, I. and Fishbein, M. (1977), "Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research", Psychological Bulletin, 84, 888-919.
- Holbrook, M. and Maier, K. (1978), "A Study of the Interface Between Attitude Structure and Information Acquisition Using a Questionnaire-Based Information-Display Sheet", Proceedings, Association for Consumer Research, Eighth Annual Conference, 93-8.
- Kassarjian, H. (1971), "Personality and Consumer Behavior: A Review", Journal of Marketing Research, 8, 409-14.

Raju, P., Bhagat, R. and Sheth, J.(1975), "Predictive Validation and Cross-Validation of the Fishbein, Rosenberg and Sheth Models of Attitudes", in M. Schlinger (ed.), Advances in Consumer Research, Vol. II, Chicago: Association for Consumer Research.

Warshaw, P. (1980), "A New Model for Predicting Behavioral Intentions: An Alternative to Fishbein", Journal of Marketing Research, 17, 153-72.

Wicker, A. (1971), "An Examination of the 'Other Variables' Explanation of Attitude-Behavior Inconsistency", Journal of Personality and Social Psychology, 19, 18-30.

Zuckerman, M. and Reis, H. (1978), "Comparison of Three Models for Predicting Altruistic Behavior", Journal of Personality and Social Psychology, 36, 498-510.

ATTITUDE RESEARCH AND BEHAVIORAL INTENTIONS:
A CRITICAL REVIEW

Jeffrey E. Danes, Virginia Polytechnic Institute and State University

Abstract

This paper reviews the work by Holbrook, Velez, and Tabouret, "Attitude Structure and Search: An Integrative Model of Importance-Directed Information Processing," Brinberg, "A Comparison of Two Behavioral Intention Models," and Minard, "Examining the Diagnostic Utility of the Fishbein Behavioral Intentions Model."

Attitude Structure and Search

Holbrook, Velez, and Tabouret begin their paper with the following statement, "In spite of extensive research on multiattribute attitude models and information-acquisition paradigms, the linkage between these two important facets of buyer behavior is not well understood." Later on they assert, "...Multiattribute Attitude Research has tended to focus primarily on the static structure of attitude while neglecting the process by which new information inputs are incorporated into the determination of affect" (emphasis added).

At the conceptual level, Holbrook, et al move away from a static description and provide a process, dynamic view of the linkages between information and attitude. One would expect, therefore, that dynamic information-attitude models would be presented to compliment their conceptual work. Unfortunately, this was not the case. The four algebraic multiattribute attitude models appear as static representations of attitude. The modeling work provided by Holbrook, et al could be improved if dynamics were explicitly incorporated into their four models. We now address this issue.

A Dynamic Representation

Holbrook, et al presented four multiattribute models to predict attitude after individuals have acquired information (messages) about a product: a full and partial additive model and a full and partial averaging model. What is the attitude value before information (messages) is received? Holbrook, et al do not address this issue, and neither do they identify or develop multiattribute models that gives us the pre-information search (pre-message) attitude. It is argued below, however, that they implicitly assume that pre-message attitude is given by Fishbein's model (or an "averaging" variant of his model). Without loss of generality, let us restrict our discussion to Holbrook's, et al partial additive model.

After individuals acquire information (receive messages) about the object; Holbrook, et al, assumed that if the message, m , associates an attribute with an object, the new belief value equals 1. Or, if the message disassociates an attribute with an object, the new belief value equals -1. Evaluation, e , is not affected by the messages. The post-message attitude as given by Holbrook's, et al, partial additive model may be expressed as:

$$\text{Post message } a = \sum_k \text{ post-message } b_k e_k \text{ acq}_k, \quad (1)$$

where, a is the attitude toward the object, b is belief, e is evaluation, and acq is the information (message) variable which is coded 1 or 0 depending upon whether the message was received.

$$\text{acq} = \begin{cases} 1, & \text{if information was acquired} \\ 0, & \text{if otherwise} \end{cases} \quad (2)$$

As stated above the attitude, a , reported in Equation 1 is a post-message or "new" attitude. Also, the belief values are post-message or "new" beliefs. With new notation, let us rewrite Equation 1:

$$\text{New } a = \sum_k \text{ new } b_k e_k m_k, \quad (3)$$

where the message, m , is used denote acquired information, acq . Both new a and new b may be expressed as their pre-message values plus the change, i.e.,

$$\text{new } a = a + \Delta a \quad (4)$$

$$\text{new } b = b + \Delta b \quad (5)$$

where a and b is the pre-message attitude and belief, and Δa and Δb is the change in attitude and belief. If we make the appropriate substitutions, Equation 3 may be expressed as:

$$a + \Delta a = \sum_k (b_k + \Delta b_k) e_k m_k \quad (6)$$

or

$$a + \Delta a = \sum_k b_k e_k m_k + \sum_k \Delta b_k e_k m_k \quad (7)$$

One perplexing problem with the Holbrook et al model is that the message, m , is entered twice in Equation 7. Note here that the pre-message attitude,

$$a = \sum_k b_k e_k m_k, \quad (8)$$

includes the message which the individual has not yet received! It makes more sense to define the pre-message attitude as:

$$a = \sum_k b_k e_k. \quad (9)$$

As implied by Holbrook, et al by Equation 7, change in attitude may be expressed as:

$$\Delta a = \sum_k \Delta b_k e_k m_k, \quad (10)$$

where, Δb stands for change in belief.

How Do Beliefs Change?

Holbrook, et al, explicitly assume that a message changes belief in the direction of the object-attribute association made in the message. They also assume maximum change; hence, the post-message belief is equal to the belief-value communicated in the message. Thus, Holbrook, et al, have simply re-stated Anderson and Hovland's (1957) "Distance-Proportional" Model:

$$\text{New } b = b + \alpha(m - b), \quad (11)$$

or

$$b + \Delta b = b + \alpha(m - b), \quad (12)$$

or

$$\Delta b = \alpha(m - b). \quad (13)$$

Additionally, since the authors assume maximum change, the proportionally constant, α , must equal 1.

$$\Delta b = m - b \quad (14)$$

Replacing $(m - b)$ for Δb in Equation 10 gives:

$$\Delta a = \sum_k (m_k - b_k) e_k m_k, \quad (15)$$

Once again there are two message terms. The first message term was derived directly from Holbrook et al.'s assumptions; the last message term appears to be unnecessary and may be deleted:

$$\Delta a = \sum_k m_k e_k - \sum_k b_k e_k. \quad (16)$$

The model presented in Equation 16 is more consistent, than is Equation 15, with Holbrook's et al. logic and assumptions. More importantly, if we began with Equation 9 and worked to build a dynamic model of the information-attitude process, we would arrive directly at Equation 16--given the belief change assumptions made by Holbrook, et al.

Holbrook, et al. was interested in attitude after messages were received. The "new" or post-message attitude may be obtained by adding Equations 9 and 16; this gives:

$$\text{new } a = \sum_k m_k e_k, \quad (17)$$

where, $m_k = \begin{cases} 1 & \text{if message associates object with attribute,} \\ -1 & \text{if message disassociates object with attribute.} \end{cases}$

Equation 17 produces a numerical result identical to Holbrook's et al. partial additive model. However, the dynamic assumptions made by Holbrook, et al. are explicitly stated in the new model.

Summary

Holbrook, et al. presented a promising conceptualization of the dynamic relation between information acquisition (or message reception) and attribute formation/change. The four algebraic models presented, however, failed to capture the dynamic assumptions asserted in their conceptual work. Our discussion centered upon translating Holbrook's et al. assumptions about the dynamics of the information (message) - attitude process into change equations isomorphic with their assumptions.

Two Behavioral Intention Models

Bringberg's paper compares two models for the prediction of behavioral intentions: the extended Fishbein model and Triandis' model. The major differences between the two models are (1) Fishbein assumes "affect" and "evaluation" are indicators of different constructs; (2) Fishbein asserts "subjective norms" influence behavioral intentions and Triandis argues for "social determinants."

To compare the two models, Brinberg compares the factor structure of the predictor variables. He argues against a standard multiple regression analysis "...since the regression technique does not allow for the effects of measurement error, a researcher may conclude that two (or more) constructs exist, when, in fact the constructs are imperfect measures of the same underlying dimension." This, indeed, is an interesting criticism, for Bringberg puts the blame on regression when the blame ought to be placed elsewhere--on measurement. If Brinberg had obtained reliability coefficients for his scales, he could have used the corrected correlations and standard deviations in his regression analyses. Correlations and standard deviations that are corrected for attenuation (unreliability) are "free" of measurement error. If only standardized regression coefficients are of interest, only the correlation matrix need be corrected for attenuation. The more obvious problems associated with multicollinearity were not addressed.

"To circumvent some of the limitations of the regression technique ..." Brinberg relies on a principle components orthogonal factor analysis (with varimax rotation) of the predictor variables. Two factor analyses were performed, one for evaluation, affect, and consequences. And, one for Triandis' social components and Fishbein's normative components. In each case, two factors were obtained. The first two-factor solution was interpreted as support for Triandis' Hypothesis that the affective and evaluative components are not the same. The second two-factor solution was interpreted as "what relevant others think" and as "what is appropriate to do." Hence, Brinberg argues that both Fishbein and Triandis have misspecified the "social influence" component of behavioral intentions. Warshaw (1981) has criticized the scales used for the measurement of "affect" and "evaluation;" the same criticism may be made for the "social influence" constructs. Brinberg does not cite any evidence for the validity of his measures.

At this point a few questions become apparent:

- (1) Why was a Principal Components Factor analysis used?
- (2) Why was the Factor Analysis constrained to be orthogonal?
- (3) Would an oblique solution yield highly correlated factors?
- (4) What are the relationships between "affect," "evaluation," "social determinants," "subjective norm," and behavioral intentions?

As to the first question ("Why was a principal components factor analysis used") a principal components analysis makes assumptions about the data which may not be true--it assumes perfect measurement. The principal components analysis operates on a correlation matrix with ones in the diagonal; i.e., it assumes the variables are measured without error. Other factor techniques do not make this assumption; perhaps, they should have been explored.

For the second and third questions ("Why was the Factor Analysis constrained to be orthogonal," and "Would an oblique solution produce highly correlated factors") there appears to be no justification for constraining the factors to be uncorrelated. We suspect that "affect" and "evaluation" are highly correlated; likewise that "social determinants" and "subjective norms" are also highly correlated. Brinberg could have performed an Oblique Multiple Groups Factor Analysis (Harmon 1967) and the "natural" factor correlations could have been found. Assuming uncorrelated factors is arbitrary and unjustified.

The last question ("What are the relationships between 'affect,' 'evaluation,' 'social determinants,' 'subjective

norms,' and 'behavioral intentions') could be found using a variety of highly related methodologies: (1) an oblique multiple groups factor analysis followed by a path analysis performed upon the factor correlation matrix of the five constructs, or (2) a multiple indicator path analysis, or to use the newer terminology, a structural equation (causal model) analysis with latent variables.

In summary Brinberg attempted to compare two models for the prediction of behavioral intentions: the extended Fishbein model and Triandis' model. For the comparison of the two models, Brinberg used an orthogonal principle components factor analysis on the predictor variables. Their results were interpreted as support for Triandis' notion that "affect" and "evaluation" are separate constructs; that the "social influence" construct consists of two dimensions: "what relevant others think," and "what is appropriate to do."

Several methodological problems were discussed, including:

- (1) Reliability and validity of measurement;
- (2) Choice of factor analysis technique;
- (3) Use of orthogonal solution;
- (4) Structural equations analysis as an alternate analytic tool.

The Diagnostic Utility Of The Fishbein Model

The purpose of Minard's paper is to assess whether the subjective norm component is detectable when it "is present" and when "it is not present." Minard reports two studies. The first experiment was designed to be void of normative influences; the primary task of this experiment was to assess whether normative influences could be detected. The results of his experiment supported the extended Fishbein model. There are, however, several methodological problems associated with the first study:

- (1) No manipulation checks on the experimental treatments were made; hence, we have no guarantee that normative influences were actually absent;
- (2) The reliability of measurement (apparently) was not assessed. Unreliability may have distorted the obtained omega squared's. (Below we demonstrate such distortion using regression.)

Minard's second study reports a reanalysis of his previous published work. In this study, he asserts that normative influences were present. Two out of four regression analyses, however, failed to detect them. There are three apparent problems with the second study:

- (1) The manipulation may not have been relevant to males.
- (2) The neighbor, husband, and parent significant others were arbitrarily selected. No evidence is given justifying this choice.
- (3) Reliability of measurement is not assessed; the regression analyses were not corrected for unreliability.

The last criticism appears to be widespread in consumer research; thus, let us focus our attention on the impact measurement error may have upon multiple regression. As stated above, two of four regression models failed to detect the expected subjective norm influence. The four models differed in their operationalization of subjective norm. One model contained a direct measure of subjective norm, SN; the others varied in terms of specificity: general, moderate, and situation specific. The direct and general

models failed to receive support. For the direct measure of subjective norm, SN, the β coefficient equaled .08; and the f^2 coefficient for attitude toward behavior, A_b , equaled .55. Unfortunately, Minard does not provide us with his correlation matrix; however, the matrix supplied in Table 1 will be used for discussion.

TABLE 1

Example correlation matrix for behavioral intention, attitude toward behavior, and subjective norm

	BI	A_b	SN
BI	1.00	.612	.463
A_b	.612	1.00	.683
SN	.463	.683	1.00

From Table 1, the following regression weights may be derived: $\beta_{A_b} = .555$ and $\beta_{SN} = .084$. It is assumed, however, that all three measures are perfectly reliable. Could measurement unreliability account for the nonsignificant weight for subjective norm? Indeed it could, and we show such an example below. Minard did not provide reliability estimates for his measures; so let us, for discussion purposes, assume that the reliabilities are: .64, .98, and .68 for behavioral intentions, BI, attitude toward behavior, A_b , and subjective norm, SN. If we replace the diagonal of ones with the reliability coefficients and correct (for attenuation) the off-diagonal correlations, we arrive at the corrected correlation matrix reported in Table 2.

TABLE 2

Corrected correlation matrix for behavioral intention, attitude toward behavior, and subjective norm. Reliabilities are in parentheses.

	BI	A_b	SN
BI	(.640)	.773	.702
A_b	.773	(.980)	.834
SN	.702	.834	(.680)

Does unreliability influence the regression weights? It certainly does! The corrected β coefficient for subjective norm now equals .19 and the corrected β coefficient for attitude toward behavior now equals .62; both coefficients may be significant using Minard's degrees of freedom. Since Minard did not estimate the reliability of measurement or correct his analyses for unreliability, we are unable to assess how much his regression weights were distorted.

In summary, Minard proposed an interesting test of the extended Fishbein model. He presented two studies. The first study was designed to be void of subjective norm information. In this experiment, the extended Fishbein model accurately detected the lack of a normative influence. The second study was designed to produce normative influences; and only two of four tests were able to detect the expected influence. Our discussion centered upon several methodological problems:

- (1) Failure to use manipulation checks or possible irrelevant manipulations;
- (2) Failure to assess reliability of measurement and correct analyses for unreliability.

References

Anderson, Norman H. and Hovland, Carl I. (1957), "The Presentation of Order Effects in Communication Research," in The Order of Presentation in Persuasion, ed., Carl I. Hovland, New Haven, CT: Yale University Press, pp. 158-69.

Brinberg, David (1981), "A Comparison of Two Behavioral Intention Models," in Advances in Consumer Research, ed., Kent B. Monroe, Washington, D.C.: Association for Consumer Research.

Harmon, Harry J. (1967), Modern Factor Analysis, Chicago: The University of Chicago Press.

Holbrook, Morris B., Velez, David A., and Tabouret, Gerard R. (1981), "Attitude Structure and Search: An Integrative Model of Importance-Directed Information Processing," in Advances in Consumer Research, ed., Kent B. Monroe, Washington, D.C.: Association for Consumer Research.

Warshaw, Paul R. (1981), "A Discussion of Attitude Research and Behavioral Intentions," in Advances in Consumer Research, ed., Kent B. Monroe, Washington, D.C.: Association for Consumer Research.

AN EXPLORATION INTO THE SCALING OF CONSUMERS' INVOLVEMENT WITH A PRODUCT CLASS¹

Peter H. Bloch, Portland State University

Abstract

Product involvement as an explanatory or moderating variable with respect to consumer behavior has recently become a topic of significant interest to marketers. However, the usefulness of the construct in consumer research is yet to be fully realized due to a lack of agreement and rigor in operationalizing product involvement. This paper addresses the need for valid measures of individual differences in involvement for particular product classes and reports on the psychometric development of a prototype scale to assess involvement with automobiles. Findings provide preliminary support as to the scale's reliability and validity.

Introduction

The ways in which consumers are involved with the products they purchase and use has, in recent years, become a subject of increasing interest to consumer researchers. This interest is evidenced by the special sessions on the topic of involvement at the American Marketing Association's 1977 Attitude Research Conference and at the 1978 Conference of the Association for Consumer Research. In addition, journal articles dealing with some aspect of involvement have become increasingly common (e.g., Rothschild 1979b, Tyebjee 1979b, Webb 1979). While in these and other studies, product involvement has been defined in various ways, the definition proposed by Day (1970) is characteristic of several in the consumer literature. Day defined involvement as "the general level of interest in the object or the centrality of the object to the person's ego-structure" (p. 45). Several writers (Bogart 1967, Mitchell 1979) have agreed that involvement reflects the amount of interest evoked by the product while, Day's centrality notion has found support from Tyebjee (1979a) and others (DeBruicker 1979, Houston and Rothschild 1978, Lastovicka and Gardner 1979) who suggested that involvement occurs when a product is related to important values, needs or the self-concept.

To date, involvement researchers generally agree that for consumers as a group, products differ in their tendency to arouse involvement. For example, Robertson (1976) has emphasized the low level of involvement generated by many consumer products. In addition to differences across products, several writers (Houston and Rothschild 1978, Lastovicka and Gardner 1979, Tyebjee 1979a) have suggested that for any particular product class, levels of involvement will differ across consumers. The focus of this research is on this between-individuals perspective.

Reviewing the research on individual differences leads to the conclusion that highly involved individuals engage in more complex purchase decision making (Houston and Rothschild 1978), are more brand loyal (Jacoby 1971), are more likely to be opinion leaders (Corey 1971) and are more likely to generate negative cognitive responses to product-related messages (Houston and Rothschild 1978, Wright 1973) than do persons with low involvement. Given this research stream, product involvement appears to be not only a potentially useful moderating or explanatory variable with respect to consumer behavior, but also a possible basis for segmenting markets.

Despite the promising nature of involvement in consumer research, several authors have noted that its explanatory

¹The author thanks Marsha L. Richins and Robert A. Peterson for their comments and contributions to this work.

potential is, as yet, unrealized due to a lack of consensus and rigor in operationalizing the construct (Mitchell 1979, Ray 1979, Tyebjee 1979b). Although many measures of individual differences in involvement have been proposed since the concept's introduction to the marketing literature, few have been subjected to standard reliability and validity testing and at the present time there is no single, well-accepted involvement measure.²

In light of the prominent measurement difficulties relating to product involvement research, Mitchell (1979) has argued that the "first priority" in future field research in involvement is the development of a scale to measure consumers' involvement with a particular product class or brand (p. 194). He went on to say that only when such a scale is developed and subjected to traditional psychometric testing can product involvement be fruitfully related to other aspects of consumer behavior.

Given this call for action, the purpose of this paper is to report on an attempt to scale consumers' involvement with a particular product class and to provide a prototype involvement scale as one step in the further refinement of the involvement construct. This paper next reports on the development of a scale to measure consumer involvement with automobiles.

Choice of Product to be Studied

In developing an involvement scale, the first critical decision that a researcher faces is whether the measure should be applicable to several products or to a single product. As part of their work on the dimensionality of a product involvement, Lastovicka and Gardner (1979) devised a scale able to tap individual differences in involvement for any product class. The benefits of their measure's generalizability appear to be balanced, however, by the imprecise character of its constituent items. Many of the Lastovicka and Gardner scale items appear ambiguous and may not capture how a consumer really feels about any one product. Examples include: "I can make many connections or associations between experiences in my life and this product" and "use of this product helps me behave in the manner that I would like to behave" (p. 63). When scale items are specific to a single product class, they would appear to be more relevant and meaningful.

In a critique of involvement measures, Ray (1979) posited that such measures should be developed "in individual consumer research application situations" (p. 198). He cited the involvement measure developed by Rothschild (1974) in a study of political races and that used by Webb (1979) in research on a television clutter as particularly effective in that they were tailored to specific research situations. In a similar vein, Mitchell (1979) suggested that future involvement research might be based on the use of a scale(s) measuring the degree of interest or arousal evoked by a particular product class or brand.

A further argument for building a product-specific measure rests on the state of current understanding of the construct of product involvement. As noted by several authors (Lastovicka and Gardner 1979, Houston and Rothschild 1978, Mitchell 1979, Tyebjee 1979b), the literature on involvement

²For a fuller treatment of issues and problems relating to the measurement of involvement, see Mitchell 1979, Houston and Rothschild 1978, and Ray 1979.

is fragmented and for the most part based on measures of dubious quality. Therefore, because involvement theory remains in a developing state, the search for relationships between involvement and other aspects of consumer behavior appears to be worthwhile even if in a product-specific context.

Once the product-specific focus was chosen, the next decision concerned the choice of the product to be studied. After reviewing several possible options, automobiles were selected as the product class of interest due to the observation that car owners exhibit wide variation in involvement. For instance, a person who has minimal involvement with cars may drive a plain model and treat the vehicle essentially as an appliance, whereas another person with high involvement (such as in the case of car enthusiasts) may spend his weekends tinkering with cars and subscribe to several automotive magazines. This potential for variance and the familiarity of the product class to most consumers made automobiles a reasonable choice for scale development efforts.

Psychometric Development

The literature on scaling presents a bewildering array of possible approaches to constructing a multi-item scale. The particular approach employed in this project is that presented by Churchill (1979) in an article on the development of measures for marketing constructs. The sections which follow discuss the steps used in specifying the domain of the construct, generating scale items, measure purification procedures, and the analysis of the scale's structure, as well as the reliability and validity of the final involvement scale.

Specifying the Domain of the Construct

The first step in developing a scale is to specify the domain of the construct under study, based on a review of relevant literature (Churchill 1979). A review of conceptual work on product involvement (e.g., Bogart 1967, Day 1970, Houston and Rothschild 1978, Lastovicka and Gardner 1979, Tyebjee 1979a), indicated the construct has generally been conceptualized in one of two ways. First, involvement has been treated as a temporary concern with high risk products which occurs during the purchase process (Bowen and Chaffee 1974, Houston and Rothschild 1978). Product involvement has also been viewed as a long-term interest in a product which is based on the centrality of the product to important values, needs, or the self-concept (Day 1970, DeBruicker 1979, Houston and Rothschild 1978, Tyebjee 1979a). According to Houston and Rothschild (1978) this second form of involvement is primarily a function of individual differences and as such, is appropriate for present scaling efforts. The temporary or situational form of involvement will not be studied here because it appears to offer little beyond another perspective on perceived risk.

As used in this research, product involvement is a construct which affects consumer behavior on an ongoing basis. Further, involvement varies across individuals, ranging from minimal levels to the extremely high levels exhibited by consumers such as car enthusiasts, wine connoisseurs, or camera buffs. Procedures used to generate scale items appropriate to the above conceptualization of product involvement are discussed in the following section.

Generating Scale Items

In order to achieve content validity for the scale, the item development process was also based on the review of the product involvement literature. In addition, interviews were held with automobile dealers and sports car enthusiasts, persons presumed to be highly interested and involved in cars and therefore, able to shed light on the nature of the construct. As a final source of scale item

ideas, advertising and editorial content of various automotive magazines were reviewed.

This exploratory work resulted in the construction of 66 statements which captured several aspects of involvement noted in the literature. Items dealt with (1) interest in cars (Bogart 1967, Day 1970, Mitchell 1979) and accompanying readiness to talk about cars (Bogart 1967, Corey 1971), (2) the relatedness of cars to important needs or values (Houston and Rothschild 1978, Tyebjee 1979a), and (3) use of one's car to express the self-concept (DeBruicker 1979, Lastovicka and Gardner 1979). Other items which were developed from the experience survey tapped highly product-specific domains such as willingness to perform car maintenance and devotion to the automobile in light of energy and ecological pressures.

The next step in item development involved editing the raw statement set. A group of six judges, consisting of both faculty and doctoral students from the areas of consumer behavior and social psychology, evaluated the 66 items for content validity (i.e., whether or not the items appeared to tap an abiding interest and concern with automobiles). Items which were deemed irrelevant were eliminated along with those which did not produce a consensus as to their direction of scoring. A 44-item preliminary scale emerged from the item development stage.

The scale consisted of items scored according to a six-point Likert-type format. This is the type of scale recommended by Houston and Rothschild (1978) for use in assessing individual differences in product involvement. Responses are scored so that low scores represent high involvement with negatively stated items to be reverse-scored prior to analysis. An individual's scale score is obtained by summing across the items.

This preliminary version of the involvement scale was administered to a group of 381 marketing students in large classroom settings. Classroom administrations were utilized for reasons of convenience and to obtain a nearly perfect response rate. This response rate benefit becomes salient in the case of research on involvement with a single product in that involvement would appear to be a determinant of response. For example, it seems reasonable that individuals highly involved with a product are more likely to complete a mail survey or personal interview that deals with that product than are subjects with low involvement. The responses to the 44-item scale were used in the next phase of scale development, measure purification.

Measure Purification

The major thrust in purifying a multi-item scale is to reduce, as needed, the number of scale items, with the goal of producing an internally-consistent instrument with highly intercorrelated items. This argument is based on the domain sampling model of measurement (Nunnally 1967) which "holds that the purpose of any particular measurement is to estimate the score that would be obtained if all the items in the domain were used" (Churchill 1979, p. 68). When all the items of a scale are highly intercorrelated, it may be concluded that they are drawn from the domain of a single construct.

The most widely accepted indicator of a scale's internal consistency is Cronbach's (1951) coefficient alpha (α) which represents the average of all possible split-half reliability coefficients. Alpha will tend to be large when scale items are highly intercorrelated and thus appear to be tapping the same construct.³

In this phase of scale development, an iterative procedure was followed where alpha was calculated, items with low

³For a fuller discussion of Cronbach's alpha, see Peter 1979.

item-total correlations were discarded, followed by a subsequent alpha calculation. The 44-item preliminary scale produced an alpha of .82. Although this alpha greatly exceeds Nunnally's (1967) criterion for acceptable internal consistency of .50 to .60, further analysis revealed that several scale items had low item-total correlations. A series of alpha calculation iterations produced a set of 17 items which attained an alpha of .83. While this appears to be only a marginal increase in internal consistency, it should be noted that alpha is affected by scale length with more items tending to increase alpha levels (Cronbach 1951, p. 313). For purposes of comparison, the alpha level of the 17-item set was corrected to make it equivalent in length to the 44 item set. Using the formula below, if the 17 item set was lengthened to 44 items while maintaining the same internal consistency, alpha would be .93 (Nunnally 1967, p. 223). This corrected alpha indicates that the

$$\alpha_c = \frac{k \cdot \alpha}{1 + (k-1)\alpha} \quad (1)$$

where: α_c = α corrected for scale length

k = number of times the scale is to be lengthened

α = uncorrected alpha

measure purification process not only produced a more parsimonious set of items, but also one which possesses a significantly higher level of internal consistency. The 17 items along with their item-total correlations are displayed in Table 1.

Scale Structure

Responses to the reduced set of items shown in Table 1 were next used in a principal components analysis to investigate the dimensionality of the scale and shed further insight on the nature of the involvement construct. Although many scale development articles perform principal components or factor analyses at the outset and prior to measure refinement work, Churchill (1979) noted that in such instances, a large number of difficult to define dimensions tend to emerge due to the presence of irrelevant items. It is Churchill's position that principal components analyses are more profitably employed after an internally consistent set of items is obtained via the assessment of coefficient alpha.

In performing the principal components analysis, the eigenvalue ≥ 1.00 criterion was used to limit the number of factors extracted. Table 2 presents the varimax rotated factor loadings matrix. After examining the content of those items with the highest loadings, the six factors which emerged may be interpreted as follows:

Factor 1: Enjoyment of driving and usage of cars

Factor 2: Readiness to talk to others about cars

Factor 3: Interest in car racing activities

Factor 4: Self-expression through one's car

Factor 5: Attachment to one's car

Factor 6: Interest in cars

These dimensions, although specific to automobiles, appear to capture various aspects of product involvement well established in the consumer behavior literature. Factors 1, 4, and 5 capture the centrality orientation of involvement, while Factors 2, 3, and 6 concern product interest. The dimensionality of product involvement originally used to generate the scale items appears to have reasonable empirical support.

TABLE 1
17 Item Involvement Scale

Item Number	Statement	Item-Total Correlation
1.	It is worth the extra cost to drive an attractive and attention-getting car	.344
2.	I prefer to drive a car with a strong personality of its own	.407
3.	I have sometimes imagined being a race driver	.426
4.	Cars offer me relaxation and fun when life's pressures build up	.566
5.	Sometimes I get too wrapped up in my car	.467
*6.	Cars are nothing more than appliances	.395
7.	I generally feel a sentimental attachment to the cars I own	.369
8.	Driving my car is one way I often use to relieve daily pressures	.511
*9.	I do not pay much attention to car advertisements in magazines or on TV	.351
*10.	I get bored when other people talk to me about their cars	.450
*11.	I have little or no interest in car races	.404
12.	Driving along an open stretch of road seems to "recharge" me in body, mind, and spirit	.458
13.	It is natural that young people become interested in cars	.348
14.	When I'm with a friend, we often end up talking about cars	.385
15.	I don't like to think of my car as being ordinary	.449
16.	Driving my car is one of the most satisfying and enjoyable things I do	.530
17.	I enjoy discussing cars with my friends	.519

*Reverse scored prior to tabulation

Reliability Assessment

Although the preceding analyses produced a scale with high internal consistency, Churchill recommends that alpha be calculated on a new sample to assess the scale's internal consistency reliability. Using a new sample of 57 students, responses to the 17-item scale produced an uncorrected alpha of .79. This figure is slightly less than that calculated with the larger sample; however, its magnitude is still well above the .50-.60 level of acceptability, particularly in light of scale length.

Internal consistency reliability, estimated by coefficient alpha, assesses the measurement error which is attributable to scale content. Alpha does not estimate measurement error or unreliability caused by factors which are external to the scale, such as differences in administration situations (Churchill 1979). As a further test of scale

TABLE 2
Varimax Rotated Factor Matrix^a

Item No.	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	.103	-.022	.033	[.702] ^b	.070	.266
2	.066	-.003	.307	[.724]	.199	.018
3	.107	.188	[.774]	.190	.165	-.077
4	[.706]	-.068	.329	.194	.157	.182
5	.145	.318	.227	-.087	[.644]	.049
6	.271	.014	-.079	.311	.137	[.553]
7	.146	-.154	.057	.155	[.695]	.194
8	[.664]	-.003	.097	.116	.452	.217
9	.057	.063	.072	.091	.081	[.815]
10	-.004	[.511]	.235	.020	.045	[.610]
11	.118	.204	[.778]	-.019	.010	.201
12	[.737]	.205	.034	.234	-.059	.045
13	.310	.359	-.231	[.500]	.033	.036
14	.123	[.788]	.111	.005	.079	.025
15	.103	.212	-.027	.308	[.697]	-.013
16	[.640]	.344	.000	.063	.337	-.042
17	.121	[.746]	.284	.074	.115	.169
Explained Variance:						
	27.7%	9.9%	7.9%	6.9%	6.3%	5.9%
Cumulative:						
	27.7%	37.7%	45.5%	52.4%	58.7%	64.6%

a - 381 cases

b - [] indicates factor loading of .50 and above

reliability the instrument was readministered to the sample of 57 students after a two week interval. Although the assessment of test-retest reliability is characterized by serious problems brought about by memory and history effects (Nunnally 1967, Peter 1979), this test was used to provide at least some indication of the scale's stability over time.

These two administrations produced a Pearson product moment correlation of .78 ($p < .001$) for summed scale scores. A correlation of this magnitude is evidence of satisfactory test-retest reliability.

Scale Validity

The final stage of the scale development process concerned the initial validation of the involvement scale. Validity deals with the extent to which a scale is measuring what it purports to measure. As Allison (1978) noted,

...for psychological traits, validity is a matter of degree rather than an all-or-none characteristic. Moreover, it must be evaluated over time and various situations, especially in the case of construct validity of a newly developed test (p. 571).

Keeping this in mind, the analyses which follow can only be regarded as providing a beginning in the validation process for the proposed involvement scale.

To assess its criterion validity, the involvement measure was administered to a group of 90 students along with a set of demographic questions and measures of reported behavior. The behavioral measures tapped the frequency of participation in a variety of activities presumed to be associated with high involvement with automobiles. The behavioral items were developed from a review of automotive magazines and depth interviews with members of sports car clubs. Frequency of participation was measured on a six-point scale, ranging from "always" to "never," with lower scores indicative of higher involvement. Frequency of pleasure driving was also investigated using a multichotomous question with six choices ranging from "more than once a week" to "never or almost never." Table 3 presents the correlations of the scale scores and the reported behavioral items. The proposed involvement scale correlated significantly in the hypothesized direction with all of the reported behavioral measures providing evidence of the scale's criterion validity.

TABLE 3
Correlations Between The Involvement Scale
And Reported Behavioral Measures^a

Reported Behaviors ^b	Pearson r
Purchase of name brand automotive supplies	.401 ($p < .001$)
Visit car dealers to see new models when they are introduced	.362 ($p < .001$)
Attend automobile races and shows	.356 ($p < .001$)
Perform own car repairs	.277 ($p < .01$)
Correct nicks and scratches in car's finish with touch-up paint	.239 ($p < .05$)
Perform periodic car maintenance earlier than required by owner's manual	.229 ($p < .05$)
Park away from other cars when in large parking lots to avoid door nicks	.185 ($p < .05$)
	<u>Spearman rho</u>
Take car on pleasure drives	.418 ($p < .001$)

^a n = 90

^b questions tapped frequency of engaging in each behavior

Several researchers have suggested that one measure of a scale's construct validity is whether it can differentiate the positions of groups known to differ on the construct of interest (Churchill 1979, Lundstrom and Lamont 1976). To further validate the scale and determine if it indeed behaves as expected, such an analysis was undertaken.

In order to perform this analysis, the involvement scale was administered to 52 members of sports car clubs. Members of such clubs typically attend ten or more automotive events (e.g., races, rallies, car shows) per year, subscribe to several automotive magazines, and have a considerable financial investment in the ownership of several automobiles. Thus, one can assume these individuals have high enduring involvement with automobiles. The scale was administered at the regular monthly meetings of the car clubs. Scores obtained from the car club members were contrasted with those of the 90 students used in the previous analysis. The student sample was assumed to possess a

moderate degree of interest and involvement in cars relative to the car club members. The mean score for the student sample was 59.01 contrasted with 37.39 (a higher level of involvement) for the car club group. Using a t-test, the difference between the two means was significant at $p < .001$. In addition, significant differences were also found for all the reported behavioral items.

Conclusions

The results presented above indicate that the construct of product involvement is scalable for a particular product class using conventional procedures. Furthermore, the scale presented is a prototype for others which may be constructed for different research applications and it appears to stand up well to traditional psychometric testing procedures. Although the scale presented is applicable only to the product class of automobiles, the product-specificity should not limit its usefulness in efforts to further refine the construct of involvement. The scale would appear to have application to a number of research situations either as a predictor variable or as a covariate. For example, involvement may be related to other consumer behavior concepts such as opinion leadership, perceived risk, innovativeness, brand loyalty, or information processing.

While the above analyses provide preliminary support for the worthiness of the proposed involvement scale, one caution is in order. This scale was developed using student samples, and further testing with other populations and survey modes is needed to further establish the scale's validity.

In summary, this paper has presented a scale to assess individual differences in consumers' involvement with the product class of automobiles as a first attempt to directly scale involvement in a particular product class. Psychometric tests of the measure's reliability and validity seem to provide adequate support that use of the scale in future research deserves consideration.

References

- Allison, Neil K. (1978), "The Psychometric Development of a Test for Consumer Alienation from the Marketplace," Journal of Marketing Research, 15, 565-575.
- Bogart, Leo (1967), Strategy in Advertising, New York: Harcourt, Brace Jovanovich.
- Bowen, Lawrence and Chaffee, Steven H. (1974), "Product Involvement and Pertinent Advertising Appeals," Journalism Quarterly, 51, 613-621.
- Churchill, Gilbert A. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," Journal of Marketing Research, 16, 64-73.
- Corey, Lawrence G. (1971), "People Who Claim To Be Opinion Leaders: Identifying Their Characteristics by Self-Report," Journal of Marketing, 35, 48-53.
- Cronbach, L. J. (1951), "Coefficient Alpha and The Internal Structure of Tests," Psychometrika, 16, 297-334.
- Day, George S. (1970), Buyer Attitudes and Brand Choice, New York: Free Press.
- DeBruicker, F. Stewart (1979), "An Appraisal of Low-Involvement Consumer Information Processing," in Attitude Research Plays for High Stakes, eds. J. C. Maloney and B. Silverman, Chicago: American Marketing Association, 112-132.
- Houston, Michael J. and Rothschild, Michael L. (1978), "Conceptual and Methodological Perspectives on Involvement," in 1978 Educators' Proceedings, ed. S. C. Jain, Chicago: American Marketing Association, 184-187.
- Jacoby, Jacob (1971), "A Model of Multi-Brand Loyalty," Journal of Advertising Research, 11, 25-31.
- Lastovicka, John L. and Gardner, David M. (1979), "Components of Involvement," in Attitude Research Plays for High Stakes, eds. J. C. Maloney and B. Silverman, Chicago: American Marketing Association, 53-73.
- Lundstrom, William J. and Lamont, Lawrence M. (1976), "The Development of a Scale to Measure Consumer Discontent," Journal of Marketing Research, 12, 373-381.
- Mitchell, Andrew A. (1979), "Involvement: A Potentially Important Mediator of Consumer Behavior," in Advances in Consumer Research, Vol. 6, ed. W. L. Wilkie, Ann Arbor: Association for Consumer Research, 191-196.
- Nunnally, J. (1967), Psychometric Methods, New York: McGraw Hill.
- Peter, J. Paul (1979), "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," Journal of Marketing Research, 16, 64-73.
- Ray, Michael L. (1979), "Involvement and Other Variables Mediating Communication Effects as Opposed to Explaining All Consumer Behavior," in Advances in Consumer Research, Vol. 6, ed. W. L. Wilkie, Ann Arbor: Association for Consumer Research, 197099.
- Robertson, Thomas S. (1976), "Low Commitment Consumer Behavior," Journal of Advertising Research, 16, 19-24.
- Rothschild, Michael L. (1974), "The Effects of Political Advertising on Voting Behavior of a Low-Involvement Electorate," unpublished Ph.D. dissertation, Graduate School of Business, Stanford University.
- _____ (1979a), "Advertising Strategies for High and Low Involvement Situations," in Attitude Research Plays for High Stakes, eds. J. C. Maloney and B. Silverman, Chicago: American Marketing Association, 74-93.
- _____ (1979b), "Marketing Communications in Nonbusiness Situations," Journal of Marketing, 43, 11-20.
- Tyebjee, Tyzoon T. (1979a), "Refinement of the Involvement Concept: An Advertising Planning Point of View," in Attitude Research Plays for High Stakes, eds. J. C. Maloney and B. Silverman, Chicago: American Marketing Association, 94-111.
- _____ (1979b), "Response Time, Conflict and Involvement in Brand Choice," Journal of Consumer Research, 6, 295-304.
- Webb, Peter H. (1979), "Consumer Initial Processing in a Difficult Media Environment," Journal of Consumer Research, 6, 225-236.
- Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-62.

CONSUMER LEARNING THROUGH EXPERIENCE:
A STUDY AND EXPERIMENTAL PARADIGM¹

Joel Huber, Duke University
Terry Elrod, Columbia University

Abstract

In this study respondents were asked to make a series of choices on a set of stimuli defined by abstract cues. After each choice a reward/penalty was given depending on the quality of the item chosen. Generally, respondents did a poor job of correctly inferring the additive values of the cues from 32 trials. An examination of the patterns of successive choices indicates that respondents were not really attempting to learn but were primarily avoiding negative feedback.

Introduction

The question this paper raises is how well people learn to use unfamiliar cues to guide choice when the only source of information is past choices. In a consumer context a distinction has been drawn between experience goods, whose worth can be ascertained only through consumption, and inspection goods, whose worth can be ascertained prior to purchase (Nelson 1970).

Suppose a consumer desires to discover the most preferred item from a set of unfamiliar experience goods. A rather laborious way to find the optimal product is to try each. Such a process would be more difficult still if feedback from each item had random error since then each brand would have to be sampled a number of times to arrive at a stable estimate of its value. If, however, there is a set of cues that are consistently related to the evaluations of the products, the task of finding the best is simplified. The consumer can sample certain items to determine the worths of the cues and then use the cues to find the most preferred item.

This paper presents a paradigm that allows an exploration of people's use of this latter strategy. It estimates the extent to which they can quickly infer the worths of unfamiliar cues from choice experiences and use these inferences to arrive at better choices. The use of the term "cues" is used rather than "attributes" to emphasize their unfamiliarity and hence the lack of priors as to their worths.

An understanding of how consumers assign worths to cues through experience is important because (1) virtually all cues are of unknown significance to a person at some point in time, and (2) the efficiency of a market economy depends upon the consumer's ability to learn preference through purchasing. Consumers who can't learn to utilize cues efficiently to guide choice need more:

- a) protection by legislation from deceptive, or irrelevant advertising claims;
- b) assistance with brand evaluation by independent testing agencies, and
- c) training in normative methods of comparison that take in-

¹
The authors wish to thank Morris Holbrook and Bob Chestnut for very helpful comments on an earlier draft.

to account the consumer's limited ability to evaluate brands.

Prior Methodology - Multiple Cue Probability Learning

The research tradition that has addressed the question of how well people can learn to use unfamiliar cues to guide choice most directly has been called Multiple Cue Probability Learning (MCPL). Relevant reviews by Slovic, Fischhoff and Lichtenstein (1977) and Schmitt, Coyle and King (1976) provide good overviews of what appears to be a relatively mature area of inquiry. This tradition has (1) controlled stimulus outcome and the relation of cues to choice outcome, (2) measured the subjects' prior beliefs about the cues' relations to choice outcome, (3) studied how rate of learning of cue worth is affected by such factors as the number of different cues, the strength and complexity of their relationship to choice outcome (e.g. nonlinearities, interactions), and the value of different types of additional feedback. A common research paradigm has been to control prior beliefs by presenting abstract stimuli to subjects in an experimental setting. The subject is shown a stimulus-object, (e.g. consisting of three bar graphs of certain heights), then the subject is asked to guess the worth of the stimulus-object, and finally the subject is told its worth (often with some stochastic error). This stimulus/guess/feedback process is repeated 100-400 times. The measure of learning is typically the degree of correspondence between guessed and actual stimulus worths.

While the above paradigm has led to a rich stream of results, it differs in two critical respects from the consumer learning about experience goods referred to earlier. First, in the multiple cue paradigm, notice the subject has no control over the selection of the stimulus object, as s/he might in choosing a soft drink or a detergent. The subject is simply shown the object and asked to assess its worth. Second, in the experimental paradigm the subject is concerned with the accuracy of the judgments rather than the actual worth of the stimulus objects. By contrast, in a consumer trial the worth of the object chosen is important and learning or accuracy may have secondary importance. Thus in consumer choice there is a tension between short term reward and long term information. This tension simply cannot be captured by the traditional multiple cue probability paradigm because choice is passive and the outcome (as opposed to accuracy) is of no material concern to the subject.

A Modified Paradigm

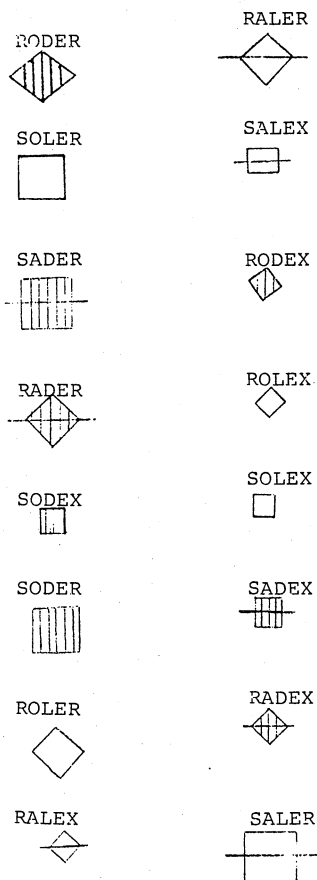
A desire to bring the task characteristics more in line with the consumer context, while still retaining control over cues and outcomes, lead to the development of the following task: (1) allow the subject to choose any one stimulus-object from a set of available ones; (2) inform the subject of the material consequence of the choice made (money won/lost), (3) have the subject repeat this choice/result process 32 times.

Two sets of data were collected. First, the choice/feedback process was administered and monitored by computer, which allowed automatic recording of the subjects' choices, the results of the choices, and the subjects' deliberation times.

Second, the subjects completed questionnaires after finishing the learning task. A first set of questions simply asked respondents to evaluate the 16 possible stimuli on a 7-point quality scale that mirrors the feedback given. A second set of questions asked respondents to directly evaluate each level of the four dichotomous cues.

The 16 stimulus objects are shown in Figure 1, and represent all possible combinations of the four two-level cues: line/no line, large/small, shaded/unshaded, and diamond/square. The name given each stimulus object has a one-to-one correspondence with its geometric configuration, e.g. large corresponds to the suffix 'er' whereas small corresponds to 'ex.'

FIGURE 1
Names and Symbols Used in the Study



Experimental Manipulations

Several factors that have been found in the MCPL literature to affect learning of cue-worths were manipulated.

Relative Importance of the Cues

Respondents in MCPL tasks have been shown to have prior biases with respect to the amount of information in different cues (Dudycha and Naylor 1966, Janke 1972, Castellan 1974).

The current experiment included four levels of dispersion of weights among the cues, ranging from equal importance to highly unequal importance (in which one cue alone accounted for 80% of explainable variance in stimulus worth).

Absolute Worth of Cues

Stated worths of chosen objects contained two error components. First, repetitive selection of the same stimulus yielded different choice outcomes (as in a consumer context, where repeat purchases of a brand do result in different degrees of satisfaction). This source of error is referred to here as 'stimulus error,' and has been found in MCPL tasks to inhibit learning of worths of both stimuli and cues. Second, the cues did not perfectly predict even the expected worth of a stimulus object. This source of error, 'systematic error,' was thought to inhibit the learning of cue, but not object.worth. Each of these sources of error was manipulated as a two-level factor. These manipulations taken together created a 2 (stimulus error) by 2 (systematic error) by 4 (relative cue weights) full factorial design of 16 cells.

The task was performed by 48 MBA students at a major university and typically took 15-20 minutes to complete. These business students were challenged by the task, responding to the experiment as a test of pertinent business-related skills. The 48 respondents allowed three replications per cell. The assignment of weights to cues was rotated across these three replications to give each cue a roughly equal chance, across subjects, of having a given weight.

Results

The main result is that the respondents did not learn the worths of either the stimuli or their cues very well. This result was independent of the experimental manipulations and is based upon analysis of both the questionnaires and the choice process.

Poor Learning as Evidenced by Questionnaires

The measures of learning, as are those in much of the MCPL literature, are based on Brunswik's (1956) lens model as formalized by Dudycha and Naylor (1966). These measures reflect correlations between several estimates of the stimulus worths: (1) Direct estimates were supplied by the questionnaire, (2) Bootstrapped estimates were the predicted worths of the stimuli provided by regressing the direct estimates against the cues, (3) Compositional estimates were derived by combining the ratings of each cue. The implied worth of an object-stimulus was then the sum of the ratings of its cue-levels. Since ratings of the geometric shapes were obtained separately from the ratings of the letters, two separate compositional indices were created.

Given these four subject-based estimates of stimulus-object worths, what measure should be used as a criterion to evaluate the degree of learning reflected in these measures? One could regress each of these measures against the true worths of the stimulus-objects but these true worths are not perfectly related to the cues due to the presence of random error. We therefore calculated a 'fairer' criterion estimate of the stimulus worths by computing those cue weights that would be inferred by a perfect OLS process given the sequence of chosen stimuli and feedback.

These optimal ratings of the stimuli were obtained by regressing the feedback levels against the cue levels scored as dummy variables. The resulting predicted stimulus worths were optimal in the sense of being the best predictors of the particular stimuli chosen. That is, if a respondent concentrated choices among five or six stimuli, the optimal values reflected the fit of the additive model from among these choices. These scores might be quite different from those generated from a broad selection of stimuli. Thus

the optimal scores are relative to the particular choices and to a certain extent control for whether a respondent used a high-learning strategy with many different items chosen or a low-learning strategy. It should also be noted that the optimal values can, in theory, be perfectly reproduced by the subject in that each respondent had the information needed to produce a scale that correlates perfectly with the optimal values.

TABLE 1

Average Correlation of Subject-Based Estimates of Stimulus Worths with "Optimal Estimates"

	<u>Direct</u>	<u>Bootstrapped</u>	<u>Shapes</u>	<u>Letters</u>
Correlation with Optimal	0.19	0.29	0.26	0.31
Standard Error	(0.05)	(0.08)	(0.08)	(0.08)

TABLE 2

Average Correlations Among Subject-Based Estimates of Stimulus Worths

Estimate:	<u>Direct</u>	<u>Bootstrapped</u>	<u>Shapes</u>
<u>Bootstrapped</u>	0.61		
<u>Shapes</u>	0.28	0.43	
<u>Letters</u>	0.34	0.56	0.55

The four subject-based estimates of stimulus worth: direct, bootstrapped, shapes and letters, were correlated with the optimal scales for each individual. The means and standard deviations across the 48 respondents are shown in Table 1. It is noted that, except for the direct estimates, which are predictably unreliable, the three scales correlate about 0.3 with the optimal model. This translates to an R^2 of less than 10%.

The first question raised by this relatively low performance is, how it can be so much lower than typical correlations found in either compositional or decompositional models? The answer to this question can be found by examining the correspondences between the various subjective scales shown in Table 2. These more respectable correlations indicate that consumers are internally consistent. But this internal consistency may greatly overstate actual association with the true worths of the stimuli. Thus it appears that respondents were consistent with their own ratings but these had relatively little correspondence with reality. In the next section we explore the pattern of actual choices made in an attempt to discover the strategies used and why they were unsuccessful in uncovering the true values.

Lack of Learning in Choice Patterns

Some of the reasons for the poor learning become apparent as one examines the patterns of choices through the 32 choice-feedback trials. While no clear criterion of learning is available in this case, one would expect that choices might reflect a partly systematic search for the best stimulus object, and that subsequent choices would be somewhat consistent with information obtained from prior choices.

Lack of Search for Structure

All the stimuli differ from each other on between one and

four cues. By choosing a stimulus that differs from the previous one on only one cue, the change in feedback is an estimate of that cue's worth. We term such switching on only one cue on two consecutive choices an experimental test. The value of such a strategy when there is no error in feedback is clear: by ignoring the cues, all 16 brands must be tried before the best brand is identified, but by correctly inferring an additive relationship between cues and stimulus worth and 'experimentally testing' the cues, only 5 choices are needed to determine the best brand. When feedback contains error, the best brand can never be identified with certainty by either strategy, but even if the subject decides to try all brands in search of the best, trying them in an 'experimental testing' order would yield valuable information. Did respondents in fact employ such a strategy? The answer is no, as is shown by Table 3, which compares the number of cues switched between consecutively chosen stimuli if all stimuli were chosen randomly (and with equal probability) with the frequency observed with our respondents. Note that the percentage of single-switch choices (experimental tests) is virtually identical to that expected under the chance model. This percentage (25%) was unexpectedly small given the high level of information that can be derived from just changing one cue. Instead, most of the switches (30%) involved changing two cues, a strategy that has limited value from an informational standpoint and is only marginally efficient as a way to find a brand that is quite different. High-aversion strategies occurred relatively rarely, perhaps because of the search difficulty in finding the item that differs over three or four of the cues. Not surprisingly, perhaps, zero switches ('brand loyalty') occurred much more frequently than expected by chance. Overall, there is little evidence that respondents acted as "intuitive scientists" by searching for structure using controlled experiments. Furthermore, as the next section shows, even when "experiments" were used their results were ignored.

TABLE 3

Relative Frequency of Number of Cues Switched

<u>Label</u>	<u>Percent of Subjects' Choice Pairs</u>	<u>Percent if Stimuli Were Chosen Randomly</u>
Brand Loyalty (no switch)	23%	6%
Experimental Test (1 attribute switched)	25%*	25%
Experiment/Aversion (2 attributes switched)	30%	38%
Aversion/Experiment (3 attributes switched)	18%	25%
Pure Aversion (all attributes switched)	4%	6%

*Read: Experimental tests, defined as choosing a product with one factor (out of four possible) differing from the previous choice, occurred in 25% of the choice occasions.

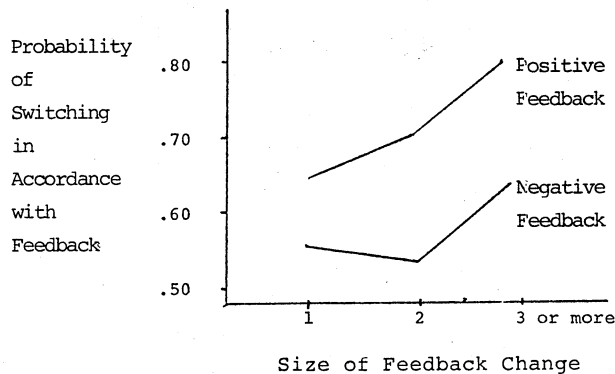
Lack of Response to Structure

Figure 2 shows that, when subjects did switch on only one cue, their next choice often conflicted with the information obtained on that cue. If switching on one cue resulted in a positive improvement in choice outcome, then one might expect the subject to retain that level on the subsequent choice. Subjects in fact did this 65%-78% of the time, depending upon the magnitude of the change in outcome. If switching a cue resulted in a negative change in choice outcome, then one would expect subjects to switch

back to the original one on their subsequent choice; they in fact did so less than two-thirds of the time.

FIGURE 2

Reactions to Positive and Negative Changes in Feedback When Exactly One Cue is Changed (Experimental Test)



Evidence of Sensitivity to Immediate Feedback

What, if anything, did influence the number of cues switched on consecutive choices? There are three likely candidates. First might be the outcome on the last item chosen--the higher its value, the greater the probability that the next choice will be highly similar. Another possibility is the change in outcome for the last two choices. One would expect smaller change following an increase in choice outcome than following a decrease. Finally, one might expect experience to result in switching on fewer cues; that is, the respondent might, with experience, eliminate some stimuli from consideration.

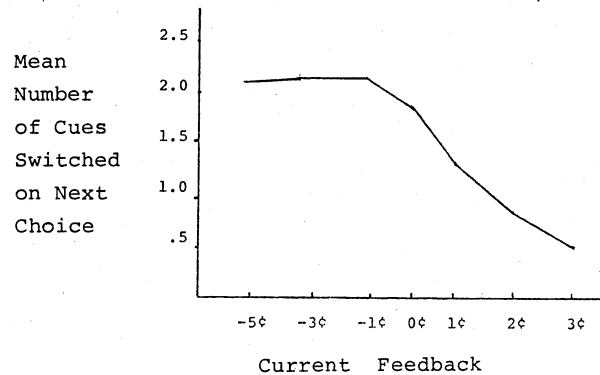
In the present study, current feedback accounted for virtually all of the accountable variance in the number of cues switched. Thus, once current feedback is accounted for, the change in feedback was not significant and task experience became much less so. The effect of current feedback is shown in Figure 3.

Summary and Directions for Future Research

The foregoing has described an experimental paradigm that modifies the multiple cue probability learning task to better approximate that aspect of consumer learning that occurs through the experience of the product. Different ways were tried to get respondents to give the values of cues that correspond to the feedback received on the 32 purchase/feedback occasions. While these different measures had reasonable correlations among themselves (internal validity), none accounted for more than 10% of the variance (external validity) that could have been realized had the respondents been OLS processors. The patterns of choices provided some explanation for the lack of apparent learning in the task. Respondents did not use what would have been the most efficient way to measure cues worths, experimental tests, any more than would be expected with random selection of stimuli. Further, where experimental tests did occur, subsequent choice was in accordance with the feedback from the

FIGURE 3

Effect of Current Feedback on the Number of Cues Switched on Next Choice



test in only 2/3 of the cases. Indeed, the best predictor of the number of cues switched was simply the value of the last time chosen--the more negative its feedback, the more dissimilar the next item chosen. Perhaps so little learning occurred because respondents were not trying to learn but simply avoiding bad products, a kind of aversive random walk.

While the results from this exploratory study are too tentative to be applied directly to consumer learning in the marketplace, the importance of learning in that context is great enough to warrant research into the following questions:

Can Subjects Be Induced to Learn Cue Worths?

Several factors would likely serve to increase the level of cue learning. First, the number of stimuli can be increased relative to the number of cues, thus forcing attention to cues rather than global judgments on stimuli. Second, the consequences of choice can be increased. This change would increase the cost of a 'try all brands and see' strategy. Finally, more cue learning might occur with messages (e.g. ads) to orient respondents to the different cues.

Does Prior Knowledge Affect Learning?

Abstract choice objects were employed in this study, as in the MCPL literature, to minimize the effects of prior learning and, most importantly, to control choice outcomes. An alternative might be to use hypothetical brands of a real product class (as in conjoint analysis), measure (rather than control) the subjects' priors concerning cue worths, and then study the change in these priors as a result of feedback on choices. Alternatively, one might use brand names and measure prior beliefs concerning both cue (attribute) and brand worths.

Thus, the current study, while tentative and largely exploratory, suggests that consumers may have difficulty in correctly inferring the values of purchase cues from product experience. It is hoped that future research will shed more light on this issue of concern to regulators, practitioners and consumer researchers.

References

Brunswik, E. (1956), Perception and the Representative Design of Experiments, Berkeley: University of California Press.

Castellan, N. John (1974), "The Effect of Different Types of Feedback on Multiple-Cue Probability Learning," Organizational Behavior and Human Performance, 11, 44-64.

Dudycha, L.W. and Naylor, J.C. (1966), "Characteristics of the Human Inference Process in Complex Choice Behavior Situations," Organization Behavior and Human Performance, 1, 110-128.

Janke, Mary (1972), "Effect of Stimulus and Instructional Variables in Ambiguous Concept-Attainment Task," Journal of Experimental Psychology, 93, 21-29.

Nelson, P. (1970), "Information and Consumer Behavior," Journal of Political Economy, 78, 40-53.

Schmitt, Neal, Coyle, B.W. and King, L. (1976), "Feedback and Task Predictability as Determinants of Performance in Multiple Cue Probability Learning Tasks," Organizational Behavior and Human Performance, 16, 338-402.

Slovic, Paul; Fischhoff, B., and Lichtenstein, S. (1977), "Behavioral Decision Theory," Annual Review of Psychology.

MEASURING CONSUMER ATTITUDES TOWARD ALTERNATIVE CHECK VERIFICATION SYSTEMS

David J. Barnaby, University of Tennessee
John W. Philpot, University of Tennessee
Richard C. Reizenstein, University of Tennessee

Abstract

While the technologies for various electronic funds transfer (EFT) systems are well developed, little attention has been directed toward the study of consumer attitudes toward and acceptance of these technologies. In particular, several check verification systems (CVS) have been introduced as an electronic mechanism to verify consumer checks and reduce the number of bad checks accepted by merchants. Since the systems provide unique benefits to different consumer and merchant user groups, an assessment of consumer reaction to CVS alternatives provides insights into consumer check cashing attitudes, activities, and preferences. Multivariate procedures are utilized to yield profiles of specific CVS preference groups.

Introduction

Relative to other consumer oriented service industries, banking has only recently begun to devote substantive attention to the application of established marketing principles. Certain financial institutions, such as City National Bank of Columbus (Ohio), have emerged as innovators in their use of such practices to develop new market segments and to better serve their existing customers. The banking industry as a whole, however, is only beginning to tap the marketing resources available to them to better assess market reaction to and promote the potential array of consumer services available under the umbrella of Electronic Funds Transfer Systems (White and Woodside 1973).

Historically, the primary thrust in what many banks have perceived as marketing, both overall and with respect to Electronic Funds Transfer Systems (EFTS), has been institutional and product-oriented, rather than market oriented (Mason and Mayer 1974). Bank marketing has historically been narrowly defined, as sales and/or promotion in the rare instances where it has been used, as opposed to the broader strategic approach emerging in other consumer-oriented industries (White and Woodside 1973). This direction is evident in the numerous articles that have appeared in the last few years which discuss the concept of EFTS as well as specific EFTS applications (Automated Teller Machines, Point of Sale Systems, Check Guarantee Systems, etc.). These generally dwell on the advantages and disadvantages of EFTS (more from the financial institution's perspective than that of the consumer or retail merchant), and costs versus benefits with relatively little emphasis on market planning, market acceptance, consumer research, or promotion to the consumer or retail sectors of either the EFTS concept or the potential variety of EFTS services (Benton 1977).

In light of the evident paucity of marketing-oriented consumer research in this field, this study will attempt to apply marketing research and analytical techniques to a specific type of EFTS: check guarantee/verification systems (CVS), systems which identify bad check writers by use of a computerized negative card file accessible via retail in-store terminals. The major objective of this research will be to apply the market segmentation concept to describe and classify potential consumer preferences for two general categories of check verification systems: consumer operated and retail employee operated.

Literature Review

As stated, the preponderance of literature related to consumer attitudes toward and usage of EFTS has focused on the advantages, disadvantages, costs, and benefits of EFTS, various delivery mechanisms, and the variety of services available under the EFT umbrella. Business Week, Fortune, and Harvard Business Review, among the general business periodicals, have featured such articles in recent years ("Electronic Banking: A Retreat from the Cashless Society" 1977, Rose 1977, Benton 1977), in addition to such financially oriented publications as The Banker, Banker's Monthly Magazine, the Journal of Bank Research, and Savings and Loan News (Simon 1977, Nadler 1977, White and Woodside 1973, "You Can Take the 'E' out of EFT" 1977). Moreover, with the experimental implementation of various forms of EFTS services in the banking community, successes and failures are increasingly reported, including the generally favorable experiences of City National Bank of Columbus and the Wilmington Savings Fund Plan as well as the less successful experiment at Glendale Federal Savings and Loan ("Electronic Banking: A Retreat from the Cashless Society" 1977, De Costa 1977).

In contrast to these are articles directed toward the application of marketing to bank services other than EFT (Levy 1973). Some of these studies attempt to apply retail marketing techniques in a segmentation context, generally describing users and nonusers of types of banks or banking services, such as commercial banks (as opposed to savings and loan institutions), bank credit cards in terms of life styles, and overdraft checking in terms of usage rates (Arbeit and Sawyer 1974, Plummer 1971, Morgan 1978). Most of the remaining marketing research in the field tends to emphasize specific kinds of analyses such as determinant attribute analysis and perceptual mapping as segmentation mechanisms (Anderson, Cox, and Fulcher 1976, Gillette and Evans 1974). Marketing-oriented research on specific EFT services has been reported only rarely, with the Vinson and McVandon study of the impact of consumer education on attitudes toward the debit card concept one of the only recent examples in the marketing literature (Vinson and McVandon 1978).

Research Methodology

During 1979, a personal interview survey of 325 consumers was conducted by means of a systematic random sample from the City Directory of a medium-sized Southeastern city (population 200,000). The primary objective of this research was to assess consumer familiarity with and attitudes toward the concept of EFTS, as well as specific EFT services; the major thrust of the research, however, was launched toward alternative forms of check guarantee/verification systems (CVS). (Coincidentally, CVS were initially introduced in the community shortly after completion of the study.) Because of respondents' potential unfamiliarity with point of sale/check verification systems (and their variations), focus groups were utilized to get an idea of the level of consumer and retailer knowledge in this field and to pinpoint possible areas for investigation. The check verification systems under study, in many cases, represented entirely new services that would be potentially available in the market areas; hence, the focus groups were followed by extensive personal interviews with respondents over a

four-month period to gauge consumer reaction to this new EFT concept.

In order to focus the information collection on the most relevant subset of the sample, i.e., those who write checks outside the home who would be the critical consumer market for check verification systems, two filter questions were posed: Respondents were first asked if they had checking accounts; of those interviewed, 21% (or 68) did not. Interviews with these individuals were terminated after collection of demographic data. Subjects who did have checking accounts were then queried as to whether they used checks for purposes other than bill paying at home; 20% (or 51) did not, and were eliminated from the remainder of the interview after classificatory information was obtained.

The remaining 206 respondents were first analyzed in comparison with updated census data, adjusted for inflation by use of the Consumer Price Index. This information, provided in Table 1, shows an excellent match in the age, income, and marital status categories between the sample and updated Census figures for the SMSA. (In the age category, children under 18 were removed from the Census figures and the remaining totals readjusted, as children did not qualify as potential respondents.) Given the close parallel in most categories between the sample and the Census data, it would seem as if the respondents are potentially representative of the market area under study.

TABLE 1
DEMOGRAPHIC PROFILE OF RESPONDENTS
COMPARED WITH CENSUS DATA
(N = 260)

Demographic Category	Respondent (%)	SMSA (Census %)
<u>Age:</u>		
Under 20	1	1
20-29	26	24
30-39	23	18
40-49	18	19
50-59	15	17
60 and over	18	21
<u>Income:</u>		
Less than \$10,000	20	20
\$10,000 - \$19,999	39	40
\$20,000 - \$29,999	25	24
\$30,000 and over	16	16
<u>Occupation^a:</u>		
Retired	13	N.A.
Professional/Technical	26	17
Manager/Proprietor	12	8
Blue Collar	22	33
Sales	7	8
<u>Sex^a:</u>		
Male	74	47
Female	26	53
<u>Marital Status:</u>		
Married	67	65

^aThe major check writer in the household was the subject of the interview, accounting for much of the variation of the sample from Census figures.

Results

Of the 206 questionnaires from respondents who completed the full forty-minute personal interview, 11 could not be utilized due to missing data; less than half of the remaining 195 had ever heard of point of sale (POS) or check verification systems. In fact, only 2% had ever used a POS

terminal. These low percentages correspond to the limited development of electronic funds transfer systems of any type in the relevant market area. Despite this lack of familiarity with EFTS, a majority of respondents were clearly interested in check verification systems that would both aid in the processing of checking payments and speed transactions. Consumers noted that of their total check cashing activities, nearly 45% required some form of check verification procedure. These transactions normally occurred in the following types of retail establishments (in decreasing order of frequency): grocery stores, department stores, drug stores, discount stores, and auto service stores. Given such results indicating a generally low level of awareness, but a strong interest in check verification systems, consumers' attitudes toward specific check cashing activities will be explored in depth in the analyses to follow.

Initial assessment of consumers' attitudes toward check cashing activities were measured by 27 Likert scale items developed from focus group research. While these questions included items that pertained to many aspects of payment for retail goods and services, a number of items were specific to check verification procedures and systems currently employed by merchants. In Table 2, a sampling of these items, representing greatest percentage agreement or disagreement, is shown along with the agreement or disagreement scaled percentage. Generally, the sample "agrees" that bad checks are a problem for retailers and empathizes with merchants regarding this problem. Respondents, however, found charge cards to be more acceptable than checks to merchants and strongly preferred charge card verification systems to present check cashing systems. (From the merchant perspective, a verification system that substantially shifts risk of nonpayment for goods or services to the credit card sponsor is preferable.) Consumers do agree substantially that retailers should cash checks, and should provide an efficient identification procedure that would provide both protection for the merchant and a nonthreatening or embarrassing transaction environment for the consumer. Interestingly, driver's license numbers are assuming an increasingly critical role in this identification procedure.

TABLE 2
CONSUMER ATTITUDES TOWARD CHECK CASHING
(N = 195)

Percentage Agreement	Attitudinal Item
55%	I prefer to pay for retail purchases by check.
43%	I would rather use a charge card for retail purchases.
78%	I do not like to carry much cash with me.
72%	Retailers should cash personal checks.
66%	Merchants are eager to accept my personal check.
73%	It is difficult to get a check cashed without a driver's license.
68%	I prefer to patronize retailers where I am known, since no ID is required for personal checks.
52%	I prefer retailers' credit card verification procedures to check cashing procedures.
86%	Bad checks are a major problem for retailers.
86%	Retailers require a lot of ID to cash checks because of bad check problems.
84%	Charge cards are more acceptable to merchants than personal checks.
Percentage Disagreement	Attitudinal Item
91%	I don't really care if merchants get stuck with bad checks.
59%	I resent a retailer who requires an ID so I can purchase.
52%	Retail check cashing procedures make me feel untrustworthy.
49%	It is embarrassing to be required to have so much ID to get my check cashed in a retail store.
66%	I have never written a check that was returned for insufficient funds.

Table 3 indicates CVS attributes deemed most important by consumers in evaluating alternative systems for check verification. These results depict consumers as stressing the importance of such status quo factors as price maintenance, account privacy, system reliability, stop payment procedures, "float" availability, and remedies for accidental overdraft. Additionally, such factors as bad debt reduction, faster identification procedures, cashing other than personal checks, and cashing checks for more than the amount of purchase are considered important components of the system.

TABLE 3

IMPORTANCE OF CHECK VERIFICATION SYSTEM ATTRIBUTES
(N = 195)

Respondents Rating as Important (%)	System Attributes
95	Mechanically reliable system.
94	No increase in retail prices with system use.
93	Guarantee of customer privacy regarding account balance.
91	Fewer bad debts for merchants.
90	Faster identification procedures for check cashing.
89	Ability to stop payment on checks.
87	Ability to remedy accidental overdraft.
75	Ability to cash checks other than personal checks.
72	Ability to cash checks for more than amount of purchase.
62	Ability to cash checks without identification.
59	System operated by consumer.
52	Ability to continue use of "float."

Prior to testing the relationship between consumer attitudes toward check cashing activities, importance ratings of attributes, and specific check verification systems, factor analyses of the set of 27 attitudinal items and the 29 importance ratings were conducted. Results of both of these analyses were quite lackluster with communalities and factor loadings being very low; interpretation of the factors proved difficult. Attempts to correlate factors with system preferences proved nonsignificant at the .05 level. These poor results imply, in part, that consumers appeared to display some confusion about CVS systems due, to a great extent, to their unfamiliarity with these systems; as a result, they showed little consistency in their replies to the items.

As an example, consumers expressed a strong preference for small retailers who knew them and could accept checks without ID; at the same time they reported no resentment of retailers requiring ID. Both items had high positive loadings on one of the attitude factors, yet the first item was positively correlated with the grouping variable and the second item had a negative correlation with the grouping variable. This led to a factor that was practically orthogonal to the grouping variable.

The diffuse nature of the factor analysis, which implies confusion in the minds of consumers about CVS, suggests that subsequent discriminant analysis results would be subject to change, particularly after a public education and information campaign. That is, the real polarization of preferences has yet to occur.

In a further effort to identify those variables related to specific CVS preferences, correlations between the larger predictor sets (attitudes, attribute importance, and demographics) and consumer operated system preferences were obtained. Those items significant at the .05 level (univariate) or better are shown in Table 4 and were selected as candidates for further study. (To preclude spurious correlation, these and subsequent tests were made using the critical value that is appropriate when all predictors are

included in the analysis, i.e., $F .05 (1,98) = 3.94$.) In total, 13 candidate variables survived this correlational screening, 10 of which were equally divided between attitudinal and importance items. Only one item each pertaining to specialty store system usage intentions, friends as a new product information source, and number of children and/or others 15 years or older living at home, comprised the final three candidates. All correlations, as shown in Table 4, were less than .20, reinforcing the relatively weak level of the relationships discovered.

TABLE 4

CORRELATIONS OF SIGNIFICANT VARIABLES WITH
CONSUMER CVS PREFERENCE

Variable	Correlation	Significance Level
I waste a lot of time waiting for merchants to cash my checks.	.133	.03
I prefer to patronize small retailers because they will accept my check without identification.	.155	.01
Retailers should accept personal checks for amount of purchase.	-.170	.01
If it were not so difficult I would consider changing banks.	-.122	.04
I am extremely satisfied with the bank in which I have my checking account.	.157	.01
Faster than normal identification procedure for check cashing is important.	.183	.01
System operated by sales clerk (instead of customer) is important.	-.114	.05
Ability to obtain cash readily in retail stores is important.	.114	.05
Ability to stop payment on approved checks is important.	.177	.05
Ability to cash nonpersonal checks (e.g., payroll, social security) in addition to personal checks is important.	.124	.04
Likely to use system if available in specialty stores.	.115	.05
Friends as important new product information source.	----	----
No. of children/others 15 or older that reside in household.	-.156	.01

Despite the weak (though significant) nature of the correlations, however, the candidate variables were included as predictors in a multiple discriminant analysis to test the ability of this multivariate function to discriminate between those consumers preferring consumer as opposed to retail employee operated check verification systems. Preference grouping for either the consumer operated check verification system or the retail employee operated check verification system were determined by categorizing preferences for specific CVS alternatives. Scenarios describing the use and function of each alternative were provided to familiarize respondents with each CVS possibility prior to measurement of consumer preferences. First choice preference served as an indicator of preference for consumer versus employee operated CVS; respondents were grouped accordingly with the great majority (73%) favoring the consumer operated version.

The results of the stepwise MDA shown in Table 5, reveal those predictors which significantly ($p = .01$) discriminate between consumer and employee operated CVS preference groups. Examination of the results (means and standardized discriminant coefficients) suggests that those respondents preferring consumer operated check verification systems can

be described as: consistently agreeing that retailers should cash personal checks for the amount of purchase, extremely satisfied with the bank handling their checking account, desiring faster than normal identification procedures for check cashing, having few children or others 15 years or older residing in the household, and counting friends as an important new product information source. Correspondingly, those respondents preferring retail employee operated check verification systems agree to a lesser extent that retailers should cash checks for amount of purchase, are less well satisfied with their bank, have a lesser desire for faster check cashing identification procedures, have barely any children or others 15 years or older residing in the household, and do not rely as much on friends as a new product information source.

TABLE 5
DISCRIMINANT ANALYSIS--CONSUMER VS. RETAILER
OPERATED CVS PREFERENCE GROUPS

Variable	Means (Std. Deviations)		Standardized Discriminant Coefficients
	Consumer CVS Preferred	Retailer CVS Preferred	
Cashing personal checks for amount of purchase is service retailers should provide.	2.00 (1.19)	1.57 (.83)	.532
I am extremely satisfied with the bank in which I have my checking account.	1.62 (.97)	2.02 (1.27)	-.426
Faster than normal identification procedure for check cashing desirable.	1.40 (.65)	1.69 (.86)	-.434
Number of children/other 15 or older that reside in household.	.70 (.96)	.33 (.68)	.406
Friends as important new product information source.	1.58 (.69)	1.90 (1.06)	-.379

A further indication of the ability of the discriminant function to statistically classify respondents based on the five significant predictors is found in the hit or miss classification results presented in Table 6. The disproportionate group sizes precluded the traditional hold-out sample analysis to test the classificatory ability of the discriminant function. The jackknife procedure was utilized to test the function, however, with results showing a 75% correct classification rate while chance classification was calculated at 61%.

TABLE 6
CLASSIFICATION OF CONSUMER VS. RETAILER
CVS SYSTEMS USING MDA^{a,b}

Actual Group	No. of Cases	Predicted	Predicted
		Consumer CVS Preferred	Retailer CVS Preferred
Consumer CVS Preferred	143	132 (92%)	11 (8%)
Retailer CVS Preferred	52	37 (71%)	15 (29%)
	195		

^aPercent of cases classified by chance: 61%.

^bPercent of cases classified by MDA: 75%.

Conclusions and Implications

From reviewing the literature, it seems evident that the cashless/checkless society and electronic funds transfer

systems (EFTS) are only in the early stages of market development. The issues and controversy surrounding EFTS has had primarily a financial and product-oriented focus, rather than a marketing-oriented perspective. Some departure from this product orientation, however, can be found in the stances of consumer groups advocating the limited adoption of EFTS because of its potentially negative impact on consumer privacy. Otherwise, EFTS innovations such as check guarantee/verification systems have received relatively little attention in what little marketing-directed financial literature exists, a major omission since transaction volume is a key variable in estimating the potential profitability of such systems, thus emphasizing the importance of marketing techniques to achieve necessary volume levels.

Check verification systems (CVS) unquestionably address a potential need evident in retail transactions. Checks continue to play a key transactional role; consequently, bad checks pose significant problems for both retailers and consumers. CVS is an EFTS alternative which has the potential to dramatically reduce bad check losses, if sufficient retailers and consumers will use such systems to make them profitable to sponsoring banks or business firms.

Since several CVS alternatives exist, research is desirable to gain insight into consumer preferences for specific CVS options and alternatives. The personal interview survey conducted in this study attempted to provide at least a partial response to this issue. Although geographically limited, and thus not generalizable to a national market, the results do provide insight into the differentiating characteristics of two distinct CVS alternative preference segments in a retail market prior the actual introduction of such systems.

Data collected by this research showed that, for the market area under examination, consumers as well as retailers definitely recognized the problems associated with bad checks. Impact on retailer costs and thus prices were understood by a large majority of respondents; the reasons for identification procedures to minimize such expenses were thus comprehended, although disliked due to their inconvenience and time consuming nature, by consumer subjects. Despite such perceived negative aspects of retail check cashing, however, availability of retail check cashing, at least for the amount of purchase, was a highly desired service in retail stores. CVS alternatives were seen as quite attractive to many respondents desiring such check cashing availability, due to the simpler and faster identification procedures inherent in the consumer operated system in particular.

Results of multiple discriminant analysis of preference for consumer versus retail employee operated CVS support several differentiating characteristics of groups favoring each type of alternative, including the greater orientation of those preferring the consumer operated system toward more check cashing and faster identification procedures. In addition, a potentially useful promotional idea for such systems is revealed in the MDA results. The greater reliance on friends as an important new product information source by those respondents favoring consumer operated CVS is an important insight for those interested in marketing such techniques. If this CVS method was launched as a new product introduction, apparently word of mouth could play a key role; thus, consideration of ad copy using a representative of friends' dialogue might be effective as a promotional mechanism.

Finally, the MDA results show that those who appear to rely more on credit cards and favor credit card transaction procedures also appear to be more interested in retail employee operated systems which parallel credit card verification procedures. On the other hand, the group favoring the consumer operated CVS seems to rely more on check payment for retail purchases and sees this self-initiated and self-operated system as preferable.

References

- Anderson, Thomas W., Jr., Cox, Eli P. III, and Fulcher, David (1976), "Bank Selection Decisions and Market Segmentation," Journal of Marketing, 40, 40-45.
- Arbeit, Stanley, and Sawyer, Alan G. (1974), "Applying Psychographic Analysis to a Retail Banking Market," The Magazine of Bank Administration, 39-45.
- Benton, John B. (1977), "Electronic Funds Transfer: Pitfalls and Payoffs," Harvard Business Review, 55, 16-17ff.
- De Costa, Sidney A. (1977), "WSFS Plan Meets Consumer Needs and Wants," Marketing Scope, 8, 22-25.
- "Electronic Banking: A Retreat from the Cashless Society" (1977), Business Week, 2479, 81-83ff.
- Gillette, William, and Evans, Richard H. (1974), "Service Analysis: Bank Marketing Example Using Perceptual Mapping" in Proceedings, Annual Conference of the Association for Consumer Research, 525-533.
- Levy, Sidney J. (1973), "Consumer Views of Banks' Services," Journal of Bank Research, 4, 100-104.
- Mason, Joseph Barry, and Mayer, Morris L. (1974), "Analyzing Factors that Affect Banking Patronage," The Magazine of Bank Administration, 42-45.
- Morgan, Fred W., Jr. (1978), "Profitability Market Segmentation: Identifying the Heavy Users of Overdraft Checking," Journal of Business Research, 6, 99-110.
- Moroney, William R. (1976), "What You Should Know About EFTS," Management World, 5, 12-13.
- Nadler, Paul S. (1977), "Bank Pricing Key to EFTS Growth," Banking Monthly Magazine, 94, 9-11.
- Plummer, Joseph L. (1971), "Life Style Patterns and Commercial Bank Credit Card Usage," Journal of Marketing, 35, 35-41.
- Rose, Stanford (1977), "Checkless Banking Bound to Come," Fortune, 95, 118-121ff.
- Simon, Leonard S. (1977), "Advances in Electronic Funds Transfer," The Banker, 127, 79ff.
- Sullivan, Barry F. (1977), "Cooperating for the Cashless Society," Executive, 3, 27-28.
- Vinson, Donald E., and McVandon, Wayne (1978), "Developing a Market for a New E.F.T.S. Bank Service," Journal of Marketing, 42, 83-86.
- White, S. Douglass, and Woodside, Arch G. (1973), "Marketing in Banking: Philosophies and Action," Journal of Bank Research, 3, 265-268.
- "You Can Take the 'E' Out of EFT" (1977), Savings and Loan News, 98, 76-81.

DISCUSSION
INVOLVEMENT, LEARNING AND ATTITUDES

David M. Gardner, University of Illinois, Urbana/Champaign

These three papers are a good example of the diversity in consumer research that exists today. The Huber and Elrod study is aimed at theory development using non-consumer stimuli. Block's paper represents an attempt to point out a missing tool in the involvement research field and develops a specific measuring tool. And different yet is the Barnaby, Philpot and Reizenstein paper that is very much an applied attitude study.

It is encouraging to see research of the type presented by Huber and Elrod. Any discipline can benefit by theory development research. Our tendency in consumer behavior is to "borrow" and therefore we often avoid some of the basic steps in theory development. Therefore, an attempt to understand choice, unencumbered by any previous experience or learning is welcome. Unfortunately, the writeup of this effort leaves out a discussion of many of the critical details that should allow others to evaluate the design and results of the study. This may be a very worthwhile study that will serve as the basis for other research, but, it is almost impossible to determine from the writeup. For instance, the authors do not specify what the seven-point quality scale was measuring (i.e., the direct measure). Also, it is not at all clear how the dependent variables were measured. The manipulation of systematic error seems questionable as many cues (attributes) predict expected worth of value. Without knowing the rationale for using only 32 trials (in another place 48 trials were specified) it seems that the results of this experiment might be very different if subjects were given more trials.

In addition to including a larger number of trials, this study would have been more useful if it had been replicated in four versions. The first replication would have been without any manipulation of either stimulus or systematic error. The second and third replications would have included one of these error manipulations. The fourth replication would have included both. In its present form, only the fourth replication has been attempted. Therefore, the effect of the manipulations are confounded and make interpretation difficult.

The reaction to Block's paper dealing with involvement with automobiles is almost certain to be mixed. On the one hand, the author is to be commended for the compact and relevant review of the literature and for the design and execution of the study. But, on the other hand, one must question the rationale for the study and ask--why do we need one more involvement scale.

In many ways, this study is a straightforward application of the standard methodology for scale development and testing. Unfortunately, there is little concrete support for asserting that the scale measures "involvement." If in fact, Block does want to build on existing work, then it is important to carefully link one's work to existing work. Therefore, a potential and serious flaw in the study is the lack of statistical correlation with other measures of involvement. It seems important to determine, if in fact, this new scale truly contributes any additional insight into our understanding of involvement. Specifically, does this new scale differentiate better than the Lastovicka and Gardner scale or even a simple question like, "How much do you care about cars" or "How important are cars to you?"

In addition, Block seems to overlook the difference between scales that measure differential involvement between consumers for a particular and specific product class and scales that measure differential involvement for an individual consumer toward different products. The study presented here deals only with the former. A brief procedural note is the seeming lack of evidence that criterion validity is adequately assessed. Pre-tests are not reported.

The Barnaby, Philpot and Reizenstein study is a relatively straightforward, practical, management oriented study. It is not exciting, and does not develop any new frontiers. Its analysis is not sophisticated nor the results particularly startling, although they may be useful as a management tool.

Two observations seem appropriate, however. First, no organizing or conceptual framework for this study is specified or evident. Therefore, it is not completely clear why the exact attitudes, attributes or demographics specified should correlate with CVS preferences. Future investigators who will look at this study for direction, may go away empty handed if they desire to learn what predicts CVS preferences from a conceptual perspective.

Second, the authors do not seem to take advantage of the "involvement" literature mentioned in the Block study. CVS is most likely a higher involving activity, at least initially. Therefore, this would suggest an analysis to sort out variables that would be more apt to predict initial behavior toward CVS based on high involvement issues.

TESTING COMPARISON LEVEL AND PREDICTIVE
EXPECTATIONS MODELS OF SATISFACTION

John E. Swan, University of Alabama in Birmingham
Warren S. Martin, University of Alabama in Birmingham

Abstract

Based on past literature and current developments in the field of consumer satisfaction, models were developed relating to the prediction of satisfaction from initial levels of expectations, disconfirmation of expectations and disconfirmations of past experience on similar products. In order to test these propositions, 8 different regression models were created. Data from 67 student subjects relating to the satisfaction of an automobile were analyzed. The disconfirmation of past performance expectations and initial expectations were the strongest predictors of satisfaction with an adjusted R^2 of .47.

Introduction

This paper briefly reviews the comparison level theory of satisfaction proposed by LaTour and Peat (1979b) and reports an empirical test of this theory along with empirical tests of other models of satisfaction. While the comparison level theory of satisfaction has intuitive appeal it has only been subject to a single empirical test (LaTour and Peat 1980, 1979a). A secondary purpose of this paper is to compare the LaTour and Peat model to other models in terms of their ability to explain satisfaction. In particular the contributions of the disconfirmation of predictive expectations and initial level of expectations as predictors of satisfaction were tested.

Models of Satisfaction

Most of the recent models of satisfaction have been oriented towards a paradigm relating initial expectations to perceived performance of the product (or service) (Day 1977). The main proposition is that satisfaction is directly related to the disconfirmation of expectations (Oliver 1977, 1979, Howard and Sheth 1969, Swan 1977). Disconfirmation is a process in which the user subjectively compares post purchase product performance to prior expectations. Positive disconfirmation results if post purchase performance exceeds expectations. Confirmation occurs if performance equals expectations and negative disconfirmation results if performance is short of expectations. Disconfirmation has been conceptualized as continuous variables that can range from a high positive value (as performance exceeds expectations) to a low or zero value (as performance approaches and equals expectations) to a high negative value (as performance decreases relative to expectations (Oliver 1979, 1977; Swan 1977). Satisfaction has also been thought of as a continuous variable that can range from high satisfaction to indifference to dissatisfaction and will be hereafter referred to as "satisfaction" (Day 1977). It should be noted that researchers have not claimed that interval levels of measurement have been achieved for disconfirmation and satisfaction. In addition, whether satisfaction is a continuous variable or whether satisfaction and dissatisfaction are two different states is a point that has not been fully resolved (Leavitt 1977). The main prediction of the Day model, that satisfaction is directly related to disconfirmation has been supported by research findings (Swan 1977; Oliver 1979, 1977; Swan and Combs 1976).

LaTour and Peat (1979b) have discussed the basic disconfirmation of expectations model along with similar ap-

proaches. They have critically evaluated much of the past work in this area and offered an alternative conceptualization of consumer satisfaction. LaTour and Peat's basic approach is that satisfaction is a function of how well a product performs on a set of attributes and for each attribute the customer compares performance to a comparison level (Thibaut and Kelly 1959). Favorable performance in relation to the comparison level for an attribute would contribute to satisfaction due to a positive disconfirmation of expectations.

The important theoretical contribution of the comparison level approach is that three possible sources are identified as forming the comparison level: (1) past experience with different levels of the attribute for similar products used for similar purposes; (2) levels of the attribute that the customer believes similar customers have received; and (3) expectations from information provided by marketers or the expected attribute level based on unique characteristics of the specific purchase situation.

As an example, an automobile owner's satisfaction with the attribute miles per gallon (MPG) achieved with a new car would be related to the difference between the obtained MPG compared to the mileage of his previous car, the mileage his friends claim for their cars, and manufacturers, dealer, and EPA mileage estimates. Overall satisfaction is predicted to be a function of the difference between obtained performance compared to the comparison level which is composed of the three sources of performance standards, past experience, experience of others, and the present situation.

Hypotheses and Models Tested in this Study

As noted above, prior research has found that satisfaction is directly related to disconfirmation. Implicitly the literature has defined expectations as predictive expectations. Miller (1977a) as well as Summers and Granbois (1977) have argued that consumers may have different kinds of expectations including what Summers and Granbois (1977) have termed predictive expectations. Predictive expectations have been thought of as preusage anticipations of how well the product will perform (Swan and Trawick 1979). The first model in this study tested the hypothesis that:

H1: Satisfaction will be directly related to the disconfirmation of predictive expectations.

Disconfirmation was measured as the after usage attribute ratings of the test product (A_i) minus predictive expectations ($PRED_i$) summed across 12 attributes. The model used to test the first satisfaction (S) hypothesis was:

$$\text{Model I. } S = \sum_{i=1}^n (A_i - PRED_i) \quad (1)$$

The set of hypotheses tested included some basic ones plus the same hypotheses with an expectation term added because Oliver (1979, 1977) and Swan (1977) have found that expectations explained some variation in satisfaction above the attributable to disconfirmation of expectations. Expectations may be related to satisfaction due to a level of performance effect. The argument is as follows. Expectations indicate how well or poorly the user anticipates the product will perform. Assume that in two cases performance

matches expectations. In the first case expectations and satisfaction were both at a high level, the users should be satisfied. In the second case both expectations and performance were low, so user satisfaction should be low. Swan and Trawick (1979) have found that performance meeting high expectations results in higher satisfaction than performance equal to low expectations. The second hypothesis and model was:

H2: Satisfaction will be directly related to both predictive expectation (EXP) and disconfirmation of predictive expectations.

$$\text{Model II: } S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PRED}_i) \quad (2)$$

The full LaTour and Peat theory explains satisfaction as follows:

$$S = \sum_{i=1}^n (A_i - W_1 \text{PRED}_i - W_2 \text{PAST}_i - W_3 R_i). \quad (3)$$

Where S = satisfaction; A_i = after usage subjectively experienced attribute level; PRED_i = the expected value of the attribute level based on the characteristics of the particular situation; PAST_i = past experience with attribute levels for similar products; R_i = attribute levels experience by reference persons; n = number of salient attributes; $W_1 W_2 W_3$ = empirically derived weights. Due to some limitations that will be explained more fully below, this study was not designed to test the LaTour and Peat model so that a reduced form of the LaTour and Peat model was tested:

$$S = \sum_{i=1}^n (A_i - \text{PRED}_i - \text{PAST}_i) \quad (4)$$

The PRED_i and PAST_i terms constitute the comparison level, the third hypothesis combined with this model was:

H3: Satisfaction will be directly related to perceived product performance (A_i) minus the comparison level ($\text{PRED}_i - \text{PAST}_i$).

$$\text{Model III: } S = \sum_{i=1}^n (A_i - \text{PRED}_i - \text{PAST}_i) \quad (5)$$

The fourth hypothesis added expectation as an independent variable:

H4: Satisfaction will be directly related to expectation and perceived product performance minus the comparison level.

$$\text{Model IV: } S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PRED} - \text{PAST}_i) \quad (6)$$

The LaTour and Peat approach raises the interesting possibility that satisfaction may be related in some way to the disconfirmation of both past performance and predictive expectations. How the variables should be related is not clear from the literature. Because satisfaction theory and research is in its formative stages, we felt that it would be useful to test for relationships between satisfaction and disconfirmation of past performance (hereafter disconfirmation-past) and predictive expectations (hereafter disconfirmation-predictive) including and excluding an expectation term as variations in the LaTour and Peat framework.

The following hypotheses and models were analyzed:

H5: Satisfaction is directly related to disconfirmation-past performance expectations.

$$\text{Model V: } S = \sum_{i=1}^n (A_i - \text{PAST}_i) \quad (7)$$

H6: Satisfaction is directly related to expectation and disconfirmation past performance expectations.

$$\text{Model VI: } S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PAST}_i) \quad (8)$$

The seventh hypothesis analyzed satisfaction as a function of the disconfirmation of both past performance and predictive expectations and the eighth model included the expectation term:

H7: Satisfaction is directly related to disconfirmation of past performance and predictive expectations:

$$\text{Model VII: } S = \sum_{i=1}^n (A_i - \text{PAST}_i) + \sum_{i=1}^n (A_i - \text{PRED}) \quad (9)$$

$$\text{Model VIII: } S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PAST}_i) + \sum_{i=1}^n (A_i - \text{PRED}) \quad (10)$$

Method

Design of the Study

The design of the study was a longitudinal approach where expectations (PRED_i , PAST_i , EXP) were measured prior to a test drive of a Pinto automobile and after the test drive ratings of perceived performance (A_i) and satisfaction (S) were obtained. The 67 subjects were all students in basic marketing classes at a State University.

An "after only" group (N = 9) was given the questionnaire only after the test drive in order to check for pretest effects in the main before-after measures. A comparison of the after scores in the before-after group to the after only group yielded no significant difference between the groups. This limited sample suggests that the experience of rating the Pinto before the test drive did not bias the after drive ratings.

The data for this paper were originally collected as part of a study of the effects of test driving three different distances on ratings of a Pinto. The three drive groups were combined for this paper as the present analysis was by subject, not by experimental group (length of the test drive) and ratings of the Pinto did not vary by experimental group. The data are useful in terms of an analysis of the hypotheses tested in this paper and fit the hypotheses well; however, it must be recognized that a limitation of the data was that it is a secondary analysis.

Conceptual and Operational Measurement

Predictive expectations (PRED_i) have been conceptually defined as the anticipated benefits from the product (Howard and Sheth 1969). The anticipated benefits were measured as the before ratings of the Pinto on 12 attributes (prestige, room of interior, engine power, number of options and accessories, attractiveness of interior styling, ride comfort, fun to drive, durability, well engineered, well made, handling, exterior styling). An example of an item appears below:

I EXPECT THE PINTO WILL:
 Roomy Small
 Interior : _ : _ : _ : _ : _ : _ : _ : Interior

Disconfirmation has been defined in two ways: perceived disconfirmation is the consumers post-usage perception of how well the product met his expectations, while the concept used in this study, inferred disconfirmation has been defined as the difference between after usage judgments of product performance compared to expectations (Oliver 1980, Trawick and Swan 1980). Disconfirmation of predictive expectations was measured as the difference between the after

minus before ratings of the Pinto on the 12 attributes noted above or n

$$\sum_{i=1}^n (A_i - \text{PRED}_i).$$

One potential problem with the disconfirmation measure is that the same scale was used twice which could lead to similar before and after ratings if the subjects attempted to be consistent. A factor that may have mitigated that problem is the time and driving experience between ratings which was about 15 minutes (short drive), 30 minutes (intermediate drive) or overnight usage of the Pinto (long drive). A second perspective on the disconfirmation measure is that Trawick and Swan (1980) have found that disconfirmation measured as in this paper and disconfirmation measured as the users perception of how well the product met expectations gave essentially the same results as correlates of satisfaction and were strongly intercorrelated.

In order to provide a measure of predictive expectations that was not a part of the disconfirmation measure, predictive expectation (EXP) was measured as the subjects anticipated overall satisfaction with the Pinto.

The LaTour and Peat (1979b) model incorporates a comparison level based on three components of which two were included in this study. One component is the anticipated attribute level, which the current authors believe is equivalent to predictive expectations, PRED_i . The second component is past experience, PAST_i , with attribute levels for similar products. This was measured by the subject's ratings of the automobile that he (she) had driven most frequently.

The adequacy of ratings of the car driven most frequently as an operational measure of PAST_i could be questioned especially if the car driven most frequently was quite different from a Pinto. About 40% of the subjects had compact or subcompact cars which would be similar to the Pinto. Another reason for using the most frequently driven car is that it should be a very salient experience.

In the LaTour and Peat model PRED_i and PAST_i are orthogonal. An examination of the data suggested that PRED_i and PAST_i were orthogonal, as the average simple correlation between the twelve attributes was $r = .08$ (without respect to sign).

The LaTour and Peat model includes attribute weights (W_1, W_2, W_3):

$$S = \sum_{i=1}^n (A_i - W_1 \text{PRED}_i - W_2 \text{PAST}_i - W_3 R_i) \quad (11)$$

The attribute weights are assumed to be equal in this paper. That assumption was made for two reasons. It seems reasonable given that past findings suggest "differential weights provide only limited benefits, compared to the use of equal weights." (Beckwith and Lehmann, 1973, pg. 141). Therefore for an exploratory examination of these models, this approximation should be adequate. In addition, a method for measurement of the weights has not been developed and to do so was beyond the scope of this paper.

The dependent variable, satisfaction has been defined in a number of ways, yet a common theme has been that satisfaction involves a post usage evaluation of a product or service (Day 1977). Satisfaction has been commonly measured using a simple rating of satisfaction and such methods have given results that agree with more complex measures (Miller 1977). Satisfaction was measured here after the Pinto was driven on a "very satisfactory" to "very unsatisfactory" 7 point semantic differential scale.

Analysis

The hypotheses predicted relationships between satisfaction and other variables. A simple correlation matrix was used to evaluate the hypotheses in a preliminary fashion and to see how the independent variables were related. Next, a set of multiple regression models were used to test the assumed predictors of satisfaction. The regression models were chosen such that no pair of predictors were correlated at a .5 level or higher. The .5 level is arbitrary but seems conservative (Green and Tull 1978, pg. 334). When the predictors were significantly intercorrelated the models were limited to an interpretation of only R^2 , not the Beta's. Intercorrelation can result in unstable Beta's, however the R^2 's are meaningful.

Results

The simple correlations in Table 1 suggest that satisfaction is most closely related to disconfirmation-past ($r = .60$) followed by initial expectation ($r = .51$), and disconfirmation of comparison level expectations ($r = .22$). A surprising finding was the lack of a significant disconfirmation of predictive expectations and satisfaction relationship. However, an examination of the data found a negative correlation between expectation and disconfirmation predictive expectations ($r = -.43$) and the positive expectation, satisfaction relationship ($r = .51$) suggests that expectation could be suppressing a disconfirmation-predictive and satisfaction relationship. Such may have been the case, as the regression analysis will show below, when expectation and disconfirmation predictive were included as predictors, both variables were important in determining satisfaction.

The first model (I) of Table 2 indicated that disconfirmation predictive was not a significant predictor of satisfaction unless included in a model, Model II, with expectation. Most of the influence on satisfaction was due to the expectation term, Beta = .70 for expectation versus .44 for disconfirmation predictive. Reasons for that finding were noted above. Disconfirmation of comparison level expectations alone (Model III), did not significantly [$P(F) = .1$] predict or explain much variation in satisfaction ($r^2 = .05$). Expectation and disconfirmation comparison exp. (Model IV) were able to explain a little less than a third ($r^2 = .31$) of the variation in satisfaction.

Model V with the disconfirmation of past performance expectations as a predictor indicated that disconfirmation-past alone explained about 36% of the variation in satisfaction. Adding expectation, Model VI, increases satisfaction explained ($R^2 = .42$). A Model (VII) with disconfirmation of both predictive and past performance expectations suggest that without expectation satisfaction is only related to disconfirmation-past. However, when expectation, disconfirmation of predictive and past are tested as predictors, all three are factors influencing satisfaction (Model VIII) and expectation is the most important predictor (Beta = .48). However a limitation of Model VIII is the high intercorrelation between expectation and the other predictors.

Discussion

The results gave very modest support of the LaTour and Peat model where satisfaction is directly related to the disconfirmation of a comparison level composed of subjectively perceived product performance from which predictive expectations and past performance on similar products are subtracted. The simple correlation ($r = .22$) between satisfaction and disconfirmation of a comparison level was significant but explained little variation in satisfaction. However, a part of the basic concept advanced by LaTour and Peat, that satisfaction is sensitive to perceived product

Table 1
Simple Correlation (r) Between Five Variables^a

VARIABLES	PAST (1)	CL (2)	S (3)	EXP (4)	PRED (5)
1. $\sum_{i=1}^n (A - \text{PAST}_i)^b$	1.00	.72	.60	.45	.14 NS
2. $\sum_{i=1}^n (A_i - \text{PRED}_i - \text{PAST}_i)^c$		1.00	.22	-.08 NS	.48
3. Satisfaction			1.00	.51	.14 NS
4. EXP ^d				1.00	-.43
5. $\sum_{i=1}^n (A - \text{PRED}_i)^e$					1.00

^aAll r's are significant p .05, unless indicated by NS

^bdisconfirmation past performance expectations

^cdisconfirmation comparison level expectations

^dinitial expectation

^edisconfirmation predictive expectations

Table 2
Regression of Satisfaction on Predictors in Eight Models

MODEL AND PREDICTOR(S)	BETAS			r ² or ADJUSTED R ^{2e}	F	P(F)	
	STANDARDIZED REGRESSION COEFFICIENT(S)	B ₁	B ₂				B ₃
I. $S = \sum_{i=1}^n (A_i - \text{PRED}_i)^a$.14	—	—	.02	1.28	.5
II. $S = \text{EXP}^b + \sum_{i=1}^n (A_i - \text{PRED}_i)$.70	.44	—	.40	23.1	.001
III. $S = \sum_{i=1}^n (A_i - \text{PRED} - \text{PAST}_i)^c$.22	—	—	.05	3.46	.10
IV. $S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PRED} - \text{PAST}_i)$.53	.27	—	.31	15.9	.001
V. $S = \sum_{i=1}^n (A_i - \text{PAST}_i)^d$.60	—	—	.36	37.5	.001
VI. $S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PAST}_i)$.30	.47	—	.42	24.8	.001
VII. $S = \sum_{i=1}^n (A_i - \text{PAST}_i) + \sum_{i=1}^n (A_i - \text{PRED})$.60	.06	—	.35	18.7	.001
VIII. $S = \text{EXP} + \sum_{i=1}^n (A_i - \text{PAST}_i) + \sum_{i=1}^n (A_i - \text{PRED})$.48	.35	.30	.47	20.9	.001

^adisconfirmation predictive expectations.

^binitial expectation.

^cdisconfirmation comparison level expectations.

^ddisconfirmation past performance expectations.

^er² is appropriate for bivariate, R² for multivariate analysis.

performance minus past performance on similar products was strongly supported. The results are congruent with findings of LaTour and Peat who created expectancies due to past experience with the product, information from the manufacturer and the experience of other users. They found a past experience effect but no effect of manufacturer or other users on satisfaction. LaTour and Peat's basic finding has been supported by a study that was quite different from theirs in the product, subjects and setting. A major conclusion is that models of consumer satisfaction need to incorporate the disconfirmation of past performance expectations.

In contrast to other research (Swan 1977, Oliver 1979, 1977) satisfaction was not related to the disconfirmation of predictive expectations unless that variable was included in a model with expectations. The negative relationship between expectations and the disconfirmation of predictive expectations shows that those with high expectations had low disconfirmation scores and vice versa. A potential problem with the operational measure of confirmation is that of a ceiling effect on the measurement scales. If an individual expected the Pinto to perform well on an attribute (before rating = 7) and it did so (after rating = 7) the disconfirmation score would be zero (disconfirmation = after-before). A low expectation allows more room on the after scale for improvement to yield a higher disconfirmation score. However, slight satisfaction could be expected from a customer who expected poor performance and received mediocre. Another potential measurement problem was the use of the same scale as the before and after measures.

It appears from this study and prior research (Oliver 1977, 1979; Swan 1977) that the initial level of expectation is a predictor of satisfaction in addition to a disconfirmation effect. This may be due to the ability of expectation to capture an effect of the level of performance that is independent of confirmation. The argument is similar to what was noted above. If an individual had past performance on an auto of 26 miles per gallon and another auto was expected to and did yield 28 miles per gallon, such a customer should be more satisfied than an auto owner who had past experience of 18 miles per gallon in contrast to another car that expected and gave 20 miles per gallon. The confirmation is equivalent, +2 miles per gallon but the level of performance is higher in the first case.

Work on expectations and consumer satisfaction is proceeding along at least three lines. Oliver has investigated components of and the process of consumer satisfaction. LaTour and Peat have suggested that expectations may involve different reference points, in particular this paper supports the hypothesis that the disconfirmation of past performance expectations is related to satisfaction. Finally, Swan and Trawick (1980) have found that satisfaction is related to the disconfirmation of two different kinds of expectations: (1) predictive expectations, which were not significant in explaining satisfaction in this paper and (2) desired expectations which represent the level of product performance that is necessary to please or satisfy the consumer. Work is needed to integrate and empirically test the different approaches to understanding consumer satisfaction.

An approach to the topic of disconfirmation that has received some attention in the literature is Assimilation/Contrast effects (Olson and Dover 1979). That approach is beyond the scope of this work. We feel that it was reasonable to omit assimilation/contrast for two reasons. First, assimilation/contrast would have their effects on the after measures of product performance and would be useful in terms of understanding why a particular objective level of product performance received a certain subjective rating. The subjective after ratings were measured and used in this study to define disconfirmation. This study was not designed to explain the after ratings rather they

were used as basic data. A second reason why assimilation/contrast was omitted is that the approach has criticized on a number of grounds which are too numerous to cover here, (Oliver 1977, Olson and Dover 1979).

Conclusions

In conclusion, this study found that the disconfirmation of expectations based on the past performance of similar product expectation were the strongest predictors of satisfaction. A surprising result was that the disconfirmation of predictive expectations (how the product was anticipated to perform) were not related to satisfaction. Except for the work of LaTour and Peat, the effect of different sources of expectations such as past experience or the experience of others has not been investigated. Sources of expectations could be an important topic for future research in consumer satisfaction.

References

- Beckwith, Neil and Lehmann, Donald (1973), "The Importance of Differential Weights in Multiple Attributes Models of Consumer Attitude," Journal of Marketing Research, 10, 141-145.
- Day, Ralph L. (1977), "Extending the Concept of Consumer Satisfaction," in Advances in Consumer Research, IV, ed., William D. Perreault, Jr., Atlanta, Ga.: Association for Consumer Research, 149-154.
- Green, Paul E. and Tull, Donald S. (1978), Research for Marketing Decisions, 4th ed., Englewood Cliffs, New Jersey: Prentice-Hall.
- Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: Wiley.
- LaTour, Stephen A. and Peat, Nancy C. (1980), "The Role of Situationally-Produced Expectations, Others' Experiences and Prior Experience in Determining Consumer Satisfaction," Advances in Consumer Research, VII, ed., Jerry C. Olson, Ann Arbor: Association for Consumer Research, 588-592.
- _____ and _____ (1979a), "Determinants of Consumer Satisfaction: A Field Experiment," Proceedings 1979 American Psychological Association Convention Division 23, forthcoming.
- _____ and _____ (1979b), "Conceptual and Methodological Issues in Consumer Satisfaction Research," Advances in Consumer Research, VI, ed., William L. Wilkie, Ann Arbor: Association for Consumer Research, 431-437.
- Leavitt, Clark (1977), "Consumer Satisfaction and Dissatisfaction: Bi-Polar or Independent," Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, ed., H. Keith Hunt, Cambridge, Mass.: Marketing Science Institute.
- Miller, John A. (1977a), "Studying Satisfaction, Modifying Models, Eliciting Expectations, Posing Problems, and Making Meaningful Measurements," Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, ed., H. Keith Hunt, Cambridge, Mass.: Marketing Science Institute, 72-91.
- _____ (1977b), "Exploring Some Alternative Measures of Consumer Satisfaction," Marketing: 1776-1976 and Beyond, ed., Kenneth L. Bernhardt, Chicago: American Marketing Association, 661-664.
- Oliver, Richard L. (1980), "Conceptualization and Measurement of Disconfirmation Perceptions in the Production of

Consumer Satisfaction," in Proceedings Fourth Annual Conference on Consumer Satisfaction/Dissatisfaction and Complaining Behavior, forthcoming.

_____ (1979), "Product Satisfaction as a Function of Prior Expectation and Subsequent Disconfirmation: New Evidence," New Dimensions of Consumer Satisfaction and Complaining Behavior, eds., Ralph L. Day and H. Keith Hunt, Bloomington, Indiana: School of Business, Indiana University, 66-71.

_____ (1977), "Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation," Journal of Applied Psychology, 62, 480-486.

Olson, Jerry C. and Dover, Philip A. (1979), "Disconfirmation of Consumer Expectations Through Product Trail," Journal of Applied Psychology, 64, 179-189.

Summers, John O. and Granbois, Donald H. (1977), "Predictive and Normative Expectations in Consumer Satisfaction and Complaining Behavior," in Advances in Consumer Research IV, Ed., William D. Perrault, Atlanta, Georgia: Association for Consumer Research, 155-158.

Swan, John E. and Trawick, I. Fredrick (1979), "Satisfaction Related to Predictive vs Desired Expectations," Proceedings Fourth Annual Conference on Consumer Satisfaction, Dissatisfaction and Complaining Behavior, forthcoming.

_____ and Longman, Douglas S. (1978), "Satisfaction with New Automobile Repair Experience: An Exploratory Study," Proceedings Southern Marketing Association 1978 Conference, eds., Robert S. Franz and Robert M. Hopkins and Al Toma, University of Southwestern Louisiana, 25-28.

_____ (1977), "Consumer Satisfaction with a Retail Store Related to the Fulfillment of Expectations on an Initial Shopping Trip," in Consumer Satisfaction, Dissatisfaction and Complaining Behavior, ed., Ralph L. Day, Bloomington, Indiana: School of Business, Indiana University, 10-17.

_____ and Combs, Linda Jones (1976), "Product Performance and Consumer Satisfaction: A New Concept," Journal of Marketing, 40, 25-33.

Trawick, I. Fredrick and Swan, John E. (1980), "Inferred and Perceived Disconfirmation in Consumer Satisfaction," in Proceedings 1980 American Marketing Association Educators' Conference, forthcoming.

COMPLAINERS AND NONCOMPLAINERS REVISITED:
ANOTHER LOOK AT THE DATA

Kjell Grønhaug, Norwegian School of Economics and Business Administration
Gerald Zaltman, University of Pittsburgh

Abstract

The point of departure for the present paper is previous attempts to profile complaining, dissatisfied consumers. By reviewing the previous literature, three "models" were identified: the "resource", the "learning" and the "personality" model respectively. An empirical test of the three models reveals only modest differences between complaining and noncomplaining consumers. Marketplace participation was found to be the most meaningful explanatory factor.

Introduction

Over the past few years, interest in and research on consumer satisfaction and dissatisfaction (CS/D) and complaining behavior have grown rapidly. Symptomatically, a recent review of 73 publications in the field noted that only 12 of them had appeared prior to 1972 (Hunt 1977).

Previous research efforts reveal the underlying, though implicit, assumption that dissatisfied complainers and noncomplainers, respectively, are different. The following quotation clearly indicates this belief: "... it was hoped that demographic characteristics of such consumers could be found that would distinguish consumers with problems who had not complained from those consumers who had complained" (Kraft 1977, p. 79). Much of the research emphasis has been directed toward profiling the complainers and noncomplainers by stable, sociodemographic characteristics such as age, education, and income. Such characteristics may definitely be useful as resource-related or social indicators (Bauer 1966), but there are definitely reasons why such indicators only exhibit modest descriptive and explanatory power in a consumer behavior context (cf. Sheth 1977).

Previous Research

At the more specific level, the following conclusions may be drawn from previous research:

- Only a fraction of the dissatisfied consumers actually takes overt action and complains. Andreasen and Best (1977) in a telephone survey conducted among 2,400 households found that the consumer (buyer) was dissatisfied in one of five purchases, but in less than 50 percent of these cases did he or she take any action. A recent review of the complaining behavior studies (cf. Robinsen 1978) strongly supports these findings. Although variations in propensity to complain are observed across studies, the overall conclusion to be drawn is that only a fraction of the dissatisfied purchasers are overt complainers.
- As noted above, various sociodemographics have been applied extensively in previous research to profile the complainers and noncomplainers, respectively. The results are, however, in no way conclusive. Liefeld, et al. (1975) found that Canadian consumers who had written complaint letters were middle-aged, better educated, earned higher incomes, and more represented by managerial and professional heads of households. Similar findings regarding the use of demanding sources of complaints have been reported by Grønhaug (1977), Zaichkowsky and Liefeld (1977), Kraft (1977) and

Warland et. al. (1975).

With regard to complaining in general (and, in particular, when looking away from complaining through the more demanding sources), the descriptive and predictive power of the various sociodemographics reveal very mixed results. By relating various sociodemographics (age, education, income, place of living, social status, occupation) to the propensity to complain due to a perceived dissatisfaction with a variety of nondurables (milk, meat, coffee, vegetables) and durables (textiles, cars), Grønhaug (1977) found the sociodemographics do possess no descriptive or explanatory power. Warland, et al. (1975) found, on the other hand, by relating the action due to the "most recent, salient negative experience in the marketplace" (p. 151), that the active complainers in general are younger and above average with respect to social status, income, education, and group membership. Zaltman et al. (1978) found age negatively related to propensity to complain. Granbois et al. (1977) found that sociodemographics overall possess almost no descriptive power. For some products, however, sex was found to possess some descriptive power. Wall, et al. (1977) found in a sample of female buyers of clothing that age was somewhat negatively related to the propensity to complain. Lawther (1978) found that consumers who are less socially integrated were less prone to make overt complaints than were the more integrated consumers.

The apparently inconsistent findings regarding the descriptive and explanatory power of the sociodemographics have led researchers to include other variables in their research. Inclusion of perceptual variables have revealed the almost self-evident finding that the more aware an individual is of defects, the higher the propensity to complain (cf. Kraft 1977, Zaltman et al. 1978). Also, attribution theory in terms of perceived causes to the dissatisfaction has been applied in the context of complaining behavior, revealing very promising and insightful results (Valle and Wallendorf 1977, Krishnan and Valle 1979). Personality characteristics have also been applied in attempt to profile complainers. Wall et al. (1977) found some items related to personality (liking/disliking, ways of perceiving things and self confidence), to possess some descriptive power among a sample of female buyers related to satisfaction/dissatisfaction with clothing. In an attempt to describe the personality profiles of consumer complaint letter writers, Zaichkowsky and Liefeld (1977) concluded however, "... that consumer complaint letters cannot be distinguished on the basis of personality types" (p. 128). In other words, the attempts to include various personality traits cannot be considered as successful.

The research on consumer satisfaction/dissatisfaction and complaint behavior is still in its infancy. Several objections regarding previous research may be made (cf. Ølander 1977, Haines 1979). No attempt will be made to discuss the inherent weaknesses in previous research. However, we would like to address the reader's attention to the lack of explicit arguments why and how the various indicators used should discriminate between complainers and noncomplainers, respectively.

Making the Assumptions Explicit

As noted above, in most of the previous research the assumptions underlying the choice of concepts and models have been more or less implicit. By reinterpreting previous research, the following assumptions (or models) can all be traced.

(1) The resource model: To make overt complaints requires resources. Resources are not equally distributed, and thus the propensity to complain will vary across the members of the society. This type of model (or argument) has determined very much of the consumer and welfare policy. Sociodemographics -- in particular, age, income and education -- have all been treated as individual resource indicators (cf. Atkinson 1975, Bauer 1966, Rein 1976).

(2) The learning model: Learning plays an important role in consumer behavior (cf. Howard 1977). Products and brands are learned; expectations and attitudes are learned; and how to handle purchases and complaints is learned. The basic underlying assumption or model in the CS/D-research is that "knowledge is power", i.e. the experienced, welltrained buyer will be better off. Learning in terms of awareness of unfair marketing practices and knowledge about consumer rights has been included in previous research (cf. Kraft 1977, Zaltman et al. 1978, Lawther 1978).

(3) The personality model: Some recent attempts have been made to incorporate personality variables into the research on CS/D (cf. Wall et al. 1977, Kraft 1977). The underlying assumption is that certain personality characteristics, such as high degree of self confidence, is associated with ability to perceive dissatisfaction and handle complaints.

In focusing on complaining behavior, we will -- as in previous research -- consider perceived dissatisfaction as a prerequisite to overt complaints. Based on the previous discussion, variables will be borrowed from the "resource", the "learning", as well as the "personality" model in order to describe and explain complaint behavior.

Methodology

Source of Data

Data used in this paper are from a large research study which investigated the consumer problems of the elderly. The data were obtained from two structured mail questionnaires which were developed from focus group interviews. The questionnaires were sent to 4,000 consumers between the ages of 25 and 80 who were members of a national panel maintained by Market Facts, Inc. Usable responses covering both waves (spaced by a six-month interval) were 2,849. The response rate to the original 4,000 questionnaires was 71 percent. The response rate for the second wave was 90 percent of the first wave sample.

The Sample

Due to the purpose of the data gathering, the elderly were deliberately oversampled. Compared to the U.S. population, the sample is skewed toward the elderly, which is due to deliberate disproportional sampling due to the focus of the study and the data gathering. Furthermore, the sample shows a slight underrepresentation of high school graduates and a moderate overrepresentation of college graduates. Compared to the U.S. population, the annual income in the sample was somewhat higher. In addition, it should also be noted that compared to the national population, the sample had a lower fraction of individuals whose spouses were no longer living, and fewer males living alone.

Measurement

Reported below are the measurements used in this study: Dissatisfaction. In order to map dissatisfaction, the following measurement was used. The respondents were confronted with a rather long list of products and services, and the following question was asked: "Please read the products and services listed below and 'X' all of those where you felt bad about your buying experience". From this operational definition, it follows that no restriction in time is given; in other words, many of the reported "bad buying experiences" may have taken place a long time ago. A summary measure for dissatisfaction across products was worked out by adding up the "bad buying experiences" (range: 0,, 21).

Complaint. If dissatisfied, the consumer (buyer) may react in several ways. This was mapped in the following way. For the "worst buying experience" a list of 12 different actions, which also included "no action", was given to the respondents, and the following question was asked: "Please 'X' below any actions you took because of the problems you had." The various actions listed may definitely put different requirements on the dissatisfied consumer. Actions requiring some sort of overt behavior directed toward the marketer or consumer agency were labeled "activist" complaint strategies.¹ In this paper the persons applying the "activist" strategy were contrasted with the "passivist", i.e., those who "took no action at all".²

Resources are often thought of in terms of money/economy. However, the individual resources should in no way be restricted to money/economy only. Here a distinction was made between the following types of resources: economy; problem solving capacity; health; time, and social resources. Several indicators were used to map the resource dimensions: economy (income, financial problems); problem solving capacity (education, decision-making skill, planning horizon); time available; health (age, various health problems); social resources (social contacts, perceived problems in getting help with consumer related problems).³

Learning was measured by mapping buying experience and knowledge, i.e. ability to handle the buying problems.

Personality. Indicators mapping involvement, self-confidence, and perceptions of being treated fairly besides perceived consumer influence were applied in order to map personality.

The measurements applied include self-reports on perceptions and behavior as well as sociodemographics. The dissatisfaction and complaint measures coincide to a substantial degree with measures used in previous research. With regard to the resource measures, the sociodemographics

¹The "activist" strategy included "complained to the person who sold me the product or service", "complained to the company or store", "complained to a consumer agency", "complained to a public agency or my congressman", "complained to a newspaper or magazine", and "consulted or hired a lawyer to protect my interests".

²Besides the "activist" strategy, strategies related to "word-of mouth", "stop buying", and "the passivist" strategy ("no action") were located. By examining the various strategies according to the variables (indicators), the "activist" and "passivist" strategies were really found to represent the extremes.

³The indicators included the mapping of several "life events" such as the experience of financial crises and health problems.

(income, education, and health) have been extensively applied in previous research (cf. Sheth 1977). The "life events" (cf. "financial crisis" and "health problems") have not been applied in previous research, and are included in order to indicate changes. The various indicators related to knowledge, experience, planning horizon, and perceived decision-making skill, have all been widely used in previous consumer research.

Findings

Reported below are the major results from this investigation: 30.3 percent of the respondents had not any "bad buying experience". Of those who had at least one such experience, 834 respondents had applied the "activist" strategy, and 432 the "passivist" strategy. These two subsets, the "activists" and the "passivists", will constitute the sample of analysis (n = 1,266).

Table 1 summarizes the results from the bivariate analysis breaking down the various variables by the two strategies applied. The statistical procedure is t-test applied to differences of means; i.e. the differences between the "passivist" and "activist"-scores. The results are surprising in term of lack of associations between complaint strategy and the various indicators. Closer inspection of Table 1 reveals that the active complainers are:

- higher in experience (i.e., buying experience);
- higher in income;
- higher in education; and
- they tend to be younger than do the noncomplainers.

However, as seen from Table 1, none of the indicators related to health (except age), time, social resources or personality revealed any differences between the complainers and noncomplainers.

Table 1

Variables Related to Resources, Learning, and Personality by Complainers and Noncomplainers

	Complaint Strategy		Total	Significance Level
	"Passivist"	"Activist"		
1. Resources				
(a) Economy:				
Income	4.3225	5.1248	4.8513	p < .001
Financial problem	.1019	.0899	.0940	n.s.
(b) Problem-Solving Capacity:				
Education	3.2689	3.2422	3.3704	(p = .0625)
Decision-making skill	1.7870	1.8861	1.8530	n.s.
Planning horizon	2.3889	2.9140	2.3276	n.s.
(c) Time: (perceived time)				
	2.5995	2.6559	2.6367	n.s.
(d) Health:				
Age	5.4014	5.2197	5.2817	(p = .0477)
Health problems:				
Vision	.4828	.4746	.4774	n.s.
Movement	.1435	.1415	.1422	n.s.
Depression	.0116	.0072	.0087	n.s.
Physical activity	2.4375	2.4868	2.4700	n.s.
(e) Social Resources				
Social Contacts	6.9630	7.1823	7.1074	n.s.
Help	2.7487	2.7846	2.7725	n.s.
2. Learning				
Experience	2.6019	3.1451	2.9597	p < .001
Knowledge	2.7176	2.7806	2.7591	n.s.
3. Personality				
Involvement	2.1505	2.1427	2.1453	n.s.
Self-confidence	1.9352	1.9017	1.9131	n.s.
Fair Treatment	2.1276	2.1609	2.1496	n.s.
Consumer Influence	1.8156	1.8107	1.8123	n.s.

Discussion

Several questions might be raised based on the preceding data presentation, such as: Why are there so few differences between the complainers and noncomplainers, and more basically, what do the results really mean? We will address our attention to the last question.

The indicators for which there are discrepancies are not necessarily unrelated. By inspecting the intercorrelation matrix, the following picture emerges:

Table 2

Correlation Matrix: "Activist", Education, Income, Age and Experience*)

	Experience	Education	Income	Age
"Activist" ^x	.1453	.0650	.1144	-.1119
Experience	x	.1467	.1964	-.2304
Education		x	.3483	-.1351
Income			x	-.3544

*) all the correlation coefficients are significant at the .001 level.

Here ordinary product-moment correlations have been applied. In doing so the nominal scaled complainer categories are turned into dummy variables (i.e. "activist" = 1, "passivist" = 0). Experience is here the summary measure of previous buying experiences. Age and education are in fact ordinal scaled variables. However, only modest differences were detected by applying nonparametric measures of association.

The correlation matrix reveals that "activists" are positively related to experience, education, and income, but negatively related to age, just as demonstrated in Table 1. However, it is also evident that the various descriptive variables are intercorrelated. By calculating the various partial correlations and controlling for the subsequent variables, the following emerges:

$r_{act., exp.}$ (controlled for education, income, age = .1130 (p < .001)).

$r_{act., age}$ (controlled for education, income, experience) = .0560 (p = .048).

$r_{act., educ.}$ (controlled for age, income, experience) = .0166 (p = .558).

$r_{act., inc.}$ (controlled for education, age, experience) = .0678 (p = .017).

In other words, the "experience" variable by far possesses the greatest descriptive power. Similar results were found by performing multiple discriminant analysis. However, the predictive power of the derived discriminant function was found to be modest as assessed by comparing the number of cases correctly classified with the by chance result, which also follows from the rather low correlation coefficients as demonstrated in Table 2, and the lack of association between complainers (or noncomplainers) and most of the predictors as shown in Table 1.

When taking a closer look at the applied "experience" measure, it obviously relates to buying activities. Or, maybe better, the measure represents a rough proxy for market-place participation. And, according to the present authors, this represents the key to the explanation. First, buying is not the sole activity with which man is preoccupied, nor does this activity have to be the most important one.

In fact, the individual member of the society is confronted with various types of constraints including limited time and span of attention.⁴ This, in turn, directs our attention to one of the almost forgotten elements in a marketing context, namely, that the consumer role may be one of the many roles played by the individual. He or she will furthermore put more or less emphasis on this role. The relative importance of this role for him or her will be closely related to the time devoted and the activities performed in the context of this role. And this role will be played across all social strata!

Why is this role (marketplace participation) related to, or better, associated with income, education, and age (cf. Table 2)? First, economic resources are a prerequisite to participate in the marketplace as a buyer. Second, income and education are positively correlated, and thus by controlling for income, the association between marketplace participation and education drops dramatically. One intriguing explanation may be proposed for the negative relationship between age and complaining, which is that this phenomenon is caused by a cohort⁵ effect. However, by controlling for marketplace participation, the effect of age on noncomplaining is (as demonstrated above) dramatically reduced.

The marketplace participation argument launched above may also shed some light on previous findings. The results reported by Warland et al. (1975) may partly be explained by variation in marketplace participation. The impact of the sociodemographics (which is in the same direction as the findings presented here) may also to some extent be explained in terms of variations in marketplace participation. Similarly, the impact of sex in findings reported by Wall et al. (1977) and Granbois et al. (1977) may be explained in terms of products primarily dealt with by women, and represent thus an artifact introduced by the products studied (cf. Rosenthal and Rosnow 1969). Symptomatically no differences were detected between sexes when applying our aggregate "marketplace participation" measure. But why are our deflections for the "marketplace participation" measure not more pronounced? The explanation is probably due to the very rough nature of this proxy measure. Refined measurements taking such factors as allocation of time, the variety of activities performed, and importance placed on such activities would probably yield much stronger results.

The findings reported and explanations provided do also have consumer policy implications. First, the elderly, the poor and the individual low in education do not necessarily react more passively to perceived dissatisfaction -- as demonstrated above -- than do the individuals higher in resources, which clearly seems to contradict public policy practice. In fact, for groceries the low income consumers were found in several studies to be higher in price knowledge than were the high income consumers (cf. Goldman 1977, Gabor and Granger 1961), which is in concordance with the fact that the lower income consumers necessarily have to be more conscious due to less economic resources. This "apparent" lack of the "expected" negative relationship between being active in complaining and low in income and education and high in age may also be due to the following explanation. The

more involved in the marketplace, the higher the probability of being exposed to buying problems ($r = .224$; $p < .001$). In addition, the following argument related to social class and consumption style seems to be relevant: Income and education are both well-known indicators of social class. Furthermore, basic values and consumption patterns vary across the various social strata. Expressive values and thus emphasis on expressive consumption is more frequent among the lower classes as compared to the predominance of instrumental values and behavior in the upper social classes (cf. Segal and Felson 1972). Intuitively, expressive consumption is less demanding, the consequences may be easier to foresee, and the purchases are thus less likely to produce dissatisfaction. Thus the "expected" higher fraction of dissatisfied, passive consumers is not showing up in the research. On the other hand, if the consumption aspiration level of the lower social segments is raised, this may lead to higher degree of perceived dissatisfaction.

References

- Andreasen, A.R. and Best, A. (1977), "Consumers Complain -- Does Business Respond?" Harvard Business Review, 55, No. 4, 98-101.
- Atkinson, A.B. (1975), The Economics of Inequality, Oxford: Clarendon Press.
- Bauer, R.A. (1966), Social Indicators, Cambridge, MA: M.I.T. Press.
- Bettman, J.R. (1973), "Perceived Risk and Its Components: A Model and Empirical Test", Journal of Marketing Research, 10 (May), 184-190.
- Gabor, A. and Granger, C.W.J. (1961), "On the Price Consciousness of Consumers", Applied Statistics, 10, No. 1 (March), 170-178.
- Goldman, A., "Consumer Knowledge of Food Prices as an Indicator of Shopping Behavior", Journal of Marketing, 41, (October), 65-75.
- Granbois, D., Summers, J. and Frazier, G.L. (1977), "Correlates of Consumer Expectations and Complaining Behavior", in Consumer Satisfaction, Dissatisfaction, and Complaining Behavior, R.L. Day, ed., Bloomington, IN: Indiana University, 18-25.
- Grønhaug, K. (1977), "Exploring Consumer Complaining Behavior: A Model and Some Empirical Results", in Advances in Consumer Research, Vol. IV, W.D. Perreault, ed., 159-165.
- Haines, G. (1979), "Three Papers on Consumer Satisfaction/Dissatisfaction: A Comment", in Advances in Consumer Research, Vol. VI, W.L. Wilkie, ed., Ann Arbor, MI: Association for Consumer Research, 450-452.
- Howard, J.A. (1977), Consumer Behavior: Application of Theory, New York: McGraw-Hill.
- Hunt, H.K. (1977), "Consumer Satisfaction and Dissatisfaction: A Developing Methodology", in The New Role of Marketing Professionals, P.J. LaPlaca, ed., Chicago: American Marketing Association, 241-244.
- Kraft, F.B. (1977), "Characteristics of Consumers as Complainers and Complaint and Repatronage Behavior", in Consumer Satisfaction, Dissatisfaction, and Complaining Behavior, R.L. Day, ed., Bloomington, IN: Indiana University, 79-84.

⁴Attention as a limited capacity is dealt with in several of the social sciences. For an interesting discussion, cf. March and Olsen (1976).

⁵The rationale behind such an effect is that when brought up in a certain time period, the individual is "stigmatized," and will thus carry with him or her the impact of such previous learning (stigma).

Krishnan, S. and Valle, V.A. (1979), "Dissatisfaction Attributions and Consumer Complaint Behavior", in Advances in Consumer Research, Vol. VI, W.L. Wilkie, ed., Ann Arbor, MI: Association for Consumer Research, 445-449.

Lawther, K. (1978), "Social Integration of the Elderly Consumer: Unfairness, Complaint Actions, and Information Usage", in Proceedings of the 1978 Educators' Conference, Chicago: American Marketing Association, 341-345.

Liefield, J.P., Edgecomb, F.H.C. and Wolfe, L. (1975), "Demographic Characteristics of Canadian Consumer Complainers", Journal of Consumer Affairs, 9, 73-80.

March, J.G. and Olsen, J.P., eds. (1976), Ambiguity and Choice in Organizations, Bergen: Scandinavian University Books.

Ølander, F. (1977), "Consumer Satisfaction -- A Sceptic's View," in Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, H.K. Hunt, ed., Cambridge, MA: Marketing Science Institute, 455-488.

Rein, M. (1976), Social Science and Public Policy, Baltimore: Penguin Books, Inc.

Robinson, L.W. (1978), "Consumer Complaint Behavior: A Review with Implications for Further Research", paper presented at the Third Annual Conference on Consumer Satisfaction/Dissatisfaction and Complaining Behavior.

Rosenthal, R. and Rosnow, R.L. (1969), Artifact in Behavioral Research, New York: Academic Press.

Segal, D.R. and Felson M. (1972), "Social Stratification and Family Economic Behavior", in Family Economic Behavior, E.B. Sheldon, ed., Philadelphia: Lippincott Company.

Sheth, J.N. (1977), "Demographics in Consumer Behavior", Journal of Business Research, 5 (March), 129-138.

Valle, V.A. and Wallendorf, M. (1977), "Consumer Attributions and the Cause of Their Product Satisfaction and Dissatisfaction", in Consumer Satisfaction, Dissatisfaction and Complaining Behavior, R.L. Day, ed., Bloomington, IN: Indiana University, 26-30.

Wall, M., Dickey, L.E. and Talarzyk, W.W. (1977), "Predicting and Profiling Consumer Satisfaction and Propensity to Complain", in Consumer Satisfaction, Dissatisfaction and Complaining Behavior, R.L. Day, ed., Bloomington, IN: Indiana University, 91-101.

Warland, R.H., Herrman, H.O. and Willits, J. (1975), "Dissatisfied Consumers: Who Gets Upset and Who Takes Action", Journal of Consumer Affairs, 9, 148-163.

Zaichkowsky, J. and Liefield, J. (1977), "Personality Profiles of Consumer Complaint Letter Writers", in Consumer Satisfaction, Dissatisfaction and Complaining Behavior, R.L. Day, ed., Bloomington, IN: Indiana University, 124-129.

Zaltman, G., Srivastava, R.K. and Deshpande, R. (1978), "Perceptions of Unfair Marketing Practices: Consumerism Implications", in Advances in Consumer Research, H.K. Hunt, ed., Ann Arbor, MI: Association for Consumer Research, 247-253.

EFFECT OF SATISFACTION AND ITS ANTECEDENTS
ON CONSUMER PREFERENCE AND INTENTION

Richard L. Oliver, Washington University
Gerald Linda, Marsteller, Inc.

Abstract

Theorists and researchers in the satisfaction area now generally agree that product satisfaction intervenes between expectancy disconfirmation and various postpurchase cognitive states including attitude and behavioral intention. Research in a variety of settings has supported the effect of expectation and its disconfirmation on satisfaction, but only a small number of studies addresses the cognitive consequences of satisfaction decisions and none reports data on choice processes such as brand selection. This study examines the influence of satisfaction and its determinants on behavioral intention and product preference in a simulated two-stage consumer usage situation. Generally it was found in both overall and summed attribute analyses that satisfaction was a function of expectation and disconfirmation, that intention was a function of satisfaction, and that preference was influenced by satisfaction and disconfirmation, the latter having the greater effect.

Introduction

The field of consumer satisfaction, long neglected by cognitive consumer behaviorists, is now beginning to develop a research tradition. Based on theoretical works by Engel, Kollat, and Blackwell (1968), Howard and Sheth (1969), Andreasen (1977), and Day (1977), research in both the laboratory and the field is proliferating. This has prompted reviewers (Day 1977, Olander 1977, Oliver 1977, LaTour and Peat 1979) to call for theoretical extensions and methodological refinements. Specifically, researchers have been asked to position satisfaction and its related constructs in cognitive models of consumer choice and to experiment with more coherent measures of the expectancy disconfirmation process.

Theoretical Perspective

A number of theories have been proposed to explain satisfaction decisions. These have been reviewed in detail elsewhere (Anderson 1973, Oliver 1977, LaTour and Peat 1979) and are not repeated here. In summing up the evidence to date, Oliver (1980b) has suggested that the most epistemologically efficient explanation derives from Helson's (1948, 1964) adaptation level theory which suggests that judgments of newly perceived stimuli are affected by prior experience with the general class of objects (the adaptation level) and the discrepancy perceived between the new stimulus and previously determined stimulus levels. Discrepancy, Helson noted, was a relative individually-specified phenomenon and has no necessary correspondence to objective differences.

Helson's (1964) theory may be viewed as isomorphic with the process of expectancy disconfirmation thought to occur in satisfaction judgments, differing only in terminology. One's expectation level, however determined, performs the function of an adaptation level. States of satisfaction/dissatisfaction are hypothesized to result, in part, from a comparison between one's perception of product performance and the expectation level. This discrepancy is known as (positive or negative) disconfirmation. Because the expectation level provides a baseline about which disconfirmation takes place, satisfaction is easily shown to be an additive function of expectation and disconfirmation. This interpretation was first suggested and empirically demon-

strated in the industrial and social psychological literature (Ilgen 1971, Weaver and Brickman 1974) and has received limited support in consumer environments (Oliver 1977, Swan 1977, Madden, Little, and Dolich 1979). None of these studies extended the analysis to actual post-exposure choice behavior, however.

Measurement Issues

Researchers remain in disagreement over the most appropriate way to measure the antecedents of satisfaction and satisfaction itself. The concept of expectations, in particular, has been somewhat evasive from a psychological standpoint. Perhaps the most general workable definition is that of Katona (1964) who referred to expectations as a "subclass of attitudes that point to the future (p. 34)." As such, they may be seen as predictions of affect.

Early satisfaction researchers pursued a cognitive approach to the measurement of expectation and used product and attribute ratings as proxy variables for one's expected performance. Later, Olson and Dover (1979) and Oliver (1980a) adapted Fishbein (1967) scaling to this same task. Olson and Dover used only the belief elements of Fishbein's *ab* scales while Oliver combined the cognitive and affective dimensions to provide an evaluative belief score. Both approaches demonstrated the intended effect, but neither was a pure measure of predictive affect.

Disconfirmation has followed a similar developmental process. The discrepancy concept was initially viewed in the objective sense. Experimenters using manipulations in the laboratory assumed that disconfirmation existed whenever subjects' product experience deviated from the materials used in the expectation-creation task. Unfortunately, manipulation checks on perceived discrepancy were not taken. As a result, neither disconfirmation nor its magnitude and direction were determinable. Later researchers used the difference scores between preexposure and post-exposure attribute ratings (Oliver 1977, Swan 1977, Madden, Little, and Dolich 1979). While this approach to discrepancy was "real" in the sense that it involved a mathematical comparison between perceived performance and some prior cognition, it proved to be confounded with expectation and, on a conceptual plane, ignored the "surprise" value of disconfirmation. To address these problems, "better than expected -- worse than expected" scales have been used by Oliver (1977, 1980a) and have shown promise as predictors of satisfaction and other postexposure cognitions. This approach will be expanded here.

Finally, satisfaction has been measured using a variety of postexposure variables having affect as a common theme. Most researchers would agree, however, that product evaluations and attitudes are not operational definitions for satisfaction. Other approaches include Likert scales and "raw" satisfied-dissatisfied items. The latter have the greatest face validity but tend to produce positively skewed distributions reflecting a halo effect (Andrews and Withey 1976, Day and Bodur 1977). In comparison, multi-item Likert or semantic differential scales present the problem of finding a sufficient number of synonyms for satisfaction without also sampling the affective like-dislike domain. Interested readers are referred to some preliminary findings reported by Westbrook and Oliver (1981).

Study Perspective and Objectives

In an effort to extend further the basic satisfaction model to consumer intention and choice, a two-stage field study of consumer reactions to the preprocessing of new apparel fabrics was performed. In accord with the theoretical perspective of expectations as predictions of affect, expectations were viewed as a priori perceptions of the amount of satisfaction to be received from the product as a whole and in terms of its salient attributes. This required that measures of expectation be obtained before the exposure situation was encountered.

To maintain a consistent perspective to the expectancy disconfirmation process, disconfirmation was viewed as the perceived satisfaction deficit (surplus) after the product experience. Measured in terms of less (more) satisfied than expected, disconfirmation as viewed here focused on the comparative process and not on actual states of performance per se. The "mathematics" of discrepancy judgments were made by the subjects who responded with net feelings of fulfillment or lack thereof. As before, both overall and attribute-specific disconfirmations were elicited.

The last part of the second study phase, in keeping with the temporal nature of the model, required that satisfaction and subsequent cognitions be obtained. Care was taken to insure that the proper sequence of variable measurement was maintained to overcome partly problems associated with the necessity for concurrent measurement in this last study phase.

The approach taken here differs in a number of respects from prior studies and may provide a different perspective to satisfaction processes. First, the measurement of satisfaction and its antecedents involved consistent referents across study stages. Specifically, operationalizations of expectations of satisfaction, disconfirmation of satisfaction, and satisfaction itself were constructed with the same semantic phrases and attribute lists. Previous studies have confounded a host of different referents and algebraic difference scores in testing the basic model. While a multiple indicator approach will eventually be required for purposes of construct validation, research on satisfaction could benefit greatly from attempts to establish a coherent framework of satisfaction-related semantic concepts.

Secondly, attribute-specific satisfaction processes were compared to an overall approach in an effort to determine if the scheme suggested here operates in both modes. Attribute-specific measurement may not fully represent all salient attributes for all subjects, but does provide diagnostic information on an individual attribute basis. In contrast, overall scales serve to capture the totality of the cognitive process at the sacrifice of detail and, in addition, can be evaluated on the basis of internal consistency reliability. To date, few studies have been conducted using attribute-specific and/or multi-item overall scales for satisfaction and expectancy disconfirmation and none has compared these two variations. Hopefully, some insight into the relationship between specific and general satisfaction processes will be evident from this analysis.

Method

Overview

A mall intercept study of male and female shoppers was conducted in a suburban Chicago shopping center. After an initial screening to determine eligibility requirements for the product group under investigation, subjects were asked to examine two pairs of men's sleeping apparel having different texture and softness. All product samples were recently manufactured and contained the sizing and other fabric conditioners used to facilitate construction

of the garment. In the first stage of the study, the subjects indicated how satisfied they expected to be with both garments and garment attributes in daily normal usage and care, their interest in buying the products, and their preference for one over the other.

In the second stage of the study, the subjects were shown the identical pair of sleeping apparel after each had been laundered five times. Laundering removes the sizing and other pretreatment chemicals and generally makes fabrics softer, although to varying degrees. Subjects now responded to overall and attribute-specific disconfirmation scales on each product, then to overall and attribute-specific satisfaction scales, and finally to purchase intention and new preference.

Subjects

Based on the sponsoring company's directive, 500 mall shoppers were approached on a quota basis to achieve a final sample of 250 males and 250 females. Males were required to be fifteen years of age or older (the target group) and to wear pajamas at least twice a week. Females, known to be the primary purchasers of men's pajamas for members of their families, were required to be 21 or older, to have males 15 years or older who wear pajamas at least twice a week in the household, and to usually buy the pajamas for the men in their family. The average age of males in the final sample was 38 years and that of females was 46.

Measures

Attribute-specific variables. Evaluative criteria for the attribute-specific measures were obtained by pretesting a list of 20 factors known to be related to judgments of wearing apparel on an independent sample of 400 consumers. Based on mean importance scores, a shorter list of nine attributes was judged sufficient by company management and is not reported here for proprietary reasons. Based on the Katona (1964) approach, expectations were measured by asking the respondent to indicate how satisfied "you or the male you would buy for" would be with regard to each attribute. Scales ranged from "very satisfied" to "very dissatisfied." The nine expectation scores were then summed to form the attribute expectation measure. Pre-exposure preference, in turn, was based on the subjects' actual preferred choice of one pair of sleeping apparel over that of the other.

Postexposure attribute disconfirmation was measured on 15-point scales where the midpoint defined each attribute as expected. The negative disconfirmation pole described attributes as worse (e.g., not as soft) than expected while the positive pole described them as better (e.g., softer) than expected. Attribute satisfaction was obtained after all disconfirmation measures were taken to insure that the measurement process maintained the temporal ordering of the underlying framework. The same items used for one's expectations of satisfaction were used here as well except that the lead-in to the satisfaction items simply asked the respondent to indicate how satisfied he/she was now. Finally, postexposure intention was measured on a five point "definitely would buy" to "definitely would not buy" scale while postexposure preference was obtained by asking the subjects to reconsider the pajamas a second time and to choose the pair they now preferred. The criterion used in the study was change (no change) in preference coded in 0-1 format.

Overall variables. A consistent three-referent scheme was used to measure overall expectation, disconfirmation, and satisfaction. Based on extensive work by Andrews and Withey (1976) and Campbell, Converse, and Rodgers (1976) on the meaning of satisfaction, two constructs having close (but not identical) semantic properties were combined with satisfaction (expressed as satisfied-dissatisfied) to form three-item scales. The two related constructs, happiness

(happy-unhappy) and pleasantness (pleasant-unpleasant) were thought to reflect emotions close to those involved in a satisfaction response. Thus, expectation was measured by asking respondents how (satisfied, happy, pleased) they expected to be; disconfirmation was obtained on a "more than-less than" (satisfied, happy, pleased) scale; while satisfaction was an outright rating on the three items. Alpha reliability coefficients over the total sample were .85, .79, and .94 for expectation, disconfirmation, and satisfaction respectively.

Analysis

Because the males in the sample were judging their own satisfaction while females were judging satisfaction vicariously for the male they would buy for, all analyses were run separately for men and women. In addition, separate analyses were performed for the overall and summed attribute scales to determine if equivalent results would be found. Finally, only the initial preferred pair of pajamas was analyzed across the two stages of the study. Results for the second (initially least preferred) pair were virtually identical to the first and are not reported here.

The models tested derive from those suggested in Oliver (1980a). His basic hypotheses were adapted to the context of the present study and appear as follows:

$$\text{Satisfaction} = f(\text{Expectation, Disconfirmation})$$

$$\text{Preference} = f(\text{Expectation, Disconfirmation, Satisfaction})$$

$$\text{Intention} = f(\text{Expectation, Disconfirmation, Satisfaction, Preference})$$

Elementary path analysis was employed to test these propositions. Of interest are the magnitudes of the coefficients of determination and the comparative magnitudes of the (standardized) regression coefficients attributable to each of the independent variables within equations for each sample taken separately and for each variable across samples. Relative stability of the weights across situations would lend further support to the basic expectancy disconfirmation model.

Results

Zero-order correlations between all variables across the two sample and measurement groups are shown in Table 1. The data are remarkably consistent across all analyses. Satisfaction was significantly correlated with its hypothesized antecedents in all cases, although disconfirmation yielded the highest correlations in the female sample. Preference (coded as change in preference) was most highly correlated with disconfirmation in all cases although significant correlations were also obtained with satisfaction and to a lesser degree, with intention. Surprisingly, preference was not correlated with expectation, a result explained more fully in the discussion section.

Finally, intention was universally correlated with all antecedents. In fact, with the exception of the preference variable, the magnitudes of the intention intercorrelations reflect the temporal nature of the proposed scheme exactly ($r_{53} > r_{52} > r_{51}$). Although no sex differences were found in this analysis, the overall scales produced higher associations, particularly for the intention-satisfaction relationship.

Table 2 shows the results obtained when the suggested models were tested in a sequential path analysis. The data show, first, that satisfaction was an additive combination of expectation and disconfirmation, reflecting the low degree of multicollinearity between these measures (Table 1). Moreover, disconfirmation appeared to be the stronger of the

TABLE I
VARIABLE INTERCORRELATIONS BY SAMPLE AND
LEVEL OF MEASUREMENT

		Male Sample			
		1.	2.	3.	4.
Summed Attribute Scales	1. Exp.	-			
	2. Disc.	.13 ^a	-		
	3. Satis.	.48 ^b	.50 ^b	-	
	4. Pref.	-.04 ^b	.33 ^b	.18 ^b	-
	5. Int.	.33 ^b	.42 ^b	.76 ^b	.17 ^b
Overall Scales	1. Exp.	-			
	2. Disc.	.13 ^a	-		
	3. Satis.	.45 ^b	.51 ^b	-	
	4. Pref.	.04 ^b	.34 ^b	.26 ^b	-
	5. Int.	.38 ^b	.42 ^b	.83 ^b	.17 ^b
		Female Sample			
		1.	2.	3.	4.
Summed Attribute Scales	1. Exp.	-			
	2. Disc.	.18 ^b	-		
	3. Satis.	.31 ^b	.55 ^b	-	
	4. Pref.	.00 ^b	.37 ^b	.34 ^b	-
	5. Int.	.30 ^b	.52 ^b	.72 ^b	.25 ^b
Overall Scales	1. Exp.	-			
	2. Disc.	.25 ^b	-		
	3. Satis.	.47 ^b	.63 ^b	-	
	4. Pref.	.06 ^b	.42 ^b	.37 ^b	-
	5. Int.	.40 ^b	.53 ^b	.83 ^b	.25 ^b

^a $p \leq .05$ ^b $p \leq .01$

two effects, particularly for women respondents. Between 35% and 50% of the variance in satisfaction was explained by this two-variable scheme.

The results for preference change were dominated mostly by disconfirmation and, to a lesser degree, by satisfaction. All analyses yielded roughly equivalent findings, explaining between 12% and 21% of the criterion variance. These figures compare to a theoretical maximum value of .52 for a dichotomous criterion having the probability of occurrence (.26) found in this study (Nunnally 1978). Female respondents appeared to place greater emphasis on satisfaction which accounts for the higher coefficients of determination for the female analyses.

Finally, intention was almost exclusively a function of satisfaction. The path coefficients were very high for the satisfaction variable and virtually zero for all other predictors including preference. All coefficients of determination were above .5 and it appears that the overall scales yield a more complete explanation than do the summed attribute measures.

Discussion

This study has addressed a number of issues which merit further elaboration. These include the structure of the theoretical framework, level of variable measurement, sex differences in response, criterion predictability (validity), limitations of the methodology, and further research directions.

Theoretical framework. The basic theoretical premise,

TABLE 2
STANDARDIZED COEFFICIENTS OBTAINED WHEN
SATISFACTION, PREFERENCE, AND INTENTION
WERE REGRESSED ON THEIR ANTECEDENTS

Criterion: Sample/Scale	Independent Variable				R ²
	Exp.	Disc.	Satis.	Pref.	
Satisfaction:					
Male Attribute	.42 ^b	.45 ^b	-	-	.43 ^b
Male Overall	.39 ^b	.46 ^b	-	-	.41 ^b
Female Attribute	.22 ^b	.51 ^b	-	-	.35 ^b
Female Overall	.33 ^b	.55 ^b	-	-	.50 ^b
Preference:					
Male Attribute	-.11	.30 ^b	.08 ^a	-	.12 ^b
Male Overall	-.07	.26 ^b	.16 ^a	-	.13 ^b
Female Attribute	-.12	.26 ^b	.24 ^b	-	.18 ^b
Female Overall	-.12	.31 ^b	.23 ^b	-	.21 ^b
Intention:					
Male Attribute	-.03	.03	.75 ^b	.02	.58 ^b
Male Overall	.00	.02	.84 ^b	-.06	.70 ^b
Female Attribute	.07	.19 ^b	.60 ^b	-.02	.55 ^b
Female Overall	.00	.04	.82 ^b	-.07	.68 ^b

^a p < .05 ^b p < .01

Satisfaction = f(Expectation, Disconfirmation) patterned after Helson's (1964) adaptation level theory, received consistent support. It appears that expectations do provide a standard against which product deviations are perceived. This finding adds further support to a growing number of studies which suggest this perspective (Oliver 1977, Swan 1977, Gilly 1979, Linda and Oliver 1979, Madden, Little and Dolich 1979, Oliver 1980a, Swan and Trawick 1980, Westbrook 1980). It should be noted that most all of these investigations were conducted in different contexts using two-stage measurement and a varied mix of products and services including automobiles, breakfast bars, flu inoculations, newly opened department stores, and fabric cleaners.

It is also of interest to note that all studies including the one described here have found that disconfirmation is the more potent of the two effects. In terms of the magnitude of the standardized regression coefficients, a ratio of one and one-half to one is not uncommon. While methodological issues provide a partial explanation in that disconfirmation and satisfaction are typically measured in close proximity after product exposure, this sequence of events conforms to the nature of the hypothesized cognitive response. Additionally, great care was taken in the present study to segregate the two measures through questionnaire sequencing, although it is admitted that separate instruments and longer time intervals between measurement will be required before this issue is resolved.

Summed attribute vs. overall measurement. Inspection of Table 2 shows that the overall bipolar adjective scales yielded marginally superior results to those obtained using summed attribute scales. The differences were most dramatic when coefficients of determination for the two scaling procedures were compared, particularly for the intention criterion. Sizable differences were also evident between the attribute and overall satisfaction path coefficients in the intention regressions.

Two explanations pertain. The first relates to current theoretical perspectives of cognition formation. If one adopts the approach taken by Fishbein and Ajzen (1975), the essence of constructs having formative elements is best

captured in summary statements (e.g., one's overall subjective norm). This appears to be true because any list of elements will in all probability be deficient for some subjects and extraneous for others. Secondly, summed attribute lists of diverse elements cannot be expected to have high internal consistency reliability. In contrast, the bipolar adjectives selected for the overall scales demonstrated adequate to high reliability for the three constructs tested here and are probably to be preferred unless diagnostic information is desired.

Sex differences. The data suggest that women responded with slightly greater magnitudes of effect in two cases. First, they placed greater emphasis on disconfirmation in the prediction of satisfaction and also on satisfaction in the prediction of preference. While the study was not designed to provide explanations for this observation, it may be that women were more involved in the decision. Involvement has been suggested as a moderator in reactions to product satisfaction situations (Olson and Dover 1979). Because the majority of men's sleepwear purchases are made by women, it is likely that their potentially greater involvement made them more sensitive to feelings of disconfirmation and satisfaction. The notion that disconfirmation perceptions are affected by product involvement has potential implications for practice and should be pursued in future research efforts.

Validity. The model as proposed quite adequately predicted intention to buy. In fact, coefficients of determination for cognitive criteria in excess of .5 are fairly common among the studies cited previously. The model can also be shown to predict (in a concurrent sense) postexposure attitude and other reactions including product ratings. Unfortunately, behavioral choice (here preference) was not predicted to the same degree, even taking into account the fact that the theoretical maximum R² value in this study was roughly .5. This is further clouded by the fact that disconfirmation and not satisfaction played a major role in one's postexposure preference. Typically, preference, like attitude, would be a significant function of satisfaction.

Further reflection suggests that the result obtained is that which would have been anticipated. Our criterion here is change in preference, not preference per se. Helson's (1964) theory suggests that expectations define the baseline for disconfirmation but that movement from level (change) is a function of disconfirmation only. Thus, postexposure attitude should be a function of expectation and satisfaction but attitude change (the difference) would be predicted best by the performance-expectation discrepancy. In fact, reanalysis of data from prior studies available to the authors suggests that this is indeed the case. This logic provides a reasonable explanation for the strong influence of disconfirmation on preference. The coefficients for satisfaction were significant, however, in three of four analyses, as hypothesized.

Limitations and future research directions. A number of limitations which detract from the contribution value of the study requires elaboration. Each, in turn, suggests an area of improvement for future investigations. First, the study did not involve an actual purchase situation with normal usage. Rather, these two necessities were "short-circuited" in the procedure. Although the product response was accurately duplicated, the lack of a myriad of extraneous factors normally present in real usage situations may have provided an overly sterile environment for respondents, contributing to an overstatement of the theory's validity.

Closely related to the preceding comment is the issue of the timing of the measures relative to their occurrence in a natural environment. Ideally, the study should have been replicated after each washing rather than after five consecutive washings. Thus, expectations would have been measured before the first wash and disconfirmation, satis-

faction, and revised expectations after the laundering. The revised expectations then become input to the cognitive response following the second washing. No study has looked at the process of satisfaction as a function of successive usage over time and work in this area is needed urgently.

A third issue surrounds the measurement of satisfaction and its antecedents. While each new study provides more focused perspective to different measurement vehicles, little consensus exists to suggest a one best approach. The three item bipolar adjective format used here demonstrated promise on two counts. First, the internal consistency appeared acceptable especially for measurement of satisfaction per se. And second, it is easily adaptable to expectation and disconfirmation items. Readers may object to the use of pleasantness and happiness as synonyms for satisfaction. The authors share similar reservations and welcome suggestions from other researchers.

Finally, the sample frame used here remains a potential source of error. A national probability sample was not used and one wonders how the results would have been affected. While the findings do not generalize beyond the mall sampling procedure used here, we have hopefully identified a close approximation to the process of satisfaction decisions used by consumers generally.

References

- Anderson, Rolph E. (1973), "Consumer Satisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance," Journal of Marketing Research, 10 (February), 38-44.
- Andreasen, Alan R. (1977), "A Taxonomy of Consumer Satisfaction/Dissatisfaction Measures," in Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, H. Keith Hunt, ed., Cambridge MA: Marketing Science Institute.
- Andrews, Frank M. and Withey, Stephen B. (1976), Social Indicators of Well-Being, New York: Plenum Press.
- Campbell, Angus, Converse, Philip E. and Rodgers, Willard L. (1976), The Quality of American Life, New York: Russell Sage Foundation.
- Day, Ralph L. (1977), "Toward a Process Model of Consumer Satisfaction," in Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, H. Keith Hunt, ed., Cambridge MA: Marketing Science Institute.
- _____ and Bodur, Muzaffer (1979), "Analysis of Average Satisfaction Scores of Individuals over Product Categories," in New Dimensions of Consumer Satisfaction and Complaining Behavior, Ralph L. Day and H. Keith Hunt, eds., Bloomington Ind.: School of Business, Indiana University.
- Engel, James F., Kollat, David T. and Blackwell, Roger D. (1968), Consumer Behavior, New York: Holt, Rinehart and Winston.
- Fishbein, Martin (1967), "Attitude and the Prediction of Behavior," in Readings in Attitude Theory and Measurement, Martin Fishbein, ed., New York: John Wiley & Sons.
- _____ and Ajzen, Icek (1975), Belief, Attitude, Intention, and Behavior, Reading MA: Addison-Wesley.
- Gilly, Mary C. (1969), "Complaining Consumers: Their Satisfaction with Organizational Response: in New Dimensions of Consumer Satisfaction and Complaining Behavior, Ralph L. Day and H. Keith Hunt, ed., Bloomington Ind.: Department of Marketing, Indiana University.
- Helson, Harry (1948), "Adaptation-Level as a Basis for a Quantitative Theory of Frames of Reference," Psychological Review, 55 (November), 297-313.
- _____ (1964), Adaptation-Level Theory, New York: Harper & Row.
- Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: John Wiley & Sons.
- Ilgen, Daniel R. (1971), "Satisfaction with Performance as a Function of the Initial Level of Expected Performance and the Deviation from Expectations," Organizational Behavior and Human Performance, 6 (January), 345-361.
- Katona, George (1964), The Mass Consumption Society, New York: McGraw-Hill.
- LaTour, Stephen A. and Peat, Nancy C. (1979), "Conceptual and Methodological Issues in Satisfaction Research," in Advances in Consumer Research, Vol. VI, William L. Wilkie, ed., Ann Arbor: Association for Consumer Research.
- Linda, Gerald and Oliver, Richard L. (1979), "Multiple Brand Analysis of Expectation and Disconfirmation Effects on Satisfaction," in Proceedings of the Division 23 Program, 87th Annual Convention of the American Psychological Association, Ivan Ross, ed.
- Madden, Charles Stanley, Little, Eldon L. and Dolich, Ira J. (1979), "A Temporal Model of Consumer S/D Concepts as Net Expectations and Performance Evaluations," in New Dimensions of Consumer Satisfaction and Complaining Behavior, Ralph L. Day and H. Keith Hunt, eds., Bloomington, Ind.: School of Business, Indiana University.
- Nunnally, Jum C. (1978), Psychometric Theory, Second Edition, New York: McGraw-Hill.
- Ölander, Folke (1977), "Consumer Satisfaction -- A Skeptic's View," in Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, H. Keith Hunt, ed., Cambridge, MA: Marketing Science Institute.
- Oliver, Richard L. (1977), "Effect of Expectation and Disconfirmation on Postexposure Product Evaluations: An Alternative Interpretation," Journal of Applied Psychology, 62 (August), 480-486.
- _____ (1980a), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," Journal of Marketing Research, 17 (November), forthcoming.
- _____ (1980b), "Theoretical Bases of Consumer Satisfaction Research: Review, Critique and Future Direction," in Theoretical Developments in Marketing, Charles W. Lamb, Jr., and Patrick M. Dunne, eds., Chicago: American Marketing Association.
- Olson, Jerry C. and Dover, Phillip (1979), "Disconfirmation of Consumer Expectations Through Product Trial," Journal of Applied Psychology, 64 (April), 179-189.
- Swan, John E. (1977), "Consumer Satisfaction with a Retail Store Related to the Fulfillment of Expectations on an Initial Shopping Trip," in Consumer Satisfaction, Dissatisfaction, and Complaining Behavior, Ralph L. Day, ed., Bloomington, Ind.: School of Business, Indiana University.
- _____ and Trawick, I. Fredrick (1979), "Satisfaction Related to Predictive vs Desired Expectations," Paper presented at the Fourth Annual Conference on Consumer Satisfaction, Dissatisfaction, and Complaining Behavior.
- Weaver, Donald and Brickman, Phillip (1974), "Expectancy, Feedback and Disconfirmation as Independent Factors in Outcome Satisfaction," Journal of Personality and Social

Psychology, 30 (March), 420-428.

Westbrook, Robert A. (1980), "Intrapersonal Affective Influences on Consumer Satisfaction with Products," Journal of Consumer Research, 7 (June), 49-54.

_____, and Oliver, Richard L. (1981), "Developing Better Measures of Consumer Satisfaction: Some Preliminary Results," in Advances in Consumer Research, Vol. 8, Kent B. Monroe, ed., Ann Arbor: Association for Consumer Research. (this volume)

DEVELOPING BETTER MEASURES OF CONSUMER SATISFACTION: SOME PRELIMINARY RESULTS

Robert A. Westbrook, University of Arizona
Richard L. Oliver, Washington University

Abstract

Increased empirical study of consumer satisfaction has not yet been accompanied by commensurate refinements in the conceptualization and measurement of the construct. A review of the related literatures in job, health (patient), marital, and life satisfaction reveals that varying degrees of methodological rigor have been used to develop satisfaction measures, resulting in instruments as long as 80 items, but that little consensus on conceptual or measurement issues is evident. In an attempt to further satisfaction measurement in the consumer domain, a number of suggested multi-disciplinary approaches were combined in a two-product satisfaction instrument to test for response characteristics, reliability, and convergent and discriminant validity. Results in two separate pilot samples indicate that the highest reliabilities were achieved with semantic differential and Likert scales. These measures along with other verbal and graphic rating scales also demonstrated a high level of convergence and acceptable levels of discriminability.

Introduction

While interest in consumer satisfaction has grown rapidly in recent years (Hunt 1977, Day 1977, Day and Hunt 1979), there is little agreement on measurement of the construct. Empirical studies reflect a variety of different measures, although typically neither the reliability nor validity of these measures has been demonstrated. Not only does this hinder the interpretation and synthesis of research findings, but it also raises the possibility that applied studies in industry and government may be obtaining inaccurate assessments of prevailing satisfaction levels. Further attention to the development and evaluation of consumer satisfaction measures is clearly warranted. Accordingly, this paper reviews alternative satisfaction measures and presents preliminary findings of a measure validation study.

Conceptualization of Satisfaction

Prior to consideration of alternative measures, attention must be given to the conceptualization of consumer satisfaction and its relation to other theoretical constructs. While satisfaction is most readily associated with the purchase and consumption of specific products or services, it may also be relevant for shopping and patronage at retail outlets, for media usage, and even overall participation in the marketplace (Czepiel, et al. 1975). Thus, consumer satisfaction refers to an evaluative response concerning the perceived outcomes of experiences in the consumer domain, comprising acquisition, consumption, and disposition activity. These outcomes are often assumed to be evaluated according to the extent to which they fulfill consumers' expectations (Howard and Sheth 1969). However, other kinds of evaluative standards are conceivable, such as correspondence of perceived outcomes to those ideally desired or to the minimum outcomes considered acceptable. The more favorable the evaluation of perceived outcomes, the greater the satisfaction.

Central to the construct of satisfaction is the presence of affect. In connection with their evaluations of outcomes, consumers may experience varying degrees of feeling

or emotion. Favorably evaluated outcomes are associated with happy, pleasant feelings, and unfavorably evaluated outcomes with unhappiness, irritation, or regret. In addition, the notion of satisfaction implies some degree of conation in that the consumer is more or less inclined to repeat the behavior in question given recurrence of the situation in which it was initially performed.

With respect to theoretical conceptions of brand purchase behavior (Howard and Sheth 1969), satisfaction with a particular product occurs after purchase commitment but prior to any revision in evoked set, brand attitude, and brand (re)purchase intentions. A high level of satisfaction is believed to increase the likelihood that the brand in question will be included in the user's evoked set, increase the favorability of brand attitude, and increase the degree of intention to (re)purchase that brand (Oliver 1980). Low levels of satisfaction, or dissatisfaction, presumably have opposite effects and tend to motivate the consumer to seek redress or remedy by complaining to the seller or third parties.

Measurement of Consumer Satisfaction

Most research on consumer satisfaction has been concerned with specific products and services (Day and Ash 1979, Swan and Combs 1976, Andreassen and Best 1977, Oliver 1980) or retail outlets (Miller 1976). Satisfaction with other facets of consumer experience have been studied only infrequently (see Lundstrom and Lamont 1976, Westbrook and Newman 1978). Typically, measurements of satisfaction with products/services and retailers are based on direct subjective estimation by consumers of the intensity or frequency of overall satisfaction experienced. Most often, simple, single-item rating scales are employed. There has been little uniformity in the number of scale steps used or nature of verbal anchoring, however; they range from 3-point fully-labelled rating scales to 10- and 11-point variants labelled only at the extremes and midpoint. Comprehensive measure comparisons have seldom been undertaken, and investigators rarely report the reliability, much less the validity, of their measures.

Multi-item rating scale measures of product/service/retailer satisfaction have found application infrequently, despite their potential to reduce measurement error. In one of the few instances, Oliver (1980) used a set of six Likert-format items dealing with overall satisfaction, for which a high level of internal consistency was observed ($\alpha = .82$). Multi-item measures based on satisfaction ratings for individual product/service/retailer attributes have been conspicuously avoided (for an exception, see Morris 1977), most likely because of uncertainty as to the functional form in which the latter should be combined into overall satisfaction judgments.

In view of the complexity of the construct of consumer satisfaction, the predominant approach to measurement reflected in the literature may be naive. It is doubtful that the cognitive-evaluative, affective, and conative elements of satisfaction can be adequately captured in a single 5- or 7-point "very satisfied -- very dissatisfied" rating scale. If anything, such simplistic measures may be biased because of too few scale increments and the absence of explicit evaluative and affective anchoring along the scale continuum. Andreassen (1977) reports some evidence that a 4-point rating scale overreported satisfaction compared to judgments based on various open-ended

questions. Moreover, satisfaction studies have typically resulted in "bunching" of respondents at the upper end of the satisfaction continuum.

More comprehensive measures of consumer satisfaction are clearly needed. In searching for such measures, it is instructive to examine methods of measurement employed in studies of satisfaction with other domains beyond consumption. Accordingly, the following section presents a brief review of the major measures of job satisfaction, life satisfaction, marital satisfaction, and patient satisfaction.

Measurement of Satisfaction in Other Domains

Job Satisfaction

Of other disciplines which have studied satisfaction, none has a longer tradition than management/industrial relations. Job satisfaction has been studied intensively for some fifty years, and a substantial literature has accumulated (for a recent review, see Locke 1974). The predominant measurement strategy involves direct verbal self-reports on rating scales of various forms. Often these are simply single- or dual-item overall satisfaction ratings, but a number of complex instruments examining the various facets of satisfaction have also emerged. In a review of the simpler measures, Robinson et al. (1969) find most promising two rating scales, neither of which involves an estimation of satisfaction per se. One is the self-anchoring Ladder Scale, which depicts a 9-step ladder whose bottom rung is labelled "worst job I could expect to have," and whose top rung is denoted "best job I could expect to have." Respondents select the rung that best describes their feelings about their job. The second measure asks "if you had the chance to start your working life over again, would you choose the same kind of work as you are doing now?" Responses may be made along a subjective likelihood continuum, or simply recorded verbatim for later categorization by coders along a satisfaction continuum. While these simple items may suffice to identify relative differences in satisfaction, accurate assessment of the absolute level of satisfaction is generally recognized as requiring complex instruments such as the Job Description Index (Smith, Kendall and Hulin 1969). The latter in particular has been impressively developed and validated by its authors.

Other measurement strategies in job satisfaction involve (1) inferring satisfaction from measurements of its presumed causes, (2) semi-structured interviews, and (3) recall of critical incidents. The inferential approach (Porter 1962) measures satisfaction as the inverse of the sum of discrepancies between how much of each job aspect an employee feels he is getting and how much he thinks he should be getting or how much he ideally wants. The semi-structured interview approach, while less efficient and objective than other approaches, has much to commend it; Locke (1974) encourages its wider use in satisfaction assessment. Finally, the critical incident recall method developed by Herzberg et al. (1959) is less useful as a measure of how much satisfaction was experienced than as an indicator of the sources of those feelings.

Marital Satisfaction

Marital satisfaction has received considerable attention from home economists and family sociologists. While simple "very satisfied" to "very dissatisfied" self-report rating scales have been used, there is widespread use of multi-item instruments of varying degrees of sophistication. Perhaps best known are the Blood and Wolfe (1960) Marital Satisfaction Index and the Locke and Wallace (1959) Marital Adjustment Test. The former is a four-item scale involving evaluations of the major domains on married life using a fully anchored, five-point affective response

scale, ranging from "pretty disappointed" to "enthusiastic, couldn't be any better." The majority of measures of marital satisfaction involve self-reports, though some are relatively indirect. A review of these measures (e.g., see Straus 1970) reveals considerable variability in the method of measure construction, including a variety of Likert-type summated rating scales, Guttman Scales, and judges' ratings of responses to unstructured projective questions. Noteworthy of marital satisfaction measurement is the common practice of obtaining an individual's evaluation of and affective responses to various component areas of marriage, e.g., spousal understanding, love and affection, companionship, etc., and combining these into an overall measure. Also evident is the attention given to considerations of measure reliability and validity (e.g., Rollins and Cannon 1974).

Patient Satisfaction

Research on patient satisfaction with health care delivery has been recently reviewed by Swan and Carroll (1979). Many investigators in this area have developed their own ad hoc measures which are often simplistic "satisfied--dissatisfied" self-report single item rating scales, either with the overall health care received or with a few selected aspects such as physician attitude, professional competence, convenience, etc. Noteworthy, however, are several major efforts to develop measures based on standard psychometric methods of attitude scale construction. Hulka et al. (1970) used Thurstone's Method of Equal Appearing Intervals for Scaling Attitudes toward physicians, but in a follow-up study (Zyzanski et al. 1974), the researchers revised the scale to a summated Likert-type to obtain improved internal consistency. The content of the 42-item scale suggests that it measures generalized satisfaction with the overall domain of primary health care rather than evaluation of a specific experience. In a similar effort, Ware and Snyder (1975) developed an 80-item Likert summated scale which has found application in several other patient satisfaction studies. One of the 4-item subscales of this measure is termed "general satisfaction" and is reported to have an alpha internal consistency of .77. Again, however, this measure is a highly generalized satisfaction indicator. Finally, Mangelsdorff (1979) developed a 19-item scale for measuring patient satisfaction with a specific health care service received. Individual satisfaction ratings for various aspects of the service are made on a five-point scale and cumulated into an overall index score. The scale has demonstrated high internal consistency and some degree of validity.

Life Satisfaction

One recent sociological study of the perceived quality of life has resulted in perhaps the most sophisticated satisfaction measure development procedures to date. Andrews and Withey (1976) identified a large number of alternative satisfaction measures, which they broadly categorized according to the perspective of evaluation (absolute vs. relative, long range vs. short range), generality (general vs. specific focus), and range (full-range of experiences vs. part-range). In the category of absolute, general, full-range measures, which are of most direct relevance to this review, a number of distinct rating scales measures were examined, including a 7-point "completely satisfied--completely dissatisfied" item, a 7-point fully anchored "Delighted-Terrible" item, and a variety of graphic or non-verbal items. The latter included the self anchoring Ladder Scale previously noted, along with the Faces Scale (Kunin 1955), the Thermometer Scale (warm to cold feelings), and the Circles Scale, which is comprised of nine circles each containing some proportion of pluses and minuses to indicate the incidence of favorable and unfavorable evaluations. In a comprehensive measure evaluation and validation effort, Andrews and Withey concluded that the Delighted-Terrible Scale was the most useful, yielding high

construct validity (estimated at .8), reasonably symmetrical response distributions, and ease of use. The Circles and Faces scales were ranked second on these criteria, followed by the Ladder and Simple Satisfaction scales. The authors suggested that selected combinations of these single item measures might profitably be employed as indexes yielding even higher validities.

Andrews and Withey also evaluated two other types of measures of interest, a Social Comparison rating, in which respondents assessed their satisfaction with life vis a vis other persons they knew, and a Peer Rating, in which three other persons' ratings of an individual's life satisfaction were averaged. Interestingly, neither of these measures attained appreciable validity coefficients and were not recommended for life satisfaction assessment.

Implications for Consumer Satisfaction

The satisfaction literatures in each of the above disciplines reflect three principal common elements: (1) the development of diversity of methods for measuring the construct; (2) widespread use of multiple item scales or index measures; and (3) consistent attention to issues of measure evaluation and validation. Consumer satisfaction research would be well advised to adopt similar measurement traditions. Many of the specific approaches to measuring satisfaction with job, spouse, health care, or life enumerated above also appear potentially suitable for application to product satisfaction assessment.

Job satisfaction measurement suggests the usefulness of two particular kinds of single-item rating scales: the self-anchoring Ladder Scale and the graphic Faces Scale. In addition, the value of a "behavioral tendency" item, in which the respondent's predisposition to repeat his previous behavior is assessed, seems apparent. Less direct self-report measures, particularly the inferential measures in which the extent to which disparities exist between desired outcomes and actual outcomes, also appear to have application. Finally, Locke's (1974) recommendation that open-ended interview data supplement the ubiquitous rating scale, is deserving of further consideration of consumer satisfaction researchers. In fact, Miller (1976) has previously advocated the use of free response formats.

Though not unique to marital and patient satisfaction research, measurement in these areas is often characterized by multi-item scales created by summing evaluations dealing particular aspects of the phenomenon. This approach might be fruitfully applied in the assessment of product/service satisfaction, provided that agreement can be reached on the basic aspects or outcomes involved in consumption. A start in this direction is the distinction between expressive and instrumental product outcomes (Swan and Combs 1976).

Life satisfaction research, in addition to providing an excellent model for measure validation studies, suggests the value of the Delighted-Terrible scale for a wide range of satisfaction assessment applications. Also promising is the Circles graphic scale. Both these rating measures were shown to be superior to simple "satisfied--dissatisfied" overall rating scale items typical of consumer satisfaction research. Life satisfaction studies also suggest the importance of temporal perspective for consumer research, which is rarely considered in the assessment process.

The Study

To appraise the suitability of selected measures from other disciplines in the assessment of consumer satisfaction with specific products and services, the authors undertook a

pilot study whose preliminary results are presented below. This effort is part of a broader investigation currently in progress to develop and evaluate improved indicators of satisfaction with products and services.

Measures

Five types of measures were considered in this study, all rating scale variants. The first is a three-item verbal scale (hereafter VERBAL) combining separate overall assessments of product satisfaction using the Delighted-Terrible 7-point rating scale, a "Completely Satisfied--Not at all Satisfied" 11-point rating scale ranging from 100% to 0%, and a behavioral tendency 11-point rating scale ranging from "Certain I'd do it again" to "No chance I'd do it again." The specific items constituting the overall scale, which can be found in Andrews and Withey (1976, Appendix A), were combined additively to product verbal. These items and all others discussed subsequently are available on request from the authors.

In contrast to the verbal orientation of the first measure, the second involved four distinct graphic rating scales: Faces, Thermometer, Circles, and Ladder. These particular items represent a desirable mixture of nonverbal content. The faces and thermometer scales are chiefly affective, while the Circles and Ladder are more cognitive-evaluative in tone. They were combined additively to yield the overall graphic scale (hereafter GRAPHIC). As before, these items can be found in the Andrews and Withey Appendix and are available from the authors.

The third measure was a Likert summated scale (hereafter LIKERT) in which 12 statements indicating varying sentiments of overall satisfaction with the product were presented to respondents for their agreement. Responses were made on a five-interval "strongly agree--strongly disagree" continuum.

The fourth measure consisted of a set of seven semantic differential items, again dealing with various means of expressing overall satisfaction judgments. To reduce the cognitive strain on respondents only five intervals were used on the semantic differential instead of the conventional seven. These items were combined into a simple additive index (hereafter S-D) after scoring responses as to their favorability.

The final measure was intended as an inferential satisfaction instrument. For each of a variety of product attributes, it assessed (a) the level currently provided by the product and (b) the level ideally desired by the consumer. A difference was computed between each of these ratings and summed over all attributes. This "disparity" figure was presumed inversely related to the level of satisfaction experienced. This scale was termed the PORTER scale.

Sources of Data

Self-administered questionnaires were completed by upper-level undergraduate students at the University of Arizona ($N_1 = 68$) and Washington University ($N_2 = 107$). Cooperation was voluntary and anonymous in both cases. The Arizona sample was surveyed in-class and all subjects agreed to participate. The Washington University survey was to be completed out-of-class; 82% of the students returned their questionnaires. Completion of the instrument took approximately 30 minutes. Respondents were questioned about their experiences with two products currently owned, automobiles and hand-held calculators. Data were analyzed separately by sample.

Results

Descriptive statistics for each measure as applied to each product are shown in Table 1. Ideally the distribution of a satisfaction measure should exhibit a high level of dispersion of responses, thus avoiding "clumping" of respondents within a narrow range of scores. At the same time, a reasonably symmetrical distribution shape is also desired. Examination of the standard deviation of each measure reveals that the LIKERT Scale achieves the greatest dispersion of individual scores for both products, in both samples. The skewness statistics indicate that the VERBAL and LIKERT measures achieve the most symmetrical distributions for automobiles, while for calculators it is the GRAPHIC and LIKERT measures.

TABLE 1
Simple Descriptive Statistics for Satisfaction Measures

Measure	Product	Sample	Mean	S.D.	Skew
VERBAL	Auto	I	17.6	5.7	-.60
		II	19.4	5.0	-.78
	Calc'tr	I	20.3	4.8	-1.62
		II	19.8	5.1	-.58
GRAPHIC	Auto	I	23.8	5.0	-1.22
		II	24.6	4.8	-.59
	Calc'tr	I	19.9	1.4	-.39
		II	24.4	5.6	-.77
LIKERT	Auto	I	40.3	12.1	-1.19
		II	44.8	8.8	-.51
	Calc'tr	I	43.3	6.1	-.95
		II	43.9	9.4	-.70
S-D	Auto	I	22.0	5.5	-1.44
		II	22.8	5.8	-.64
	Calc'tr	I	23.1	4.9	-1.44
		II	21.5	6.8	-.72
PORTER	Auto	I	-10.1	7.0	.81
		II	-9.6	4.9	-1.01
	Calc'tr	I	-4.2	4.5	-1.20
		II	-5.0	4.7	-1.44

Note: Sample I refers to Arizona ($N_1=68$);
Sample II refers to St. Louis ($N_2=107$).

The reliabilities of the various measures as estimated by their internal consistencies are shown in the diagonals of the multitrait-multimethod matrices of Tables 2a and 2b. For the various automobile satisfaction measures, the LIKERT and S-D Scales both attain very high alpha coefficients in both samples, exceeding .93. The VERBAL Scale reliability is also satisfactory in both samples, $\alpha = .72$. In contrast, the PORTER Scale had the lowest internal consistency of all measures in both samples ($\alpha = .68$ and .46). For calculators, the highest internal consistency is attained by the GRAPHIC, LIKERT and S-D Scales, while the others yield somewhat lower though roughly equivalent reliabilities.

Campbell and Fiske's (1959) criteria of convergence and discriminability represent necessary though not sufficient conditions for measure validation. Convergence is demonstrated by high correlations between alternative measures within a given trait. Tables 2a and 2b indicate that there is a high level of convergence among satisfaction measures for automobiles with the exception of the PORTER Scale, which in both studies failed to correlate highly with all

of the other measures. With regard to calculators, all measure intercorrelations are significant, although there is some variability in the strength of the relationships. In contrast to the automobile ratings, the PORTER Scale does indicate convergence with the other satisfaction measures when applied to calculators.

Discriminability also requires that different measures of different traits not correlate. In Sample I (Table 2a), none of the heterotrait-monomethod correlations and only two of the 20 heterotrait-heteromethod correlations reach significance. The significant trait-method pairs are the VERBAL--PORTER and the LIKERT--PORTER.

In Sample II (Table 2b), none of the heterotrait-monomethod or heterotrait-heteromethod correlations reach significance. In addition to the lack of convergence displayed by the PORTER Scale for automobile satisfaction, its lack of convincing discriminability opens its validity to question.

Conclusions

The results of this pilot study suggest that selected satisfaction measures derived from parallel disciplines may have merit as indicators of the level of consumer satisfaction, thereby warranting further attention by researchers in this area. Perhaps most importantly, however, the results also provide much needed evidence as to the validity of satisfaction measures for products and services. Overall, the various multi-item rating scale measures examined in this study appeared to perform reasonably well. All but one of the measures (PORTER) clearly met Campbell and Fiske's (1959) criteria of convergence and discriminability and, in addition, attained high levels of internal consistency reliability. The foregoing, however, may be viewed as necessary though not sufficient conditions for inferring measure validity. Further analysis of the multi-trait-multimethod matrix based on path-analytic conceptualizations and confirmatory factor analysis would be helpful by providing specific estimates of proportions of valid (i.e. trait) variance and correlated errors (i.e. method) variance.

While it is not appropriate to attempt to identify a single "best" measure given the limitations of the samples in this pilot study, the LIKERT, S-D, and VERBAL measures appear promising for automobile satisfaction measurement. Their internal consistency is high, they converge with other measures, and they succeed in discriminating between unrelated constructs; at the same time, they also yield fairly symmetrical, dispersed distributions of individual responses. The shorter length of the VERBAL Scale (3 items) makes it attractive from the standpoint of administrative efficiency. For calculator satisfaction, however, the VERBAL Scale does not appear particularly as promising as LIKERT and S-D measures. Whether this finding indicates that different classes of products or services may require different methods of measurement or whether it simply reflects the vagaries of a small pilot test must await further data collection and analysis.

Perhaps the most problematic of the measures studied was the PORTER Scale, an inferential satisfaction measure based on a summation of disparities between product outcomes and those ideally desired. Its lack of convergence for automobiles is troubling, and suggests that not all of the relevant product outcomes may have been identified for inclusion into the scale. Of course, this potential limitation is not necessarily unique to the PORTER Scale, but rather applies to all attribute-based composite satisfaction measures.

TABLE 2a
 Sample I - Multitrait Multimethod Matrix:
 Satisfaction Measures for Automobiles and Calculators

Method		VERBAL		GRAPHIC		LIKERT		S-D		PORTER		
		Auto	Calc	Auto	Calc	Auto	Calc	Auto	Calc	Auto	Calc	
VERBAL	Auto Satisfaction	(.88)										
	Calc Satisfaction	-.28	(.68)									
GRAPHIC	Auto Satisfaction	.80 ^a	-.21	(.72)								
	Calc Satisfaction	-.27	.39 ^a	-.25	(.90)							
LIKERT	Auto Satisfaction	.87 ^a	-.31	.71 ^a	-.16	(.96)						
	Calc Satisfaction	-.29	.69 ^a	-.29	.46 ^a	-.25	(.75)					
S-D	Auto Satisfaction	.77 ^a	-.19	.69 ^a	-.12	.91 ^a	-.24	(.95)				
	Calc Satisfaction	-.15	.67 ^a	-.15	.41 ^a	-.06	.78 ^a	-.22	(.91)			
PORTER	Auto Satisfaction	.17	-.11	-.06	.12	.15	-.17	.23	-.28	(.68)		
	Calc Satisfaction	-.33 ^a	.55 ^a	-.28	.32 ^a	-.36 ^a	.57 ^a	-.22	.63 ^a	-.08	(.70)	

NOTE: Parenthesized values represent reliability estimates (Cronbach's alpha). Correlations between automobile satisfaction measures are based on N=58 or 59 (difference due to item nonresponse). Those between calculator satisfaction measures are based on N=55 to 59, and those between automobile and calculator satisfaction measures are based on N=47 to 51.

^aSignificant at .01 level

TABLE 2b
 Sample II - Multitrait Multimethod Matrix:
 Satisfaction Measures for Automobiles and Calculators

Method		VERBAL		GRAPHIC		LIKERT		S-D		PORTER		
		Auto	Calc	Auto	Calc	Auto	Calc	Auto	Calc	Auto	Calc	
VERBAL	Auto Satisfaction	(.76)										
	Calc Satisfaction	-.10	(.52)									
GRAPHIC	Auto Satisfaction	.82 ^a	-.08	(.87)								
	Calc Satisfaction	-.15	.84 ^a	-.10	(.93)							
LIKERT	Auto Satisfaction	.87 ^a	-.10	.79 ^a	-.13	(.93)						
	Calc Satisfaction	-.06	.87 ^a	-.01	.78 ^a	-.02	(.95)					
S-D	Auto Satisfaction	.86 ^a	-.11	.82 ^a	-.17	.88 ^a	-.05	(.94)				
	Calc Satisfaction	-.03	.83 ^a	.02	.81 ^a	-.02	.90 ^a	-.02	(.95)			
PORTER	Auto Satisfaction	.32 ^a	-.10	.47 ^a	.01	.27 ^a	-.02	.29 ^a	.04	(.46)		
	Calc Satisfaction	-.05	.65 ^a	.03	.63 ^a	-.08	.67 ^a	-.02	.68 ^a	.08	(.72)	

NOTE: Parenthesized values represent reliability estimates (Cronbach's alpha). Correlations between automobile and calculator satisfaction measures are based N=101; those between automobile and calculator satisfaction measures are based on N=95.

^aSignificant at .01 level

This research has focused on more or less explicit rating scale methods for satisfaction measurement. Overall, these measures have been observed to work reasonably well in two separate samples. Future research comparing these methods to less structured methods of measurement, notably those based upon open-ended questions, would be especially helpful. Such data realistically will require collection by personal or telephone interview rather than self-administration. However, as Locke (1974) has argued in the context of job satisfaction research, they may provide considerably deeper insight into the meaning of consumers' evaluation and sentiments. Andreassen (1977) has indicated that consistently lower estimates of satisfaction are obtained from free-response data. Which measurement technique is the more accurate indicator of consumer satisfaction, however, remains a question for further research.

References

- Andreassen, Alan R. (1977), "A Taxonomy of Consumer Satisfaction/Dissatisfaction Measures," in H. K. Hunt, ed., Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, Cambridge, Mass.: Marketing Science Institute.
- _____ and Best, Arthur (1977), "Consumers Complain - Does Business Listen?," Harvard Business Review, (July-August), 93-101.
- Andrews, Frank M. and Withey, Stephen B. (1976), Social Indicators of Well-Being, New York: Plenum Press.
- Blood, Robert O. and Wolfe, Donald M. (1960), Husbands and Wives: The Dynamics of Married Living, Glencoe, Ill.: Free Press.
- Campbell, Donald T. and Fiske, Donald W. (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56, 81-105.
- Czepiel, John A., Rosenberg, Larry J. and Akerele, Adelbayo (1975), "Perspectives on Consumer Satisfaction," Proceedings of the 1975 Conference of the American Marketing Association, Chicago: American Marketing Association, 119-23.
- Day, Ralph L. (1977), ed., Consumer Satisfaction, Dissatisfaction and Complaining Behavior, Bloomington, IN: Indiana University Press.
- _____ and Ash, Stephen B. (1979), "Comparisons of Patterns of Satisfaction/Dissatisfaction and Complaining Behavior for Durables, Non-Durables, and Services," in R. L. Day and H. K. Hunt, eds., New Dimensions of Consumer Satisfaction and Complaining Behavior, Bloomington, IN: Indiana University School of Business, 190-195.
- _____ and Hunt, H. Keith (1979), eds., New Dimensions of Consumer Satisfaction and Dissatisfaction, Bloomington, IN: Indiana University School of Business.
- Engel, James F., Blackwell, Roger D., and Kollat, David T. (1978), Consumer Behavior, 3rd edition, New York: Holt, Rinehart, and Winston.
- Herzberg, Frederick, Mauser, Bernard, and Snyderman, Barbara B. (1959), The Motivation to Work, 2nd edition, New York: Wiley.
- Howard, John A. and Sheth, Jagdish (1969), The Theory of Buyer Behavior, New York: John Wiley and Sons.
- Hulka, Barbara S., Zyzanski, Stephen J.; Cassel, John C., and Thompson, Shirley J. (1970), "Scale for the Measurement Attitudes Toward Physicians and Primary Medical Care," Medical Care, (September/October), 429-36.
- Hunt, H. Keith (1977), ed., Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, Cambridge, Mass.: Marketing Science Institute.
- Kunin, Theodore (1955), "The Construction of a New Type of Attitude Measure," Personnel Psychology, 8, 65-77.
- Locke, Edwin A. (1974), "The Nature and Causes of Job Satisfaction," in Marvin D. Dunnette, ed., Handbook of Industrial and Organizational Psychology, New York: Holt Rinehart, 1297-1349.
- Locke, H. J. and Wallace, K. M. (1959), "Short Marital Adjustment and Prediction Tests: Their Reliability and Validity," Marriage and Family Living, 21, 251-55
- Lundstrom, William J. and Lamont, Lawrence, M. (1976), "The Development of a Scale to Measure Consumer Discontent," Journal of Marketing Research, 13 (November), 373-81.
- Mangelsdorff, David A. (1979), "Patient Satisfaction Questionnaire," Medical Care, 17 (January), 86-90.
- Miller, John A. (1976), "Exploring Some Alternative Measures of Consumer Satisfaction," in K. L. Bernhardt, ed., Marketing: 1776-1976 and Beyond, Chicago: American Marketing Association, 661-64.
- Morris, Earl W. (1977), "A Normative Deficit Approach to Consumer Satisfaction," in H. K. Hunt, ed., Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction, Cambridge, Mass.: Marketing Science Institute, 240-74.
- Oliver, Richard L. (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," Journal of Marketing Research, 17 (November), forthcoming.
- Porter, Lyman W. (1962), "Job Attitudes in Management: I. Perceived Deficiencies in Need Fulfillment as a Function of Job Level," Journal of Applied Psychology, 46, 375-84.
- Robinson, John P., Athanasiou, R., and Head, Sandra B. (1969), Measures of Occupational Attitudes and Occupational Characteristics, Ann Arbor: Survey Research Center, University of Michigan.
- Smith, Patricia C., Kendall, L. M. and Hulin, Charles L. (1969), The Measurement of Satisfaction in Work and Retirement, Chicago: Rand McNally.
- Straus, Murray (1970), Family Measurement Techniques, Chicago: Rand McNally.
- Swan, John E. and Carroll, Maxwell G. (1979), "Patient Satisfaction: An Overview of Research, 1965-1978," paper presented at the Fourth Annual Conference on Consumer Satisfaction, Dissatisfaction, and Complaining Behavior, Bloomington, Indiana.
- _____ and Combs, Linda J. (1976), "Product Performance and Consumer Satisfaction: A New Concept," Journal of Marketing, 40 (April), 25-33.
- Ware, John E. and Snyder, Mary K. (1975), "Dimensions of Patient Attitudes Regarding Doctors and Medical Care," Medical Care, 13 (August), 669-82.
- Westbrook, Robert A. and Newman, Joseph W. (1978), "An Analysis of Shopper Dissatisfaction for Major Household Appliances," Journal of Marketing Research, 15 (August), 456-66.

HIERARCHICAL COGNITIVE CONTENT: TOWARDS A MEASUREMENT METHODOLOGY

Elizabeth C. Hirschman, New York University
Susan P. Douglas, New York University

Introduction

In recent years increasing interest has arisen in developing methodologies for measuring cognitive phenomena. In particular, growing use has been made of information processing approaches to understand consumer decision-making (Bettman 1980), product perception (Olson 1978) and attitude structure (Capon and Lutz 1979). Attention has thus been drawn to identifying what information relating to products is used by consumers, and how that information is organized in memory.

Research has focused primarily on examining product attribute information, and on assessing how this is used in evaluating brands and products. Use has been made of protocol techniques (for a summary see, Douglas, Craig and Fairre 1980), or less commonly free recall prompted by an external cue or cover story (Russo and Johnson 1980, Olson and Muderrisoglu 1977). This research has, however, largely ignored the broader semantic network surrounding the product, namely product-variant linkages to form categories or product classes and brand-product hierarchical linkages to form general to specific orderings of content (Lachman, Lachman, Butterfield 1979).

One exception to this is the work of Gutman (1980) which has explored the broadness/narrowness of categories created by consumers from given product stimuli. The hierarchical character of product-brand-attribute categories has, however been largely ignored.

For example, whether the consumer uses the product-concept primarily as a node (storage area) for accumulating and organizing other types of information, or as an item of knowledge moderating (i.e. qualifying, or describing) a category. Take, for example, the concepts "Levi's" and "jeans". For some consumers, the primary cognitive category is "jeans", for which Levi's is a moderator (i.e. a qualifier). For other consumers "Levi's" is the node (category), and "jeans" is a moderator.

Such information is, nonetheless, of interest to assess how consumers categorize and classify products, product classes, brands and domains, and how this information is maintained in memory. It might also be useful in developing advertising copy, by indicating the scope of the semantic image, i.e. the combinations of products, and brands, which are stored together by consumers.

An issue which arises in this context is the appropriate methodology to examine the hierarchical content of information in memory. This paper reports an attempt to apply free-association techniques to examine cognitive content and, given the limitations of this methodology, to suggest an alternative approach.

Free Association Methodologies

Free response and word association techniques are widely used in cognitive psychology (Rosch 1978) and in cross-cultural research (Szalay and Deese 1978, Rosenzweig 1961) to identify domains, examine product categorization, and the meanings associated with products and objects and to compare these across cultures. Their use in marketing has been somewhat more limited, largely due to difficulties and tediousness of analysis (Green, Wind and Jain 1973).

The advantage of using consumers' free associations to a product concept, as opposed to the multi-attribute approach, is that all types of semantic associations regarding a product can be measured, not just the attributes. Hence, the measurement of consumers' product cognitions is extended beyond attribute structure to the total semantic associational structure.

The free association technique can, for example, provide valuable information about brands perceived as competing. For example, in response to the stimulus brand "McDonalds", responses of "Burger King" and "Wendy's", may be obtained, indicating that these are perceived as similar and hence potential competitors. Similarly, in response to "jeans" the responses "Levi's", "Gloria Vanderbilt", "Sasson" and "Jorache" may be obtained. These, therefore, represent the consumer's evoked set for this product category. Finally, free associations may provide information on the uses to which a product/brand could be put by the consumer. For example, in response to the stimulus product "shoes", associations such as "running", "soccer", "basketball", and "jogging" may be obtained. These show both the degree of differentiation within the product category perceived by the consumer and also the possible activities for which s/he might utilize the product (i.e. shoes).

Extension to a Cognitive Hierarchy

Free association techniques have been widely used in psychology to examine the hierarchical structure of information content. This perhaps best exemplified by the work of psychologist Eleanor Rosch. Rosch (1973, 1974, 1975, 1978) maintains that concepts may be arranged along a continuum from abstraction/generality to concreteness/specificity. Three levels along this continuum are identified: (1) a superordinate level of the most general or abstract concepts (e.g. furniture), (2) a basic level of more concrete/specific concepts (such as chair) and (3) a subordinate level of most concrete/specific concepts (such as dining room chair).

Rosch has focused primarily on studying horizontal categorization at each level of the category by asking respondents to indicate attributes associated with the level. Here, a key finding has been that the number of attributes associated with a concept decreases as a function of generality. Thus, fewer attributes are cited at higher levels of abstraction in the hierarchy. This does not, however, imply that the total number of cognitive associations is lower for more general concepts. Rather the character of these associations changes, with concrete nouns instead of attributes being associated with more general concepts.

This was demonstrated in consumer research by Hirschman (1980). Free association techniques revealed no relation between the absolute number of word associations to a product and its level in the hierarchy. Four levels were examined: domain-level, (entertainment), type-level (movie), generic-level, (adventure Movie) and brand-level (James Bond movie). The associations made at the general domain level (entertainment) tended to be predominantly concrete nouns (e.g. a movie); whereas the associations made at the specific generic and brand levels were more likely to be descriptive qualifiers and attributes (e.g. soft, delicious). This is shown in Table 1. Thus, use of a free response word association approach not only confirms

TABLE 1

MOST FREQUENTLY OCCURRING RESPONSES: DOMAIN OF FOOD

<u>FOOD</u>	Meat (17)	<u>ICE CREAM</u>	Baskin-Robbins (22)
	Eat (15)		Haagen-Dazs (21)
	Fruits (12)		Carvel (17)
	Hunger (11)		Vanilla (16)
	Dinner (9)		Delicious (16)
	Lunch (8)		Cold (15)
<u>DAIRY PRODUCTS</u>	Restaurant (8)	<u>BREYERS ICE CREAM</u>	Breyers (13)
			Fattening (13)
	Milk (53)		Natural (33)
	Cheese (44)		Vanilla (25)
	Butter (33)		Chocolate (23)
	Eggs (26)		Delicious (18)
	Yogurt (24)	Cold (13)	
	Cow (21)	Creamy (13)	
	Ice Cream (19)	Flavors (12)	

Rosch's finding that more attributes were associated with more specific concepts, but also demonstrated that different types of associational content (e.g. nouns were associated with different levels.

The use of a free association approach in examining hierarchical cognitive content has, however, some disadvantages. One inherent disadvantage of free association data are that they provide no information on the nature of the association, and, in particular, on the propositional relationship of an evoked response to the product stimulus. For example, does the association of "sneaker" and "running" mean that a sneaker is seen as a running shoe, or that sneakers are used in running?

A further disadvantage is that a stimulus list (i.e. set of product/brand names) has to be developed a priori and arranged in a specific hierarchy. For example, in one experiment, the following hierarchy was selected: Apparel, Pants, Jeans, and Levi's Jeans. This set was then presented (along with several other sets in randomized order) to the respondent and words associated with each level were recorded. In analyzing the data, it was assumed that this provided an appropriate general-to-specific ordering.

This assumption, although it appears logical and possesses "face validity" may, however, not be appropriate for some or all consumers or for some or all product classifications. A final, and perhaps most critical limitation of the free association technique in studying cognitive hierarchies, was that the associations tended to be predominantly horizontal rather than vertical. This was particularly the case in relation to the more specific concepts where modifiers were commonly provided. Consequently, while it has initially been hoped to capture more of the vertical semantic network surrounding the concepts, this was only partially achieved, with the stimulus set used.

The Hirschman-Douglas Card Sorting Task

Given the limitations of the word association task, an

alternative procedure was developed for examining hierarchical cognitive structure. This procedure, termed the Hirschman-Douglas Card Sorting Task (or H-D CST), consists of listing on cards the names of products/brands hypothesized to represent different levels of generality-specificity in a given consumption domain. These cards are then given to subjects who are asked to sort them into piles or groups based on perceived similarity.

In a trial experiment of this method, subjects were provided with six product concepts from two of three different consumption domains (e.g. food, clothing and entertainment). The six product concepts within each domain were selected a priori from three hierarchical levels: Type (television, meat), Generic (talk show, hamburger), and Brand (Johnny Carson, McDonalds). Five pairs of adjectives or modifiers were also included. Three of these were the pairs found by Csgood to be universal qualifiers of meanings (Csgood, May and Miror, 1975). Two more product specific adjectives, i.e. cheap/expensive, and hard/soft, were added. Thus, the subject was given a total of 32 cards representing products/brands from two different consumption domains and five pairs of adjectives. A complete listing of all stimuli is given in Table 2.

In performing the task, subjects were instructed first to look at the cards and then arrange them into "groups that make sense to you" or "piles of things that are closest." Subjects were told that they could make as many or as few groups as they wanted and that there was no time limit and no right or wrong answer to their response. Half of the subjects were required to perform the task twice, being presented first, with one combination of two domains e.g. food and clothing and then another, e.g. food and entertainment. These were systematically rotated to test for order or domain effects.

A convenience sample (n=40) was used for the trial experiment and subjects were drawn from a variety of socioeconomic backgrounds: students, secretaries, guards, professors, businessmen and so forth. The sample is, therefore, not representative of any population. It does, however, provide sufficient variety to make some comments on the methodology.

The first general observation is that the task elicited a high level of involvement and interest on the part of the subjects. In contrast to the disinterest and fatigue often displayed by subjects responding to a questionnaire, the card sorting task generated enthusiasm. The majority of subjects stated that they found the task to be "fun" and "interesting" and completed it in five to fifteen minutes. Thus the recruitment and involvement of subjects does not appear to be a problem with this particular data-gathering technique.

Results

These data were then analyzed by three judges to identify the patterns used by subjects in categorizing the stimulus set. Two major dimensions were identified: 1) whether subjects organized the cards vertically or horizontally, and 2) the degree of tightness, complexity or order subjects used to form the groupings. A high degree of reliability was obtained in the analysis, -disagreement arising in only 5 out of 60 cases.

A key finding is that many people do appear to use a hierarchical framework in organizing their knowledge within consumption domains, though not the tidy and neat structure assumed by the Rosch model. The majority, of subjects (80%) recovered the hierarchies in the stimulus set. They differed, however, in the extent that to which they added adjectives or pairs of adjectives to qualify hierarchies (Figure 1).

TABLE 2

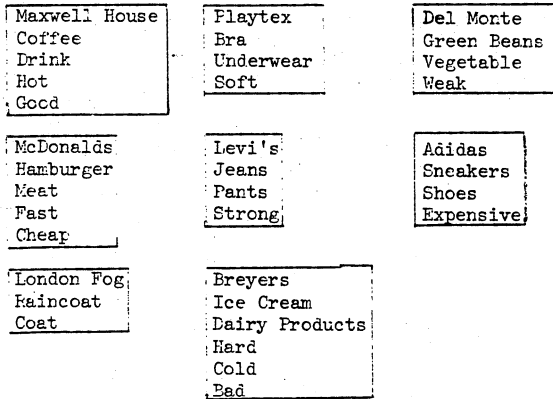
THE PRODUCT CLASS STIMULI AND QUALIFYING PAIRS OF ADJECTIVES

	Superordinate	Ordinate	Brand
<u>FOOD</u> <u>CATEGORY</u>	Drink Vegetable Dairy Products Meat	Coffee Green Beans Ice Cream Hamburger	Maxwell House Del Monte Breyers McDonalds
<u>CLOTHING</u> <u>CATEGORY</u>	Pants Coat Shoes Underwear	Jeans Raincoat Sneakers Era	Levis London Fog Adidas Playtex
<u>ENTERTAINMENT</u> <u>CATEGORY</u>	Television Sports Music Movie	Talk Show Boxing Rock and Roll Western Movie	Johnny Carson Muhammed Ali Elvis Presley John Wayne
		Source	
<u>ADJECTIVES</u>	Good Strong Fast Hard Hot Expensive	Bad Weak Slow Soft Cold Cheap	Osgood's Evaluation Dimension - Potency - Activity Product Class Attribute Product Class Attribute Product Class Attribute

FIGURE 1

FIGURE 2

A VERTICAL BRAND-ORDERED PATTERN



A HORIZONTAL ORDERED PATTERN

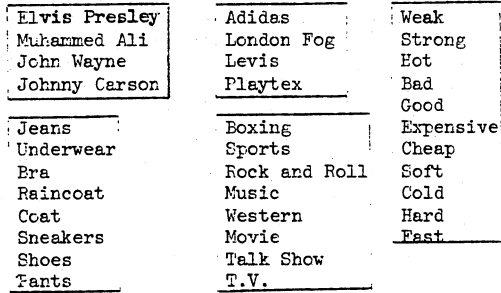
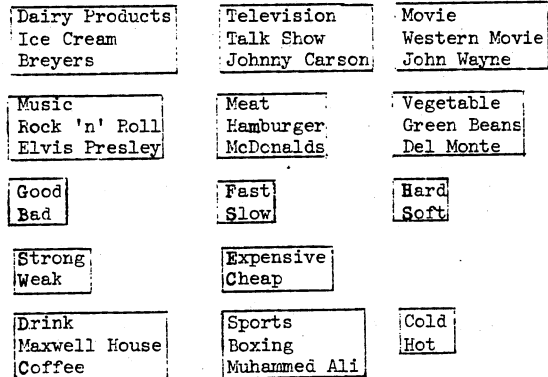


FIGURE 3

A VERTICAL TYPE ORDERED PATTERN



Some subjects (19%), particularly those who expressed low involvement with the stimulus set, grouped the cards into three piles - adjectives, brand names and products (Figure 2).

Another aspect was how the hierarchies were organized. Some subjects adopted a predominantly brand-organized structure. For example, in Figure 1 the groupings are all headed by a brand-level concept. Others adopted a type-level structure. This is shown in Figure 3. Within each grouping the Rosch-like hierarchical structure is maintained (i.e., Type, Generic, Brand) with only one interesting exception: drink, Maxwell House, coffee. Some subjects declared themselves unable to order within piles. This occurred primarily when somewhat complex structures were developed initially.

The other dimension - tightness vs. looseness - refers to how highly organized the structure created by the subject was. A very "tight" structure was one in which all the cards were grouped into a maximum of three piles (10%) (Figure 4). This occurred predominantly where subjects appeared to have a strong need for order. 16% had a moderately tight structure i.e. four to five piles. The majority (73%), adopted a loose structure in which several cards were grouped into pairs (representing a two-level, or binary hierarchy) and/or were left unrelated to one another (that is, were "grouped" as individual units).

FIGURE 4
A TIGHT STRUCTURE

Underwear	Music	Cheap
Bra	Rock and Roll	Expensive
Playtex	Elvis Presley	Hot
Pants	Sports	Cold
Jeans	Boxing	Bad
Levis	Muhammed Ali	Good
Coat	Movie	Weak
Raincoat	Western Movie	Soft
London Fog	John Wayne	Hard
Shoes	T.V.	Slow
Sneakers	Talk Show	Fast
Adidas	Johnny Carson	

A critical assumption underlying this procedure is that the groupings created by the subject with the cards correspond to the way information is structured in his/her memory. Whether this assumption is valid or not remains open to some question. However, subjects did not appear to have any difficulty with the task, or find it artificial, and the differences in the resulting structures suggest that the assumption may be appropriate.

It should be noted that the card sorting task methodology is still in the early stages of development. Further work in devising appropriate stimulus sets is required. The stimulus hierarchies were arbitrarily defined by the researchers and imposed on the subject. It is therefore difficult to assess how much of the structure obtained is a result simply of the specific stimulus set. One alternative which might partially eliminate this bias is to provide the subject with a more ambiguous and complex set of words from varied domains of consumption. If the task is repeated for several domains the consistency of organizational patterns can be examined. Furthermore, the same set of words can be given at various time intervals to measure the stability of response.

The card sorting task may not be appropriate for all consumers. Tests conducted with a number of children indicated that children do not respond to the stimuli in the same way as adults. Children, for instance, tended to respond to the words not as representing products, but rather as semantic entities. For example, a typical child's approach to sorting the cards is to alphabetize the words, or to group them into piles according to word length, number of words on the card, and so forth.

This is consistent with studies of cognitive development (Gelman 1978) which indicate that younger children are unable to reason abstractly using semantic stimuli or handle hierarchical classification, because they have not arrived at the level of what Piaget terms symbolic operations (Piaget 1976). Hence, they do not perceive the words as representing products, but rather as semantic entities.¹

¹The card sorting task may be similarly inappropriate for psychometricians, consumer researchers and marketing model-builders. For example, one well-known marketing model-builder who, while complaining that it was trivial and meaningless, proceeded to construct a geometric con-

Additional Applications of the Free Association and Card Sorting Techniques

Free Association

The type of free association techniques discussed here also have a number of other potential applications. Perhaps one of the most valuable applications is in the area of subcultural and cross-cultural consumer research (Wind and Douglas 1980). The free association approach is essentially an unstructured and hence predominantly emic method for data gathering and analysis (e.g. Szalay and Deese 1978, Pike 1966, Ember 1977, Triandis 1972).² It does not imply the imposition of a construct identified in one culture on another culture, where it may not be relevant. Rather it allows consumers in various cultures or subcultures to respond freely in their own terms with whatever verbal or written signs they feel most appropriate.

Some cultural self-referent bias (Lee 1966) may however be introduced by the researcher given his specific cultural referent in the choice of the stimuli and in interpretation of the data. Hence, careful attention to "decentering" the stimulus set will be required (Werner and Campbell 1973). In addition, since the onus of interpretation is placed on the researcher in determining where there are similarities and differences between countries, use of multiple judges from different cultures will be desirable.

One potential use of free association data in this context is to compute "shared meaning" (i.e. similarity) ratios between cultures or subcultures, or to identify to which culture or subculture a consumer belongs (Szalay and Deese 1978). If the same or similar responses are associated with the same product, the cultures may be classified as similar (Hirschman 1980). Additionally, where consumers belong to more than one subculture, for example, a Jewish American teenager, the associations made in relation to a particular product will indicate which subculture dominates in relation to that particular product.

The methodology can also be extended to examine the meaning attached to different colors, shapes or to composite objects in different cultures and countries (Deregowski 1980, Pick 1980). Consumers could, for example, be given a set of colors or shapes and asked to provide associations to them. Similarly more complex stimuli sets such as actual products and brands could be used.

figuration atop his desk with precise euclidean distances separating the cards. Unfortunately, the author-administrator neglected to copy down the resulting multi-dimensional cognitive map. On the other hand, perhaps we are better off left in ignorance.

²The "emic" and the "etic" distinction is an important one in cross-national and cross-cultural research (Ember 1977, Pike 1954, Triandis 1972). The "emic" school holds that attitudinal or behavioral phenomena are expressed in a unique way in each culture, and are best understood in their own terms. Following this point of view, measures specifically adapted to each cultural context will be required. The "etic" school, on the other hand, is primarily concerned with identifying universal or "culture-free" measures. This facilitates comparisons but potentially leads to loss of some precision or accuracy of measurement in each national context (Elder 1976, Przeworski and Teune 1966/7, 1970).

Such procedures might also be appropriate to study children's responses to products, brands and advertisements. Since small children have difficulty grasping abstract concepts, use of visual stimuli may aid considerably in increasing comprehension.

Card Sorting

The card sorting task would also appear to have potential applications in cross-cultural and subcultural research. Again, the same caveats apply as for the free association task in that care must be taken to avoid a pseudo-etic bias in the design of the product set and in its interpretation. In many cases it may be desirable to include in the set, products specific to a particular culture, as well as those common across cultures.

If a set of products common across cultures is identified, the card-sorting task may be used to assess whether certain cultures are characterized by predominantly product-type, brand-type or other cognitive organization structures. The level of advertising may, for example, affect information organization. The relative looseness or tightness of the organizational pattern and complexity of organization also may vary across cultures depending upon the level of education or societal development.

As in the case of free association techniques, the Hirschman-Douglas sort task can be extended by using more complex stimuli. These might, for example, include use of black and white or colored pictures and photographs of products and brands or actual objects. Consumers would then be requested to sort these into piles based on similarity, and differences and similarities in these patterns between cultures and countries could be examined. Again, the technique might be highly appropriate with children.

Finally, both the free association and card-sorting tasks may be used to examine the relation between cognitive structure and respondent characteristics. The number of responses given to a stimulus product may be related to high levels of cognitive complexity and intelligence. The tightness and complexity of the organizational pattern may also reflect intelligence and cognitive complexity, and may be related to such personality traits as desire for closure and tolerance of ambiguity. Other factors such as media usage patterns and relative level of information seeking may also relate to cognitive structure and, hence, to a consumer's performance on both the free response and card-sorting tasks.

Conclusions

The purpose of this paper is to suggest the use of two approaches - free association and card sorting - in investigating hierarchical cognitive content. These approaches provide a useful extension to the type of information obtained from traditional approaches. In particular, they may aid in extending knowledge of the content and structure of consumer product cognitions and information storage and have potential applications in cross-cultural and cross-national research.

The results of the present experiment indicate that the majority of subjects organize the stimulus set into hierarchies, generally of more than five groups. The test re-test reliability of this finding was also high, indicating stability in results, and specific that the patterns obtained were not specific to a particular product class. The patterns may however in part due to the nature of the stimulus set, although alternative groupings were feasible.

It should, however, be noted that both approaches, and in particular, the card sort task, are still in the early stages of development. Considerable further work in

refinement of the method, particularly with regard to the development of the stimulus set, notably a more ambiguous set and appropriate analytic procedures will clearly be required. Furthermore, adaptation relative to specific research objectives is also needed.

References

- Bettman, James R. (1979), "Memory Factors in Consumer Choice: A Review," Journal of Marketing, (Spring), 37-53.
- Caron, Noel and Lutz, Richard J. (1979), "A Model and Methodology for the Development of Consumer Information Programs," Journal of Marketing, (January), 58-67.
- Deregowski, Jan B. (1980), "Perception," in Harry C. Triandis and Walter Lonnes (eds.), A Handbook of Cross-Cultural Psychology, Vol. 3, Basic Processes (Boston: Allyn and Bacon, Inc.) 21-117.
- Douglas, Susan P., Craig, Samuel and Faivre, Jean-Philippe (1961), "Protocols in Consumer Research: Problems, Methods and Uses," to appear in Jagdish Sheth (ed.), Research in Marketing, Vol. V.
- Ember, Carol R. (1977), "Cross-Cultural Cognitive Studies," Annual Review of Anthropology, 6, 33-56.
- Elder, Joseph W. (1976), "Comparative Cross-National Methodology," Annual Review in Sociology, Palo Alto, California.
- Gelman, Rachel (1978), "Cognitive Development," in Annual Review of Psychology, 29, 297-332.
- Green, Paul E., Wind, Yoram and Jain, Arunk (1973), "Analyzing Free-Response Data in Marketing Research," Journal of Marketing Research, X (February), 45-52.
- Gutman, Jonathan (1980), "Equivalence Range in Categorization," in Jerry Olsen (ed.) Advances in Consumer Research, Vol. VII, (Ann Arbor: Association for Consumer Research), 411-416.
- Hirschman, Elizabeth C. (1980), "Commonality and Idiosyncrasy in Popular Culture," in Consumer Aesthetics and Symbolic Consumption, E.C. Hirschman and Morris Holbrook (eds.), Association for Consumer Research, forthcoming.
- Lachmar, Roy, Lachman, Janet and Butterfield, Earl C. (1979), Cognitive Psychology and Information Processing, Hillsdale, N.J.: Lawrence Erlbaum and Assoc., 1979.
- Lee, James A. (1966), "Cultural Analysis of Overseas Operations," Harvard Business Review, 44 (March-April), 106-114.
- Olson, Jerry C. and Muderrisoglu, Aydin (1977), "The Stability of Responses Obtained by Free Elicitation" in Advances in Consumer Research, Vol. 6, W.C. Wilkie (ed.), Ann Arbor, MI, Association for Consumer Research, 269-275.
- Olson, Jerry C. (1978), "Theories of Information Encoding and Storage: Implications for Consumer Research," in Effects of Information on Consumer and Market Behavior, A.A. Mitchell (ed.) Chicago: American Marketing Association.

- Osgood, Charles E., May, William H. and Miron, Murray S. (1975), Cross-Cultural Universals of Affective Meaning Urbana: University of Illinois Press.
- Piaget, Jean (1970), Genetic Epistemology, New York: Norton.
- Pick, Anne D. (1980), "Cognition: Psychological Perspectives," in Harry C. Triandis and Walter Lonner (eds.), A Handbook of Cross-Cultural Psychology, Vol. 3, Basic Processes, (Boston: Allyn and Bacon, Inc.) 117-154.
- Pike, Kenneth (1966), Language in Relation to a Unified Theory of the Structure of Human Behavior, the Hague: Mouton.
- Preworski, Adam and Tenue, Henry(1970), The Logic of Comparative Social Inquiry, New York: Wiley Interscience.
- Rosch, Eleanor (1973), "Natural Categories," Cognitive Psychology, 4, 328-350.
- _____ (1974), "Universals and Cultural Specifics in Human Categorization," in Cross-Cultural Perspectives On Learning, R. Breslin, W. Conner and S. Eachner (eds.) London: Sage Press.
- _____ (1975), "Cognitive Representations of Semantic Categories," Journal of Experimental Psychology, 104, 192-233.
- _____ (1978), "Principles of Categorization" in Cognition and Categorization, E. Rosch and B. Lloyd (eds.), Hillsdale, N.J., Lawrence Erlbaum and Associates.
- Rosenzweig, Mark R. (1961), "Comparisons of Word Association Responses in English, French, German and Italian," American Journal of Psychology 74, 347-360.
- Russo, Edward J. and Johnson, Eric J.(1980), "What Do Consumers Know About Familiar Products?" in Jerry Olson (ed.) Advances in Consumer Research Vol. VII (Ann Arbor: Association for Consumer Research), 417-423.
- Szalay, Lorand B. and Deese, James(1978), Subjective Meaning and Culture, Lawrence Erlbaum Associates: Hillsdale, N.J.
- Triandis, Harry C. (1972), The Analysis of Subjective Culture, New York: John Wiley.
- Werner, Oscar and Campbell, Donald T(1973), "Translating, Working Through Interpreters' and the Problems of Decentering", in Faoul Naroll and Ronald Cohen (eds.) A Handbook of Method in Cultural Anthropology, New York: Columbia University Press, 398-420.
- Wind, Yoram and Douglas, Susan(1980), "Comparative Consumer Research: The Next Frontier," Wharton School Working Paper, University of Pennsylvania, Philadelphia, PA, (February).

USAGE-SITUATIONAL INFLUENCES ON PERCEPTIONS OF PRODUCT-MARKETS:
THEORETICAL AND EMPIRICAL ISSUES

Rajendra K. Srivastava, University of Texas at Austin

Abstract

Cognitive approaches for the modeling of consumer choice and product-market structures have generally been based on the perspective where product-markets may be defined as consisting of those products that are purchased by the same consumers who presumably desire the same benefits/costs that products may offer. However, products and consumers are embedded in an environment which may influence choices. This paper reviews the affect of environmental (in particular, usage-situational) conditions on perceptions of product-markets and raises a number of research issues.

Introduction

Deterministic brand preference/choice models are generally based on the premise that products are valued for the attributes they possess and that customers seek to maximize their "utility" by choosing/purchasing desired combinations of attributes. Thus products/brands offering similar combinations of (levels of) attributes are likely to be more substitutable/competitive. This perspective is widely maintained in consumer/marketing research as evidenced by the formulation of multi-attribute attitude/preference models (see Wilkie and Pessemier 1979) and associated methodologies such as perceptual mapping (see Shocker and Srinivasan 1979). These models/techniques portray differences in preference among consumer segments by means of their objectives (ideal-points) and/or trade-offs (importance weights) for the benefits/costs associated with the dimensions of the attribute space. Though the conceptualization is intuitively appealing, the validity of the approach has been questioned by several researchers who have noted the discrepancy between attitude/preference and actual behavior/choice.

First, the proponents of the notion that customer choices are the result of multiple stage choice processes (Bettman 1979, Wright and Barbour 1977, and Payne 1976) would contend that the importance weights, and indeed the set of items under consideration for usage or purchase, would depend on the stage of information processing. In the initial stages of information processing, the emphasis is on screening alternatives in order to minimize information overload, there is a tendency for individuals to use lexicographic rules. In subsequent stage(s) there is an increasing tendency to employ compensatory strategies. The implication of this notion is that the attribute weights and ideal points that are derived could be a function of items included in the preference elicitation task (structured by the research) and the subsequent modeling of these data as a single-stage choice process could be misconceived.

Second, there is the contention that the first choice (most preferred item) attitude or preference models used for predicting customer behavior fail to recognize the customers often choose multiple items from a product class. For example, if a subject furnishes a preference ranking of Coke, Pepsi and Seven-up, yet purchases only Coke and Seven-up for consumption, the discrepancy between preference and choice could be explained by the notions of: (1) dominance--Coke dominates Pepsi but not Seven-up on all attributes (see Huber and Reibstein 1977), and/or (2) attribute satiation--where customers

are seen as seeking balance among the attribute levels of a set of items (Farquhar and Rao 1976, McAlister 1979).

Finally, recent research has clearly illustrated the importance of environmental influences in determining consumer choice (Belk 1974, 1979, Srivastava, Shocker and Day 1978, Srivastava 1979, 1980, Kakkar and Lutz 1975). At a conceptual level, situational influences may be seen as moderating the importance of benefits that customers seek. For example, within the broad markets for "beverages," the usage context "at a picnic" may make the attribute "refreshing ability" more salient than in the usage context "after dinner."

The concepts of environmental influences, multiple choice and multistage processing appear to be quite consistent. In the initial stage(s), simplifying strategies may be used to reduce the number of alternatives to be examined at subsequent stages. A customer may first develop consideration sets of products based on their functional attributes, i.e., those characteristics that are in some sense "required" in the environmental context. Subsequently, they may examine this reduced set of alternatives on the basis of attributes still considered important, but along which the alternatives in the reduced set vary in terms of the "level of attribute" present in order to determine their final choice (see the concept of "determinant" attributes--Myers and Alpert 1968). Different environmental contexts are likely to lead to varying consideration sets and consequently multiple choices. Thus, rather than merely contend that customers seek a balance among attributes presumably due to either variety seeking or satiation, an alternative (and perhaps more explanatory) perspective is that environmental requirements dictate the importance of product attributes, and that consumers develop a product portfolio that reflects the set of situations they expect to encounter.

Though environmental influences seem to provide a vital link between preference and choice, there are unfortunately numerous types of such factors that could be relevant (e.g., physical and social surrounding, task definition, antecedant conditions, and temporal perspective; see Belk 1975). This paper attempts to provide a conceptual framework where the purpose of the research, or the decision context within which product market structure is to be used, is seen as moderating the importance of the types of environmental variables.

A Contingency Model For Market Relationships

The derivation of product market structures is dependent on two considerations: (1) the choice of a measure of the degree of competition, or its surrogate substitutability, between products (for example, perceived similarity between product-pairs), and (2) the choice of an appropriate data-reduction technique for analyzing the inter-product similarities (for example, multidimensional scaling, cluster analysis). In their comprehensive review of customer-oriented (based on demand, rather than supply considerations) approaches for identifying product-markets, Day, Shocker and Srivastava conclude that "the suitability of different methods (measure-technique combination) is strongly influenced by the character of the market environment" (1979, p. 18, words within paren-

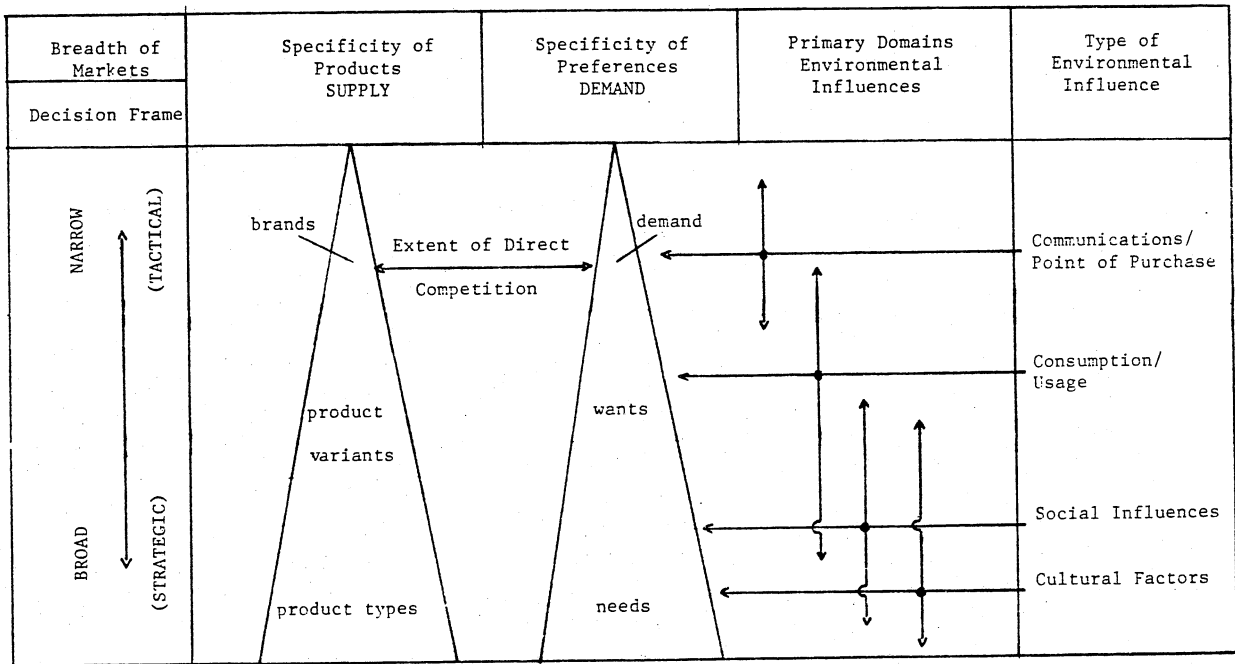
theses, ours). Further they conclude that "the questions of how to identify product-market boundaries cannot be separated from the way results are to be used."

Exhibit 1 presents a framework that relates the market environment (types of environmental influences) and the way results are to be used (strategic versus tactical decisions) to the concept of the product hierarchy (Lunn 1972). This framework is partially adapted from Arndt's (1976) conceptualization of the relationship between competition at various levels (brand, product-variants, product types) of the product hierarchy, specificity of customer needs, and types of situational influences.

as nutritional value and keeping awake. Product-variants are available within the same overall product type and are likely to be more competitive in the short run. For example, ground and instant coffee can be expected to compete more directly. Finally, brands within the same product variant compete even more directly. For example, brands of instant coffee are very similar in the purposes they serve. The specificity of customer preferences may be related to the notion of the product hierarchy via the conceptualization that consumer decision-making may be a hierarchical or multi-stage process where simplifying strategies (rules of thumb) may be utilized to eliminate product types, variants, and brands sequentially. For example, a person considering

EXHIBIT 1

A CONTINGENCY MODEL FOR MARKET RELATIONSHIPS



The "decision frame" (column 1) is related to the breadth of market that is likely to be of managerial interest. Very narrowly-defined boundaries appear adequate for short-run, tactical decisions (Day, Shocker, and Srivastava 1979) for which management is mostly concerned about the current and direct competition from the brands of rival firms. On the other hand, the long-run strategic decisions, management is concerned not only about the current and direct competition from very similar brands but also the potential, less direct competition from other product-variants/types which may be more substantial threats in the future due to changed economic conditions, regulations or shifts in consumer needs. Therefore, for strategic decisions, broader market definitions would appear to be more appropriate.

Columns 2 and 3 reflect the relationship between the levels of the product hierarchy and the specificity of consumer preferences. It must be emphasized that these levels are not discrete but represent points along a continuum such that the extent of direct competition decreases as one comes down the hierarchy (from brands to product categories). Totally different product-types exist to satisfy significantly different purposes in different situations. For example, in the market for hot beverages, coca and coffee may serve purposes such

the choice of a beverage may discard cocoa since his primary objective might be to stay awake, then ground coffee, as it may require too much effort to prepare and finally choose among various brands of instant coffee on the basis of flavor. This example is quite simplistic, yet it serves to illustrate the process that is likely to occur in a more subtle manner.

Though environmental factors influence customer choice, different types of situational factors are likely to affect competition at various levels of the product hierarchy through their influence on the customer choice process (specificity of preference). For example, adopting Hansen's (1972) categorization of the types of environmental influences again considering the beverage example, social influences (peer pressure) and cultural norms may affect the choice between tea and coffee. If the coffee route was chosen, the consumption or usage-situation (e.g., while entertaining guests versus grabbing a quick cup) may determine the choice between instant and ground coffee. Finally, if instant coffee were the chosen route, communication (advertisements, word-of-mouth communication regarding flavor) or point-of-purchase factors (coupons, displays) may influence the choice among brands of instant coffee. This is not to say that the different types of

environmental influences affect discrete levels of the hierarchies. Column 4 depicts the likely domains of influence of the various types of environmental influences (column 5). Further, it is important to realize that the distinctions drawn in the previous paragraphs do not categorically imply, for instance, that cultural factors do not shape brand preferences. Though cultural factors primarily affect competition between product types, their influence at the higher (e.g., brand) levels of the product hierarchy will be indirectly based on: (1) the sequential conversion of diffused goals/preferences into more specific ones, and (2) the dependence of the less stable factors on the more stable ones (for example, cultural norms may determine the relevance of usage-situations as well as the frequency with which they arise for different individuals).

The relationship in Exhibit 1 may appear to represent an "all-encompassing" model of customer behavior and certainly cannot be verified within the scope of a single research. However, scattered pieces of previous research verify portions of the framework. For example, Miller (1975) and Srivastava, Shocker and Day (1978) observe that anticipated usage situations appear not to affect preferences at the brand level.

Perceptions of Product Markets: Research Issues

Response Homogeneity and Deterministic Choice Models

Belk (1979), Srivastava, Shocker and Day (1978) and Srivastava (1979) observe a high degree of response homogeneity among subjects when the effects of the usage-situation are controlled for, while previous researchers which ignore situational influences note the discrepancy between preference and choice ("error variance") as well as the differences among individuals ("heterogeneity"). Some of the error variance and heterogeneity may be due to the fact that choices are situation-specific (while the elicited preferences were not obtained as such), and different individuals may have given their preferences with varying usage or consumption situations in mind. For example, if a respondent is provided several brands of instant and regular coffee and asked to rank order preferences, she/he may do so keeping "flavor" in mind and provide higher ranks to the regular coffee brands. Another respondent may perform the task with "ease of preparation" in mind and provide relatively higher ranks to the instant brands. However, both may actually use instant coffee when in a hurry and regular coffee while entertaining. The above self-serving and somewhat simplistic example illustrates that both error variance and respondent heterogeneity may be attributable to the ambiguity of the response task. Accordingly, as argued by Rokeach and Kliejunas, we must elicit attitudes toward objects within a situation if we are interested in predicting situation specific behaviors (1972).

The adoption of situational influences in consumer choice models is equivalent to contending that the usage situations affect importance weights and/or ideal points (Srivastava 1979, Hagerty 1980, Miller and Ginter 1979, Pekelman and Son 1976). It is much less likely that perceptions of attribute levels would be affected. The specification of the situational variables may be expected to more clearly define the attributes that consumers seek, and further, simplify the judgmental task. Consequently one may expect greater homogeneity between respondents and less error variance. That is, the explicit consideration of situational influences should allow for a more deterministic view of customer behavior (Srivastava 1980). Bass notes that "it will never be possible to prove conclusively that behavior is fundamentally deterministic since it will never be possible to measure all the variables which influence choice"

(1974, p. 2). While it is hard to dispute this statement, the explicit consideration of situational influences enables a more deterministic (and correspondingly more explanatory) view of customer behavior.

Situational Taxonomies

Though this deterministic conceptualization is appealing in terms of its ability to account for discrepancy between preference and choice, the estimation of the distribution of attribute weights remains a very real problem. Additionally, even if the distributions were estimable, we would not know the relationship between attribute weights and different types of situational influence. This kind of understanding is more probable if a taxonomy of situational influences (which enables the explicit, systematic consideration of environmental factors) is developed.

In general, two approaches have been followed in an attempt to develop situational taxonomies. Neither is totally satisfactory. The first has sought to classify situations by the nature of the psychological processes which they lead to. For example, Kakkar and Lutz (1975) attempt to classify situations in terms of Mehrabian and Russell's (1974) dimensions of pleasure, arousal and dominance. The advantage of this approach is that the focus is on the situation as perceived by the individual, and that is parsimonious and general (applies to all product contexts). Unfortunately, the psychological taxonomic dimensions appear to have little systematic effect on preferences (Kakkar and Lutz 1975).

The concept of a general taxonomy of situational influences for the moment, appears unlikely to be realized. This is because situational influences that affect some consumer behaviors (e.g., the choice of "high-involvement" products such as automobiles) could be entirely different from those that affect other behaviors (e.g., the choice of "low-involvement" products such as soft drinks) (Belk 1979). However, product-specific taxonomies based on objective situations (factors external to consumers that may lead to differences in elicited behavior) may be easily constructed by following procedures outlined in Fredriksen's (1972) article and as implemented by Srivastava, Shocker and Day (1978) in the "breath freshener" market and Belk (1979) in the "clothing" market. While the objective taxonomies appear to be better descriptors of consumer behavior compared to psychological taxonomies, they present a major drawback in that situations may have to be defined in great detail to the point where the number of variables is massive enough to make reliable measurement and meaningful analysis impossible (Lutz and Kakkar 1975, Russell and Mehrabian 1976). However, the efforts devoted to developing objective situational taxonomies (Belk 1975, 1979, Srivastava, Shocker and Day 1978, Srivastava 1979) have generally resulted in two or three taxonomic dimensions that are adequate for explaining most of the situational variance.

The key point is that in the creation of situational taxonomies we are interested in those aspects of the environment that affect consumer behavior. Although the total number of situations that persons encounter is enormous, and each situation is unique and the likelihood of exact replication is exceedingly small, it is undoubtedly also true that people do not behave differently under all such changed circumstances. Prior learning and usage experience are integrated into relatively routinized responses (Howard and Sheth 1969) in many of these instances. Many authors have argued that humans have relatively limited capability for processing information and problem solving (see Slovic 1972, Bettman 1979). They develop heuristic or simple rules of thumb that enable them to deal with complex situations. They appear to categorize prior learning and

experience in memory and base choices and decisions on such categorizations (Rosch 1978). Consequently, respondents may evoke similar categorizations or stereotyped uses that are represented in a simple way in memory.

Situational/Lifestyle Segmentation

The apparent homogeneity of consumer response when situational influences are controlled for suggests the possibility of situational segmentation. That is, product-markets and competitive effects can be examined within "situational submarkets". This line of inquiry may lead to interesting research issues as certain situational types are likely to occur with differing frequencies for different consumer segments. Subsequently, the identification of lifestyle, cultural, demographic or other factors which effect the frequency of occurrence of situational types can offer better opportunities for targeting marketing efforts to those people who experience relevant (for use of specific products/brands) with greater frequency.

Measures of Substitutability

Traditionally, marketing researchers have focused their attention towards examining competition between brands, perhaps partially due to the demands placed by the utilizers of research who may be besieged with the more immediate (tactical) decision-contexts. Within these contexts, where narrowly-defined markets (e.g., brands of instant coffee may be adequate, the traditional focus on point-of purchase (display, shelf-level/space) and communication (advertising, promotional activities) influences certainly encompass the primary environmental factors affecting customer choice. Moreover, these brands are likely to be very similar (in terms of physical characteristics/functions served/benefits provided) and are likely to be used for the same usage/consumption situations(s) by the same customers (with similar social/cultural backgrounds). In this context, one cannot question the adequacy of the more familiar measures of inter-product similarity (a surrogate for substitutability) such as: (1) similarity of brand attributes - correlation or "matching" coefficients between brands across attributes, (2) overall perceived similarity - as measured by similar-dissimilar brand-pair ratings (Jain and Etgar 1975) or by sorting brands into groups and subsequently obtaining a similarity measure based on the proportion of times brand-pairs were sorted into the same group (Bourgeois and Haines 1975), and (3) direct measures of substitutability - such as Pessemier's (1975) "dollar-metric" measure or other measures eliciting the degree of substitutability between brand-pairs on a rating scale.

However, when the interest lies in developing market structures for strategic decisions, it is desirable to not only understand the immediate competition (i.e., at the brand level) but also the nature of the competition between product variants/types that may be potentially competitive. When the interest lies in examining the competition between product variants/types, usage-situational influences begin to play an important role because: (1) products may be substitutable under some circumstances (usage situations), yet may not compete under others, i.e., products may have multiple uses which may not overlap completely between product pairs and the appropriateness for use of a product may depend on the "match" between product benefits and the requirements of the usage-situation (Belk 1974, 1979, Steffire 1971, Srivastava, Shocker, and Day 1979), and (2) customers may develop "assortments" or collection of products in order to be prepared to meet the future contingencies across diverse, anticipated usage occasions (Wind 1977).

It would appear from the previous paragraph that for strategic marketing decision-making where broader market structures are desirable, "substitution-in-use" or the similarity of product usage patterns would provide a managerially relevant measure of inter-product substitutability. Further, since product variants/types need not have all attributes (benefits/characteristics) in common, the more traditional measures (e.g., similarity of product attributes) may be "discontinuous" when applied to study competition at the lower levels of the product hierarchy (this problem can be alleviated to some extent by the use of discrete multivariate methods to analyze the data). This does not imply that the more traditional measures are irrelevant but merely that substitution-in-use appears to be better-suited for developing broader market structures. Indeed, it may be useful to adopt a two-stage methodology where substitution-in-use is first used to develop the broader market structure and to identify the discontinuities in this space. Then, within each situationally defined submarket, the traditional measures may be employed to obtain a better understanding of the more direct competitive relationships (Shocker and Srinivasan 1974, 1979). Additionally, if the relative frequencies of taxonomic cells are obtained it should be feasible to predict the assortment of products that individuals are likely to purchase to serve their needs across anticipated usage-situations. As mentioned earlier, this offers an alternate explanation to attribute satiation for modeling multiple choices.

Discussion

Throughout this paper there has been an emphasis on the matchings between situational requirements and product benefits/attributes. This interaction is to be expected and indeed has on occasion explained more variance than some main effects (see Kakkar and Lutz 1979 for details). However, the close correspondence between situational factors and product attributes leads to the question: "What are situational factors?" Belk (1974, 1975) proposes that environmental factors should include all variables not included in the description of persons or products. This exclusion rule is debateable, especially given the differences in training of researchers. For example, social norms may be seen as an environmental influence by sociologists, yet modeled as person factors in the prediction of behavioral intentions (Fishbein and Ajzen 1975). However, as long as relevant influences are included, no matter how they are modeled, we should develop a better understanding of consumer behavior.

The response homogeneity between individuals under situational control appears to indicate the absence of the person-situation interaction. This may be an artifact of the research samples (primarily students) which may a priori have been expected to be homogeneous. Other evidence suggests that individuals may be expected to respond differently to the same situation. For example, "Internals" may be expected to be influenced less by environmental circumstances compared "Externals" (Rotter 1966).

In summary, this paper has provided a review of environmental influences on consumer behavior. Situational influences were seen as moderating consumer choice. A contingency model showing the dependency between the specificity of competition (direct to less direct, based on the level of the product hierarchy) and the types of situational influences was proposed. This was used to suggest that usage-situational influences were likely to directly affect non-brand competition. However, if consumer choices were indeed made through multi-stage processes, usage situations may indirectly affect competition at the brand level. A variety of

implications of usage situational influences were discussed including response homogeneity across persons, multiple choice or variety seeking, measures of product substitutability, deterministic choice models, segmentation, and types of situational taxonomies. The validity of each of these implications is researchable. It is hoped that this paper will lead to further exploration of environmental factors in consumer/marketing research.

References

- Arndt, J. (1976), "The Marketing Thinking of Tomorrow: Beyond the Marketing Concept Toward New Dignity and Responsible Freedom for Managers and Consumers." Norwegian School of Economics and Business Administration, (February).
- Bass, Frank M. (1974), "The Theory of Stochastic Preference and Brandswitching," Journal of Marketing Research, 11 (February), 1-20.
- Belk, Russell W. (1974), "An Exploratory Assessment of Situational Effects in Buyer Behavior," Journal of Marketing Research, 11 (May), 156-163.
- Belk, Russell W. (1975), "Situational Variables and Consumer Behavior," Journal of Consumer Research, 2 (December), 157-164.
- Belk, Russell W. (1976), "Situational Mediation and Consumer Behavior: A Reply to Russell and Mehrabian," Journal of Consumer Research, 3 (December), 175-177.
- Belk, Russell W. (1979), "A Free Response Approach to Developing Product-specific Consumption Situation Taxonomies," in Allan D. Shocker, ed. Analytical Approaches to Product and Marketing Planning. Cambridge, MA: Marketing Science Institute, (April), 177-196.
- Bettman, James R. (1979), An Information Processing Theory of Consumer Choice. Reading, MA: Addison Wesley.
- Bourgeois, J.C. and G.H. Haines. (1975), "Defining an Industry," Paper Presented at the Joint National Conference of ORSA and TIMS, Las Vegas, Nevada (November).
- Day, George S., Allan D. Shocker, and Rajendra K. Srivastava. (1979), "Identifying Competitive Product Markets: A Review of Customer Oriented Approaches," Journal of Marketing, 43 (Fall), 8-19.
- Farguhar, P.H. and V.R. Rao. (1976), "A Balanced Model for Evaluating Subsets of Multi-Attributed Items," Management Science, 22, 528-539.
- Fishbein, Martin and Izek Azjen. (1975), Belief, Attitude, Interaction and Behavior: An Introduction to Theory and Research, Reading, Mass.: Addison-Wesley.
- Fredriksen, Norman. (1972), "Towards a Taxonomy of Situations," American Psychologist, 27, (February), 114-124.
- Hagerty, Michael R. (1980), "A Deterministic Model of Consumer Choice," in Jerry C. Olson (ed.), Advances in Consumer Research, 7, Ann Arbor, Mich: Association of Consumer Research, 129-133.
- Hansen, F. (1972), Consumer Choice Behavior: A Cognitive Theory. New York: The Free Press.
- Howard, John A. and Jagdish Sheth. (1969), The Theory of Buyer Behavior, New York: John Wiley.
- Jain, Arun K. and Michael Etgar. (1975), "How to Improve Antitrust Policies with Marketing Research Tools," in 1975 Combined Proceedings of the American Marketing Association, Edward M. Mazze, ed., Chicago: American Marketing Association, 72-5.
- Kakkar, P., and Lutz, R.J. (1975), "Toward a Taxonomy of Consumption Situations." Combined Proceedings. American Marketing Association, 206-210.
- Kakkar, Pradeep and Richard J. Lutz. (1979), "Situational Influence on Consumer Behavior: A Review," Working Paper #80, University of California, Los Angeles (October).
- Lunn, Tony. (1972), "Segmenting and Constructing Markets," in R.M. Worcester (ed.) Consumer Market Research Handbook, Maidenheas, Berkshire: McGraw Hill.
- Lutz, Richard J. and Pradeep Kakkar. (1975), "Situational Influence in Interpersonal Behavior," in M.J. Schlinger (ed.), Advances in Consumer Research, Vol. 2, Chicago: Association for Consumer Research, 370-378.
- McAlister, Leigh. (1979), "Choosing Multiple Items From a Product Class," Journal of Consumer Research, 6, 3 (December), 213-224.
- Mehrabian, A., and Russell, J.A. (1974), An Approach to Environmental Psychology, Cambridge: M.I.T. Press.
- Miller, K.E. (1975), "A Test of Situational Multi-Attribute Attitude Model," in M.J. Schlinger (ed.), Advances in Consumer Research, Vol. 2, Chicago: Association for Consumer Research, 455-464.
- Miller, Kenneth E. and James L. Ginter. (1979), "An Investigation of Situational Variance in Product Choice Behavior and Attitude," Journal of Marketing Research, 16 (February), 111-123.
- Myers, J.H. and M.I. Alpert (1968), "Determinant Buying Attitudes: Meaning and Measurement," Journal of Marketing, 32 (October) 13-20.
- Payne, John W. (1976), "Heuristic Search Processes in Decision Making," in Beverlee B. Anderson, ed., Advances in Consumer Research, Vol. 3, Chicago: Association for Consumer Research, 321-27.
- Pekelman, Dov and Subrata Sen. (1976), "Stochastic Weights in Multi-Attribute Decision Making," in Beverlee B. Anderson (ed.) Advances in Consumer Research, Vol. 3, Chicago: Association for Consumer Research, 81-87.
- Pessemier, E.A. (1975), "Marketing Structure Analysis of New Products and New Opportunities," Journal of Contemporary Business, (Spring) 35-67.
- Rokeach, M. and P. Kliejunas. (1972), "Behavior as a Function of Attitude-Toward-Object and Attitude-Toward-Situation," Journal of Personality and Social Psychology, 22, 194-201.
- Rosch, Eleanor. (1978), "Principles of Categorization," chapter 2 in Eleanor Rosch and Barbara J. Lloyd (Eds.), Cognition and Categorization, Hillsdale, NJ: Lawrence Erlbaum Association, 27-48.
- Rotter, J.B. (1966), "Generalized Expectancies for Internal vs. External Control of Reinforcement." Psychological Monographs.
- Russell, J.A. and A. Mehrabian. (1976), "Environmental Variables and Consumer Behavior," Journal of Consumer Research, 3 (June), 62-63.

Shocker, A.D. and V. Srinivasan. (1974), "A Consumer-Based Methodology for the Identification of New Product Ideas," Management Science, 20 (February), 921-38.

Shocker, A.D. (1979), "Multi-Attribute Applications for Product Concept Evaluation and Generation: A Critical Review," Journal of Marketing Research, 15 (May), 159-80.

Slovic, Paul. (1972), "From Shakespeare to Simon: Speculations--and Some Evidence--About Man's Ability to Process Information," Research Monographs, 12 #5, Eugene, OR: Oregon Research Institute, University of Oregon.

Srivastava, Rajendra K. (1979), Situational Influence on Competitive Market Structure: Implications for Marketing Planning, Doctoral Dissertation, Graduate School of Business, University of Pittsburgh, (July).

Srivastava, Rajendra K. (1980), "Usage-Situational Influences on Perceptions of Product Markets: Response Homogeneity and Its Implications for Consumer Research," in Jerry Olson, (ed.), Advances in Consumer Research, Vol. 7, Chicago: Association for Consumer Research, 644-9.

Srivastava, R.K. and A.D. Shocker. (1979), "The Reliability and Validity of Usage-Situational Taxonomies: Implications for Marketing Research," Working Paper, Graduate School of Business, University of Pittsburgh, (November).

Srivastava, Rajendra K., Allan D. Shocker, and George S. Day. (1978), "An Exploratory Study of the Influences of Usage-Situation on Perceptions of Product-Markets," in H.Keith Hunt (ed.), Advances in Consumer Research, Vol. 5, Chicago: Association for Consumer Research, 32-7.

Stefflre, V.J. (1971), New Products and New Enterprises: A Report of an Experiment in Applied Social Science, Irvine, CA: University of California, (March).

Wilkie, W.L. and E.A. Pessemier (1973), "Issues in Marketing's Use of Multi-Attribute Attitude Models," Journal of Marketing Research, 10 (November), 428-441.

Wind, Yoram. (1977), "Toward a Change in the Focus of Marketing Analysis: From a Single Brand to an Assortment," Journal of Marketing, 41 (October), 12 and 143.

Wright, Peter and Frederick Barbour. (1977), "Phased Decision Strategies: Sequels to Initial Screening," in M.K. Starr and M. Zeleny, eds., Multiple Criteria Decision Making: TIMS Studies in the Management Sciences, Vol. 6, Amsterdam: North Holland Publishing Co., 91-109. (available as a special issue of Management Science).

CONTEXT EFFECTS IN PRODUCT PERCEPTION¹

Michael E. Johnson, The University of Chicago

Abstract

An experiment demonstrates the existence of context effects in consumer perceptions of product similarity. Such effects violate the measurement assumptions of familiar models of product perception such as multidimensional scaling. The observed violations are explained by a theory that represents products as sets of qualitative features.

Introduction

Through years of advertising and use occasions, Budweiser beer has come to be associated with such distinctive features as "Clydesdales", "beechwood aged", and so on. When marketers ask consumers to rate the similarities among beers, their dimensional models of similarity assume that such nondimensional features have no influence. The resulting dimensional representation or "perceptual space" of the brands is assumed to reflect only the differences along uniformly relevant dimensions, such as a beer's "lightness".

The goal of this study is to show that product perceptions are affected by unique features as well as by shared dimensions. Such a demonstration contradicts the measurement assumptions of the familiar models of product perception (multidimensional scaling, multidiscriminant analysis, and conjoint measurement). More specifically, my aim is to show that the salience of different features (and dimensions) will vary with context. Where context can be controlled, such features should influence product comparisons in predictable directions.

After outlining a psychological theory that describes similarity judgments as the result of a feature matching process, the generality of the theory is tested in a consumer products context. The results support such a feature based approach to product perception and attest to the sensitivity of perception to changes in context.

Similarity Judgment as a Feature Matching Process

A dimensionally based technique such as multidimensional scaling (MDS) presumes a full analogy between the cognitive concepts of similarity and dissimilarity on the one hand, and the (Euclidean) geometry of spatial proximity and distance on the other (Cunningham and Shepard 1974). Specifically, MDS postulates that perceived similarity among objects is a monotonically decreasing function of the distances between those objects represented as points in n-dimensional (metric) space. Therefore, such a method is based on both dimensional and metric assumptions.

However, Tversky (1977) has demonstrated that these assumptions may be inappropriate in describing people's perception of similarity. The metric assumptions require that similarity be symmetric. That is, the similarity of a, the subject, to b, the referent, is necessarily identical to the similarity of b to a. The metric models also assume that perceived similarities among objects should be perfectly negatively correlated with the corresponding perceived dissimilarities. In many cases, however, neither of these assumptions holds. For example, Tversky (1977) found subjects' ratings of the similarity of North Korea to Red China to be greater than the similarity of Red China to North Korea. He also showed that when two stimuli in a group appeared to have both more common and more distinctive features than any other pair of stimuli in the group, these two were often judged to be both the most similar and the most dissimilar pair in the group. For instance, the USSR and the United States received both the highest similarity and the highest dissimilarity ratings among the countries judged by Tversky's subjects. Such findings lead one to question the general applicability of both the dimensional and metric assumptions of MDS. Indeed, dimensional interpretations may be inappropriate for a large class of product categories and, if applied, will produce misleading product spaces.

Tversky offers an alternative description of how people make similarity judgments. Judging similarity is a feature matching process.

It seems more appropriate to represent faces, countries, or personalities in terms of many qualitative features than in terms of a few quantitative dimensions. The assessment of similarity between such stimuli, therefore, may be better described as a comparison of features rather than as the computation of metric distance between points (Tversky 1977, p.328).

When faced with a similarity task, people extract and compile from remembered information a limited list of relevant features on the basis of which they perform the required task. As formally stated by Tversky, the similarity between two objects, $s(a,b)$, where a and b are associated with feature sets A and B respectively, is:

$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

The similarity between two objects is a function of their common features ($A \cap B$), features of a but not of b ($A - B$), and features of b but not of a ($B - A$). Similarity can be a function of just common or just distinctive features or both depending on the values of the parameters θ , α , and β .

Because format can influence the value of these parameters, Tversky's theory has the power to account for diverse empirical observations. For instance, the North Korea/Red China asymmetry can be explained by a small value of β (in conjunction with Red China's having more distinctive features than North Korea). In other words, when one stimulus is the subject and the other the referent, the distinctive features of the referent do not

¹
The author thanks J. Edward Russo and the Consumer Behavior Laboratory of the Graduate School of Business of the University of Chicago for their assistance in conducting the experiments and preparing the paper. Financial assistance was provided by the Red and White Foundation and the Center for Research in Marketing at the University of Chicago.

receive as much weight in the overall similarity judgment as the distinctive features of the subject. When, in addition, one stimulus has more distinctive features than the other, an asymmetry occurs. Thus, when North Korea is the subject its features map quite well into those of Red China. But when Red China is the subject its features will not map as well into those of North Korea because of the greater number of distinctive features associated with Red China. Similarly, the difference between similarity and dissimilarity judgments can be accounted for by a greater emphasis on common features in the former case (let $\theta >> \alpha = \beta$ in Equation 1) and a greater emphasis on distinctive features in the latter case (let $\alpha = \beta >> \theta$).

The power of the feature model to explain context effects lies in the weights given to the different sets of features. By controlling the context or format within which judgments are elicited, one should be able to strategically manipulate the relative size of these parameters and, as a result, influence product perception.

It should be noted that the empirical violations of MDS that Tversky reported may be explained by other than his feature based set theoretic model. Attempts have been made to reconcile these violations with both metric (Cooper 1979, 1980) and antimetric (Krumhansl 1978) distance functions. The important issues here, however, are the generality of Tversky's findings in a consumer products context and the ability to predict such contextual effects. As long as Tversky's model is a good predictor of contextual effects it will be of obvious usefulness to marketing researchers.

Testing The Theory Using Consumer Products

Method

Experiment 1 tested the asymmetry hypothesis for consumer products. Five product categories were used. Some of these categories were selected because they contain several homogeneous brands with widely varying market shares, such as soft drinks (colas and non-colas) and beers. The others represent product names within a larger, heterogeneous product category where imitation or substitute products must compete with established or prototype products. These products include frozen desserts (frozen yogurt versus ice cream), appliances (toaster versus waffle iron), and fruits (orange versus cangerine).

Assuming that high share or established products are associated with more distinctive features than their counterparts, consumer ratings of similarity in a subject/referent format should be asymmetric. For example, the similarity of Shasta Cola to Coke should be larger than the similarity of Coke to Shasta Cola. Given the nature of these product categories, the predictions that follow are made relative to the specific stimulus pair involved. For example, Coke is the market leader for colas followed by Pepsi and R.C. with brands such as Shasta and Canfield (a brand local to the Chicago area) bringing up the rear. If the features sets associated with the various products roughly correspond to their relative market shares, Tversky's theory predicts that asymmetry should hold, for example, in such cases as R.C. versus Coke and Shasta versus R.C..

Experiment 2 tested the relationship between similarity and dissimilarity judgments among soft drinks (colas and non-colas), beers, and fruits. These categories were chosen because they represent cases where there exist more than one low share or lesser known product and more than one relatively high share or established product.

Assuming that high share products are associated with both more distinctive and more common features, then both the similarity and the dissimilarity ratings among them should be enhanced. This, in turn, should reduce the negative correlation between similarity and dissimilarity. For example, Coke and Pepsi should be both relatively more similar and more dissimilar than other pairs of colas.

Subjects

Participants included students from the University of Wisconsin and residents of the Madison area recruited from ads in the city newspaper and posters on campus. All of the assumptions concerning market shares and associated features were relative to this target market. For example, because Old Style beer is a market leader in this area, it is assumed to be associated with relatively more distinctive features.

Procedure

Stimulus pairs within each product category were randomized and divided into two stimulus sets, (A,B) and (C,D). Subjects were divided into four treatment groups. Treatment group I received subject/referent questions of the form "How similar is A to B?". Group II received questions of the form "How similar is C to D?". Group I was also asked to judge the similarity of pairs of products from stimulus set (C,D) with no subject/referent distinction while group II did likewise for stimulus set (A,B). Group III received questions of the form "How similar is B to A?" and then was asked to rate dissimilarities among stimulus set (C,D). Finally, group IV was given questions of the form "How similar is D to C?" and then was asked to rate dissimilarities among stimulus set (A,E). These four treatment groups were used to avoid having subjects rate subject/referent similarities on the same stimulus pairs they faced in the similarity versus dissimilarity condition.

All of the judgments were elicited in a questionnaire format. Responses were made by placing an X on 125mm rating scales below each question. The scale ranged from "Not At All Similar" to "Very Similar" or from "Not At All Dissimilar" to "Very Dissimilar". After obtaining the judgments, memory probes were obtained on all of the products used in the study. The purpose of the probes was to check on the validity of the assumptions made concerning associated features.

Results

Subject-Referent Asymmetry

The results support the generality of Tversky's criticisms in a subset of the product categories. Significant asymmetry was found to hold for both types of soft drinks and for beers. For example, the similarity of Shasta Lemon Lime Soda to Seven Up was significantly greater than the similarity of Seven Up to Shasta ($t=2.15, P<.05$). Both Pepsi and R.C. were found to be more similar to Coke than vice versa ($t=1.42$ and $t=1.40$ respectively, $P<.10$). Meanwhile, Shasta Cola was more similar to both R.C. and Pepsi than R.C. and Pepsi were to Shasta ($t=2.22$ and $t=1.72$ respectively, $P<.05$). A sample result for beers finds both Huber and Leinenkugel (two local Wisconsin beers) being more similar to Old Style than vice versa ($t=1.95, P<.05$, and $t=1.55, P<.10$, respectively).

A summary of the results for experiment 1 is presented in Table 1. Column 2 presents the average t value of stimulus pairs for which asymmetry was predicted in each

product category. Columns 3, 4, 5, and 6 present the number of stimulus pairs involved, the standard deviation on the t's, a summary t value for each product category, and the significance of each summary value respectively.

TABLE 1
Average Asymmetry Within Product Categories

Product Category	Mean t	n	S.D.	t	significance
Colas	1.39	8	0.54	7.23	sig. α = .005
Non-Colas	1.46	7	0.58	6.64	sig. α = .005
Beers	0.72	8	0.90	2.27	sig. α = .05
Fruit	0.09	8	0.93	0.28	not sig.
Appliances	-0.74	3	0.84	-1.53	not sig.
Desserts	0.97	3	1.25	1.34	not sig.

Similarity versus Dissimilarity

The results of experiment 2 are remarkably consistent with the results of experiment 1. In those cases where asymmetry was significant, a relatively poor relationship was shown to hold between judgments of similarity and judgments of dissimilarity. Likewise, where subject/referent judgments were symmetric, strong negative relationships existed between similarities and dissimilarities. Table 2 presents the least squared regression results which compare the similarity judgments to the dissimilarity judgments for each of the product categories. Columns 2, 3, 4, and 5 present the regression constants, beta coefficients, r2 statistic, and r2 adjusted for the number of independent variables, respectively.

TABLE 2
Similarity vs Dissimilarity Regression Results

Product Category	α	β	r^2	r_a^2
Colas	110.13	-.46	.179	.076
Non-Colas	135.57	-1.15	.585	.499
Beers	131.46	-1.13	.626	.579
Fruit	127.79	-1.06	.849	.830

The cola and beer categories provide the most interesting results. The adjusted r2 for colas is a dismal .076. This small value is due for the most part to the fact that, as predicted, Pepsi and coke were relatively more similar and more dissimilar than any other pairs of colas (see Figure 1). Non-colas do moderately well, obtaining a fit of .585 (see Figure 2). The market leaders in beer, Old Style and Budweiser, were also relatively more similar and more dissimilar than other pairs of beers (see figure 3). The best fit of all occurs for fruits (see Figure 4). Such results are consistent with those in experiment 1 where fruits failed to demonstrate asymmetry.

A preliminary analysis of the thought listings supports the assumptions made concerning associated features. A detailed analysis of the nature of the associations is, however, pending.

FIGURE 1
Colas

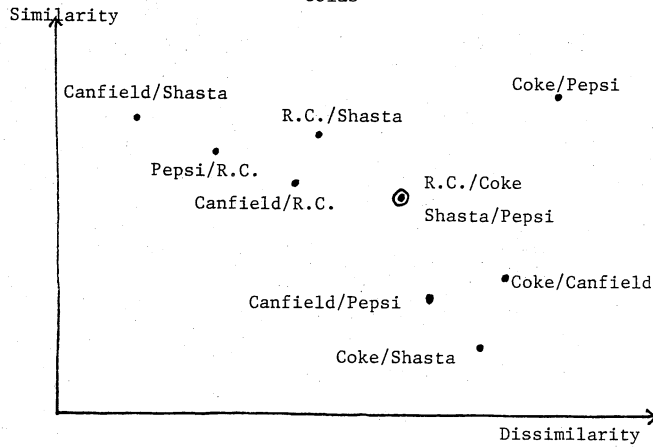


FIGURE 2
Non-Colas

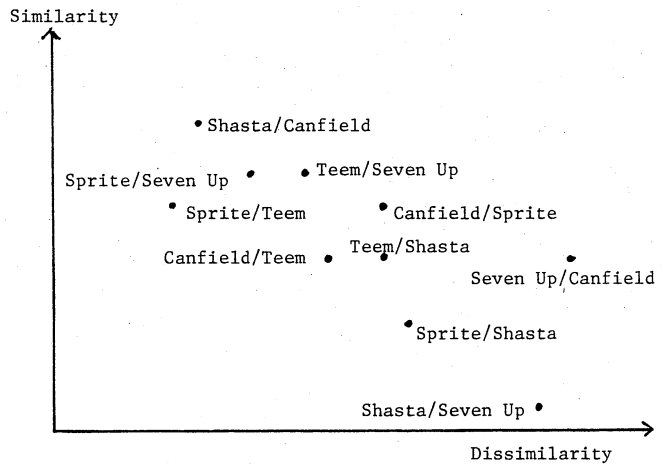


FIGURE 3
Beers

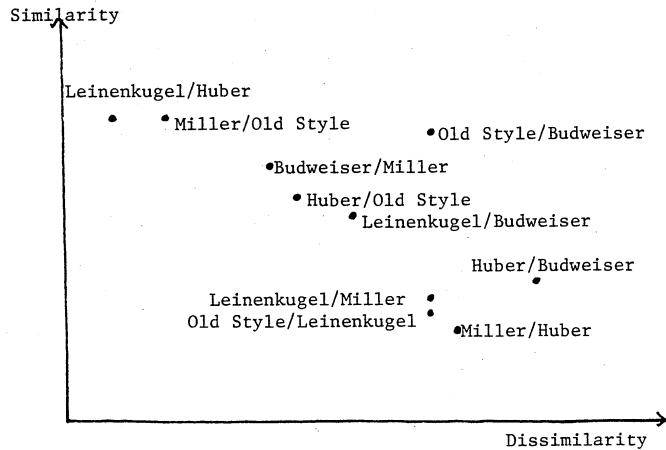
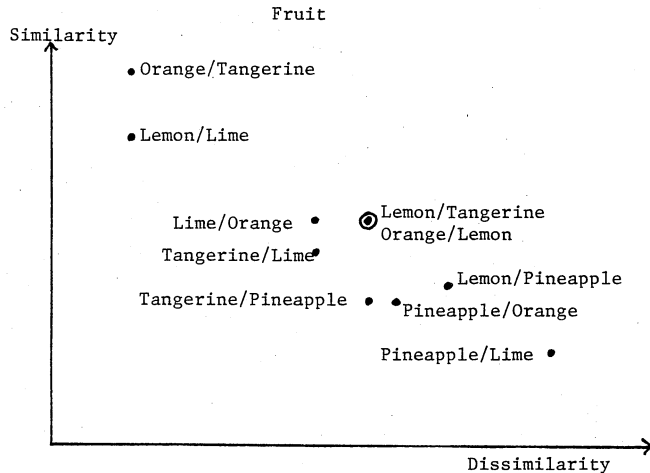


FIGURE 4



Summary and Conclusions

The results of this study support the generality of Tversky's theory in a consumer products context. Contrary to the dimensional assumptions of existing market segmentation techniques, context has a significant and predictable influence on judgments of similarity among products within certain product categories.

Judgments of similarity were found to be poorly related to judgments of dissimilarity for certain products. For example, Coke and Pepsi were found to be both more similar and more dissimilar than other pairs of colas. Therefore, if one builds a product space for colas using similarity judgments, Coke and Pepsi will wind up being very close in proximity. But if one uses dissimilarity judgments as inputs to the construction of such a space, Coke and Pepsi will be relatively distant. Obviously this will have important implications with respect to the interpretation of the market and its use in the development of marketing strategies.

Whether a product is used as a subject or referent in a comparison may also influence the degree of similarity among certain products. If a product is associated with more distinctive features than other products in the same market, such features will detract more from judgments of similarity when that product is the subject rather than the referent in the comparisons. For example, because Coca-Cola is associated with so many distinctive features, a low share cola attempting to position itself close to Coke must be careful not to use Coke as a subject in any direct comparisons of the two products.

The consistency of the results in experiment 1 with those in experiment 2 provide convergent validity for the results as a whole and for Tversky's theory. In many cases it is simply more appropriate to view a product as being associated with sets of features than as varying on a small number of shared dimensions. As to why the homogeneous product categories were the only ones prone to such effects is the subject of the author's current research.

References

Cooper, Lee G. (1979), The Reference Space Underlying Judgments of Similarity and Preference and Multiattribute Measures on Alternatives, working paper, Graduate School of Management, University of California, Los Angeles.

Cooper, Lee G. (1980), Discovering Homogeneous Groups from Comparative Judgments, working paper, Graduate School of Management, University of California, Los Angeles.

Cunningham, James P., and Shepard, Rodger N. (1974), "Monotone Mapping of Similarities into a General Metric Space," Journal of Mathematical Psychology, 11, 335-363.

Krumhansl, C. L. (1978), "Concerning the applicability of geometric models to similarity and spatial density," Psychological Review, 85, 445-463.

Tversky, A. (1977), "Features of Similarity," Psychological Review, 84, 327-352.

A MEANS-END MODEL FOR FACILITATING
ANALYSES OF PRODUCT MARKETS BASED ON CONSUMER JUDGEMENT

Jonathan Gutman, University of Southern California

Abstract

Day, Shocker, and Srivastava (1979) have discussed a number of customer-oriented approaches for identifying product markets. This paper presents a means-end chain model for facilitating analyses of product markets using consumer judgements. The model is based on categorization processes which consumers use to create sets of substitutable products.

Introduction

Day, Shocker, and Srivastava (1979) have discussed a number of customer-oriented approaches for identifying product markets. The approaches include decision sequence analysis, perceptual mapping, and several variations for obtaining direct consumer judgements of substitutability (free response data, direct grouping into categories, products by uses analysis, and substitution-in-use analysis).

Although these approaches differ, they share a number of related aspects. These aspects can all be embodied in a model of consumer categorization processes that relates "ends" in terms of consumer values to "means," which often are the products consumers consume. Such a model can provide a basis for increasing our understanding of how consumers cognitively define competitive sets of products.

This paper will briefly outline and point out the critical aspect each of these approaches. Then the model will be presented. Lastly, some examples of relevant research will be presented that suggest how the model facilitates research on product markets.

Approaches for Obtaining Consumer
Judgement of Substitutability

Decision Sequence Analysis

Bettman (1971) and Haines (1974) have discussed the use of protocols of consumer decision making which indicate the sequence in which various criteria are employed in reaching a final choice. The data are typically gathered by asking shoppers to verbalize their train of thought as they make actual shopping decisions. The critical aspect of this approach is the specification of attributes used in the choice process and the sequence in which they come into play.

Perceptual Mapping

Perceptual maps are created by a variety of techniques that create geometric representations of consumer's perceptions of product markets. Such maps require that the relevant product market be already defined before the start of the study. However, it is possible to create maps for "different levels of product competition to explore competitive relations at the level of product types, variants, or brands" (Day, Shocker, and Srivastava 1979 p.14).

Consumer Judgements of Substitutability

The Free Response approach (Green, Wind, and Jain 1973) allows the respondent to decide how similar brands must be

before they can be considered as substitutes. Consumers are presented with target brands and asked to free associate the names of similar or substitute brands. Frequency and order of mention are used as indications of similarity.

Direct grouping into categories (see Gutman 1979, Gutman & Reynolds 1979, and Bourgeois, Haines, and Sommers 1979) involves asking consumers to directly assign products to categories based on their judgments as to their degree of substitutability. Products-by-uses analysis (Stefflre 1979, Myers & Tauber 1977) involves ascertaining which usage situations are similar by virtue of their evoking the same benefits. Competitive products are determined by their being appropriate in equivalent usage situations. Substitutability-in-use analysis (Srivastava, Shocker, and Day 1977) is a refinement of the above approach brought about by a three-stage approach to refine the set of usage situations to be used.

Each of the above approaches requires that we know something different about consumers' cognitive structure. Decision sequence analysis requires that we understand the hierarchical ordering of elements in consumers' cognitive structure. The use of perceptual mapping requires that we be capable of representing the levels of inclusiveness at which products may be represented within a consumer's cognitive structure. The free-response approach emphasizes the individual nature of consumers' categorical systems. The direct grouping approaches emphasize the need to understand equivalence ranges (when are products substitutable or how similar do they have to be to be substitutable). And the products-by-uses analysis suggests that all of the above aspects vary by circumstance or situation.

Development of a Means-End Chain Model

It is possible to develop a model of consumers' cognitive structure that will provide for representation of all of these aspects. Such a model offers a theoretical and conceptual structure for connecting consumers' values to their behavior (Rokeach 1968, Howard 1977, Vinson, Scott, and Lamont 1977). Models of this type can be subsumed under the rubric of means-end chains. Means are objects (viz., products) or activities that people engage in (viz., running, reading, etc.). Ends are valued states of being, such as happiness, security, accomplishment. A means-end chain is defined as consisting of an interconnected set of cognitive elements that allows a person to select objects or activities that enable him to achieve his desired end states.

Thus, the means-end chain permits us to focus on the basic aims consumers have in life while not losing sight of how these aims influence choices in specific situations. The process of categorization is proposed as the mechanism by which consumers organize thinking about specific product alternatives so as to create arrays of products which will be instrumental in helping them achieve their values. If this connection can be made, the bases consumers have for creating sets of substitutable products will be understood more thoroughly.

Categorization processes, then, represent ways people segment their environments into meaningful groups by creating equivalences among nonidentical stimuli. Thus, for example, the categories of paper towels, toothpaste,

or coffee have meaning to people because the objects in these categories all have something in common.

This process of categorization takes place at many levels of inclusiveness, which refers to the degree of similarity among products in a category. Brands of paper towels grouped together would represent a category of low inclusiveness. If consumers distinguish between plain paper towels and those with a pattern, two product categories of lower inclusiveness would be created. If paper towels were grouped with paper napkins, cups, and tissues, a more inclusive category of "paper products" would be created.

Categorization as a Basis for Means-End Chain

Groups or categories of products (product classes, for example) have to be related systematically to the higher-level ends if the "chain" is to serve its instrumental purpose of enabling the person to achieve his or her values. This means that values have to be translated from their context at the more inclusive levels of the chain to the less inclusive levels where products are categorized into product classes, thus establishing sets of products which will be competitive (alternative satisfiers of the same ends).

The basis of this clustering is mutually determined by the object properties and value-determined criteria for goal instrumentality. This categorization process takes place at each level of the means-end chain as categories of greater inclusiveness are formed at higher and higher-levels of the means-end chain. For example, some consumers may treat the entire category of soft drinks as a single category of sugary substances that ought not to be consumed, while others regard them as tasty and delicious. Others may make a distinction between fruit drinks and carbonated drinks. Still others may make a distinction between drinks that are "fun to drink" (soft drinks) and those that are "good for you." Most consumers will be in agreement as to the identity of products--what is a soft drink and what isn't. Differences in consumer categorization are less at low levels of inclusiveness (where assignment to categories is based on product attributes) than at high levels of inclusiveness (where categories are based on function or value instrumentality). Consumers group products in different categories, depending on which features they emphasize and which features they ignore.

These categories are created by the application of a cognitive mechanism called a "distinction." Distinctions are dichotomies that represent the end points of dimensions along which products may be compared. If two products can be compared using a hot/cold distinction, this indicates that they may be compared by temperature, although they may differ with respect to temperature. An array of products can be created by applying this distinction to any of a limitless number of products. However, many products may not be comparable on this distinction. If a consumer doesn't think of a chain saw, a dry cleaner, or a box of paperclips as being either hot or cold, then none of these objects would be part of the array. Other products, such as orange juice, coffee, tea, milk, water, corn flakes, oatmeal, etc., could be compared on this basis.

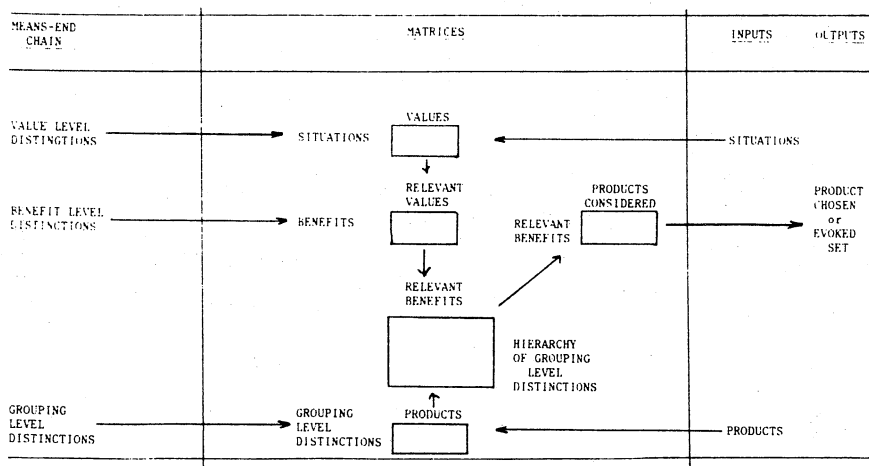
The resulting array can contain as many categories as products or just one category (if all products were thought to be the same temperature). The essential point is that the distinction being dichotomous has nothing to do with the nature of the array of products created by its application. A distinction is not an array of products--it is a dichotomous cognitive element in which the two parts of the dichotomy have some meaningful relation to one another. What is meaningful is a personal decision for each consumer to make.

Description of Means-End Chain Model

Two major events occur in the development of a means-end chain. One is the crystallization of values to the point where they become capable of articulation into goals or valued states. This event occurs at the top of the chain. At the bottom of the chain, individual products have to be arrayed into product classes to limit the complexity of the choice situation to acceptable levels. Products have to be given meaning in terms of the internal coding system established by the individual for making sense of the world (Olson and Muderrisoglu 1978).

These two issues define the basis for the model shown in Figure 1. The figure is divided into three parts: the means-ends chain, matrices, and inputs and outputs. The figure shows that the means-end chain is comprised of three levels of distinctions--values, benefits, and grouping. The focus of distinctions at the values and benefits

FIGURE 1
Means-End Chain Model



levels is on what products can do for the consumer. Distinctions at the grouping level focus on properties of the product or evaluations of properties of products. Inputs to the system are products and situations; outputs are products chosen for final consideration. Matrices are intersections between inputs and means-end chain distinctions or between outputs from such matrices.

Situations encountered by consumers are assessed in terms of values to be sought in each. This occurs in the situations-by-values matrix and results in a set of relevant values emerging that will guide behavior in a given situation. These relevant values are arrayed in a matrix with potential benefits. Analysis of the instrumentality of the benefits for achieving relevant values results in a set of relevant benefits which guide the choice of product alternatives.

At the bottom of the figure, grouping level distinctions form a hierarchy wherein the identity and evaluation of product alternatives are determined at different levels of inclusiveness.

At the lowest levels of inclusiveness, products are grouped together in accordance with Howard's notion of semantic structure or Lunn's (1972) organization of ways of satisfying consumer needs by product types, product variants, and brands. Product types satisfy different needs beyond generic needs (tooth paste vs. tooth power; hot vs. cold cereal). Product variants are variations within types (with or without fluoride, presweetened vs. unsweetened). And brands represent minor differences within variants. Thus, brands as specific alternatives are at the bottom of the hierarchy; those are subsumed by product variants, and product variants are grouped by product type.

At the upper levels of the hierarchy, evaluative categories subsume identity categories. Thus, for cereal, nutritious vs. unnutritious or tastes good vs. doesn't taste good might subsume distinctions based more on physical properties of cereals. The distinctions at the top of the grouping hierarchy are implied by, or predicted from, the distinctions at the bottom of the hierarchy. Relevant benefit distinctions for the situation are applied to the hierarchy of grouping distinctions. This determines the level in the hierarchy at which a set of products would be selected for further processing--perhaps the level at which the relevant product market would be defined for that consumer. Products at this level of grouping would themselves be compared in the "relevant benefit distinctions by products" matrix. This comparison would result in a product to be chosen for that situation or a set of products from which a choice would have to be made on a nonvalues-related basis.

Values, Situations, and Value-Situation Matrices

Consumers obviously encounter many potential product-use situations. There are many ways of defining and describing such situational variables (Belk 1975). But for the level of analysis referred to in this context we will follow Fennell (1978), who defines a situational unit as corresponding "to the activities and conditions for which products are created and marketed, such as doing the laundry, feeding the dog, having a headache."

Consumers have valued ends they are trying to achieve. And consumption situations provide them with an opportunity to achieve these values. Each consumer learns over time which choices in a given situation are instrumental to achieving values and which are not. Thus, consumers learn to perceive certain situations as opportunities for achieving certain of their values.

The situations by values matrix is a means by which the accumulated experience of the consumer can be represented. It illustrates the process by which a particular set of values comes to guide the selection of benefits sought in a given consumption situation. All of a consumer's values can't be achieved in one situation, and, frequently, achieving some values means not achieving other values. This point of view is consistent with the conclusion that situational influence is a pervasive factor in consumer behavior (Belk 1975, p. 161).

Grouping-Level Distinctions

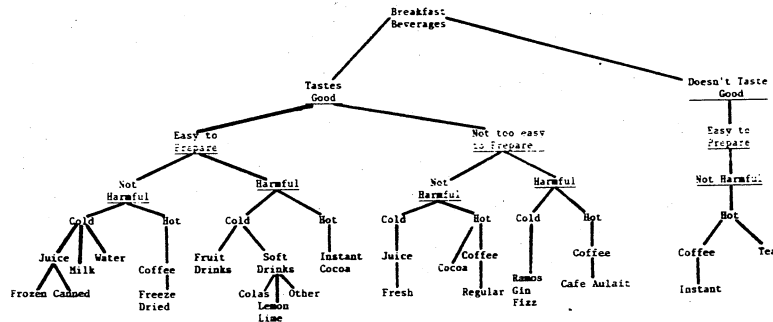
The grouping level is the level at which meaning is ascribed to products by grouping them together with other similar products and giving them a group label. Physical and surface properties of products are the bases for grouping products at the grouping level. As the most immediately perceived aspects of products, they serve to identify them in their most fundamental sense. Cars are to drive; refrigerators refrigerate; stoves cook; peas, green beans, carrots, etc., are all vegetables. It is possible to discriminate products more finely within each of these categories. There are many kinds of vegetables (root crops, green leafy, squash, etc.), many ways of classifying cars, and many types of refrigerators.

Finer and finer discriminations within categories and broader and broader discriminations of subsuming categories (cars are a means of locomotion; vegetables are food; refrigerators are appliances) create product hierarchies. The number of levels in the hierarchy is a function of the knowledge of the consumer and his or her interest in maintaining a cognitive structure of varying amounts of complexity. Someone who doesn't like coffee would have a much flatter hierarchy of types of coffee than a coffee connoisseur would. Obviously, then, product markets would be defined differently for different types of consumers.

Distinctions at this level often have the form of logical negation, which means that one side of the dichotomy is "not" (" \sim " in Boolean algebra) what the other side "is." Soft drink vs. not a soft drink in effect creates one group of products--the other "group" representing all other objects not in the first group (e.g., fruit juice, water, elephants, etc.). Obviously, such a distinction could be applied to all objects, whereas "hot vs. cold" implies that objects be comparable along a temperature dimension. Distinctions in the form of logical negation create the simple identity groupings that are characteristic of the lower levels of the grouping hierarchy.

At low levels in the grouping hierarchy, distinctions serve to identify products; at higher levels within the grouping hierarchy, evaluative distinctions subsume those identity distinctions. At the lower levels, where identity is the focus, level of inclusiveness accounts for the level of a distinction within the hierarchy. Thus, in Figure 2, which shows one person's grouping hierarchy for breakfast beverages, soft drink subsumes colas, lemon-lime, and other flavors. At higher levels in the grouping hierarchy, the level of a distinction is determined by its ability to separate products into preferred and nonpreferred sets (see Tversky's (1972) theory of sequential choice). If "not easy to prepare" meant a product was not acceptable for use regardless of any other properties, it would be more functional to have it subsume "good for you"--"not good for you." However, if a consumer would be more likely not to consider a product if it were "not good for you," even if it were easier to prepare than "good for you," "not good for you" would be at a higher level in the hierarchy. The goal of categorizing products at the grouping level is to identify products and array them into

FIGURE 2
Grouping Hierarchy for Breakfast Beverages



sets based on their possession of attributes that imply the ability to provide desired benefits for the consumer.

Discussion

The model presented above includes all the relevant aspects of the measures of consumer judgement that are used to measure product markets. It allows for the situational selection of relevant values that in turn select appropriate benefits to be sought in the consumption situation. Benefits, in turn, determine the level in the hierarchical structure at which products will be aggregated for further consideration. Thus, there are families of product markets at different levels of inclusiveness, each level of which is appropriate in different sets of circumstances. The criterion of substitutability of alternatives for defining product markets is consistent with the notion that values guide behavior and determine relevant objects for consideration in those situations.

The model also suggests that categorizing products at the values level leads to the development of a wider, more complete definition of relevant product markets than starting with the products themselves and trying to establish categories. Several studies the author has done bear on these issues. The first study (Gutman, 1978) involved eliciting distinctions from respondents and having them sort products into categories based on these distinctions. Twenty triads (sets of three cereals in this instance) were presented to each respondent, who was asked to indicate a way in which all three products could be compared. Then the respondent was to specify how two products are the same and different from the third. This allowed us to define the two poles or ends of the dimension defined by the distinction being made. Respondents were then asked to sort 23 brands of cereal into categories on the basis of their distinctions. They could use as many or few categories as desired.

Another group of respondents was asked to sort the cereals into categories without having been asked to elicit any distinctions they might have. After sorting the cereals, these respondents were asked to label the categories they had created.

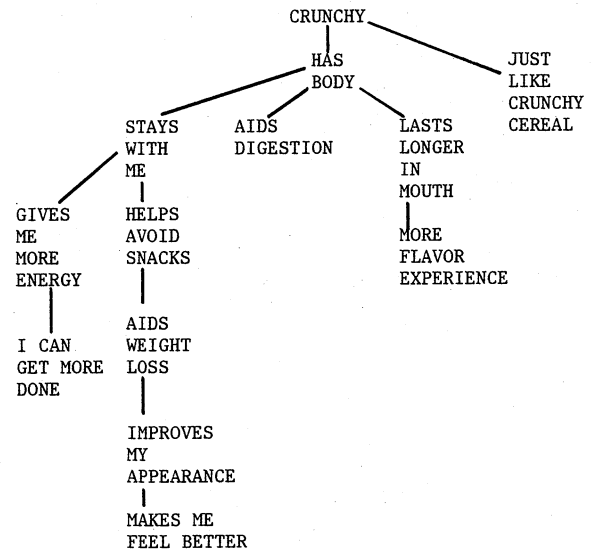
The results indicated that it is difficult to get many distinctions from people. People don't have too many ways of thinking about products (at least for breakfast cereals which may not be too complex). The free sortings were different from the sortings based on the distinctions in that the former yielded sets of nominal categories rather than dimensional arrays. People tended to see products as

organized wholes and didn't look for ways to link dissimilar products to each other.

This might suggest narrow product markets definitions where only highly similar products are considered as substitutes. If consumers are concrete and incapable of thinking abstractly about products, different types of products are not likely to be seen or thought of as alternatives.

It is possible to get consumers to operate at higher levels of inclusiveness, but they have to be "encouraged" to do so. Respondents can be asked to state their preferred poles for each distinction they elicit. Then they are asked why they prefer that pole (or that end of the dimension). The answer to the "why" question becomes the basis for asking another "why" question, and so on until the respondent can answer no further "why" questions. The results of such a procedure (referred to as laddering--see Bannister, (1970) are shown in Figure 3. The figure can

FIGURE 3
Higher Order Implication Chain
Extending from "Crunchy"



be interpreted as providing a chain of implication between the distinctions at various levels in the hierarchy from means to ends.

Another study (Gutman and Reynolds 1979) focused on getting respondents to categorize products (drinks in this instance) at higher levels of inclusiveness. Here again, laddering was used to get distinctions at higher and higher levels of inclusiveness. A content analysis of distinctions at each level showed initial distinctions referred to the contents of the drinks or related to their immediately observable properties. Distinctions at higher levels referred to functions (what the drinks did to/for the person) or end states of being achieved through consumption of the drinks. An example of one respondent's hierarchy of responses to the laddering process is shown in Table 1.

TABLE 1

Example of Laddering Derived from Repertory Grid Procedure Using the Triad: Soft Drink, Fruit Drink, Fruit Juice

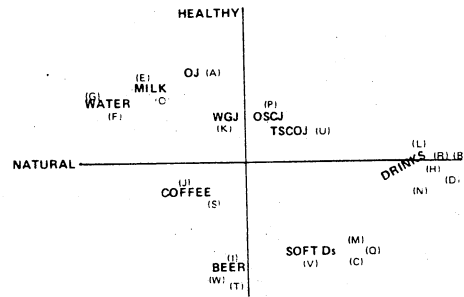
Level	Distinction
1	Fresh vs Processed
2	More Vitamins vs Less Vitamins
3	Lower Sugar vs Higher Sugar
4	Low Calories vs High Calories
5	Good for Weight Control vs Not Good for Weight Control
6	Smaller Size Clothes vs Larger Size Clothes
7	Husband at Home vs Husband Out

The sorting data consisting of a "persons-by-drinks" matrix, with category membership nominally coded within, served as the basis for generating the distance data for input to POLYCON, a multidimensional scaling program (see Rosenberg and Kim, 1975 for the procedure used). Three maps are shown in Figure 4. They show the product space in two dimensions for sorts based on distinctions at the first, third, and fifth levels (as obtained through the laddering procedure). The changes in groupings of drinks can be readily seen. The juices, fruit drinks, and soft drinks show movement across levels. In effect, these figures represent cognitive maps of product markets at different levels of inclusiveness. One has to get the consumer to consider the benefits derived and values satisfied to develop the proper context for arraying products into meaningful groups. Otherwise, the analogy is similar to that of the problem with SIC codes in that grouping is based on properties other than those related to consumption.

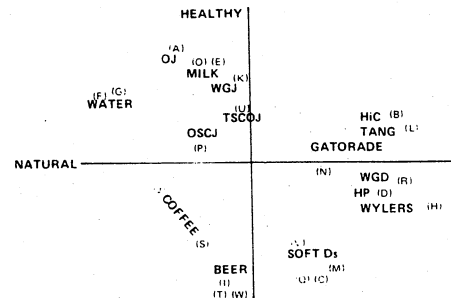
The last study (Gutman 1980) shows some evidence of preferred ranges for equivalence in sorting products into categories. Consumers differ in terms of the limits of what they consider similar. Some people will accept a slight basis of comparability as enough to group two products together. Other people would only do so if the two products were highly similar. This tendency may be related to actual product market limits as well as to cognitive mappings of such product markets. If people are narrow and concrete in terms of what they consider equivalent, then their cognitive view of the world will restrict their breadth of their solutions in terms of searching for means to satisfy their ends.

FIGURE 4

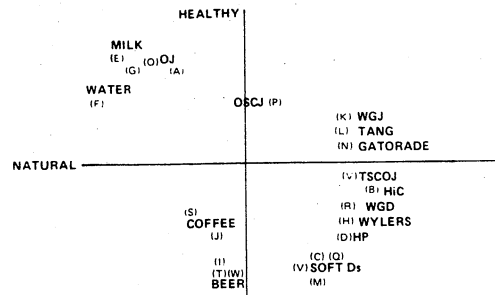
PRODUCT SPACE IN 2-DIMENSIONS BASED ON 1ST-LEVEL DISTINCTIONS



PRODUCT SPACE IN 2-DIMENSIONS BASED ON 3RD-LEVEL DISTINCTIONS



PRODUCT SPACE IN 2-DIMENSIONS BASED ON 5TH-LEVEL DISTINCTIONS



DRINKS USED IN SORTING TASK

- | | |
|--------------------------|-----------------------------------|
| (A) Fresh Orange Juice | (M) 7-Up |
| (B) Hi-C Orange Drink | (N) Gatorade |
| (C) Coca-Cola | (O) Non-Fat Milk |
| (D) Hawaiian Punch | (P) Oceanspray Cranberry Juice |
| (E) Regular Whole Milk | (Q) Dr. Pepper |
| (F) Tap Water | (R) Welch's Grape Drink |
| (G) Bottled Water | (S) Brim Decaffeinated Coffee |
| (H) Wyler's Lemonade | (T) Pabst Beer |
| (I) Lowenbrau Beer | (U) Treenweet Canned Orange Juice |
| (J) Yuban Regular Coffee | (V) Diet Pepsi |
| (K) Welch's Grape Juice | (W) Miller's Lite Beer |
| (L) Tang | |

Conclusion

The paper has presented a means-end chain model for facilitating the analysis of product markets based on consumer judgement. The levels represented in the model-extending from values to benefits to product attributes-provide for an integrated set of product groupings at different levels of inclusiveness. This family of product groupings should permit managers to relate levels of market aggregation to the criteria used by consumers at those levels. Furthermore, operating at the values level can provide the most comprehensive view of the market in terms of learning about competition from outside traditional market boundaries.

The model has the advantage of letting the consumer decide the criteria to be used in grouping products. It may exhibit the linkages between physical aspects of products (which may offer a misleading view of product markets from a consumer perspective), benefits provided by these products, and values consumers are striving for in their consumption of these products.

References

- Bannister, Donald (1970), "Science Through the Looking Glass." In D. Bannister (Editor), Perspectives in Personal Construct Theory, London: Academic Press, 47-62.
- Belk, Russell (1975), "Situational Variables and Consumer Behavior," Journal of Consumer Research, 2, 157-164.
- Bettman, James R. (1971), "The Structure of Consumer Choice Processes," Journal of Marketing Research, 8, 465-471.
- Bourgeois, Jacques, Haines, George H., and Summers, Montrose S. (1979), "Defining An Industry." Paper presented to the TIMS/ORSA Special Interest Conference on Market Measurement and Analysis, Stanford, Ca., March 26.
- Day, George S., Shocker, Allan D., and Srivastava, Rajendra, K. (1979), "Consumer-Oriented Approaches to Identifying Product-Markets," Journal of Marketing, 43, 8-19.
- Fennell, Geraldine (1978), "Consumers Perceptions of the Product-Use Situation," Journal of Marketing, 42, 38-47.
- Green, Paul E., Wind, Yoram, and Jain, Arun K. (1973), "Analyzing Free Response Data in Marketing Research," Journal of Marketing Research, 10, 45-52.
- Gutman, Jonathan (1978), "Uncovering the Distinctions People Make Versus The Use of Multi-Attribute Models: Do a Number of Little Truths Make Wisdom." In Proceedings of the 23rd Annual Conference, New York: Advertising Research Foundation.
- Gutman, Jonathan (1980), "Equivalence Range in Categorization," in Advances in Consumer Research, Vol. VII, Olson, Jerry C. (Editor), Ann Arbor: Association for Consumer Research, 411-416.
- Gutman, Jonathan and Reynolds, Thomas (1979), "An Investigation of the Levels of Cognitive Abstraction Utilized by Consumers," in Research Under the Sun, Eighney, J. (Editor), Chicago: American Marketing Association.
- Haines, George H. (1974), "Process Models of Consumer Decision Marketing," in Buyer/Consumer Information Processing, Hughes, G. D. and Ray, M. L. (Editors), Chapel Hill, NC.: University of North Carolina Press.
- Howard, John A. (1977), Consumer Behavior: Application and Theory, New York: McGraw HM Book Company.
- Myers, James H. and Tauber, Edward (1977), Market Structure Analysis, Chicago: American Marketing Association.
- Rokeach, Milton J. (1968), Beliefs, Attitudes, and Values, San Francisco: Jossey Bass.
- Rosenberg, Seymour and Lin, M. P. (1975), "The Method of Sorting as a Data-Gathering Procedure in Multivariate Research," Multivariate Behavioral Research, 10 (October), 489-502.
- Srivastava, Rajendra, K., Shocker, Allan D., and Day, George S. (1978), "An Exploratory Study of Situational Effects on Product Market Definition," in Advances in Consumer Research, Vol. V, Hunt, H. Keith, (Editor), Ann Arbor: Association for Consumer Research, 32-38.
- Stefflre, Volney (1979), "New Products: Organizational and Technical Problems and Opportunities," in Analytic Approaches to Product and Marketing Planning, Shocker, A. D. (Editor), Cambridge, MA: Marketing Science Institute, April Report 79-104, 415-480.
- Tversky, Amos (1972), "Elimination by Aspects: A Theory of Choice," Psychological Review, 281-299.
- Vinson, Donald E., Scott, Jerome E., and Lamont, L. M. (1977), "The Role of Personal Values in Marketing and Consumer Behavior," Journal of Marketing, 41 (April) 44-50.

TOWARD CONCEPTUALIZING AND MEASURING COGNITIVE STRUCTURES¹

Rajesh Kanwar, The Pennsylvania State University
Jerry C. Olson, The Pennsylvania State University
Laura S. Sims, The Pennsylvania State University

Abstract

Three characteristics of cognitive structure, considered to be the coded representations of information in memory, are identified and defined--dimensionality, abstraction, and articulation. Two measurement procedures, free elicitation and the repertory grid, were hypothesized to provide indicators of these three constructs. Reasonable levels of convergent and discriminant validity were obtained for the proposed construct measures in a study of consumers' cognitive structures regarding nutrition. Suggestions are offered for further research which can more firmly establish the construct validity of these characteristics of cognitive structure.

Introduction

Although much has been written about consumers' cognitive structures for brands and product categories, little direct research has been undertaken. Related work has focused on attitudes, where the major interest concerns belief structures and their relationships to attitude (e.g., Lutz 1975, Olson and Dover 1978). Although some researchers have been concerned with consumers' stored information about brands or products (cf. Woodruff 1972), few have tried to measure memory structures directly. Research by Bettman (cf. 1979), Olson and Mudderisoglu (1979), and Russo and Johnson (1980) are among the few exceptions.

Clearly, however, cognitive structure issues are involved in many research topics (Olson 1978b), although this has not always been explicitly recognized. For instance, information acquisition (cf. Jacoby, Szybillo and Busato-Schach 1977) is essentially concerned with the formation of cognitive structures. Attempts to measure various effects of product familiarity (e.g., Park 1976, Raju and Reilly 1980) can be seen from a cognitive structure perspective (Marks and Olson 1981). Advertising effects research (e.g., Wright 1973) can be interpreted as examining the interaction between advertising and cognitive structure (Calder 1978, Mitchell 1980, Olson 1980, Olson and Dover 1978). Indeed, the increasing interest in the cognitive processes involved in consumers' processing of information, especially as they are influenced by organized structures of knowledge stored in memory, forces us to realize that much (if not all) of consumer research involves the concept of cognitive structures (Bettman 1979, Olson 1978b).

Thus, it seems likely that the concept of cognitive structure--as it refers to the encoded representations of information in memory--will become a central concept in models of consumer behavior. This importance reflects the trend in cognitive psychology where it is now widely recognized that one's acquired knowledge about specific domains has very powerful effects on a variety of cognitive processes and outcomes (see Lachman, Lachman and Butterfield 1979). Many cognitive theorists are actively considering the

content and organization of the knowledge held in cognitive structure and its effects (e.g., Anderson 1976, Kintsch 1974, Anderson, Spiro and Montague 1977, Tulving and Donaldson 1972, Shank and Abelson 1977, among many others). A variety of concepts such as schemas (Norman and Bobrow 1975, Norman 1979), scripts (Shank and Abelson 1977), frames (Minsky 1975), and hierarchical semantic memory structures (Collins and Quillian 1969), to name only a few, have been proposed as models of or metaphors for cognitive structure.

Interestingly, the published literature contains many conceptualizations of cognitive structure, but relatively few attempts at operationalization. Many researchers are theorizing about the form and effects of cognitive structures, but few have tried to measure them directly (see Puff 1979). As one example, the cognitive structure concept of memory schemata or schemas has become popular in social psychology (Taylor and Crocker 1980). Yet to our knowledge, no one has attempted to measure the content and organizational structure of a schema directly, although the research of Markus (1977) comes close. Typically, researchers have been content to predict certain behaviors (e.g., reaction time to questions) that should occur if a certain type of schema is indeed present, and then create an experimental setting to demonstrate that effect (cf. Tesser 1978). The schema itself--i.e., content and internal organization--is usually not measured directly, or even relatively directly.

Moreover, there has been little effort directed toward developing hypothetical constructs which describe broad, generalizable characteristics of cognitive structures. Except for Scott's (1963, 1969, 1974) work dealing directly with cognitive structure, most researchers have theorized at a very abstract level about the ways in which knowledge representations may be coded and/or organized in memory. Few have been concerned with developing constructs at an intermediate level of abstraction that define the specific characteristics of cognitive structure that should influence other cognitive processes, and ultimately, overt behavior. This intermediate level is a necessary compromise in order to move away from the highly abstract theoretical level to a more operational level, yet avoid the idiosyncratic level of measuring the unique contents of each person's schemas.

The present research can best be understood against this background of developing interest in cognitive structures. We have three objectives for this paper: (a) to present our initial conceptualizations of three major characteristics of cognitive structure, (b) to briefly describe the empirical indicators we have developed to measure each characteristic, and (c) to present results from an initial study intended to establish the convergent and divergent validity of these indicators. The long-term goals for our research program, the first construct validity study of which is detailed here, are to develop a clearly conceptualized model of cognitive structure and a set of reliable and valid indicators of the major components of that model. Once reliable and valid measures are identified, we hope to learn more about how differences in cognitive structures influence a variety of interesting information processing behaviors such as product evaluations, purchase decision-making, responses to advertising, and information search, to name but a few.

Basic Assumptions

Before presenting specific details, it is first necessary to

¹This research was supported by the Science and Education Administration of the U.S. Department of Agriculture under Grant No. 5901-0410-8-0151-0 from the Competitive Grants Office to the second and third authors. The data collection assistance of Elizabeth Leung and Roslyn Weinstock is much appreciated.

describe several basic assumptions that provide the theoretical foundations for this research. Essentially, these are metatheoretical ideas, so abstract that they can not be easily subjected to empirical verification. Still, it is valuable to identify the conceptual framework which led to this work so that the reader can trace the logic underlying our approach.

Perhaps the most basic assumption is that people, through various encoding processes, cognitively represent information from the external world in symbolic form. We term these coded representations "knowledge," after Russo (1978). Our use of the term is broad; knowledge can refer to semantic, belief-like representations as well as affective and emotional, perhaps even visual, representations. Another basic assumption is that these coded representations are stored in memory in an organized or structured manner. A useful metaphor for conceptualizing such a cognitive structure is a network of interrelated, associated concepts. The links between concepts might be based on the meaningfulness of the association or the semantic relatedness between the concepts (see Collins and Loftus 1975). A third important idea is that a concept or group of concepts can be activated or "retrieved" from memory merely by attention to external or internally generated cues (Collins and Loftus 1975). In summary, then, information is presumed to be encoded and stored in organized networks of representations that we can call structures of knowledge, or cognitive structures (cf. Hayes-Roth 1977). Such structures presumably can be activated from memory, either in part or perhaps as a "unit," and used in various cognitive processes (e.g., Olson 1978a).

Dimensions of Cognitive Structure

What are the major characteristics of the network of knowledge that constitutes a cognitive structure? The cognitive psychology literature provides little help in answering this question. Instead, we turned to the cognitive style literature (see Goldstein and Blackman 1979 for review), and especially to the work of William Scott (1963, 1969, 1974), a social psychologist. During the early 1960's there was substantial interest in "cognitive structure" concepts such as cognitive complexity. Much of this work was "stylistically" oriented in that it attempted to describe and measure regular patterns or styles of behavior. For the most part, little attention was paid to the cognitive representations in memory that, with the hindsight provided by our information processing perspective, we now see as responsible for the observed behavior.

We have modified selected concepts from this earlier work in adapting them to our cognitive structure perspective. Actually, this process was less procrustean than it may appear. Scott's work in particular was fairly easy to translate into a memory structure orientation. Essentially, we selected major behavior patterns that previous researchers felt to be important, and we attempted to define the characteristics of a cognitive structure that could or should produce such behavior. Then, we developed conceptual definitions for each major characteristic of cognitive structure. In our scheme, then, the observed behaviors which previously had been the major focus of interest are now treated as outcomes of the cognitive structure constructs, and therefore, as potential indicators of these characteristics. We recognize, of course, that the constructs and measures of this initial nomological framework are likely to undergo substantial changes during the course of the next few years' research. They are offered here as a potential stimulus to that research effort, not as the final product.

Dimensionality

The notion of cognitive complexity provides the clearest example of our approach. Goldstein and Blackman (1979,

p. 115) state that "The cognitively complex individual is one who differentiates among concepts and is able to make discriminations on each of these conceptual dimensions." Note that this definition is strictly operational. It specifies two behaviors to be observed, based on which one can label an individual as more or less cognitively complex. Following Scott's (1969) lead, we conceptualized two aspects of cognitive structure that might be expected to produce these behaviors. One is dimensionality, defined as the number of salient, activatable (see, Olson, Kanwar and Muderrisoglu 1979) concepts stored in memory. Here we are concerned with the number of activatable concepts associated with a particular content domain.

Articulation

The second cognitive structure characteristic is articulation, defined as the number of category representations or levels for each salient dimension in memory. People who possess larger numbers of activatable concepts for a domain have a cognitive structure that is higher in dimensionality. Presumably, they are capable of using more concepts to perform evaluation or choice tasks regarding that domain, although whether or not they do so may be influenced by other factors. Likewise, people who are higher in articulation should, in general, be able to make finer distinctions between stimuli in terms of that dimension than those whose cognitive structure dimensions are less articulated. Thus, people high in both dimensionality and articulation should evidence "cognitively complex behavior" in certain tasks, because of these characteristics of their cognitive structures.

Abstraction

Abstraction is the third construct of interest in this paper.² As a characteristic of cognitive structure, abstraction concerns the degree of abstractness of the salient, activatable dimensions stored in the structure. A cognitive structure's abstractness is necessarily a relative concept, based on the mix of more and less abstract and concrete representations in the structure. This concept seems highly related to ideas such as hierarchical groupings of concepts (Collins and Quillian 1969), chunking (Simon 1974), unitization (Hayes-Roth 1977), and categorization (Rosch 1975). Abstraction can also be considered as a recoding process in which a new code is assigned to represent several other usually less abstract or more concrete codes. The abstraction process is functional since recoding reduces the number of salient concepts in a structure. Because larger amounts of knowledge are now represented by a smaller number of more abstract dimensions, using more abstract cognitive structures should require less of a person's limited cognitive capacity compared to using less abstract, more concrete structures.

These three characteristics--dimensionality, articulation, and abstraction--are proposed as critical features of cognitive structure. These (and perhaps other) features of someone's stored knowledge are expected to affect how that person processes information relevant to that domain. Of course, it is relatively easy to propose constructs. Compelling evidence for the validity of these constructs requires independent measures of the characteristics, in addition to accurate predictions of their presumed effects. Measures of the three characteristics are discussed below.

Measurement Approaches

The three cognitive structure constructs are defined conceptually in terms of the coded representations organized

²We have proposed two additional characteristics of cognitive structure--interrelatedness and centrality--but these are not discussed here.

in network-like memory structures. Therefore, measures of constructs must tap into memory, at least relatively directly. Most of the procedures used in past research to measure cognitive complexity are inappropriate for this purpose. For instance personality-like measures such as the scale used by Kasulius and Zaltman (1977) provide no indication of the content or organization of cognitive structure. Instead, we developed two procedures to indicate the proposed characteristics of cognitive structure more directly--free elicitation and a modified repertory grid task.

Free Elicitation. Free elicitation is the "newer" procedure, having been used relatively little in cognitive research. The conceptual basis for free elicitation is activation theory (Collins and Loftus 1975). The basic idea is that exposure to a cue activates the cognitive representation of that stimulus. Then, that activation spreads to other related concepts via the associations or linkages between representations. Thus, an entire structure (if one exists) might be activated in a relatively short time. Once a structure is activated and available for processing, the subject should be able to verbally report some, perhaps much of the structure content. By this logic, then, the elicited responses may be considered to reflect the characteristics of the cognitive structure.

The present research concerned peoples' nutrition knowledge structures. For each subject, we used four different probe cues to initiate the elicitation. In fact, we used a multiple elicitation procedure in which each elicited concept was itself subsequently used as a probe to stimulate additional elicitation. In the present study we used three or four "layers" of probes which produced a large number of idiosyncratic responses for each subject.³ Based on Epstein's (1979) arguments, we combined all the elicitation for the four probes to produce indices of dimensionality and abstraction. The total number of unique (multiple mentions of a concept were counted only once) nutritional concepts that are elicited were considered an indicator of the dimensionality of nutrition knowledge. And, the number of those nutrition responses that were relatively abstract and concrete were considered indicators of the abstractness and concreteness of a person's nutrition cognitive structure.

Repertory Grid. The repertory grid is a more structured elicitation procedure than the free elicitation technique. The repertory grid was originally developed by Kelley (1955) to identify the constructs people use to structure their perceptions of the social world. In adapting the repertory grid techniques to our purposes, we modified the original procedures somewhat.

Basically, our repertory grid methodology begins by presenting subjects with several sets of food concepts, three at a time. We use types of food as stimuli since we are

interested in nutrition knowledge structures. For each triad, subjects are to state all the "ways" in which any two of the foods are similar and different from the third. When no new concepts can be elicited, another triad is presented, and so on. We have found that after 10 to 15 triads most subjects produce few new concepts. The next step is to select the specific concepts of interest--here, nutrition concepts. For each unique nutrition concept elicited, subjects are asked to sort 20 or 30 types of food (each food printed on a card) into "piles," so that the foods in each pile are similar with respect to that particular dimension. Subjects are free to use as many or as few groupings as they wish. After the free sort, subjects are asked to name each group (pile) and describe why those foods were clustered together on that nutrition dimension.

The data produced by the modified repertory grid procedure provides possible indicators of all three knowledge structure characteristics. The number of unique nutrition concepts elicited may indicate the dimensionality of the nutrition knowledge structure. The elicited nutrition concepts can be coded in terms of their abstractness/concreteness. And, the average number of categories or levels produced in the sorting tasks may indicate the degree of articulation of the nutrition cognitive structure.

Knowledge Test. The two methods described above were developed to provide relatively direct indicants of stored cognitive structures. It was considered valuable to compare these measures with a more traditional method of measuring knowledge. A common method of measuring consumers' knowledge regarding a domain is to have them complete a paper-and-pencil test consisting of several specific, usually factual questions about the topic (cf. Edell and Mitchell 1978). This is the dominant technique used by nutrition researchers to measure people's nutrition knowledge (see Cosper and Wakefield 1975, Fusillo and Beloian 1977, Grotkowski and Sims 1978).

For the present study, we constructed a 23-item scale containing a variety of factual questions about nutrition. These items were selected from a larger set of questions commonly used in nutrition research and additional items constructed by the authors. First, items identified as ambiguous in small-scale pretests were removed or rewritten. Then the revised scale was given to a group of 250 respondents, 150 college students and 100 adult participants in a continuing education course. Each of the final 23 items had positive item-total correlations (average $r = .23$). A factor analysis produced no clear factors beyond a general factor, and Cronbach's coefficient alpha, an indicator of internal consistency, was .68 for the entire test. Scores on this knowledge "test" were related to the measures of dimensionality, abstractness, and articulation.

Methods

Subjects

The data collection procedures used in this research required substantial experimenter effort (as much as 2-1/2 hours per subject). Therefore, final sample size was restricted to 18 women who had the primary responsibility for food shopping and meal preparations in their households. Because the research purpose was to establish the convergent and divergent validities of the possible indicators of cognitive structure and their interrelationships, any person with cognitive knowledge about nutrition (nearly everyone, probably) was an appropriate subject. Therefore, because sample representativeness was not a major concern, subjects were graduate student wives and secretaries at Penn State University, selected on a convenience basis. All voluntarily agreed to participate by signing an informed consent form and were paid for their participation at the rate of \$4.00 per hour.

³We consider free elicitation to be subtly different from the free recall paradigm commonly used in cognitive and to some extent in consumer psychology (e.g., Johnson and Russo 1978). In free recall, the focus tends to be on a particular learning situation and on the specific episodic memory traces about that event. Thus instructions tend to be like, "Tell me which words (or ads) you remember seeing yesterday." In free elicitation, the researcher usually seeks probe cues that will trigger or activate a particular cognitive structure in semantic memory. Although the content of that structure probably was acquired over many learning occasions, these specific occasions are of little or no interest. It is interesting to note that the free elicitation procedure, as derived from activation theory and associationist network theory, is essentially identical to the method advocated by Fishbein (1967) for identifying salient beliefs.

Procedures

Two stages of data collection are relevant for this paper. At the first stage, 29 subjects went through the free elicitation procedures described earlier. Each subject was run individually. Four initial probe cues were used for each subject. Fifteen subjects received as probes the phrases, "good health, balanced diet, nutritious food, and vitamins," while the other 14 subjects responded to "having good health, obtaining a balanced diet, eating nutritious food, and getting enough vitamins." For present purposes this manipulation is ignored. As described earlier, elicited responses to the four original probe cues were used in turn as probes and so on again, "down" to four levels. Interviewers recorded subjects' responses as they were elicited on special scoring sheets. This multiple probe procedure was intended to maximize the chances of activating parts of each subjects' nutrition knowledge structure(s). It resulted in a substantial number of elicited concepts per subject ($M = 47.1$). Following the elicitation, subjects completed the knowledge questionnaire, plus several other questions. On average, this session required from 1 to 1-1/2 hours.

The repertory grid was conducted at a second session about two months later. Due to some attrition and one refusal, only 18 of the original 29 subjects participated in the second session. The repertory grid procedures were followed as described above. An average of 34.1 concepts were produced for each subject. This session required about one hour.

Results

Our major interest is in the convergent validity of the elicitation and rep grid procedures for measuring dimensionality and abstraction, and in the relationships (a) between these two constructs and articulation and (b) between all three concepts and the scores on the questionnaire knowledge test. To briefly review, dimensionality was measured by the number of unique nutrition concepts produced by the elicitation ($M = 6.6$) and repertory grid ($M = 5.7$) procedures. Possible indices of abstraction were generated by classifying the unique nutrition responses from the elicitation and grid procedures as relatively abstract (M 's = 1.9 and 1.3) or concrete (M 's = 4.6 and 3.4). Two independent judges were reasonably consistent in their categorizations of the concepts as nutritional or not ($r = .85$) and as abstract/concrete ($r = .87$). Unfortunately, we had only a single index of articulation, the average number of category levels per nutrition concept ($M = 4.4$), as evidenced by the rep grid free sorting task.

Table 1 presents the relevant correlations between these constructs, in a multimethod-multitrait matrix format. First, note that concreteness, the number of concrete nutrition concepts, is highly correlated with dimensionality, the total number of nutrition concepts elicited ($r = .96$ for the free elicitation method and $r = .95$ for the rep grid procedure). Apparently, subjects who produced more unique nutrition concepts also produced many concrete concepts. Therefore, "concreteness" may not be useful for indicating the abstractness of a cognitive structure. Certainly, the number of abstract concepts has more face validity. It is possible that if finer distinctions between levels of abstraction/concreteness could have been made, the apparent redundancy would be eliminated or reduced. However, criteria for more finely discriminating between levels of abstraction are not clear to us. Because concreteness is essentially totally redundant with the dimensionality measure, in the present data at least, we will ignore most of the correlations with the concrete measure, although they are shown for completeness.

Next, consider the convergent and discriminant indices. The rep grid and free elicitation methods seem to converge

TABLE 1
Correlations Between Measures of Hypothetical
Characteristics of Cognitive Structure

	Measures/Methods						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>ELICITATION METHOD</u>							
Dimensionality	(1)	--					
Abstraction	(2)	.48	--				
Concreteness	(3)	.96	.25	--			
<u>REP GRID METHOD</u>							
Dimensionality	(4)	.56	.29	.54	--		
Abstraction	(5)	.02	.65	-.16	.28	--	
Concreteness	(6)	.58	.08	.62	.95	-.04	--
Articulation	(7)	-.26	-.32	-.20	-.29	-.46	-.15
<u>QUESTIONNAIRE METHOD</u>							
Knowledge Score	(8)	.11	.10	.09	.31	.18	.26

Notes: $N = 18$
Correlations greater than .40 are significant
 $p < .10$.

fairly well on the concepts of dimensionality and abstraction (r 's = .56 and .65, respectively, $p < .05$). The low off-diagonal values of .02 and .29 provide evidence of discriminant validity. Note also the dimensionality and abstract construct constructs are apparently positively related, as one might expect (r 's = .49 and .28 for free elicitation and rep grid, respectively). That is, as the dimensionality of cognitive structure increases, there is a tendency for it to become more abstract. Because only a single measure of articulation has been developed thus far, estimates of convergent and divergent validity are not possible. However, the low, negative relationships between articulation and the dimensionality and abstraction measures are quite consistent for both measurement methods. Although the small magnitude of these correlations do not allow confident theoretical interpretations, it seems logical to suppose that as the dimensions of a structure become more abstract, they also become less articulated. In sum, both the magnitude of these relationships and their general consistency, especially given the long time period between measures, provide support for the construct validity of the three characteristics of cognitive structure and the validity of the measures developed as indicators of those concepts.

What about the traditional paper-and-pencil knowledge test? As can be seen in Table 1, knowledge test scores are only weakly related to the other measures of cognitive structure. The test scores seem to relate more strongly with the rep grid measures, although this is only a weak tendency. At least all the relationships are positive, as would be expected if each measure is tapping nutrition knowledge structures. Clearly, however, the knowledge test is not measuring the same thing as the free elicitation and repertory grid methods.

Conclusions and Implications

In our opinion, sufficient evidence of convergent and

discriminant validity has been demonstrated to warrant additional work on these hypothesized cognitive structure constructs. The magnitude of the obtained relationships and their consistency are reasonable, although certainly not outstanding. The effect sizes obtained can be seen as a bit more encouraging when one considers that the subject sample was probably more homogeneous with respect to nutrition knowledge than the general population. Given the probable restriction of range in our indicators produced by this homogeneity, the obtained coefficients can be seen as even more respectable. In conclusion, then, the repertory grid and free elicitation procedures appear to be capable of producing valid indicators of two potentially important characteristics of cognitive structure--dimensionality and abstraction. Also, the other hypothesized attribute of cognitive structure, articulation, evidenced consistent relationships with the measures of both dimensionality and abstraction.

We are encouraged by these results and are conducting further investigations concerning the construct validity of these characteristics of cognitive structure. What should be the next steps in such a research program? Obviously, a replication of this study should be conducted. To increase power, such a replication should use more subjects, if possible. Moreover, it would be valuable to obtain subjects who possess wide ranges of nutrition knowledge. Increasing the variance for measures of nutrition cognitive structures should enhance the predicted relationships, compared to the present results, if the basic model is essentially valid.

A slightly different approach to construct validation could provide a somewhat stronger "test" of the validity of these measures. One could select subjects who on some a priori basis comprise two (relatively) extreme groups in terms of their knowledge about nutrition. For instance, one might use senior undergraduate nutrition majors and seniors who had never taken nutrition courses. Or, one could examine persons known to be highly concerned about their eating habits (health food advocates, e.g.) versus less concerned people (ordinary citizens). Presumably, such groups would differ dramatically in terms of their cognitive structures regarding nutrition. If so, valid measures of the major characteristics of knowledge structures should reflect these differences.

Eventually, future research must predict the effects that differing cognitive structures have on other cognitive variables and processes and must test those hypotheses empirically. For example, if someone has a complex cognitive structure for nutrition (high dimensionality, relatively abstract, well articulated), we might predict certain types of cognitive responses to a series of food ads containing nutrition information. Presumably, such a person's responses should differ from those produced by a different consumer with a shallow, simple cognitive structure for nutrition. Minimally, the former consumer should produce more nutritional cognitive responses. Probably, these responses would be more abstract or deeper in a semantic sense (Olson 1980). In contrast, the effects of differing cognitive structures on persuasibility (e.g., attitude and intention change) are less clear (see Marks and Olson 1981). To make such predictions, one must carefully examine the idiosyncratic content of stored knowledge and determine, among other things, the degree of discrepancy between the individual's pre-exposure beliefs and those presented in the message. Finally, differences in cognitive structures should be reflected by variation in decision-making processes. Here we might examine the amount of time taken, the number of attributes considered, and the optimality of the selected alternative, among other factors.

In conclusion, we hope to have stimulated at least some readers to undertake research that deals more directly with peoples' cognitive structures than typically has been done. Issues regarding cognitive structure are intrinsically

interesting, and they are important. Developing clearer ideas regarding the content and organization of cognitive structures seems likely to enhance our understanding of the other cognitive processes that are influenced by cognitive structure, and of cognitively-mediated behavior, too.

References

- Anderson, John R. (1976), Language, Memory, and Thought, Hillsdale, NJ: Lawrence Erlbaum.
- Anderson, R. C., Spiro, R. J. and Montague, W. E., eds. (1977), Schooling and the Acquisition of Knowledge, Hillsdale, NJ: Lawrence Erlbaum.
- Bettman, James R. (1979), An Information Processing Theory of Consumer Choice, Reading, MA: Addison Wesley.
- Calder, Bobby J. (1978), "Cognitive Response, Imagery and Scripts: What is the Cognitive Basis of Attitude," in Advances in Consumer Research, 5, Keith Hunt, ed., Ann Arbor, MI: Association for Consumer Research, 630-634.
- Collins, Alan M. and Quillian, M. R. (1969), "Retrieval Time from Semantic Memory," Journal of Verbal Learning and Verbal Behavior, 8, 240-247.
- _____, and Loftus, E. F. A. (1975), "A Spreading-Activation Theory of Semantic Processing," Psychological Review, 8, 240-247.
- Cosper, B. A., and Wakefield, L. M. (1975), "Food Choices of Women: Personal, Attitudinal, and Motivational Factors," Journal of the American Dietetic Association, 66, 152-155.
- Edell, Julie A. and Mitchell, Andrew A. (1978), "An Information Processing Approach to Cognitive Responses," in Advances in Consumer Research, 5, Subhash C. Jain ed., Chicago, IL: American Marketing Association, 178-183.
- Epstein, Seymour (1979), "The Stability of Behavior: I. On Predicting Most of the People Much of the Time," Journal of Personality and Social Psychology, 37, 1097-1126.
- Fishbein, Martin (1967), "A Consideration of Beliefs and Their Role in Attitude Measurement," in Readings in Attitude Theory and Measurement, Martin Fishbein, ed., New York: Wiley.
- Fusillo, A. E., and Beloian, A. M. (1977), "Consumer Nutrition Knowledge and Self-Reported Food Shopping Behavior," American Journal of Public Health, 67, 846-850.
- Goldstein, Kenneth M., and Blackman, Sheldon (1978), Cognitive Style, Five Approaches and Relevant Research, New York, NY: John Wiley & Sons.
- Grotkowski, M. L., and Sims, L. S. (1978), "Nutrition Knowledge, Attitudes and Dietary Practices of the Elderly," Journal of the American Dietetic Association, 72, 499-506.
- Hayes-Roth, Barbara (1977), "Evolution of Cognitive Structures and Processes," Psychological Review, 84, 260-278.
- Jacoby, Jacob, Szybillo, George C. and Busato-Schach, J. (1977), "Information Acquisition Behavior in Brand Choice Situations," Journal of Consumer Research, 3, 206-216.
- Johnson, Eric J. and Russo, J. Edward (1978), "The Organization of Product Information in Memory Identified by Recall Times," in Advances in Consumer Research, 5, Keith Hunt, ed., Ann Arbor, MI: Association for Consumer Research, 79-86.

- Kasulius, Jack J., and Zaltman, Gerald (1977), "Message Reception and Cognitive Complexity," in Advances in Consumer Research, 4, William D. Perreault, ed., Atlanta, GA: Association for Consumer Research, 93-97.
- Kelly, G. A. (1955), The Psychology of Personal Constructs, 1 and 2, New York: Norton.
- Kintsch, W. (1974), The Representation of Meaning in Memory, Potomac, MD: Lawrence Erlbaum.
- Lachman, Roy, Lachman, Janet L. and Butterfield, Earl C. (1979), Cognitive Psychology and Information Processing: An Introduction, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lutz, Richard J. (1975), "Changing Brand Attitudes Through Modification of Cognitive Structure," Journal of Consumer Research, 1, 49-59.
- Marks, Larry J. and Olson, Jerry C. (1981) "Toward a Cognitive Structure Conceptualization of Product Familiarity," in Advances in Consumer Research, 8, Kent Monroe, ed., Ann Arbor, MI: Association for Consumer Research.
- Markus, Hazel (1977), "Self Schemata and Processing Information About the Self," Journal of Personality and Social Psychology, 35, 63-78.
- Minsky, M. (1975), "A Framework for Representing Knowledge," in The Psychology of Computer Vision, New York: McGraw Hill.
- Mitchell, Andrew A. (1980), "The Use of an Information Processing Approach to Understand Advertising Effects," in Advances in Consumer Research, Vol. 7, J. C. Olson, ed., Ann Arbor, MI: Association for Consumer Research, 171-177.
- Norman, Donald A. and Bobrow, Daniel G. (1975), "On the Role of Active Memory Processes in Perception and Cognition," in The Structure of Human Memory, Charles N. Cofer, ed., San Francisco: W. H. Freeman.
- _____ (1979), "Perception, Memory and Mental Processes," in Perspectives on Memory Research: Essays in Honor of Uppsala Universities 500th Anniversary, L. Nilsson, ed., Hillsdale, NJ: Lawrence Erlbaum Associates.
- Olson, Jerry C. (1978a), "Inferential Belief Formation in the Cue Utilization Process," in Advances in Consumer Research, Vol. 5, Keith Hunt, ed., Ann Arbor, MI: Association for Consumer Research, 706-713.
- _____ (1978b), "Theories of Information Encoding and Storage: Implications for Research," in The Effect of Information on Consumer and Market Behavior, Andrew Mitchell, ed., Chicago: American Marketing Association, 49-60.
- _____ (1980), "Encoding Processes: Levels of Processing and Existing Knowledge Structures," in Advances in Consumer Research, 7, J. C. Olson, ed., Ann Arbor, MI: Association for Consumer Research, 154-160.
- _____ and Dover, Philip A. (1978), "Attitude Maturation: Changes in Related Belief Structures Over Time," in Advances in Consumer Research, 5, Keith Hunt, ed., Ann Arbor, MI: Association for Consumer Research, 333-342.
- _____ and Muderrisoglu, Aydin (1979), "The Stability of Responses Obtained by Free Elicitation: Implications for Measuring Attribute Salience and Memory Structure," in Advances in Consumer Research, 6, William Wilkie, ed., Ann Arbor, MI: Association for Consumer Research, 269-275.
- _____ Kanwar, Rajesh and Muderrisoglu, Aydin (1979), "Clarifying the Confusion Regarding Salience, Importance and Determinance Concepts in Multi Attribute Attitude Research," in Educators' Conference Proceedings, N. Beckwith, M. Houston, R. Mittelstaedt, K. Monroe, and S. Ward, eds., Chicago: American Marketing Association, 286-290.
- Park, Whan C. (1976), "The Effect of Individual and Situation Related Factors on Consumer Selection of Judgmental Models," Journal of Marketing Research, 13, 144-151.
- Puff, C. Richard, ed. (1979), Memory Organization and Structure, New York: Academic Press.
- Raju, P.S. and Reilly, Michael D. (1980), "Product Familiarity and Information Processing Strategies: An Exploratory Investigation," Journal of Business Research, 8, 187-212.
- Rigney, J., Bieri, J., and Tripoli T. (1964), "Social Concept Attainment and Cognitive Complexity," Psychological Reports, 15, 503-509.
- Rosch, E.H. (1975), "Cognitive Representations of Semantic Categories," Journal of Experimental Psychology: General, 104, 192-233.
- Russo, J. Edward (1978), "Comments on Behavioral and Economic Approaches to Studying Market Behavior," in The Effect of Information on Consumer and Market Behavior, A.A. Mitchell, ed., Chicago: American Marketing Association.
- _____ and Johnson, Eric J. (1980), "What do Consumers Know About Familiar Products" in Advances in Consumer Research, 7, Jerry C. Olson, ed., Ann Arbor, MI: Association for Consumer Research, 417-423.
- Scott, William A. (1963), "Cognitive Complexity and Cognitive Balance," Sociometry, 26, 66-74.
- _____ (1969), "Structure of Natural Cognitions," Journal of Personality and Social Psychology, 12, 261-278.
- _____ (1974), "Varieties of Cognitive Integration," Journal of Personality and Social Psychology, 30, 563-578.
- Shank, R. and Abelson, R. (1977), Scripts, Plans, Goals and Understanding, Hillsdale, NJ: Lawrence Erlbaum.
- Simon, Herbert A. (1974), "How Big is a Chunk," Science, 183, 482-488.
- Taylor, Shelley E. and Crocker, J. (1980), "Schematic Bases of Social Information Processing," in Social Cognition: The Ontario Symposium on Personality and Social Psychology, T. Higgins, C. A. Herman, and M. P. Zanna, eds., Hillsdale, NJ: Lawrence Erlbaum.
- Tesser, A. (1978), "Self Generated Attitude Change," in Advances in Experimental Social Psychology, Vol. II, L. Berkowitz, ed., NY: Academic Press.
- Tulving, E. and Donaldson, W. (1972), Organization of Memory, New York: Academic Press.
- Woodruff, Robert B. (1972), "Measurement of Consumers' Prior Brand Information," Journal of Marketing Research, 9, 258-263.
- Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance to Advertising," Journal of Marketing Research, 10, 60-67.

COGNITIVE MAPS AND SHOPPING CONVENIENCE

Sanford L. Grossbart, University of Nebraska
Balusu Rammohan, University of Nebraska

Abstract

The nature and functions of cognitive mapping and maps of large scale shopping areas are reviewed. A method of measuring cognitive maps is employed to study the possible bases for customers' beliefs regarding the convenience of shopping downtown. Results are presented which indicate that mapped information is related to various forms of shopping convenience.

Introduction

The central thesis of this paper is that beliefs regarding the convenience of shopping downtown are related to customers' cognitive maps of the downtown spatial environment. While multiple factors influence customers' patronage decisions, the issue of convenience is singled out for several reasons. Observers frequently mention the inconvenience of traveling to the area, parking and moving between stores as important reasons for the decline of downtown business in cities of all sizes. Moreover, the provision of "more convenient" physical plans and parking amenities represents a costly undertaking which is apparently beyond the capabilities or willingness of many individual businesses and city planning bodies. As such redevelopment designs have been drawn and debates concerning the sharing of costs required to provide increased convenience have continued, there has been a tendency to overlook the fact that customer beliefs about convenience are apt to be highly subjective and related to a set of largely unidentified determinants. The exploratory research described here was undertaken with the expectation that differences in customer's spatial images of the downtown, termed cognitive maps, would be related to beliefs about different forms of convenience. Evidence which supports this expectation, while obviously not addressing all aspects of downtown deficiencies, carries the important implication that providing customers with more adequate information about retail identities and functions may alter their beliefs about shopping convenience and, therefore, pose a less costly alternative than large-scale physical changes.

The Nature and Functions of Cognitive Mapping and Maps

Cognitive mapping is the mental process involved in acquiring, storing and manipulating information about spatial environments. As organized representations of portions of the environment, these maps are simplified models or mental images of complex physical reality (Downs and Stea 1977). Maps are formed in terms of places within places, e.g., stores within a downtown, whose perceived prominence, functional and symbolic meaning and use increase the likelihood of their mental representation. Information concerning these stores aids customers in understanding the spatial dimension of the shopping area. There are three basic sources for mapped information (Downs and Stea 1973). Direct experience, e.g., shopping and moving about the area, provides the opportunity for the integration of kinesthetic, visual, tactile and olfactory sensory information about store identities, activities, distances, locations and other aspects of a downtown. Vicarious sources which provide second hand spatial information may include friends, salespersons, media, street maps, Chamber of Commerce brochures, etc.

Finally, persons may form inferences by elaborating upon and augmenting the direct or vicarious information which they have acquired. For instance, information about the location and density of certain types of stores (such as fashion boutiques or antique shops), and may lead to inferences about the existence and character of other stores in that particular sub-area.

In a strict sense, the term cognitive mapping is somewhat misleading but is used because of its widespread employment in environmental psychology, geography and other disciplines involved in the study of environmental behavior. Recognizing the potential for confusion, theorists have made an effort to point out that affective and conative influences impact on the cognitive mapping process and on interpretations of cognitive maps (Downs and Stea 1973; Kaplan 1973). While considerable research interest has been devoted to the nature and function of cognitive maps (Stokols 1978; Moore 1979), with few exceptions (Mazze 1974; Mackay and Olshavsky 1975; Mittelstaedt, Grossbart and Curtis 1977) these subjects have received little attention in the field of consumer behavior. Consequently, it is hardly surprising that there has been limited recognition of the likelihood that shopping behavior, like other forms of spatial behavior, is a function of persons' cognitive maps of the spatial environment. Similarly, and of more immediate concern, there has been little acknowledgement of the influence of cognitive maps on different dimensions of retail images, such as those relating to convenience. These assertions develop from a perspective which suggests that shopping areas should be treated as places rather than objects (Mittelstaedt, Grossbart and Curtis 1977; Canter 1977).

As a place, the downtown represents a large scale surface whose complexity is a function of the number, frequency and distribution of information categories which are presented to customers. The physical and functional features of retail establishments are neither ubiquitous nor are they uniformly distributed -- they have instead locations on the surface. In every sense the area engulfs customers whose limited mobility, sensory capability, information processing ability, storage capacity and available time require them to develop simplified abstractions of the downtown. Individuals interact with the downtown in a distinctive and selective manner and with each transaction learning is likely to occur (Rosenberg 1965-66; Fatouros 1968). Mapping of the downtown occurs, in part, as customers receive imperfect multisensory information (Mehrabian and Russel 1974; Grossbart, Amedeo and Chinchin 1979; Downs and Stea 1977) from an uncertain and changing retail environment. Since shopping occurs at different times and varies in duration, the direct acquisition of information for mapping takes place over varying times and intervals.

From the diversity of direct experience, the vicariously related information and inferences, customers aggregate limited and imperfect impressions to form comprehensible representations of the downtown. The product of this process, the cognitive map, has been described as a network (Kaplan 1973) of organized beliefs (Downs and Stea 1973) about a large scale environment. The absence of a map, or, of certain stores within the map, may necessitate a series of personal inquiries, checking of printed information (e.g., the Yellow Pages), and/or the expenditure of the additional time and effort involved in search. In contrast, the existence of cognitive maps serve a variety of purposes. Maps allow customers to find things (e.g.,

merchandise, stores, sales, restaurants, etc.) as well as people (e.g., favored salespersons, providers of specific services, etc.) by providing indications of direction and distance. They also allow customers to gain an impression of where they are within the area and thus predict the stores, merchandise and activities which will be encountered as they move in different directions. As a result this network of beliefs can provide a basis for evaluating the consequences of various shopping behaviors (e.g., coming to the area, choosing routes for multipurpose trips, visiting certain stores, finding certain merchandise or services, etc.) Thus references to cognitive maps and consequent evaluation of maps allow customers to make a variety of shopping decisions.

Customer Convenience

Customer convenience has long been recognized as having an important influence on shoppers' evaluations of stores and shopping areas (Kelley 1958; Cox 1959; Downs 1961; Bender 1964). As a shopping cost, inconvenience has been identified as an economic and psychological expenditure and source of frustration and annoyance which is both traded-off with the price of goods and services and added to them to form the total price paid by customers (Crafton 1979). Given the acknowledged role of convenience in influencing customer decision making (Markin 1977), it is not surprising to find theoretical and empirical support for the inclusion of convenience as one of the evaluative criteria composing retail images (for a review, see Lindquist 1974-75). Beliefs regarding customer convenience have been shown to differ when retailer and customer perceptions are compared (Pathak, Crissy and Sweitzer 1974-75) and have been identified as being dependent on a set of largely unknown factors of spatial cognition and urban form (Downs 1970).

A review of the literature indicates support for the existence of five forms of shopping convenience. The most well known form is travel convenience, i.e., the perceived ease of making a trip to a shopping area (Kelley 1958; Downs 1961; Bender 1964; Rich and Portis 1964; Kunkel and Berry 1968). Parking convenience (Kelley 1958; Downs 1961; Rich and Portis 1964; Kelly and Stephenson 1967; Kunkel and Berry 1968; Downs 1970; Lindquist 1974-75) is frequently mentioned in discussions comparing the competitive strengths of downtowns and outlying shopping centers. Customer movement convenience (Cox 1959; Downs 1961; Kunkel and Berry 1968; Downs 1970) involves the ease of movement between stores while shopping. Aggregate convenience is a term used by Cox (1959, also see Bender 1964) to refer to the ease with which consumers come into physical contact in succession with a large variety of goods and services. Properly aggregated, clusters of goods and services follow the principle of efficient congestion by concentrating retailing offerings while avoiding dysfunctional traffic and congestion. Finally, merchandise display convenience (Cox 1959; Bender 1964; Rich and Portis 1964) provides customers with necessary cues which minimize undesirable search by enabling shoppers to locate merchandise for comparison, inspection or purchase. The minimization of undesirable search can be a critical factor in shopping since customers are likely to engage in more product evaluation when shopping effort is perceived as low (Harrell and Hutt 1976).

Cognitive Maps and Perceived Convenience

Given the selective spatial character of shopping behavior and the limitations inherent in customers' mapping capabilities, three pertinent characteristics should be evident in cognitive maps of the downtown. These maps are likely to be more selective and therefore less extensive than retailer-defined definitions of the area. Second, within the cognitively defined boundaries of the area, maps are likely

to be incomplete in that the identity, function and locations of all stores will not be represented. Finally, imperfect encoding of available sensory information, incorrect and misinterpreted vicarious information, and erroneous inferential augmentation and elaboration are likely to result in cognitive mapping errors. Of course, it would be naive to expect shopping behavior to be totally confined or limited by material included in cognitive maps. While there is sufficient evidence to suggest that map accuracy does vary among individuals (Stokols 1978; Moore 1979), the "goodness" of a cognitive map is best judged in terms of its usefulness as a guide to its possessor rather than in terms of its accuracy or detail (Kaplan 1973). Shoppers can, after all, supplement mapped information or correct misinformation by inquiry, search or trial and error. Prior to such supplementation or correction, however, they are likely to form beliefs about shopping convenience based on their beliefs regarding the identities, functions or offerings, and spatial distribution of retail facilities in the downtown. As they shop, they may supplement or correct their maps or even find desired facilities, merchandise or locations without altering their maps. Any of these acts, however, are likely to involve added effort and therefore decrease the perceived convenience of shopping.

Previous comments refer to general relations between mapped retail information and perceived convenience. Expectations for specific relationships are presented in Table 1. Absolute identity level refers to the number of correctly named and located facilities in a cognitive map while relative identity level refers to the proportion of correctly named and located facilities. Absolute and relative functional level have a similar meaning but refer to knowledge of the nature, predominant offerings, or activity of mapped establishments. The five forms of shopping convenience have been defined earlier.

TABLE 1
Pattern of Expected Relationships Between Forms of Shopping Convenience and Cognitive Map Characteristics

Cognitive Map Characteristics	Forms of Shopping Convenience				
	Travel	Parking	Customer Movement	Aggregate	Display
Absolute identity level	X			X	X
Relative identity level	X	X	X		X
Absolute functional level	X			X	X
Relative functional level	X	X	X		X

Identity and functional levels of maps should influence beliefs regarding travel convenience in two respects. As the absolute knowledge of identity and function increases, the general attractiveness of the area and the perceived convenience of making a shopping trip should also tend to increase. As relative knowledge of function and identity increase, there should be less probability of customers confusion and frustration during shopping trips, less perceived effort in shopping downtown and a consequent increase in perceived travel convenience.

Relative levels of identity and functional knowledge are associated with parking and customer movement convenience. As the proportion of correctly mapped retail information increases, customers are likely to be able to better plan their spatial movements and parking activities. Conversely, parking convenience is likely to decrease if customers' limited relative knowledge of the locations of facilities and functions leads them to make less efficient parking decisions. Likewise, decreasing relative knowledge of

identities and functions is likely to lead to increasing customer confusion and frustration and/or increased shopping efforts to locate stores, merchandise or services. Thus perceived ease of movement should tend to covary with both types of relative knowledge.

Absolute levels of identity and functional knowledge are expected to be related to perceived aggregate convenience. Under normal circumstances, increasing knowledge of facilities and their offerings should lead customers to infer the existence of a wide range of goods and service offerings in the downtown.

Finally, all four map characteristics should be related to beliefs regarding display convenience. As the absolute and relative knowledge of the identity and function of retail sites within the downtown increases, customers should be expected to have a clearer conception of a wider range of offerings, be less confused about the spatial distribution of available goods in the area and, therefore, be less dependent on display cues for this information. In contrast, those whose cognitive maps possess lower levels of knowledge should tend to rely more on display cues to supplement and correct mapped information. In the process they are likely to expend greater effort in utilizing display information. Consequently, customers who possess greater levels of knowledge and therefore place less frequent reliance on display cues should judge them to be easier to use than shoppers whose lack of knowledge leads them to substitute display cues for incomplete and incorrect mapped information.

Method

Site

This study dealt with the downtown area of a small mid-western city. The city's well established downtown retail base had operated successfully for years but it currently faces competition from outlying shopping centers. In recent years some downtown retailers have chosen to relocate in these shopping centers or move elsewhere.

Retailer Definition of Downtown

Interviews with representatives of the trade council and Chamber of Commerce provided information concerning the characteristics of downtown shoppers and a geographic definition of the shopping area in terms of the center and boundaries (i.e., the extent). Following these delineations an audit was done to ascertain the names, locations and functions of all facilities in and immediately surrounding the defined area. Using this information, two maps were prepared. The first was a base map showing possible streets and blocks that might be included in the downtown. The second was a reference map which indicated the names, functions and locations of facilities in addition to the information provided in the base map.

Sample

Prior market research indicated that the downtown's primary trade area included customers within a sixty mile radius of the city. Based on earlier estimates of the proportion of shoppers residing in different parts of the trade area, a geographically stratified sample of fifty-one adults was drawn from within the community and four other randomly selected areas within fifteen, thirty, forty-five and sixty miles of the city. Persons within these five designated areas were randomly selected and contacted by telephone to arrange inhome interviews. Respondents included adults who shopped in the downtown in the last year, lived at least two years in the trade area (thus affording them the opportunity to learn about the downtown), and were members of households in which no person had been employed by a downtown establishment. The

refusal rate among those who met the stated criteria was approximately ten percent and a check of respondent profiles indicated a sample which was representative of downtown shoppers.

Instrument

In addition to screening questions, which ensured that respondents met the criteria for inclusion in the sample, and demographic items to allow estimates of the sample's representativeness, the instrumentation was divided into two parts.

Scales. The first part contained a set of self-administered scales measuring convenience beliefs, shopping evaluation and shopping frequency. Statements regarding the convenience of traveling to the downtown to go shopping, finding convenient parking, the convenience of getting from one store to another, finding a variety of goods and services, and displays making it easy to see the goods and service offerings were ordered in a randomized manner. Each statement was accompanied by a seven-point scale (ranging from very unlikely to very likely) indicating the likelihood that the statement was true. Overall evaluation of shopping was measured in terms of a seven-point scale (ranging from very pleasant to very unpleasant) indicating the pleasantness of shopping downtown. Average monthly shopping frequency was also recorded.

Mapping Task. Part two contained the base map. Respondents mapping task initially consisted of drawing a line indicating the boundaries of the shopping area and placing an "X" at the intersection which they considered to be the center of the downtown. The last phase of mapping involved respondents description of what they believed to be on each blockside within the downtown in terms of name, function and location of places. The task represented a variation of the previously employed sketch map procedure (Lynch 1960; Downs and Stea 1977; Mittelstaedt et. al. 1977; Holahan and Dobrowolny 1978), which has been suggested as the most appropriate method of measuring cognitive maps (Jackle, Brunn and Roseman 1976).

The entire instrument was pretested prior to employment. The interviewer also recorded any comments made by respondents and, as a further precaution, began and ended each interview asking respondents to indicate their general thoughts about the downtown. Interviews averaged eighty minutes in length.

Coding and Analysis

Preliminary coding of maps involved recording the extent of the defined area (number of blockside), the number of facilities included and the perceived center of the area. Next, each facility located on a respondent's map was compared to the corresponding site on the reference map and coded in terms of whether or not the identification and indicated function were correct. Incomplete responses, i.e., those which mentioned only a facility's name or function or indicated that either was "unknown," were coded as incorrect with respect to the omitted or unknown information. Correctly identified facilities were counted to form the absolute identity level score for each map. Similarly, the correctly designated functions were counted to form the absolute functional level score. Relative identity and functional levels were computed, respectively, by dividing the number of correctly identified facilities and the number of correctly designated functions by the total number of perceived facilities.

The analysis was designed to reaffirm the general salience of convenience in customer evaluation and behavior and the selectivity of customer cognitive mapping prior to testing the expected associations between convenience beliefs and map characteristics. Analysis focused first on the correlation between beliefs regarding different forms of shopping

convenience and overall evaluation of shopping and shopping frequency. Significant, although not necessarily high, correlations were expected to reaffirm the general salience of convenience beliefs. A map summarizing customer definition of the area and the spatial distribution of perceived activity was then prepared. Next, retailer definition of the downtown was compared with cognitive map definitions in terms of extent, contents and perceived center. Finally the correlations between map characteristics and forms of shopping convenience were examined and compared to the expected pattern in Table 1. The nature of the analysis in this exploratory research was deemed to be appropriate prior to future multivariate investigations of the relationships:

Results

Respondents' comments before and after each interview made it clear that they were not unduly sensitized by the instrument, had not found it difficult to understand or follow instructions, and had not deduced the purpose of the study. Prior to the presentation of scales, when asked about their general thoughts regarding the downtown, almost all respondents made unaided references to multiple forms of shopping convenience.

Convenience, Overall Evaluation and Shopping Behavior

Beliefs regarding the five forms of shopping convenience were examined for their assumed relationships with customer evaluation and frequency of shopping downtown. Given other goods and service mix attributes (price, quality, etc.) which may influence shopping evaluation and behavior, it would be unreasonable to envisage extremely high correlations. However, it would be logical to expect more association between beliefs regarding convenience and evaluation of shopping experiences than with shopping frequency because of the number of additional variables, such as situational factors, which may influence shopping frequency. Results are presented in Table 2. All five forms of convenience are significantly related to overall evaluation of shopping downtown. As might be expected, beliefs concerning travel convenience have a lower correlation with evaluation than beliefs regarding other forms of convenience, which refer in a more direct sense to the experience of shopping in the area. Beliefs regarding each form of shopping convenience are also associated with customers' frequency of shopping downtown. In general, assumptions about the role of convenience related beliefs were substantiated.

TABLE 2
Correlations of Overall Evaluation of Shopping and Shopping Frequency with Forms of Shopping Convenience

Forms of Shopping Convenience	Overall Evaluation of Shopping Downtown	Frequency of Shopping Downtown
Travel	0.24 ^c	0.26 ^c
Parking	0.39 ^b	0.28 ^c
Customer Movement	0.66 ^a	0.23 ^c
Aggregate	0.42 ^a	0.21 ^d
Display	0.54 ^a	0.22 ^d

^a p < .001 ^b p < .01 ^c p < .05 ^d p < .10

Retailer Versus Customer Definitions of Downtown

The selective nature of cognitive maps is reflected in the summary map in Figure 1 which presents the percentage of

respondents perceiving activities on each blockside in the downtown area. This summary map does not indicate the identity or functional level of customer maps; it merely shows the degree of general definition of perceived activities on the illustrated blocksides. The amount of general definition on blocksides along the edges of the map is either nonexistent or very low. The sub-area with the highest degree of definition is along 3rd Street bounded by Walnut and Locust. Moving out from this sub-area, the degree of definition declines slowly initially and then rapidly in all directions. From this summary map it is apparent that the downtown is perceived by customers as having a central core surrounded by increasingly declining activity (for further discussion of this phenomenon, see Mittelstaedt, Grossbart and Curtis 1977). While the summary map includes all blocksides on which customers perceived activities, it is actually somewhat smaller than the definition of the area provided by retailers. Retailers defined the downtown as extending over 280 blocksides, containing 311 facilities and having a center located at 3rd and Locust. An analysis of customer versus retailer definitions of the area revealed that (1) the ratio of the mean extent of customer maps to retailer defined extent was .28, (2) the ratio of the mean number of facilities perceived by customers to number defined by retailers was .16 and (3) thirty-seven percent of customers perceived the center of the downtown to be located at the intersection designated by retailers. Thus, on the average, customers defined the downtown as covering twenty-eight percent of the area as defined by retailers and containing sixteen percent of retailer-defined facilities while almost two-thirds of customers regarded the center of the downtown as being somewhere other than that indicated by retailers.

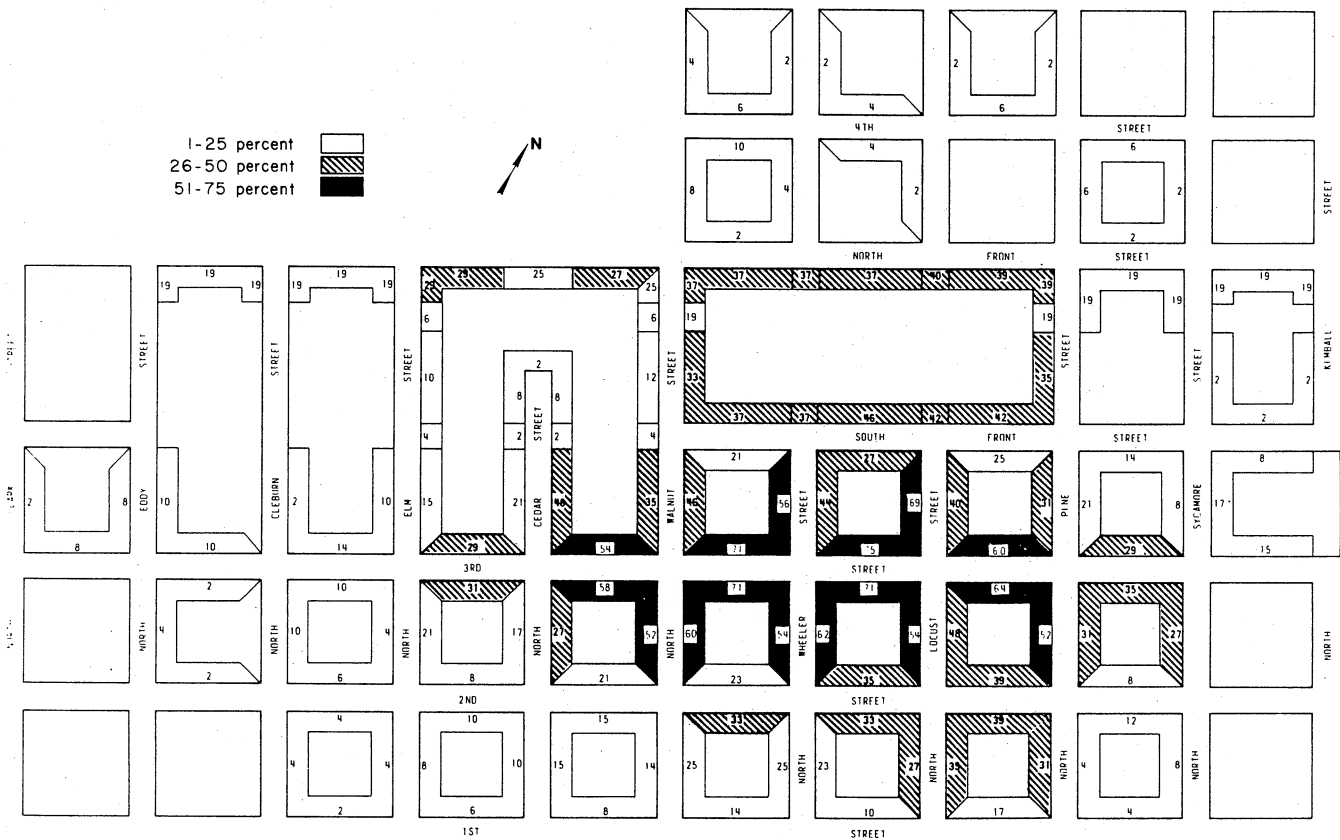
Map Characteristics and Beliefs Regarding Convenience

Correlations between map characteristics and beliefs regarding shopping convenience are presented in Table 3. The expected pattern of associations may be found in Table 1. Three of four predicted associations between map characteristics and beliefs concerning travel convenience are significant. All cognitive map measures, rather than just the two relative measures, are related to beliefs regarding parking convenience. The expected pattern of results are present for beliefs about customer movement, aggregate and display convenience. In the case of aggregate convenience, however, the relationships are inverse rather than direct. Comments before (by 50 percent of the sample) and after (by 79 percent) the structured portion of the interview made reference to downtown stores relocating elsewhere and thereby reducing the mix of available goods and services. It may be that general awareness of a diminution of the retail base coupled with absolute knowledge of the identity and functions of downtown establishments leads shoppers to believe that there is less likelihood of being provided with aggregate convenience.

Conclusions and Implications

In general, the results are consistent with predicted outcomes and the underlying rationalization regarding the function of cognitive maps in shopping. Nevertheless, simple correlations do not provide indications of causality. Further research in other cities with larger samples appears justified. As such investigations are conducted it will be possible to develop and test a multivariate causal model of cognitive mapping, shopping convenience and customer behavior. Still, the present findings underscore the potential usefulness of cognitive mapping theory and techniques to study customer beliefs about shopping areas. Moreover, they raise a fundamental question concerning whether the provision of more adequate retail information to shoppers might not alter their beliefs about convenience and, therefore, be a more efficient alternative than costly physical changes in downtown areas. While information may

FIGURE 1
Percentage of Respondents Perceiving Activities on Each Blockside in the Downtown Area



References

Bender, Wesley C. (1964), "Consumer Purchase-Costs -- Do Retailers Recognize Them?" *Journal of Retailing*, 40 (1), 1-8, 52.

Canter, David (1977), *The Psychology of Place*, New York: St. Martin's Press.

Cox, Reavis (1959), "Consumer Convenience and the Retail Structure of Cities," *Journal of Marketing*, 23, 355-362.

Crafton, Steven M. (1979), "Convenience Store Pricing and the Value of Time: A Note on the Becker-DeVany Full Price Model," *Southern Economic Journal*, 45, 1254-1260.

Downs, Anthony (1961), "A Theory of Consumer Efficiency," *Journal of Retailing*, 37 (1), 6-12, 50-51.

Downs, Roger M. (1970), "The Cognitive Structure of an Urban Shopping Center," *Environment and Behavior*, 2, 13-38.

_____, and Stea, David (1973), "Cognitive Maps and Spatial Behavior: Process and Products," in *Image and Environment: Cognitive Mapping and Spatial Behavior*, eds. Roger M. Downs, and David Stea, Chicago: Aldine, pp. 8-26.

_____, and Stea, David (1977), *Maps in Minds: Reflections on Cognitive Mapping*, New York: Harper & Row.

TABLE 3
Correlations of Forms of Shopping Convenience With Cognitive Map Characteristics

Cognitive Map Characteristics	Forms of Shopping Convenience				
	Travel	Parking	Customer Movement	Aggregate	Display
Absolute identity level	0.18 ^c	0.29 ^b		-0.29 ^b	0.26 ^b
Relative identity level	0.37 ^a	0.40 ^a	0.19 ^c		0.25 ^b
Absolute functional level		0.33 ^a		-0.25 ^b	0.26 ^b
Relative functional level	0.33 ^a	0.38 ^a	0.21 ^c		0.25 ^b

^ap ≤ .01 ^bp ≤ .05 ^cp ≤ .10

be only a partial substitute for alteration, its potential use certainly merits further study and experimentation.

- Fatouros, D. A. (1968), "The Physical Environment as a Field of Learning," Ekistics, 26, 520-521.
- Grossbart, Sanford L., Amedeo, Douglas, and Chinchon, David T. (1979), "The Influence of Retail Environments on Customer Cognitions and Feelings," in Proceedings: 1979 Educators' Conference of the American Marketing Association, eds. Neil Beckwith, Michael Houston, Robert A. Mittelstaedt, Kent B. Monroe, and Scott Ward, Chicago: American Marketing Association, pp. 268-273.
- Harrell, Gilbert D., and Hutt, Michael D. (1976), "Crowding in Retail Stores," MSU Business Topics, 24 (1), 33-39.
- Holahan, Charles J., and Dobrowolny, Mirilia B. (1978), "Cognitive and Behavioral Correlates of the Spatial Environment: An Interactional Analysis," Environment and Behavior, 10 (3), 317-333.
- Jakle, John A., Brunn, Stanley, and Roseman, Curtis C. (1976), Human Spatial Behavior: A Social Geography, North Scituate: Mass.: Duxbury Press.
- Kaplan, Stephen (1973), "Cognitive Maps in Perception and Thought," in Image and Environment: Cognitive Mapping and Spatial Behavior, eds: Roger M. Downs, and David Stea, Chicago: Aldine, pp. 63-78.
- Kelley, Eugene J. (1958), "The Importance of Convenience in Consumer Purchasing," Journal of Marketing, 23, 32-38.
- Kelly, Robert F., and Stephenson, Ronald (1967), "The Semantic Differential: An Information Source for Designing Retail Patronage Appeals," Journal of Marketing, 31 (4), 43-47.
- Kunkel, John H., and Berry, Leonard L. (1968), "A Behavioral Conception of Retail Image," Journal of Marketing, 32 (4), 21-27.
- Lindquist, Jay D. (1974-1975), "Meaning of Image: A Survey of Empirical and Hypothetical Evidence," Journal of Retailing, 50 (4), 29-38, 116.
- Lynch, K. (1960), Image of the City, Cambridge, Mass.: MIT Press.
- MacKay, David B., and Olshavsky, Richard W. (1975), "Cognitive Maps of Retail Locations: An Investigation of Some Basic Issues," Journal of Consumer Research, 2, 197-205.
- Markin, Rom J. (1977), Retailing Management, New York: Macmillan.
- Mazze, Edward M. (1974), "Determining Shopper Movement Patterns by Cognitive Maps," Journal of Retailing, 50 (3), 43-48.
- Mehrabian, Albert, and Russell, James A. (1974), An Approach to Environmental Psychology, Cambridge, Mass.: MIT Press.
- Mittelstaedt, Robert A., Grossbart, Sanford L., and Curtis, William W. (1977), "Consumer Perceptions and Retail Mapping: Research Findings and Preliminary Theory," in Consumer and Industrial Buying Behavior, eds. Arch G. Woodside, Jagdish N. Sheth, and Peter D. Bennett, New York: North-Holland.
- Moore, Gary T. (1979), "Knowing About Environmental Knowing: The Current State of Theory and Research on Environmental Cognition," Environment and Behavior, 11, 33-70.
- Pathak, Dev S., Crissy, William J. E., and Sweitzer, Robert W. (1974-1975), "Customer Image versus the Retailer's Anticipated Image," Journal of Retailing, 50 (4), 21-28, 116.
- Rich, Stuart U., and Portis, Bernard D. (1964), "The Imageries of Department Stores," Journal of Marketing, 28 (2), 10-15.
- Rosenberg, Gerhard (1965-1966), "City Planning Theory and the Quality of Life," The American Behavioral Scientist, 9, 3-7.
- Stokols, Daniel (1978), "Environmental Psychology," Annual Review of Psychology, 29, 253-295.

A METHOD FOR DETERMINING THE SEQUENCING OF COGNITIVE PROCESSES IN JUDGMENT:
ORDER EFFECTS ON REACTION TIMES¹

John G. Lynch, University of Florida

Abstract

This paper proposes a method for testing hypotheses asserting that to make a judgment, *y*, the consumer must first make another implicit judgment, *x*. The method examines reaction times for subjects to explicitly judge both *x* and *y*, varying the order and the temporal spacing of the 2 judgments. An illustrative experiment is reported that used the method to study the "false consensus effect"--the tendency for persons to attribute their own opinions and behavior to others. Because of an apparatus problem, faith should not be placed in the data. They are presented only to demonstrate the data analysis procedures employed by the method.

Testing Models of Cognitive Process

It is often the case that theories in consumer behavior and related disciplines postulate that one judgment, *y*, is mediated by another implicit judgment, *x*. For example, attribution theory (Calder & Burnkrant 1977) postulates that to judge whether a consumer's purchase of a particular brand tells us something about his or her personal dispositions, one must first make a subjective assessment of whether most consumers would have made a similar choice in that situation. However, a judgment of the commonness of choosing a particular brand does not necessarily lead to a subsequent judgment of the personal characteristics of those who buy it.

A question arises as to how a researcher might test a theory that postulates the sort of fixed sequence of mental processes described above (*y* does not necessarily follow *x*, but *y* is necessarily preceded by *x*). Typically, the best one can do is to marshal evidence that *x* and *y* are highly correlated--an unsatisfactory solution to the problem. One could try to improve upon correlation by establishing that *x* precedes *y* in time. Unfortunately, if *x* and *y* are steps in a sequence of cognitive processes, they may occur too close together in time to be able to easily establish the temporal precedence of *x*. An alternative approach would be to manipulate *x* and show predicted effects on *y*. However, this would show only that *x* has effect on *y*. It would not show that a judgment of *x* is autonomously made whenever *y* is judged, so that *x* is a necessary precondition of *y*. In short, the standard tools of casual inference fail to provide the leverage one needs to test hypothesized relationships between mental events *x* and *y* such that a judgment of *y* implies that one must first make an implicit judgment of *x*--yet such theories are not at all uncommon in information processing research. This paper will propose a method for dealing with this general class of problems by examining order effects on latency to respond to questions about *x* and *y*. To illustrate the application of the

¹The preparation of this manuscript was supported by the Center for Consumer Research, the Center for Econometrics and Decision Sciences, and the College of Business Administration, University of Florida. Thanks are extended to Ed Shoben for valuable contributions to the conceptualization and design of this research, and to Dipankar Chakravarti and Joel Cohen for helpful comments on earlier drafts of this manuscript.

method, a flawed experiment is presented that concerns the mental processes underlying judgments of the opinions and preferences of "most people" or "the average person."

Self vs. Other in Consumer Inference-making and Judgment

A pervasive theme in theories of consumer behavior is that a consumer's perceptions and behavior are affected by two broad classes of variables: a) his or her personal beliefs, attitudes, and preferences, and b) perceptions of the beliefs, attitudes, and preferences of generalized and specific others. For example, Burnkrant and Cousineau (1975) contrast informational social influence, in which a person is privately persuaded by the arguments of others, with normative social influences, in which one complies with the judgment of others while remaining privately unconvinced, because of one's concern for how the others would react to one's compliance or noncompliance. A similar self/other duality appears in attribution theory (Calder & Burnkrant 1977), which claims that the perceived cause of a consumer's behavior depends upon how consistent her actions are with the sort of behavior one would expect of most people or the "average person" in a similar situation. Fishbein and Ajzen's (1975) model asserts that a consumer's intention to buy a product is influenced by his personal attitude toward the act of buying the product, and his perception of whether important others think he should or should not make the purchase. Belk's (1976) analysis of gift-giving could be loosely construed as suggesting that when a consumer has no explicit knowledge of the recipient's preferences, he might first decide whether he personally liked the gift, then try to assess whether his tastes and the recipient's were likely to be similar in the product class being considered.

False Consensus

Germane to the discussion above is work by Ross and his colleagues (Ross 1977; Ross, Greene, and House 1977), showing that individuals tend to over-estimate the proportion of others who would make the same behavioral choices they would make (e.g., watch a favorite TV show or prefer a particular brand of coffee), and to underestimate the commonness of choices contrary to their own. Ross et al have dubbed this tendency the "false consensus effect." These researchers placed subjects in a variety of imaginary and real situations in which they were forced to choose between two alternatives, A and B. Subjects were asked to estimate the percentage of persons in the general population who would choose each option, and to indicate which choice they themselves would make. Results showed that both persons who chose A and those who chose B believed themselves to be in the majority. Other experiments showed that the false consensus effect extended into the belief domain. Subjects attributed opinions to the "average person" that tended to agree with their own.

Possible Sources of False Consensus

One might be tempted to interpret the false consensus effect as evidence that judgments of what others think are based upon one's own opinions--to judge what the average person would say, one considers one's own opinion, then generalizes from oneself to the average person. Such a conclusion would be unwarranted, since previous research on the false consensus effect was purely correlational. The correlation between (*x*) one's own choice of A or B and (*y*) the perceived commonness of choosing A might, like any correlation, be

taken as evidence that a) x caused y, that b) y caused x, c) that both were influenced by some third factor, z, or d) that x and y are part of some bi-directional causal system. Below are some of Ross et al's (1977) speculations about possible sources of false consensus. In the discussion that follows, it should be remembered that "cause" refers to "implicational causation"--x causes y if a judgment of y implies that the person must first activate knowledge structures pertaining to x.

"Independent Judgment" Explanations ($z \rightarrow x$, and $z \rightarrow y$). A number of explanations of the false consensus effect are consistent with an "independent judgment" model. One such account involves selective exposure. Ross et al note that we tend to associate with friends who share our beliefs and values to a greater extent than these are shared in the general population. Our friends do behave as we do in many instances. The false consensus effect, then, may reflect the fact that our experience leads us to see greater homogeneity in the world than actually exists, rather than any intentional or unintentional distortion of one's social reality to agree with one's own opinions or behavior. This explanation is consistent with an "independent judgment" model because a judgment of one's own opinion and a judgment of the "average person's" opinion activate independent cognitions. A person need not consider his or her own opinion to judge that of the average person, nor does a judgment of one's own opinion necessitate the activation of knowledge structures about what others think.

"Self-referent Inference" Explanations ($x \rightarrow y$). A second family of causal accounts of the false consensus effect suggests that perceptions of the commonness of various behaviors and opinions are the product of the subject's own choice of behavior or opinion. In general, information about oneself is more extensive and better integrated than information about others, especially if it is tied to an elaborate self-schema (Markus 1977). Self-schemata facilitate both the encoding of new information and the retrieval of information already stored in memory. For these reasons, it may well be that consumers use their own behavior, opinions, etc., as a basis for making judgments of the opinions of generalized and (probably to a lesser extent) of specific others.

"Other-referent Inference" Explanations ($y \rightarrow x$). A third possibility is that judgments of one's own opinions or preferences are based upon a consideration of what generalized or specific others think. For example, in a choice situation that is fairly novel, a consumer may search for information in memory about what relevant others thought or did in similar situations. Or, if questioned about one's opinion about a topic about which little is known, one may rely on information in memory about what some other person had said about it. Thus, a judgment of one's own opinion might require the activation of cognitions pertaining to what others have said or would say.

A Method for Assessing Implicational Causation

Based upon the evidence presented by Ross et al (1977), any one of the causal mechanisms discussed above might plausibly explain the false consensus effect. Of course, a complete explanation might include several of these mechanisms, and bi-directional causality may be involved. This paper proposes a method for deducing the primary direction of causality underlying the correlation between subjects' choices and their consensus estimates. The method is generally applicable to problems in which two judgments, x and y (in this case, one's own agreement or disagreement with opinion statements and estimates of whether others would agree) are found to be correlated, and one is concerned with whether one is causal in the sense that an implicit judgment of x must be made in order to make a judgment of y. The method involves an analysis of the time it takes a subject to answer questions about x and y as a function of the order in which the questions and the delay

between the two questions.

In the experiment to be reported, subjects were shown a series of opinion statements, one by one, on a computer terminal. Each was presented twice during the experiment, a) once preceded by an instruction to judge one's personal opinion (which will be hereafter called a "Self" judgment) and b) once preceded by an instruction to judge the average University of Illinois student's opinion (which will be called an "Other" judgment). Subjects responded to each statement by pressing either a key marked "AGREE" or one marked "DISAGREE." Let us consider what processes of judgment would be predicted in the various conditions by our 3 causal "theories," and how reaction times would be affected by these processes.

Components of Reaction Times for 3 Theories

The "self-referent inference" model decomposes the time to make a Self judgment into 2 parts: time to read and Comprehend the opinion statement (designated c) and time to retrieve and integrate knowledge relevant to determining Self's opinion (designated s).² A judgment of Other's opinion is assumed to require both of these same stages (c + s), plus a third stage during which the subject Generalizes from Self to Other (designated g_{os}). This last stage may involve testing for instances of disagreement, or deciding whether one's opinion about the relevant issue is likely to be typical.

The "other-referent inference" model follows a similar logic. Here a judgment of Other's opinion is assumed to involve a 2-stage process of reading and Comprehending the opinion statement (c), and retrieving and integrating knowledge pertaining to Other's opinion (designated o). The judgment of Self's opinion is assumed to involve a 3-step process: the same 2 stages (c + o) assumed to be required to judge Other's opinion, plus a 3rd stage involving Generalization from Other to Self (g_{os}).

The "independent judgment" theory concurs with the self-referent theory's account of Self judgments, and with the other-referent theory's account of Other judgments. That is, Self judgments require one to read and Comprehend the opinion statement, then retrieve and integrate knowledge relevant to judging Self's opinion (total reaction time = c + s). Other judgments also require 2 stages: a stage entailing the reading and Comprehension of the opinion statement, and one involving the retrieval and integration of knowledge pertaining to Other's opinion (reaction time = c + o).

Given the 3 different process accounts of how one judges one's own or the average person's opinion, a method to competitively support one of the models was sought. The problem was attacked by measuring latencies to respond "agree" or "disagree" to opinion statements. a) Subjects made both "Self" and "Other" judgments of each statement, but b) for half the statements, Self judgments were made prior to Other judgments, and the remaining statements were judged in the order Other First, Self Second, and c) the First and Second judgment of each opinion statement were separated by either a Short Delay or a Long Delay. As will be seen, the use of this experimental design with a dependent measure of reaction time permits a clear discrimination among the self-referent, other-referent, and independent judgment accounts of Self and Other judgments. Following is a discussion of the rationale underlying the "order effects or reaction times" paradigm in the general case.

²A third component, motor response time to physically execute a keypress, is assumed by all 3 theories to be constant across experimental conditions, so it will be ignored.

Logic of the "Order Effects on Reaction Times" Paradigm in the General Case

Suppose that a judgment of y requires an implicit determination of x, plus an additional stage like our Generalization stage, (g_{xy}), whereas a judgment of x requires only the former stage. If so, judging either x or y at time 1 (First judgment) will facilitate (i.e., reduce the latency of) making the opposite judgment at time 2 (Second judgment, Short Delay) or at time 3 (Second judgment, Long Delay). The facilitating effect results because, in making the Second judgment, the subject need not retrieve and integrate information pertaining to x, but may simply retrieve the integrated determination of x from memory. The average magnitude of the facilitating effect depends upon the likelihood that the subject will be able to (a) successfully retrieve the integrated determination of x, rather than (b) repeat the entire process of retrieving and integrating knowledge about x. The likelihood of being able to follow the shorter path (a) is an interactive function of 2 factors: 1) The delay between the First and Second elements of x-y or y-x pairs. If the delay is short, the integrated determination of x is virtually certain to be in short term memory, enabling the subject to follow path (a). In Long Delay conditions, though, it is less than certain that subjects will be able to retrieve the integrated determination of x from long term memory. 2) In Long Delay conditions, then, the probability of following the shorter path (a) on the Second judgment is positively related to the degree of elaboration of x required by the First judgment. A y-first judgment requires one to judge x, then process x further, whereas an x-first judgment requires no further processing after x has been determined. By a depth of processing rationale (Craik & Lockhart 1972; Craik & Tulving 1975), an x-First judgment should not be as helpful to a y-Second Long Delay judgment as a y-First judgment would be to an x-Second Long Delay judgment. The upshot of this is that the difference between Short and Long Delay versions of an x-Second judgment should be smaller than the difference between Short and Long Delay versions of a y-Second judgment.

Complicating this analysis somewhat is the fact that Second judgments would be expected to be faster than First judgments solely on the basis of practice effects on the speed of the reading and Comprehension (c) stage. This stage would be postulated to be faster for Second judgments made at time 2 (c_2 , occurring in Short Delay conditions) than for First judgments made at time 1 (c_1). By time 3, though, some of the benefit of prior exposure to the opinion statement should have dissipated, so c_3 (in Second judgments, Long Delay conditions) should be of intermediate length (i.e., $c_1 > c_3 > c_2$). The assumptions above allow us to derive predictions for the 3 theories of the false consensus effect.

Self-referent Theory. The self referent theory's predictions can be summarized briefly by referring to Table 1:

1. A main effect of judgment type should occur. Other judgments should take longer than Self judgments, mainly because the former involve an additional processing step of Generalizing from Self to Other (g_{so}).

2. A main effect of ordinal position should occur. First judgments of an opinion statement should take longer than Second judgments for two reasons: because of the greater Comprehension time (c) and the greater time to determine Self's opinion (s) associated with First judgments in a (Self-Other or Other-Self) pair.

3. If only Second judgments are analyzed, a significant interaction of opinion judged (Self vs. Other) and Delay should be observed. Based upon a depth of processing rationale, Other Second judgments should be more adversely affected by a long delay than should Self Second judgments, so $s_1 > s_3'' > s_3' > s_2$ in Table 1.

TABLE 1

Components of Reaction times predicted by "self-referent" "other-referent" and "independent judgment" models of implicational causation underlying the false consensus effect.

Self-referent

	Short Delay		Long Delay	
	Self Judgment Time 1	Other Judgment Time 1	Self Judgment Time 1	Other Judgment Time 1
First Judgment	$c_1 + s_1$	$c_1 + s_1 + g_{so}$	$c_1 + s_1$	$c_1 + s_1 + g_{so}$
Second Judgment	Time 2 $c_2 + s_2$	Time 2 $c_2 + s_2 + g_{so}$	Time 3 $c_3 + s_3'$	Time 3 $c_3 + s_3'' + g_{so}$

- c_1 : reading and comprehension time, first presentation.
- c_2 : reading and comprehension time, second presentation after short delay.
- c_3 : reading and comprehension time, second presentation after long delay ($c_1 > c_3 > c_2$).
- s_1 : time to retrieve and integrate self-relevant information to make Self judgment.
- s_2 : time to retrieve integrated Self judgment from short term memory.
- s_3' : average length of a stochastic process whereby subject either retrieves integrated Self judgment from long term memory, or repeats process s_1 .
- s_3'' : same as s_3' , except probability of having to repeat process s_1 is higher ($s_1 > s_3'' > s_3 > s_2$).
- g_{so} : time to generalize from Self to Other (infer implications of Self judgment for Other judgment).

Other-referent

	Short Delay		Long Delay	
	Self Judgment Time 1	Other Judgment Time 1	Self Judgment Time 1	Other Judgment Time 1
First Judgment	$c_1 + o_1 + g_{os}$	$c_1 + o_1$	$c_1 + o_1 + g_{os}$	$c_1 + o_1$
Second Judgment	Time 2 $c_2 + o_2 + g_{os}$	Time 2 $c_2 + o_2$	Time 3 $c_3 + o_3'' + g_{os}$	Time 3 $c_3 + o_3'$

- c_1, c_2 , and c_3 : same as above.
- o_1 : time to retrieve and integrate other-relevant information to make Other judgment.
- o_2 : time to retrieve integrated Other judgment from short term memory.
- o_3' : average length of a stochastic process whereby subject either retrieves integrated Other judgment from long term memory or repeats process o_1 .
- o_3'' : same as o_3' , except probability of having to repeat process o_1 is higher, ($o_1 > o_3'' > o_3 > o_2$).
- g_{os} : time to generalize from Other to Self.

Independent Judgment

	Self Judgment Time 1	Other Judgment Time 1	Self Judgment Time 1	Other Judgment Time 1
	First Judgment	$c_1 + s_1$	$c_1 + o_1$	$c_1 + s_1$
Second Judgment	Time 2 $c_2 + s_1$	Time 2 $c_2 + o_1$	Time 3 $c_3 + s_1$	Time 3 $c_3 + o_1$

- c_1, c_2 and c_3 : same as above
- o_1 : same as above (time to retrieve and integrate other-relevant information to make Other judgment).
- s_1 : same as above (time to retrieve and integrate self-relevant information to make Self judgment).

Other-referent Theory. Based upon a logic similar to that used to derive predictions of the self-referent theory, the following predictions of the other-referent model can be deduced, as shown in Table 1:

1. A main effect of opinion judged should occur. Self judgments should take longer than Other judgments, mainly because the former require an additional step of generalizing from Other to Self (g_{os}).
2. A main effect of ordinal position is predicted. First judgments of an opinion statement should take longer than Second judgments because of the greater Comprehension time (c) and greater time to determine Other's opinion (o) associated with First judgments.
3. By a depth of processing rationale, one can predict that the difference in RT between Self Second, Long Delay and Self Second, Short Delay cells should be greater than the difference between Other Second, Long Delay and Other Second, Short Delay cells ($o_1 > o_3^1 > o_3^2 > o_2$).

Independent Judgment Model. According to the independent judgment theory, knowledge relevant to determining one's own opinion and that relevant to determining Other's opinion exist in separate locations in memory, although both may have been influenced by some common third factor or complex of factors when originally stored. Self judgments are assumed to require subjects to Comprehend the opinion statement (c), then retrieve and integrate knowledge pertaining to Self's opinion (s). Other judgments are assumed to require a Comprehension stage (c), plus a stage involving the retrieval and integration of knowledge pertaining to Other's opinion (o). Retrieving and integrating knowledge relevant to making a Self judgment would not facilitate a later Other judgment, nor would the reverse be true. As shown in Table 1, Second judgments would be faster than First judgments simply because of the shorter time it takes to read and Comprehend (c) the opinion statement the second time around.

1. It is unclear whether a main effect of judgment type should occur. If time to retrieve and integrate information pertaining to Self's opinion (s) is less (or greater) than the time to retrieve and integrate knowledge pertaining to Other's opinion (o), a main effect will result.
2. A main effect of ordinal position will result, solely because of the greater time needed to read and Comprehend (c) the opinion statement during First judgments than during Second judgment.
3. When only Second judgments are analyzed, there should be no interaction between judgment type (Self vs. Other) and Delay.³

³These assertions assume that knowledge structures pertaining to Self's opinion and those pertaining to Other's opinion are completely independent. It is possible, though, that the basic idea underlying the independent judgment model is correct--Self judgments require no implicit consideration of Other's opinion, nor is the reverse true--but cognitions about certain facts must be activated whether one is judging one's own opinion or that of the average student. Such a quasi-independent judgment model makes relatively complicated predictions (see Lynch and Shoben, in preparation) not germane to the application of the order effects on reaction times paradigm in most consumer information processing settings. Hence, it will not be discussed further.

Method

Overview

In the present experiment, 66 undergraduate subjects were seated at cathode ray tube (CRT) computer terminals, on which a series of 32 opinion statements was presented. The statements concerned a variety of topics: energy (e.g., "Nuclear power is too dangerous to be a desirable source of energy"), consumer products ("Most breakfast cereals are worthless from a nutritional standpoint"), leisure activities ("It is more enjoyable to read a book than to watch television"), etc. Each statement was presented twice, a) once preceded by an instruction to judge "YOUR PERSONAL OPINION" ("Self" judgment), and b) once by an instruction to judge the "AVERAGE U OF I STUDENT'S OPINION" ("Other's judgment), for a total of 64 judgments. Subjects responded to each statement by pressing either a key marked "AGREE" or one marked "DISAGREE" as quickly as possible while still maintaining a high level of accuracy. Both latency to respond and the response itself were recorded for each trial. Half of the opinion statement pairs were presented in the order Self First, Other Second, and the other half were presented in the order Other First, Self Second. Additionally, the Second judgment in a given pair was separated from the First by either a Short Delay (1-5 trials) or a Long Delay (separated from the First member of the pair by an interpolated task).

The design, then was an Opinion Judged (Self vs. Other) x Order of Judgment (Self First, Other Second vs. Other First, Self Second) x Delay (Short vs. Long) within subjects design, with 8 items randomly nested in each Order x Delay condition. A different random nesting of items was used for each subject. To simplify the exposition of the experiment, it will be treated as an Opinion Judged x Ordinal Position (First vs. Second) x Delay factorial.

Procedure

Subjects, run in groups of 3 to 5, were told that the experiment was concerned with a) whether they personally agreed or disagreed with each of a set of opinion items, and b) whether they thought that the average student at the University of Illinois agreed or disagreed with the same statements. It was explained that 32 opinion statements would be judged once in each condition, for a total of 64 judgments, but that the order of these judgments would be "scrambled."

Each trial was structured so that an instruction phrase flashed onto the screen, telling the subject to judge either "YOUR PERSONAL OPINION" or "AVERAGE U OF I STUDENT'S OPINION" on the upcoming trial. After subjects had read the instruction phrase, they pressed the space bar on their keyboard. This action caused the instruction phrase to be erased, and an opinion statement to be written on the screen. Subjects were told to read the opinion statement as quickly as possible, to determine what they believed to be the correct response, and to press the appropriate key. Subjects were told that not only were their answers being recorded, but also the time it took them to respond. They were instructed to "respond as quickly as possible without sacrificing accuracy."⁴

⁴Cognitive psychologists employing reaction time as a dependent variable typically give subjects similar instructions to discourage them from trading off speed of response for accuracy. In applications in which one can verify the correctness of a subject's responses, it is imperative that differences in mean RT among experimental conditions are not spuriously caused by differential levels of accuracy (Pachella 1974). In the present context, however, there is no objective basis for assessing the "accuracy" of subjects' judgments, nor is it clear what "accuracy" might mean if it could be assessed.

The experimental session began with 8 practice trials, during which subjects were given feedback about their reaction times. If no questions arose, the practice trials were followed by 48 experimental trials, during which no RT feedback was given. After these trials (all short delay trials, plus the Self First and Other First trials in the long delay condition), subjects performed a 15 minute interpolated task, then completed the final block of 16 long-delay trials, again with RT feedback. Upon completion of the experiment, subjects were thoroughly debriefed and informally questioned about their processing strategies.

Procedural limitations. The experiment described was flawed by a problem with the apparatus. Although the computer was programmed to generate a different random assignment of opinion statements to experimental conditions for each subject, there is evidence that randomness was not achieved. A chi-square test of association between the 32 opinion statements and experimental conditions was significant at $p < .001$. Thus, differences in cell means may be attributable to differences in the content of the opinion statements that appeared in those cells. The results of the experiment are presented for illustrative purposes only, and should not be taken to be empirically valid.

Results and Discussion

Reaction times, with the additive effects of items and answers partialled out, were analyzed in a Subjects x Opinion Judged x Order (Self First, Other Second vs. Other First, Self Second) x Delay x Items (nested in Subjects, Order, and Delay) ($66 \times 2 \times 2 \times 2 \times 8$) completely within subjects analysis of variance. Items as well as subjects were considered random, in keeping with Clark's (1973) contention that linguistic materials must be treated as random. Data will be reported, however, as an Opinion Judged x Ordinal Position (First vs. Second) x Delay completely within S factorial. Mean reaction times in milliseconds are shown in Table 2.

TABLE 2

Effects of Opinion Judged (Self vs. Other), Ordinal Position (First vs. Second), and Delay (short vs. long) on Reaction Times (in msec) to answer "agree" or "disagree" to opinion statements.

	Short Delay Between 1st & 2nd Judgment		Long Delay Between 1st & 2nd Judgment	
	Self Judgment	Other Judgment	Self Judgment	Other Judgment
First Judgment	3922	4064	3982	4083
Second Judgment	3130	3134	3171	3292

Overall, Self Judgments were faster ($M = 3551$ msec) than Other judgments ($M = 3642$ msec), $F(1,65) = 8.84$ $p < .01$. Thus the data appear to be consistent with either the self-reference model or the independent judgment model, but not with the other-referent model. Thus the data appear to be consistent with either the self-reference model or the independent judgment model, but not with the other-referent model. Even so, one might ask whether averaging across subjects obscures individual differences in processing strategies. Perhaps some small subgroup of subjects made judgments consistent with an other referent process, using their knowledge of what the average Other might say to make Self judgments. If so, ordinal violations of the self-reference theory's

prediction that Self First RTs should be faster than Other First RTs should be associated with ordinal violations of the prediction that Self Second RTs should be faster than those for Other Second judgments. A chi-square test of the independence of these violations in individual subject data revealed no evidence of contingency ($\chi^2(1) = 1.36$, $p > .10$). Ordinal patterns in the data, then, suggest that subjects used a common cognitive process, although they apparently differ in the time they devoted to the various stages in this process. Significant Subjects x Ordinal Position (First vs. Second), and Subjects x Ordinal Position x Delay interactions ($F(65,1848) = 2.70$ and 1.56 ($p < .001$), respectively), suggest such individual differences. The data discussed up to this point allow the rejection of the other referent theory and seem to support either the self-reference model or the independent judgment model.

Inspection of the data in Table 2 reveals that Other Second RTs show a 158 msec difference between the short and long delay conditions, while Self Second RTs show only a 41 msec increase, consistent with the self-reference, depth of processing theory. A statistical test of this difference requires that only Second judgments be analyzed in a Subjects x Opinion Judged x Delay x Items (nested in Subjects, Opinion Judged and Delay) ANOVA. The critical interaction does not reach statistical significance, $F(1,65) = 2.45$ $p > .05$. If this experiment were not flawed by non-random assignment of items to experimental conditions, one would probably prefer the independent judgment model to the self-referent model based on these data.

Methodological Issues in Employing the Order Effects Paradigm

One problem of the "order effects on reaction times" paradigm as proposed is that a causal theory (in this case, self-reference) and an independent judgment model make different predictions in only one cell out of 8. Given this high variability inherent in reaction times to consumer stimuli, it is difficult to generate the kind of statistical power that is needed to discriminate two theories whose predictions are so similar. This would have been a problem in the experiment reported, had the data been taken seriously.

A second drawback of the paradigm is conceptual. To distinguish a causal theory (like self-reference) from an independent judgment model, one has to assume a "depth of processing" notion. Cognitive theorists have recently expressed doubts about the usefulness of Craik & Lockhart's (1972) levels of processing construct (cf. Baddeley 1978), and therefore, the use of the construct in the paradigm proposed by this paper may be unwarranted. To generate the predictions of the self-reference theory, it was assumed that deeper processing of one's own opinion occurred in Other First judgments (in which one implicitly judges Self's opinion, then decides whether it would generalize to others) than in Self First judgments (in which one implicitly judges Self's opinion, then overtly reports one's judgment). Lynch and Shoben (in preparation) discuss methods for testing such assumptions.

Conclusions

Despite the unresolved problems associated with using the "order effects on reaction times" paradigm, it seems uniquely suited to testing process models that postulate

that a judgment of y is mediated by an implicit judgment of x. Following are some examples of issues in consumer information processing that could be investigated through the use of the paradigm: (a) Fishbein and Ajzen's (1975) assumption that a judgment of behavioral intention implicitly involves the consideration by subjects of whether they think the behavior is good or bad, and of whether most people who are important to them think that they should perform the behavior. (b) The assumption of certain "brand-based" choice theories (see Bettman 1979, for examples) that choice from among a set of alternative brands requires overall evaluations of the brands to be made. (c) Hypotheses about the effects of inferential beliefs on evaluations of products (cf. Cohen, Miniard and Dickson 1979).

Researchers who attempt to use this general paradigm will encounter complications that are unique to the particular cognitive process and structure assumptions of the theory being tested. If one exercises care and creativity, though, the general approach of analyzing RTs to answer questions about each stage in a hypothesized chain of mental events as a function of the order of the questions and the delay separating the questions promises to be a powerful aid to the study of consumer information processing.

References

- Baddeley, A.D. (1978), "The Trouble with Levels: A Re-examination of Craik and Lockhart's Framework for Memory Research," Psychological Review, 85, 139-152.
- Belk, R.W. (1976), "It's the Thought that Counts: A Signed Digraph Analysis of Gift-giving," Journal of Consumer Research, 3, 155-162.
- Bettman, J.R. (1979), An Information Processing Theory of Consumer Choice, Reading, Mass.: Addison-Wesley.
- Burnkrant, R.E. & Cousineau, A. (1975), "Informational and Normative Social Influence in Buyer Behavior," Journal of Consumer Research, 2, 206-215.
- Calder, B.J. & Burnkrant, R.E. (1977), "Interpersonal Influence on Consumer Behavior: An Attribution Theory Approach," Journal of Consumer Research, 4, 29-38.
- Clark, H.H. (1973), "The Language as Fixed Effect Fallacy: A Critique of Language Statistics in Psychological Research," Journal of Verbal Learning and Verbal Behavior, 12, 335-359.
- Cohen, J.B., Miniard, P.W., & Dickson, P. W. (1980), "Information Integration: An Information Processing Perspective," in Advances in Consumer Research, Vol. 7, ed., J. C. Olson, Ann Arbor, MI: Association for Consumer Research, 161-170.
- Craik, F.I.M. & Lockhart, R.S. (1972), "Levels of Processing: A Framework for Memory Research," Journal of Verbal Learning and Verbal Behavior, 11, 671-684.
- _____ & Tulving, E. (1975), "Depth of Processing and the Retention of Words in Episodic Memory," Journal of Experimental Psychology, 104, 268-294.
- Fishbein, M. & Ajzen, I. (1975), Belief, Attitude, Intention, and Behavior, Reading, Mass.: Addison-Wesley.
- Markus, H. (1977), "Self-schemata and Processing Information about the Self," Journal of Personality and Social Psychology, 35, 63-78.

Pachella, R. (1974), "The Interpretation of Reaction Time in Information-Processing Research," in Human Information Processing: Tutorials in Performance and Cognition, ed. B. Kantowitz, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Ross, L. (1977), "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in Advances in Experimental Social Psychology, ed. L. Berkowitz, Vol. 10, N.Y.: Academic Press.

_____, Greene, D., & House, P. (1977), "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes," Journal of Experimental Social Psychology, 13, 279-301.

Reference Note

Lynch, J. G. & Shoben, E., "A Causal Analysis of the False Consensus Effect Using Order Effects on Reaction Times." Manuscript in preparation.

THE MODERATING EFFECTS OF COGNITIVE COMPLEXITY AND PRIOR PRODUCT FAMILIARITY ON THE PREDICTIVE ABILITY OF
SELECTED MULTI-ATTRIBUTE CHOICE MODELS FOR THREE CONSUMER PRODUCTS

Chin Tiong Tan, University of Singapore
Ira J. Dolich, University of Nebraska-Lincoln

Abstract

Respondents were classified according to their levels of cognitive complexity and product familiarity with three consumer products: automobiles, apartments, and toilet soaps. Using repeated measures and the repertory grid technique with ten brands and ten dimensions for each product class, five multi-attribute models are tested for predictive accuracy of rank order brand preferences. Three sets of 2x2x5 ANOVA studies showed significant systematic relationships between type of model and level of cognitive complexity and significant differences attributed to models across all products and to cognitive complexity for two of the three products.

Introduction and Theoretical Background

Multi-attribute choice models have been used in marketing to understand a host of buying related behavior such as attitude structure and judgment. Considerable effort has been devoted to using these models to predict brand preferences or choices.

Several exploratory studies (Park and Sheth 1975, Park 1976, and Menasco 1976) have attempted to identify selected individual characteristics or situational variables that may influence model usage and to investigate how these characteristics and variables affect judgment. Such a research direction should be encouraged, as it enhances an in-depth understanding of the interaction between brand choice, individual characteristics and situational differences.

The notion that a person acquires product familiarity while progressing through different stages of the learning curve is well documented in the consumer behavior literature (Howard and Sheth 1969). Sheth and Raju (1973) further proposed that different choice models, some more complicated than others, may be used at these different learning stages. In other words, consumers with different levels of familiarity with a product may rely on different types of brand choice models. This issue of product familiarity and use of choice model is a major consideration of the present study.

Marketers have made numerous attempts to link individual characteristics such as personality variables to buying behavior. In this study, the personality variable of interest is the complexity-simplicity of one's cognitive structure. Treating one's cognitive style as a personality structure is common in psychology (Schroder 1971), and its relationship with processing behavior is well documented in the works of Schroder, Driver and Streufert (1967). In that tradition, this personality characteristic is used as a moderating variable to examine its effect on the prediction of choice models.

This study centers on three specific consumer behavior issues: personality traits, choice models and familiarity with products. The major research objectives are to test propositions about the moderating effects of cognitive complexity and product familiarity on brand preferences derived from multi-attribute choice models.

Cognitive Complexity

Cognitive structure is viewed as a system of attributes or dimensions which form the basis of one's perceptual and judgmental processes. Kelly (1955), Bieri et al. (1966), and Schroder et al. (1967) showed that persons differ in cognitive structure and that such differences result in response variations. The relationships and interdependencies of dimensions within these systems bring about certain properties or characteristics that are considered unique to a person. Cognitive complexity is one such characteristic.

Cognitive complexity has been defined differently by psychologists (refer to Bieri (1971) for a review). The present study adopts the definition of Bieri (1955)-namely, the extent to which a person uses dimensions in a differentiated manner to construe objects. A complex person evaluates objects in a more multi-dimensional way and has a more differentiated set of dimensions available than does a simple person.

Psychologists have found cognitive complexity to mediate a wide range of behaviors such as attitude change (Lundy and Berkowitz 1957), information processing (Petronkon and Perlin 1970), salience of dimensions (Mueller 1974), and stereotyping (Koenig and King 1962). Marketers have recently investigated the construct. Among them, Park and Sheth (1975) found mixed results with the effect of cognitive complexity on choice behavior. The present study re-examines how cognitive complexity affects accuracy of prediction for several judgment models.

Product Familiarity

The extent to which a person is familiar with a product class depends largely on factors such as prior knowledge, usage, and purchase of brands in the product class. Marketers conceptualize one's familiarity to be a multi-dimensional concept comprised of elements such as awareness, knowledge, and actual usage of brands from the product class. Thus far, empirical studies on prior familiarity with products have been limited. Park (1976) found it to influence types of judgmental rules. However, another study showed familiarity to have little effect on brand preferences (Monroe 1976). The present study investigates how one's prior product familiarity may affect multi-attribute choice model efficiencies in predicting brand preferences.

Choice Models

Most of the marketing studies utilizing multi-attribute choice models disregarded individual differences and situational variations which might affect their usage. It is only recently that marketing researchers have begun to look to individual and situational differences as causes of the poor results found in some studies. Sheth and Raju (1973) have postulated that different processing rules for brand choices are used in different contexts. If such is the case, multi-attribute models, such as Fishbein's (1967) and Rosenberg's (1956), when used across different product classes for all types of consumers, would inevitably lead to weak results in some instances. The weighted linear-compensatory rule of their models cannot be assumed to hold for all buyers for all purchases.

The generalized model assumes that evaluations of a brand are based on multiple attributes that vary in importance. Thus, one's brand choice is related to his overall evaluation which is determined by aggregating individual evaluations adjusted for differences in importance.

This research effort utilizes four versions of the summated model and the lexicographic model as described below:

1. Unweighted summative model I (full version) -- subject's ratings on ten attributes are added.
2. Unweighted summative model II (partial version) -- subject's ratings on the five most important attributes are added.
3. Weighted summative model I (full version) -- subject's ratings on ten attributes are weighted by importance values and summed.
4. Weighted summative model II (partial version) -- subject's ratings on the five most important attributes are weighted again by importance ratings and summed.
5. Lexicographic model -- a sorting procedure in which the attributes are ranked according to importance, and alternative brands evaluated sequentially starting with the single most important attribute downward. If a tie arises, successive dimensions are used. The sorting process continues until a rank ordering of alternatives is completed.

This study, then, investigates how these various choice models may provide different predictive accuracy for persons with different characteristics and familiarity with products. Specifically, the major research objectives are to test the following hypotheses:

- H₁: The predictive accuracy of selected choice models differs between cognitively complex and cognitively simple persons.
- H₂: The predictive accuracy of selected choice models differs between persons with different levels of product familiarity.

Research Methods

Product Selection

Three products were required to differ in importance and to have multiple brands or options that were generally known to the respondents. Automobiles, rental apartments and toilet soaps were shown to meet the above criteria using a group of respondents similar to the final sample.

Research Instruments

Respondents indicated brand knowledge by the percentage of brands checked on a list of brands provided for the purpose. A person's familiarity with a product class was measured as the proportion of brands in the product class that one knew something about.

A new research approach was adopted to operationalize the choice models. Instead of imposing a standard set of attributes (dimensions) for evaluation of brands, each respondent was asked to provide his/her own attributes. Such a procedure should partially eliminate the problem of non-salient attributes. The brands (stimuli) to be evaluated were also self-selected from the larger list of brands. Arbitrary responses to unknown brands were thereby avoided.

Kelly's Repertory grid (Kelly 1955) was used to organize attributes and then to rate brands. In essence, it is a matrix which captures a person's repertoire of self-generated dimensions used in the evaluation of objects. Table 1 shows such a grid.

Cognitive complexity is computed using the matching of dimensions procedure as proposed by Bieri and colleagues

(1966). In brief, a cognitively simple person, when using the grid, would not be able to use many of the dimensions in a differentiated manner, and hence would evaluate the objects identically. The ratings in the grid are compared with one another for agreement on objects to determine the number of identical ratings. Every dimension is compared with all others. In such a manner, the entire repertoire of dimensions is examined to determine one's relative complexity. A simple person will have a high number of matches (similar to identical semantic differential profiles). A low score therefore denotes a more differentiated structure and cognitively more complex. Table 2 summarizes the results of this procedure using the data from Table 1.

Data Collection

Respondents were upperclass undergraduate students and each was paid two dollars for participation. Ninety-six persons formed the final sample. They were arranged in small group settings and detailed instructions were presented verbally and in writing with special care taken to explain the Repertory Grid routines. A repeated measure design was used in that each subject provided evaluations for all three product classes. Product presentation in the questionnaire booklets was randomly distributed to prevent order effects. Three Repertory Grids were included in each booklet.

The grid was a 10 X 10 matrix. Each respondent was instructed to pick the ten most familiar brands from a list of twenty-two and to write the brand names at the column headings. Starting from row one to row ten, the respondent was to list ten different attributes/dimensions he or she commonly used to evaluate brands of that product. Each dimension was then to be used with a six-point Likert type bipolar scale having values of +3, +2, +1, -1, -2, -3 to rate all ten brands on the grid. To assist in the generation of dimensions, three randomly pre-selected cells were circled on each row of the grid. The subject was told to look at the three brands and to consider a way, dimension, or characteristic for which two of them were similar and yet different from the third. Each dimension generated was written down on the side of the row and the procedure continued until ten bipolar dimensions were elicited. (See Kelly 1955, and Bieri et al. 1966 for details on the procedures).

In addition, preference rankings for the ten brands and importance rankings for the ten self-generated dimensions were collected. The sequence of tasks was identical for each product class.

Analyses and Results

Each subject's rank ordered brand choice prediction was derived for each model and correlated with the independently obtained preference ranking. The Spearman Rank Order Correlation was used as a measure of the accuracy of prediction of the models. In other words, the "goodness of fit" criterion was used to compare predictive ability of competing models. These correlation coefficients were then used as the dependent variables in the various analysis of variance studies.

For each product class, frequency distributions of the subjects' cognitive complexity scores and product familiarity indices were determined. The distributions were then dichotomized into the high (upper) and low (lower) halves for analyses. The final ANOVA model for each product class is a 2 X 2 X 5 repeated measure design with two fixed factors (2 levels of cognitive complexity and 2 levels of product familiarity) and one repeated factor (5 levels, one for each choice model).

TABLE 1
A Completed Rep Grid for Automobiles

Brand	1	2	3	4	5	6	7	8	9	10	
	Nova	VW Beetle	Impala	Torino	Dodge Dart	Pinto	Corvette	Mustang	LaSabre	Skylark	
<u>Dimension Positive Pole</u>											<u>Negative Pole</u>
1. Not expensive	+2	+2	+1	-1	+1	+3	-3	+1	-3	+2	Too expensive
2. Nice body shape	+1	-3	+2	+3	+2	-2	-1	+3	+1	+2	Ugly shape
3. Good gas mileage	+3	+3	+1	+1	+2	+2	-3	+2	-3	+2	Bad gas mileage
4. Lots of space	+1	-3	+3	+2	+2	-2	-2	+2	+3	+2	Little space
5. Compact in size	+2	+3	-1	-1	-1	+3	+2	+1	-3	+1	Too big
6. Smooth ride	+2	-3	+2	+2	+2	+1	+1	+1	+2	+2	Rough ride
7. Well manufactured	+1	-1	+2	+2	+2	-2	-3	+1	+2	+1	Cheaply made
8. Dependable	+2	-2	+2	+2	+2	-2	-3	+1	+3	+2	Not dependable
9. Engine starts easy	-1	-2	+2	+2	+2	-2	-1	+1	+3	+2	Hard to start
10. Safe	+1	-3	+3	+2	+2	-1	-3	+2	+3	+2	Unsafe

TABLE 2
Computation of Cognitive Complexity Score From the Completed Grid

Matching Scores

1-2 1, 2-3 2, 3-4 3, 4-5 0, 5-6 2, 6-7 5, 7-8 6, 8-9 8, 9-10 4,
1-3 4, 2-4 5, 3-5 2, 4-6 4, 5-7 2, 6-8 5, 7-9 5, 8-10 5,
1-4 1, 2-5 0, 3-6 2, 4-7 4, 5-8 2, 6-9 5, 7-10 4,
1-5 5, 2-6 4, 3-7 2, 4-8 5, 5-9 1, 6-10 4,
1-6 3, 2-7 4, 3-8 3, 4-9 5, 5-10 0,
1-7 2, 2-8 4, 3-9 2, 4-10 8,
1-8 4, 2-9 5, 3-10 4,
1-9 2, 2-10 4,
1-10 2,

Note: Matching scores were computed by comparing each dimension in the grid with all possible dimensions. When comparing dimension 1 with dimension 2, the number of identical ratings was one, hence a matching score of 1. The number of matches between dimension 1 and dimension 3 was four, between dimension 1 and dimension 4 was one, and so forth. The larger the matching score between two dimensions, the greater the similarity in ratings between them. Cognitive complexity score is the summation of all the matching scores. In this case, it is 154.

Table 3 summarizes the main and interaction effect significance tests for each product class. Mean scores are shown in Table 4.

Product Familiarity

Product familiarity as defined herein was not a statistically significant main effect. The overall group means for the high and low familiarity subjects are rather close. However, the high familiarity group consistently shows slightly higher mean scores than the low familiarity group across all three product classes. No further individual comparison is performed.

TABLE 3
Analysis of Variance Summary Table of Significance Tests For Automobiles, Apartments and Toilet Soaps

	Automobiles	Apartments	Toilet Soaps
Complexity (C)	.01	.05	NS
Familiarity (F)	NS	NS	NS
Model (M)	.001	.001	.001
C X F	NS	NS	NS
M X C	.05	.05	NS
M X F	NS	NS	NS
M X C X F	.05	NS	NS

TABLE 4
Mean Correlation Scores for High and Low Complexity Subjects for Five Models and Three Products

Cognitive Complexity	Automobiles		Apartments		Toilet Soaps	
	High	Low	High	Low	High	Low
Model 1 Unweighted (10)	.70	.42 ¹	.66	.46 ²	.65	.62
2 Unweighted (5)	.58	.46 ²	.58	.43 ¹	.63	.58
3 Weighted (10)	.69	.51 ¹	.65	.51 ²	.69	.63
4 Weighted (5)	.56	.48	.58	.47 ²	.66	.63
5 Lexicographic	.50	.36 ¹	.31	.33	.55	.54
Total	.61	.46 ¹	.55	.43 ²	.63	.60

¹p = .01

²p = .05

Cognitive Complexity

It was previously proposed that the predictive accuracy of selected choice models differs between cognitively complex and cognitively simple persons. And, the ANOVA results shown in Table 3 indicate significant complexity main effects for two of the three product classes. Since familiarity with product class was nonsignificant for all three products, this main effect was eliminated for presentation in Table 4.

The significance tests on cognitive complexity can be seen more clearly from Table 4 where overall mean differences for automobiles and apartments are shown to be statistically significant at 0.01 and 0.05 levels, respectively. These individual comparisons highlight differences among the five models and explain the significant main effect shown in Table 3. In the two product classes where the complexity factor is significant, four out of five individual comparisons are significant as well. All of the comparisons for toilet soaps are insignificant.

Choice Models

All three product classes show significant main effects on choice models (Table 3). This issue has been considered by many researchers and evidence of variability in predictive accuracy of different models has been well documented (Wilkie and Pessemier 1973). In general (Table 4), the weighted summative models (3 and 4) provide higher correlations than the unweighted models, (1 and 2) the full version models (1 and 3) show higher correlations than the partial version models (2 and 4), and the summative models (1 through 4) give more accurate results than the lexicographic model.

Interactions

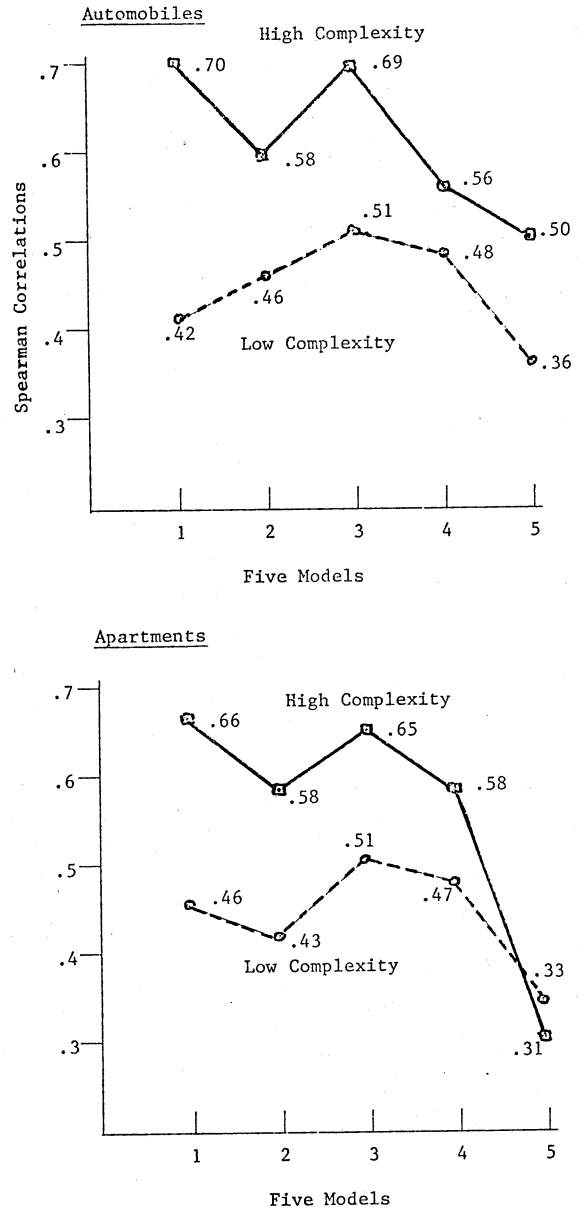
Of the various interactions, the one that shows up well is between type of model and level of complexity. Both automobiles and apartments show significant interactions. In addition, the three-way interaction is significant for the automobiles. These model-complexity interactions are presented graphically in Figure One. As can be seen, the high and low complexity subjects do not provide similar results across the five models.

Discussion

The present study found respondents differing in cognitive complexity to provide differences in accuracy of prediction for five choice models. The cognitively complex subjects appear to have consistently better predictions than the simple subjects, irrespective of the types of choice model considered. This may be due to the complex persons' more differentiated cognitions which enable them to see things in a more multi-dimensional fashion. For them, the more refined and differentiated evaluations of brands (full models) resulted in choice rankings which corresponded closely to actual preferences. Conversely, cognitively simple persons may lack the ability to see differences in brands. Brand choice predictions based on the evaluation of such dimensions may not realistically reflect actual preferences for this latter group.

For the cognitively simple subjects, the weighted summative models consistently out-perform the others. In the case of the cognitively complex subjects, the unweighted model appears to be the best predictor. In general, the differences between the full and the partial models appear to be more substantial for the cognitively complex subjects. One explanation may be that the additional attributes are meaningless to the cognitively simple subjects due to their simplistic cognitive processes, whereas, more attributes enable the cognitively complex subjects to further differentiate and thus, predict more accurately.

FIGURE 1
Model X Complexity Interaction



Cognitive complexity as a moderating variable is found here to be significant in the two product classes of highest importance. When the product is less important and rather simple such as toilet soap, no difference in predictive accuracy is found. In other words, the cognitively simple subjects seemed to be as capable as the cognitively complex subjects in evaluating stimulus objects that are inherently simple. Of the three products, the former subjects have the highest predictive accuracy in toilet soap ($r_s = 0.60$) which is significantly higher than the other two product classes ($r_s = 0.46$ and 0.43 for automobiles and apartments respectively). For the latter subjects, differences among the three products are not statistically significant ($r_s = 0.61, 0.55$ and 0.63 for automobiles, apartments and toilet soaps, respectively).

Prior familiarity with the product class was not shown to be a moderating variable in choice model predictability in this study. Persons with high and low product familiarity did not show any significant differences in the predictive accuracy of the models studied. An alternative explanation is that evaluation of brands is related more to cognitive structure than to product familiarity. Hence, a person, regardless of his/her level of brand familiarity, can still generate meaningful brand evaluations through unique cognitive capacities. Such a theoretical rationale has support. Bieri (1971) and Hall (1966) have claimed one's cognitive style to be a consistent trait manifested in different situations rather than domain specific (experience related) characteristic.

Conclusion

This research, although exploratory in nature, does provide a new perspective to a high interest area in marketing. Introducing moderating variables to the study of multi-attribute models and brand choice unveils significant details which are otherwise unnoticed.

This research is an attempt to refine our use of multi-attribute models. Of two variables with high potential as moderators, one -- cognitive complexity -- was found to be a significant factor. Persons of different levels of cognitive functioning were found to provide differences in predictive accuracy for five multi-attribute choice models.

References

- Belk, R. (1975), "Situational Variables and Consumer Behavior," Journal of Consumer Research, December.
- Bieri, J. (1955), "Cognitive Complexity-Simplicity and Predictive Behavior," Journal of Abnormal and Social Psychology, 51.
- _____, (1971), "Cognitive Structures in Personality," in H. M. Schroder and P. Suedfeld, (ed.) Personality Theory and Information Processing, New York: Ronald Press.
- _____, Attins, A. L., Briar, S., Leaman, R. L., Miller, H., and Tripodi, T. (1966), Clinical and Social Judgement: The Discrimination of Behavioral Information, New York: Wiley and Sons.
- Fishbein, M. (1967), "Attitudes and Prediction of Behavior," in M. Fishbein (ed.), Readings in Attitude Theory and Measurement, New York: John Wiley and Sons, Inc.
- Hall, M. F. (1966), The Generality of Cognitive Complexity-Simplicity, an unpublished Ph.D. Dissertation, Department of Psychology, The Vanderbilt University.
- Higgins, J. C. (1959), "Cognitive Complexity and Probability Preferences," unpublished manuscript, Department of Psychology, University of Chicago.
- Howard, J. and Sheth, J. (1969), The Theory of Buying Behavior, New York: Wiley and Sons, Inc.
- Kasulis, J. and Zaltman, G. (1976), "Message Reception and Cognitive Complexity," in Perreault, W. (ed.), Advances in Consumer Research, Vol. IV.
- Kelly, G. A. (1955), Psychology of Personal Constructs, New York: Norton Press.
- Koening, F. W. and King, M. (1962), "Cognitive Simplicity and Prejudice," Social Forces, 40, March.
- _____, and _____ (1964), "Cognitive Simplicity and Out-Group Stereotyping," Social Forces.
- Lundy, R. and Berkowitz, L. (1957), "Cognitive Complexity and Assimilative Projection in Attitude Change," Journal of Abnormal and Social Psychology, 55.
- Meansco, M. (1976), "A Further Exploration of the Modern Effects of Cognitive Complexity Upon Consumer Choice Behavior," Working Paper on the Bureau of Business and Economic Research, University of Iowa, July.
- Monroe, K. (1976), "The Influence of Price Differences and Brand Familiarity on Brand Preferences," Journal of Consumer Research, June.
- Mueller, W. W. (1974), "Cognitive Complexity and Salience of Dimension in Person Perception," Australian Journal of Psychology.
- Nakanishi, M. and Bettman, J. (1974), "Attitude Models Revisited: An Individual Level Analysis," Journal of Consumer Research, December.
- Park, W. and Sheth, J. (1975), "Impact of Prior Familiarity and Cognitive Complexity on Information Processing Rules," Communication Research, Vol. 2, No. 3, July.
- Petronko, M. and Perin, T. (1970), "A Consideration of Cognitive Complexity and Primacy - Recency Effects in Impression Formation," Journal of Personality and Social Psychology, Vol. 15.
- Rosenberg, M. J. (1956), "Cognitive Structure and Attitudinal Effect," Journal of Abnormal and Social Psychology, Vol. 53.
- Schroder, H. M. (1971), "Conceptual Complexity and Personality Organization," in Schroder and Suedfeld, (eds.), Personality Theory and Information Processing, New York: Ronald Press.
- Schroder, J., Driver, M., and Streufert, S. (1967), Human Information Processing, New York: Holt, Rinehart and Winston.
- Sheth, J. and Raju, P.S. (1973), "Sequential and Cyclical Nature of Information Processing Models in Repetitive Choice Behavior," in Wards, S. and Wright, P. (eds.), Advances in Consumer Research, Vol. 1.
- Stiles, G. (1974), "Determinants of Industrial Buyer's Level of Information Processing: Organizations, Situations and Individual Differences," in D. Hughes and M. Ray (eds.), Buyer/Consumer Information Processing, Chapel Hill, North Carolina: University of North Carolina Press.
- Tripodi, T. and Bieri, J. (1966), "Cognitive Complexity, Perceived Conflict, and Certainty," Journal of Personality, 34.
- Wilkie, W., and Pessemier, E. (1973), "Issues in Marketing's Use of Multi-Attribute Attitude Models," Journal of Marketing Research, November.
- Wilson, D. T. and Tan, C. T. (1977), "Dimensional Complexity of Cognitive Structure: A Personality Trait in Decision Making," in Stolen, J. D. and Conway, J. J. (eds.), Proceedings of the American Institute of Decision Sciences.
- Wright, P. (1975), "Consumer Choice Strategies: Simplifying Vs. Optimizing," Journal of Marketing Research, February.
- Zanjonc, R. B. (1968), "Cognitive Theories in Social Psychology," in The Handbook of Social Psychology, Vol. 1, Reading, Mass.: Addison-Wesley.

TOWARD A COGNITIVE STRUCTURE CONCEPTUALIZATION OF PRODUCT FAMILIARITY¹

Larry J. Marks, Pennsylvania State University
Jerry C. Olson, Pennsylvania State University

Abstract

Product familiarity is conceptualized in terms of the cognitive structures of knowledge concerning the product that are stored in memory. Such knowledge could have been derived from direct or indirect experiences with the product in question (e.g., product use vs. advertisements or word-of-mouth). Different amounts and types of product experience should be reflected by variation in cognitive structure. Theoretically, differences in cognitive structures should influence the cognitive processes and outcomes that involve those cognitive structures and, thereby, should also affect overt behavior. Data from an existing study are used to illustrate these ideas. Suggestions for future research are offered.

Introduction

Concepts such as product familiarity, product experience or expertise, and/or prior information have been popular as mediating variables in many models of consumer behavior. However, despite the frequency with which product familiarity has been proposed as an individual difference variable affecting choice behavior, the concept has not yet been clearly conceptualized in a compelling manner.

The resulting confusion is apparent in the variety of terms researchers have used to describe this general concept. For example, Berelson and Steiner (1964) found that "pre-existing information" was one of several predisposing factors in determining audience receptivity to "congenial and noncongenial messages". Park (1976) measured product familiarity in terms of subjects' agreement with statements designed to operationalize Howard and Sheth's (1969) concepts of extensive, limited, and routinized problem solving. Based on the Bayesian concept of prior distribution, Woodruff (1972) used subjects' evaluations of a brand-attribute combination and their uncertainty about this rating to operationalize "prior information" about brand attributes. Lastovicka (1979) measured "knowledge about the product class" by asking subjects if they could "talk about a general group of products for a long time." He also measured "remembered personal experience" by subjects' responses to "I can remember having purchased something in this general group of products." Raju and Reilly (1979) measured product familiarity in terms of subjects' self-reported "frequency of use, overall familiarity, and knowledge of how to select the best brand." These studies exemplify the diverse and occasionally vague approaches taken in much of the past research on "familiarity." Thus far, this research has emphasized developing predictive models of consumer choice, where measures of product familiarity are one element in such models. Unfortunately, this approach has not produced a clear conceptualization of the construct of product familiarity.

The purposes of this paper are to propose a broad concep-

tualization of product familiarity, defined in terms of the cognitive structures of product knowledge that are derived from past product-related experiences, and to present initial data regarding its usefulness and validity. Our broad objectives are to begin to develop a better understanding of how product experiences are represented in cognitive structure, and of how cognitive structures influence the cognitive processes involved in handling product information (e.g., making purchase choice decisions).

Product Familiarity as Cognitive Structure

Product familiarity has often been operationalized in terms of product-related behavior, perhaps most often in terms of self-reported product usage experience or frequency of use (cf. Anderson, Engledow and Becker 1979, Jacoby, Chestnut and Fisher 1978). An information processing approach, however, provides a different perspective. From this point of view, the construct of interest is not the past product experience *per se*, but the cognitive representations of that experience that are stored in memory (see Bettman 1979, Olson 1978b, Russo and Johnson 1980). These representations can be considered to be organized in memory as a product-related cognitive structure or schema. Product schemas contain knowledge in the form of coded representations of brands, product attributes, usage situations, general product class information, as well as evaluations and choice rules. This coded information can be considered to be organized or linked together and stored in memory as a structural framework of knowledge (Norman and Bobrow 1975). Thus, product cognitive structures or schemas may contain not only factual knowledge about the product, but also evaluations or affect, and in well developed schemas, purchase criteria and even decision rules or strategies (Olson 1978a).

Given this information processing perspective, product familiarity can be defined in terms of consumers' cognitive structures for a product that were presumably caused by various product-related experiences. Consumers with different amounts or types of product-related experience probably have acquired cognitive structures with differing characteristics. These variations in cognitive structure are assumed to cause the observed behaviors previously attributed to the vague concept of "product familiarity." Specific features or characteristics of cognitive structure are addressed later.

Although this line of reasoning may be conceptually compelling, it will not have much impact on our research unless we can develop reliable and valid measures which tap, at least relatively directly, these cognitive structures for products and brands. Before addressing operational issues, however, consider the developmental aspects of product familiarity.

Effects of Product-Related Experiences on Cognitive Structure

Several authors have called for a theoretical explanation regarding how product-related experiences create cognitive structures and/or influence existing cognitive structures (cf. Olson and Mitchell 1975, Russo and Johnson 1980). Hayes-Roth (1977) presented such a conceptual framework which seems useful for considering the effects of product

¹This research was funded by a corporate grant to Peter D. Bennett, whose inputs are gratefully acknowledged along with the insightful comments of two anonymous ACR reviewers. We also thank Aydin Muderrisoglu and Robert Humberston for conducting the experimental sessions.

experience on cognitive structure (see Olson and Dover 1978). According to her model, processing information about a product creates the elementary coded representations of knowledge--the concept nodes or cogits--of a product schema. As more product-related experiences occur, these coded representations are "strengthened" in memory and associations between the knowledge nodes develop and become stronger (i.e., learning occurs). Eventually, presumably after multiple experiences with the product, the configuration of associated representations has developed to a point that it can be considered a well-developed, reasonably stable structure.

Moreover, substructures of the now "well-developed" schema may be recoded, a chunking or "unitization" process (cf. Miller 1956, Simon 1974) by which several separate components of the cognitive structure come to be treated (cognitively) as a single representation. Thus, this newly-created concept or chunk represents several other concepts. Chunking can be considered as a process of forming abstract concepts by creating more general concepts to represent several more concrete concepts. According to Hayes-Roth (1977), a chunk usually can be cognitively decomposed into its "original" components, given appropriate instructions and processing, although under ordinary circumstances it would be treated as a unitized representation.

Consumers who are quite familiar with a product due to multiple past experiences might be expected to have formed a stable, complex cognitive structure (or set of structures) of product knowledge, perhaps containing multiple chunks (Olson and Dover 1978). At least one published study provides indirect support for the chunking propositions discussed above. Edell and Mitchell (1978) found that subjects with experimentally-manipulated prior experience regarding a previously unfamiliar product category generated fewer total thoughts in response to an advertisement for a brand in that product class than did subjects with little or no prior experience. From the perspective discussed above, the experienced subjects may have generated ad-stimulated thoughts about more abstract concepts--chunks if you will, "bigger" concepts with "extra" meaning--and thus they produced fewer total responses. Certainly this interpretation is only speculative; however, as a partial test of this notion, subjects' cognitive responses could be analyzed in terms of their level of abstractness or concreteness.

Effects of Cognitive Structure on Information Processing

It is generally accepted that people use information from long-term memory to perceive, interpret, and store new information (cf. Bettman 1979, Mitchell 1978, Olson 1978b, 1980). From the perspective advocated here, product-related schemas are activated and provide the necessary cognitive structure for interpreting product information or new product experiences (see Olson 1978a). In this sense, cognitive structures have a fundamental influence on the processes of attention, encoding, evaluating, storing and using information. For instance, differences in memory schemas should be evidenced by variations in how consumers process information for which those cognitive structures are relevant, whether the information is contained in ads, for instance, or is derived from product use experiences. Thus, consumers' cognitive responses to an advertisement (Wright 1973) could provide clues about how cognitive structures influence encoding processes (Olson 1980). In sum, peoples' existing cognitive structures should exert a powerful influence on their cognitive processes and on their subsequent overt behavior.

Perhaps the major stumbling block in doing research on cognitive structures is the measurement problem. How does one measure the content and organization of a memory schema for a product? This issue is too large for a thorough discussion here. Briefly, however, Olson (1978b) has advocated "free elicitation" as a possible measurement technique. In this procedure, subjects are given cue stimuli or memory probes (usually words) and asked to "Tell me everything that comes to mind when I say..." Presumably, the verbal responses elicited in response to product and/or situation-specific probe cues can provide (admittedly imperfect) indicators of the content and perhaps the organization of cognitive structures.² In our research program thus far, we have had some success with this technique (see Kanwar, Olson and Sims 1981, Olson and Muderrisoglu 1979). A simple free elicitation procedure was used in this research to provide indices of two characteristics of a product cognitive structure.

Research Purpose

The study described below serves only as an initial, exploratory investigation of the ideas discussed above. It is exploratory because the research was designed primarily for other purposes; however, part of the original design produced data that are relevant for our objectives. Specifically, there are two broad issues of major interest. First, what differences in cognitive structures result from varying levels of product experience or familiarity? That is, do our presumed measures of cognitive structure reflect differences in product familiarity? Second, how do these cognitive structure variables affect subsequent information processing operations and the formation of product attitudes and purchase intentions?

Methods

Product and Subjects

The product was an innovative office chair. The chair had been designed to orthopedically accommodate and support the user's body in a manner that would be both healthful and comfortable. Moreover, the chair was intended to enhance the worker's task performance while working at a desk. The specific chair was totally unfamiliar to all subjects, although it was being marketed in other regions of the country. Subjects consisted of 34 business students (mostly MBA's with some senior undergrads) and 20 secretaries from the University. Subjects were volunteers selected on a convenience basis. They were paid \$4.00 for their participation.

Design

The part of the original study examined here can be represented as a simple 2 x 2 between-groups design. Two levels of product experience or familiarity (lower and higher) were represented by the student and secretary groups. The second factor was manipulated by exposing half of the subjects in each group (on a random basis) to a fairly elaborate sales promotion message for the desk chair; the remaining subjects received no product message. An after-only design was used because taking pre-exposure measures was considered to be too sensitizing, and besides, subjects

²Russo and Johnson (1980) use a thought protocol task similar to free recall tasks and the free elicitation task used here, but that may provide a somewhat different perspective on product cognitive structures.

had no pre-experimental knowledge of this particular chair that could be measured.

Product Familiarity. The main focus of interest is the cognitive structure representation of product familiarity. To maximize inferential rigor, it might have been preferable to experimentally manipulate familiarity, and thus cognitive structure. However, it is difficult to create major differences in product familiarity and related cognitive structures in the context of a controlled experiment. These variables develop naturally in the real world over relatively long periods of time and many experiences. Moreover, for the initial studies in a program of research on cognitive structure, it seemed most efficient to first establish whether differences in existing cognitive structures could be measured. Later, after we learn more about cognitive structures, we can make better attempts to manipulate them. Therefore, we chose to use the secretary and student groups as a classification variable, a surrogate "manipulation" of familiarity with office desk chairs. It seemed reasonable to assume that secretaries have greater amounts of experience and familiarity with office desk chairs than do students.

Exposure to Product Message. The second independent variable was exposure or not to a sales presentation for the new chair. Half of the subjects in each group saw a 6-minute slide show of professional quality, produced by the manufacturer, and accompanied by a synchronized tape-recorded sales presentation. In the promotion message, the product was presented as of high quality and innovative in design. Four major sales points were described verbally and portrayed visually (the chair is accommodating to one's body, health giving, task motivating, and comfortable).

Procedures

Subjects were run individually. To enhance their motivation and to provide a context for the experimental procedures, subjects were told to imagine that they were part of an employee committee at their company which was to recommend the type and style of furniture to be used in a remodeling of their offices. In fact, the chair is sometimes sold in this way. After completing an informed consent form, subjects in the exposure condition were shown the slide show presentation. Immediately after, they were asked to record "all the thoughts you had while watching the presentation." Then, concepts were elicited in response to three probe cues that were presented via the instruction, "tell me all the characteristics that come to mind when you think about buying an office desk, ... an office desk chair, ... a filing cabinet." Desk chair was always the second probe. The interviewer recorded all responses verbatim. Following the third probe, subjects rated their attitudes (A_0) and purchase intentions (BI) regarding the advertised chair.

The nonexposed subjects served as an internal control group to provide an index of peoples' normative cognitive structure regarding the three probe cues, uninfluenced by any persuasive message. After completing the informed consent form, these subjects received the free elicitation procedures. No other measures were taken (no cognitive responses, A_0 or BI).

Dependent Variables

Cognitive Structure. To measure differences in cognitive structure due to variation in product familiarity, we examined the results of the free elicitation instruction, "Tell me all the characteristics that come to mind when you think about purchasing an office desk chair." Empirical indicators of two cognitive structure characteristics were developed (see Kanwar, Olson and Sims 1981). The number of concepts elicited served as an in-

dicator of cognitive structure dimensionality and the percentage of the elicited concepts judged more "abstract" (rather than more concrete) was used as a measure of abstractness of cognitive structure.³ Whether a concept was considered abstract or concrete was based on experimenter judgment. Significant differences in these indicators were expected between the secretary and the student groups due to their presumed differences in product experience. However, exposure to the promotional message might also produce some variation.

Cognitive Processing. Immediately following the message subjects were given four minutes to write down "the thoughts they had while watching the presentation." These cognitive responses were coded in terms of the numbers of counterarguments, support arguments, and total thoughts produced. As discussed earlier, the cognitive responses of the more and less experienced groups were expected to vary because of their different product-related cognitive structures.

Attitude and Intentions. Attitude toward the chair (A_0) was measured by the mean of three bipolar, 5-point scales labeled good-bad, high quality-low quality, and like-dislike (coefficient alpha=.82). Behavioral intention (BI) was measured on a 7-point bipolar scale by asking the subjects how likely they would be to recommend purchase of the chair (not at all likely-very likely).

Results and Discussion

Check on Familiarity Classification

Intuitively, it seemed reasonable to assume that secretaries, by virtue of their daily experiences with office desk chairs, would be more familiar with the general product class than students. Thus, secretaries' cognitive structures regarding desk chairs should be different from those of students. To establish that their past experiences with desk chairs did indeed differ, we asked both secretaries and students to report several past behaviors. As expected, 100% of the secretaries had worked at a job "that involved desk work," while only 65% of the students had. Moreover, secretaries reported that 85% of their work time was spent working at a desk, compared to 67% for those students who had ever had a "desk job."

Differences in Cognitive Structure

Next, it is important to establish that the secretary-student distinction, as a surrogate "manipulation" of product familiarity, produced variation in our measures of cognitive structure. Empirically, differences between secretaries and students might be reflected in (a) the total number of salient characteristics of desks elicited, and (b) the level of abstraction of the elicited concepts. Moreover, we need to determine whether or not exposure to the persuasive sales material influenced these elicitation measures. To enhance the study's low power caused in part by the relatively small cell sizes (n 's = 17 and 10), and because the exploratory nature of this research makes us relatively more sensitive to Type II than Type I errors, we adopted an alpha level of .15 for statistical significance (cf. Cohen 1977).

³In another study, we obtained adequate levels of agreement for these two judgments between two independent coders ($r=.85$ for total number of thoughts; $r=.87$ for judgment of abstract vs. concrete; see Kanwar, Olson and Sims 1981)

⁴Two independent judges agreed on 85% of these judgments.

A 2-way, Familiarity by Exposure ANOVA showed that exposure to the sales promotion did not affect the number of concepts elicited ($p > .50$, M 's = 6.1 for exposed and 5.8 for not exposed, respectively), or the abstractness of the elicited concepts ($p > .20$, M 's = 44% and 36%). Although not statistically reliable, the modest increase in abstract concepts due to exposure was to be expected, since the sales promotion message did describe and portray several abstract ideas (regarding the comfort, healthfulness, and accommodating properties of the chair) that subjects probably were not aware of prior to message exposure. The ANOVA also revealed that the total number of elicitations for students was somewhat greater than that produced by the secretaries (M 's = 6.2 vs 5.2, $F = 6.60$, $df = 1/50$, $p = .01$). This difference is consistent with the proposition that greater product familiarity results in the chunking of cognitive elements, which in turn causes fewer verbalized responses to the elicitation probe. Implicit in this notion, however, is the requirement that those fewer, chunked concepts represent higher levels of abstraction. In fact, the presumably less familiar students did produce a lower percentage of abstract concepts than did the secretaries, but not significantly so (M 's = 41% and 34%, $p > .20$). In sum, the present data provide weak support that the surrogate "manipulation" of familiarity produced differences in cognitive structure.⁵

Effects of Cognitive Structure and Product Message

Next, we examine the effects that subjects' cognitive structures had on their responses to the promotional material. The following results are based only on those 17 students and 10 secretaries who were exposed to the sales promotion message.

Cognitive Responses. Consistent with Edell and Mitchell's (1978) results and the chunking proposition, we found that the presumably more experienced secretaries produced marginally fewer cognitive responses to the product promotion than the less experienced students (M 's = 7.0 vs 8.0, $F = 1.51$, $df = 1/26$, $p = .23$). Secretaries produced no more support arguments than students (M 's = 2.2 versus 2.0, $p > .70$), but did have fewer counterarguments (M 's = .4 versus 1.8, $F = 6.16$, $df = 1/26$, $p = .02$) than the students.

These results provide weak support for the notion that cognitive processing of a product message is influenced by one's product-related cognitive structures, which in turn reflect differences in product familiarity. By this perspective, secretaries may possess better integrated, more abstract, and presumably more effective product structures. Therefore, they might be expected to better understand and appreciate the rather complex information presented in the sales promotion message. Since the product was convincingly described as of high quality, secretaries would be expected to perceive it as such and support argue for the message, which they did, although not significantly more so than students. Moreover, secretaries would not be expected to find as many points with which to disagree in the apparently accurate product message. In contrast, students with less-well developed, product-related schemas, may not have clear criteria on which to judge the product. Thus, they might be expected to have some difficulty in relating the information in the adver-

⁵ Parenthetically, it might be noted that an office chair is not a highly complex product. Thus the degree of abstraction that logically can be expected may be somewhat limited. If so, the variance in the percentage of abstract concepts to be discovered might be rather small for a product of this type. For more complex products, differences in level of abstraction might be more apparent.

tisement to their personal experiences. If so, they would be more likely to disagree with the information and produce more counterarguments, which they in fact did. Alternatively, however, students may simply have been more cynical or critical than the secretaries, or may have felt greater "demands" to appear critical during the experimental task.

Attitude and Intentions Toward the Chair. Cognitive responses, especially counterarguments, have been shown to mediate the formation of attitudes and intentions (e.g., Olson, Toy and Dover 1978, Wright 1974). Therefore, it was expected that the secretaries would be more likely than the students to perceive the chair as being of high quality and to recommend its purchase. This was indeed the case. Secretaries had more favorable product attitudes than the students (M 's = 4.6 versus 4.1, $F = 3.03$, $df = 1/26$, $p < .10$). And, consistent with their more positive attitudes toward the chair, secretaries were more likely than students to recommend purchase of the chair (M 's = 6.0 vs 5.1, $F = 2.93$, $df = 1/26$, $p < .10$). These results lend additional credence to the presumed causal flow between cognitive responses, attitudes, and intentions (cf. Edell and Mitchell 1978, Olson, Toy and Dover 1978, Wright 1974).

Conclusions and Suggestions

Like many initial investigations of new conceptualizations and measures, the present research generates more questions than it answers. The paper's basic value is probably in raising the issue of how "product familiarity" is to be treated as a construct in consumer behavior theory. Although weak, we believe that the pattern of obtained results are sufficiently interesting to warrant further investigation.

Although the present data were derived from a larger study primarily intended to address other issues, this does not invalidate the obtained results. In particular, the elicitations were carried out about as we would do in a study specifically designed to investigate cognitive structure issues. There are, however, problems in interpreting the product familiarity/experience, student-secretary factor. Because levels of familiarity were not manipulated, a variety of other differences exist between the two groups besides their past experiences with desk chairs. These confoundings become serious if one or more of these other factors can account for the observed pattern of results.

A series of internal analyses revealed that neither age, nor sex, nor marital status could consistently account for the obtained results. A stronger rival explanation involves the experimental demand characteristics that may have been felt by the secretary subjects. Perhaps the secretaries felt that they were "expected" to react favorably to the positive information provided by the advertisement because they had job experience with the product class. While this could account for their favorable attitudes and purchase intentions, it is not clear why students would not also feel "obligated" to react favorably. Additionally, if the secretaries felt that their "expertise" was an issue, it seems likely that they would have sought other ways to demonstrate their expertise. Thus, compared to students, the secretaries should have produced more, not fewer, verbal elicitations, and more, not less, counterarguments in response to the product message. Perhaps the most compelling alternative explanation for the elicitation results is that students--especially MBA students--are rather verbal, and therefore likely to produce more concepts in a free elicitation task than secretaries. This explanation, however, cannot account for the other effects that were obtained. For instance, why should "high verbalness" have a negative effect on A_0 or BI ?

In summary, although of course it cannot be proposed with high confidence, the most parsimonious explanation for the entire range of obtained results is that members of the two groups had different cognitive structures. Interestingly, a similar study, recently published by Anderson and Jolson (1980), obtained some of the same patterns of results as found here. They found that advertisements for 35mm cameras containing relatively technical information produced more favorable beliefs and attitudes for more experienced consumers than for less experienced consumers. We would hypothesize that greater product experience produced more complex cognitive structures containing abstract concepts (chunks) and strong, clear interrelationships between concepts. Such memory schemas seem necessary in order to comprehend complex information in a meaningful way. Stated simply, complex information is more persuasive for people who are able to more fully comprehend it.

Future Research Directions

This study introduces the idea of conceptualizing product familiarity, product experience, and product expertise in terms of the cognitive structures produced by past product experience. Clearly, additional theorizing and empirical research is necessary before this conceptualization can be generally accepted for use in other research. In particular, we need a more explicit model of cognitive structure. What are the relevant characteristics of a product cognitive structure or memory schema? What measures can serve as empirical indicators of these constructs? How do these aspects of cognitive structure affect other cognitive processes such as the encoding/comprehension process or the information integration processes involved in judgment or decision making? In short, how is past experience represented in a cognitive structure? Answers to such questions will require an extensive research program focused on the construct validity of a model of cognitive structure.

We have begun such a program (cf. Olson and Dover 1978, Olson and Muderrisoglu 1979, Olson and Sims 1980) which recently has produced signs of progress (Kanwar, Olson and Sims 1981). Thus far, we have proposed three characteristics of cognitive structure: dimensionality, the number of salient concepts stored in memory, articulation, the number of discriminable categories along each dimension, and abstraction, the level of abstractness of the dimensions. In addition to free elicitations, the repertory grid procedure and a knowledge test appear to provide potential measures of these cognitive structure elements (see Kanwar, Olson and Sims 1981). These developments also indicate a need for further improvements in methodology in order to facilitate further research on cognitive structure.

The first stage of future research could use the three cognitive structure characteristics just mentioned to establish the aspects of cognitive structure that are related to varying degrees of product familiarity. It does not seem important at this time to attempt the actual manipulation of product familiarity. Rather, we should establish the differences in structure that are associated with differences in experience. Presumably, the constructs of dimensionality, articulation and abstraction vary not only across levels of familiarity, but also for types of products. If so, of course, the data patterns obtained in this preliminary research may not hold for very simple or very complex products.

Future research should attempt to replicate the present design in order to establish the generalizability of these results. For instance, differences in cognitive structures should be found between groups of subjects who

a priori have different, hopefully more extreme levels of product familiarity for several products of varying complexity. Once this step is accomplished, it is necessary to demonstrate that the measured differences in cognitive structure are related to differences in information processing behavior, to attitudes and behavioral intentions, and to overt behavior. While experimental simulations are useful for much of this research, actual field observations in a natural environment will be necessary to establish external validity.

Finally, research should focus on the developmental aspects of product familiarity created by product experience or consumer socialization or consumer education. Essentially, in a longitudinal study, subgroups of subjects could be brought to various levels of product familiarity through product experience. Both cognitive and behavior measures, which presumably would possess known reliabilities and validities by this time, should reflect the levels of product familiarity.

In sum, a wide variety of interesting questions regarding the cognitive structure representation of product familiarity can be addressed, once we have suitable criterion measures. Of course, these questions have not been answered by the present research. We have merely introduced the idea of conceptualizing product experience and product familiarity in terms of cognitive structure and have produced some data that point to the potential usefulness of this point of view. We hope that others will become interested in developing these concepts and measurement procedures further.

References

- Anderson, Rolph E. and Jolson, Marvin A. (1980), "Technical Wording in Advertising: Implications for Market Information," Journal of Marketing, 44, 57-66.
- Anderson, Ronald D., Engledow, Jack L., and Becker, Helmut (1979), "Evaluating the Relationships Among Attitude Toward Business, Product Satisfaction, Experience, and Search Effort," Journal of Marketing Research, 16, 394-400.
- Berelson, Bernard and Steiner, Gary A. (1964), Human Behavior: An Inventory of Scientific Findings, New York: Harcourt, Brace and World, Inc.
- Bettman, James R. (1979), An Information Processing Theory of Consumer Choice, Reading, Mass.: Addison-Wesley.
- Cohen, Jacob (1977), Statistical Power Analysis for the Behavioral Sciences, rev. ed., New York: Academic Press.
- Hayes-Roth, Barbara (1971), "Evolution of Cognitive Structures and Processes," Psychological Review, 84, 260-78.
- Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: John Wiley.
- Jacoby, Jacob, Chestnut, Robert W., and Fisher, William A. (1978), "A Behavioral Process Approach to Information Acquisition in Nondurable Purchasing," Journal of Marketing Research, 15, 532-44.
- Kanwar, Rajesh, Olson, Jerry C. and Sims, Laura (1981), "Toward Conceptualizing and Measuring Cognitive Structures," in Advances in Consumer Research, Vol. 8, ed. Kent Monroe, Ann Arbor: Association for Consumer Research.
- Lastovicka, John L. (1979), "Questioning the Concept of Involvement Defined Product Classes," in Advances in Consumer Research, Vol. 6, ed. William L. Wilkie, Ann Arbor: Association for Consumer Research, 174-9.

Miller, George A. (1956), "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," Psychological Review, 63, 81-97.

Mitchell, Andrew A. (1978), "An Information Processing View of Consumer Behavior," in Research Frontiers in Marketing: Dialogues and Directions, ed. S. C. Jain, Chicago: American Marketing Association, 178-83.

Norman, Donald A. and Bobrow, Daniel G. (1975), "On the Role of Active Memory Processes in Perception and Cognition," in The Structure of Human Memory, ed. Charles N. Cofer, San Francisco: W. H. Freeman.

Olson, Jerry C. (1978a), "Inferential Belief Formation in the Cue Utilization Process," in Advances in Consumer Research, Vol. 5, ed. H. Keith Hunt, Ann Arbor: Association for Consumer Research, 706-13.

_____ (1978b), "Theories of Information Encoding and Storage: Implications for Consumer Research," in The Effect of Information on Consumer and Market Behavior, ed. Andrew Mitchell, Chicago: American Marketing Association, 49-60.

_____ (1980), "Encoding Processes: Levels of Processing and Existing Knowledge Structures," in Advances in Consumer Research, Vol. 7, ed. J. C. Olson, Ann Arbor: Association for Consumer Research, 154-60.

_____ and Dover, Philip A. (1978), "Attitude Maturation: Changes in Related Belief Structures Over Time," in Advances in Consumer Research, Vol. 5, ed. H. Keith Hunt, Ann Arbor: Association for Consumer Research, 333-42.

_____ and Mitchell, Andrew A. (1975), "The Process of Attitude Acquisition: The Value of a Developmental Approach to Consumer Attitude Research," in Advances in Consumer Research, Vol. 2, ed. M. J. Schlinger, Chicago: Association for Consumer Research, 249-64.

_____ and Muderrisoglu, Aydin (1979), "The Stability of Responses Obtained by Free Elicitation: Implications for Measuring Attribute Salience and Memory Structure," in Advances in Consumer Research, Vol. 6, ed. William L. Wilkie, Ann Arbor: Association for Consumer Research, 269-75.

_____, Toy, Daniel R. and Dover, Philip A. (1976), "Mediating Effects of Cognitive Responses to Advertising on Cognitive Structure," in Advances in Consumer Research, Vol. 5, ed. H. Keith Hunt, Ann Arbor: Association for Consumer Research, 72-8.

Park, C. Whan (1976), "The Effect of Individual and Situation-Related Factors on Consumer Selection of Judgmental Models," Journal of Marketing Research, 13, 144-51.

Raju, P. S. and Reilly, Michael D. (1979), "Product Familiarity and Information Processing Strategies: An Exploratory Investigation," Journal of Business Research, 8, 187-212.

Russo, J. Edward and Johnson, Eric J. (1980), "What Do Consumers Know About Familiar Products?" in Advances in Consumer Research, Vol. 7, ed. Jerry C. Olson, Ann Arbor: Association for Consumer Research, 417-23.

Simon, Herbert A. (1974), "How Big is a Chunk?" Science, 183, 482-8.

Woodruff, Robert B. (1972), "Measurement of Consumer's Prior Brand Information," Journal of Marketing Research, 9, 258-63.

Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance to Advertising," Journal of Marketing Research, 10, 53-62.

_____ (1974), "Analyzing Media Effects on Advertising Responses," Public Opinion Quarterly, 38, 192-205.

PRODUCT FAMILIARITY AND LEARNING NEW INFORMATION¹

Eric J. Johnson, Carnegie-Mellon University
J. Edward Russo, University of Chicago

Abstract

What is the relationship between product familiarity and the ability to learn new product information? An experiment shows that product familiarity can lead to increased learning during subsequent purchase decisions. However, this ability strongly interacts with the specific decision task: the monotonic relationship between familiarity and learning holds for judgment but not for choice. The results also show that judgment and choice strategies leave different information in memory. We believe that phased choice strategies account for this difference.

will then be brand-based or attribute-based. Examples of the interaction between inferential level and inferential basis are found in the cells of Figure 1.

Figure 1
A Taxonomy of Product Knowledge

Introduction

Product familiarity is an important factor in explaining consumer behavior. Its rule is recognized in both traditional theories (Howard and Sheth 1969, Hansen 1972, Howard 1977) and newer information processing approaches (Bettman 1979, Olson 1978). Familiarity has also received recent attention in empirical research (Olson 1979, Bettman and Park 1980a). With the exception of Bettman and Park's paper, however, little research explores the effect of familiarity on consumer information processing. The present paper reports an empirical investigation of the effect of familiarity on one important consumer behavior, learning new product information during purchase decisions.

Consumer Knowledge

In this paper, product familiarity refers to prior knowledge of the brands within a product category. Obviously, all consumers start as novices at some point before their first purchase in a product class. As the consumer gains experience, product familiarity grows, and this knowledge affects the acquisition of new product knowledge. To understand this relation between familiarity and learning, a theory of product knowledge is essential. We use a taxonomy of knowledge developed by Russo and Johnson (1980).

The two central components of this taxonomy are shown in Figure 1. The level of inference refers to the relationship between external information and knowledge in memory. This resembles a hierarchy. New knowledge starts with externally available information, the brand-attribute values at Level 5 in Figure 1. The purchase decision process generates additional intermediate knowledge in the natural course of comparisons among brands. Finally, there is knowledge of the best brand, shown as Level 1 at the top of the hierarchy.

Knowledge can also be classified by its inferential basis, the dimension across the top of Figure 1. Consumers can make decisions using either brand-based or attribute-based strategies (Bettman 1979). The resultant product knowledge

INFERENTIAL LEVEL	INFERENTIAL BASIS	
	ATTRIBUTE	BRAND
1. Best Brand		"My favorite is Excedrin."
2. Whole Comparisons	"The main thing most people think about ... is the cost of gas."	"I've always wanted a Cadillac, much more so than a Lincoln."
3. Whole Evaluations	"I never worry about calories in any product."	"Liver is excellent."
4. Single Comparisons	"The next cheapest, I think, would be tuna."	"The best thing about chicken is its price."
5. Single Evaluations	"Beef liver is another very economical food."	

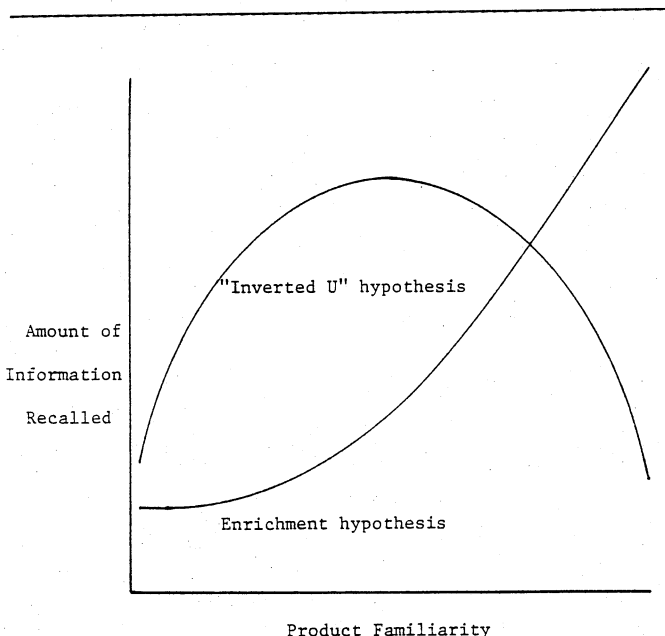
With this scheme for characterizing product knowledge, both old and newly acquired, we can ask how product familiarity affects learning additional product information. There are two plausible hypotheses.

The first, which might be called the "enrichment hypothesis", derives from cognitive psychology. It has been repeatedly demonstrated that greater prior knowledge facilitates learning. One classic example is the research on chess performed by Chase and Simon (1973). They showed a number of board positions to both chess masters and novice players for five seconds apiece. Knowledgeable subjects remembered much more than novices. However, when shown random patterns of chess pieces, the masters' recall was no

¹ The authors thank Jim Bettman for his comments on an earlier draft. The first author is now at the Department of Psychology, Stanford University, Stanford, CA 94305. This work was supported by National Science Foundation Grant #DAR76-81806 to the second author.

better than the novices'. This, Chase and Simon argued, was because the recall of random positions did not use the superior knowledge of the chess masters. Thus, the "enrichment hypothesis" argues that prior knowledge provides experienced consumers with better encoding and recall skills. Therefore, when confronted with new information, greater experience facilitates greater learning, a "rich get richer" hypothesis. This pattern would generate data similar to the exponential curve in Figure 2.

Figure 2
Alternative Hypothetical Relations
Between Product Familiarity and Learning



In contrast, Bettman and Park (1980a) claim that prior knowledge has the "inverted U" effect shown in Figure 2. They are concerned with external information search, but it seems natural to extend their hypothesis to the knowledge that remains in memory after search. In their view, very inexperienced consumers have difficulty understanding new information, and therefore search less. Consumers with a moderate level of knowledge search widely; they can both understand the new information and also benefit from its retention in memory. Finally, very experienced consumers search less. Although they can understand new product information, they have little need of it. This hypothesis was confirmed by Bettman and Park's search data, and the same "inverted U" effect might describe newly acquired knowledge in memory.

Experimental Rationale

To test these two hypotheses, we performed an experiment that used a brand-by-attribute display of information. The matrix contained new information about sub-compact cars that had just become available at the beginning of the model year. Consumer subjects, differing in their familiarity with automobiles, were instructed to evaluate the cars using only this information. A surprise recall of this information provided us with evidence of the role of prior knowledge in learning (remembering) new product knowledge.

Choice Versus Judgment

Besides the effect of different levels of familiarity, we added a second manipulation. We suspected that the strategy used to evaluate products would influence the product knowledge that was acquired and, therefore, remembered. First, consider a conventional choice task, i.e. choosing the best of several alternatives. Bettman and Park (1980b), Wright and Barbour (1977), and Johnson and Russo (1978) all find clear support for the use of phased decision rules in consumer choice. If subjects used phased strategies, some alternatives should be eliminated quickly, on the basis of only one or two pieces of information. Less knowledge of these eliminated brands should be generated and retained in memory. Some support for these claims was found by Johnson and Russo (1978).

In contrast with a choice task, subjects can also judge the overall quality of each alternative. This second task should force subjects to analyze, and remember, equal amounts of information for every alternative. In this experiment, half the subjects performed an ordinary preferential choice task, while the other half rated each alternative on a scale of attractiveness.

Method

Subjects

The 54 subjects participating in this experiment were students in the evening master's program of the Graduate School of Business of the University of Chicago. They completed the task as part of a classroom demonstration during the first meeting of a consumer behavior course. The product category, new sub-compact cars, was selected so this subject population would represent at least one segment of typical purchasers. Their demonstrated interest in the task partially confirmed this.

Stimulus

The brand-by-attribute matrix shown to the subjects was an edited copy of an advertisement placed by General Motors' Oldsmobile Division. Originally the advertisement compared one model of Oldsmobile to a number of imported sub-compact automobiles. By removing the Oldsmobile from the choice set, we created a matrix that had no dominant alternative, yet preserved the interrelationships among the attributes that naturally occurred in the marketplace. Additionally, we deleted anything beyond one model per manufacturer to avoid possible confusion during recall. These actions left a matrix of eight alternatives, all smaller imported cars. They were rated on ten attributes: model, transmission, engine (number of cylinders), EPA estimated miles per gallon, fuel tank capacity, estimated driving range, passenger capacity, interior volume, trunk volume, and the manufacturer's suggested retail price.

Procedure

Subjects were run in a single group. At the beginning of the session, they were given a booklet containing all experimental materials. The first page presented the appropriate task instructions and requested that they not proceed until asked to do so by the experimenter.

Judgment Task. In the judgment condition, the instructions informed the subjects that they would be given a brand-by-attribute matrix of information relevant to automobiles. They were asked to judge each of the automobiles on a seven-point scale, based only on the information provided, and not prior knowledge.

Choice Task. Subjects in the choice task were similarly informed of the brand-attribute matrix, but were asked to choose the most preferred of the eight automobiles rather than make individual judgments.

When all subjects finished their respective tasks, they were asked to turn the page to a "Demographic Questionnaire." This page had two purposes. First, it was used to gather demographic information, including a self-rating of product familiarity. Subjects were asked to "rate your previous knowledge of automobiles, compared to the rest of the population", on a five-point scale. Second, the recall of demographic facts prevented the retention of product information in short-term memory.

Subjects were then instructed to turn the page and read the instructions for the unexpected recall task. They were told to "try to recall as much as you possibly can." This included not only the information that they were given but also any "observations and judgments about the alternatives and the attributes." These instructions were designed to encourage complete recall, including information at the higher levels of inference. Subjects were then given an unlimited time to record their recall on blank sheets of paper. After the recall was completed, subjects were asked to report the car selected as best, any past experience with each of the eight automobiles, and the strategy they used to perform the choice or judgment task.

Analysis

The written protocols were divided into a series of "complete thoughts" (Newell and Simon 1972) and then coded according to the scheme shown in Figure 1 and described in Russo and Johnson (1980). This scheme is similar to the one used by Bettman and Park (1980a).

Familiarity with the product category was measured by subjects' self-reports on the scale described above. The distribution of these self-ratings was concentrated at the middle value. Self-ratings at or below 2 were classified in the low familiarity group, between 3 and 4 as the moderate familiarity group, and above 4 as the high familiarity group. The number of subjects in each familiarity condition was 12, 27 and 16, respectively.

The factor of task, judgment versus choice, was crossed with the three levels of familiarity to create a 2 X 3 factorial design (with unequal cell sizes). This two-factor design served as the basis for all analyses reported.

Results

We have claimed that the information presented to subjects represented new knowledge, even to the subjects with greater product familiarity. This was reinforced by task instructions asking subjects to limit their attention to the information presented. However, to check the validity of this claim, we examined the recall protocols for accuracy and intrusions. Inaccurate information could come from two sources: either outside knowledge or a faulty remembrance of the presented information.

A rater blind to our hypotheses, coded each brand-attribute value on a five-point scale. Inaccurate recall was defined as those statements rated 4 (somewhat inaccurate) and above, and accounted for less than 10% of the information recalled. When the numbers of intrusions and inaccuracies are included as covariates in the analyses of variance, our results remain unchanged. Thus, the results that follow are probably not due to the recall of previously remembered information by the highly familiar consumers.

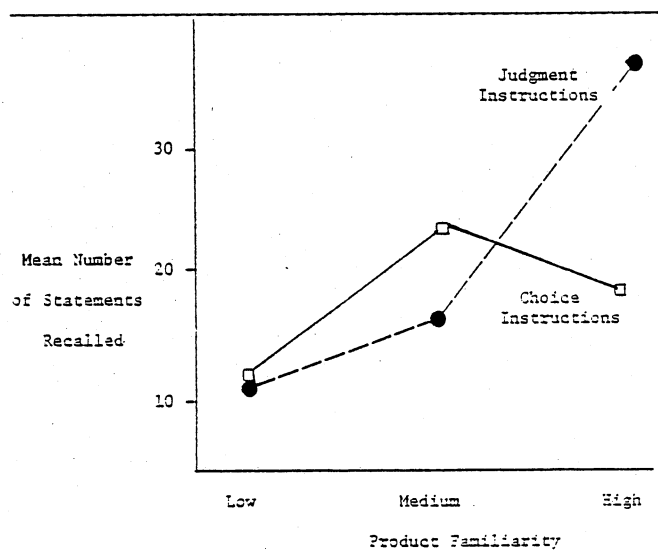
Familiarity

What is the relationship between existing product knowledge and the ability to learn new information? We have considered two possible relationships: the "enrichment hypothesis" and the "inverted U" hypothesis, shown in Figure 2. A test of these two plausible relationships between familiarity and learning can be based on the total amount of knowledge recalled as a function of familiarity.

The mean number of statements recalled increases with the level of familiarity. For the low, medium and high levels, the mean frequencies were 12.0, 16.6 and 21.1, respectively. However, an analysis of variance shows that this effect is only marginally significant ($F(2,49) = 2.825, p < .07$). Nonetheless, the pattern of results mildly supports the "enrichment" hypothesis: increased familiarity does lead to the recall of more new information.

A closer look at these data reveals that the effect of familiarity interacts with the evaluation task. The mean number of statements recalled is plotted separately for choice and judgment tasks in Figure 3. For judgment, the enrichment effect is large and consistent, with highly familiar consumers recalling two and one-half times as much information as the low familiarity consumers (28.6 versus 11.6 statements per protocol). The consumers who were asked to choose one brand, however, show an "inverted U" relationship. The subjects moderately familiar with automobiles exhibit the greatest recall. The ANOVA shows that this familiarity by task interaction is reliable, $F(2,49) = 3.20, p < .05$. A priori contrasts indicate that the judgment condition contains a significant linear relationship with familiarity, ($F(1,49) = 8.32, p < .001$). In contrast, the group making choices exhibits the inverted U or quadratic relationship ($F(1,49) = 4.11, p < .06$).

Figure 3
Learning as a Function of Familiarity and Task



Thus, while the enrichment hypothesis holds for consumers making judgments, it does not describe the memory of consumers making choices. As Bettman and Park predict, moderately familiar consumers search and remember more information.

Choice versus Judgment

How do these decision strategies affect newly acquired product knowledge? An important manipulation in this study concerns the task instructions. The consumers either chose one alternative or judged the overall quality of each alternative. We suspect that these two tasks require different patterns of information processing.

If, as we believe, consumers use phased choice rules, they first examine the alternatives in an attribute-based manner, using this information to eliminate alternatives. Eventually, the basis of the decision process shifts, and

the remaining alternatives are examined more intensively. This later processing is by brand (Bettman and Park 1980b, Bettman 1979, Johnson and Russo 1978). As a consequence of this phased strategy, more information is examined and recalled for the chosen alternatives (Payne 1976, Johnson and Russo 1978). This differential recall is also affected by the level of inference (Russo and Johnson 1980). If a brand is eliminated early in the choice process, few higher level inferences should be made. Thus, the higher the inferential level, the greater the recall advantage of the chosen brand (Johnson and Russo 1978).

In contrast, the judgment task does not permit the use of phased strategies. Because subjects rated all eight automobiles, they were required to examine information for every alternative. Thus, in the judgment task we expect equal recall for all brands, preferred or not.

To test these predictions, we tallied the number of statements in each protocol that referred to the automobile selected by the subject. In the choice condition, this was the most preferred alternative. In judgment, subjects reported which car they judged as the best. The proportion of recalled statements referring to the preferred automobile is .35 for the choice task and .18 for judgment. This difference was highly significant ($F(1,49) = 13.84, p < .001$). In addition, both values, including the 18% observed in the judgment task, are significantly greater than the 12.5% expected by chance ($p < .001$). These data provide clear evidence of the use of phased strategies in the choice task, and identify much smaller effects for judgment instructions.

To further examine memory for chosen products, we broke down the proportions of statements referring to chosen products by task and level of familiarity. The means are shown in Table 1.

Table 1
Proportion of Statements Referring to
the Preferred Alternative

<u>Low</u> <u>Familiarity</u>	<u>Moderate</u> <u>Familiarity</u>	<u>High</u> <u>Familiarity</u>	<u>Mean</u>
<u>Choice Task Instructions</u>			
.32	.27	.54	.35
<u>Judgment Task Instructions</u>			
.13	.23	.12	.18
<u>Mean</u>			
.24	.25	.36	.27

An ANOVA confirms that there is a task-by-level interaction. For judgment there is no effect of familiarity, but the choice task shows a large effect ($F(2,49) = 5.82, p < .01$). Half the recall of high familiarity subjects in the choice condition referred to the car they chose. This compares to less than a third for the low and moderate familiarity groups. A priori comparisons confirm the reliability of this pattern ($F(1,49) = 11.20, p < .005$).

Overall, the data indicate that the consumers who made choices used phased rules which eliminate alternatives, while consumers making judgments did not. Familiarity leads

to an increasingly selective memory for information about the chosen product caused, we suspect, by the use of highly selective search. The most experienced consumers do not look at all the new information, only that which they consider useful.

Types of Product Knowledge

Let us return to the taxonomy of product knowledge discussed earlier and summarized in Figure 1. Do the above results hold over all levels of inference? To answer this question the recalled statements were partitioned into five inferential levels. The same two-factor (familiarity and task) ANOVA was performed on the number of statements at each level. The only differences attributable to the main effect of familiarity occurred at Level 5, brand-attribute values, ($F(2,49) = 3.46, p < .05$). This lowest level of product knowledge consists simply of the values in the brand-by-attribute matrix. This is also the only inferential level that shows a significant familiarity-by-task interaction ($F(2,49) = 3.65, p < .05$). Thus, differences in recall can be attributed entirely to knowledge at the lowest level of processing, the acquisition of the presented brand-attribute values. No other inferential level shows these effects.

How Does Familiarity Work?

Can these results help explain the differential effects of familiarity on choice and judgment? Recall that Bettman and Park explain their "inverted U" results by claiming that experienced consumers use internal search as a substitute for external search. Because of our earlier analysis of intrusions into recall, we suspect that experienced subjects did not use stored knowledge of the brand-attribute values. Since little of the information recalled deviates from the information presented (less than 10%), internal search of existing knowledge was probably not a factor. Thus, we need another explanation for these results.

An alternate explanation of increased familiarity relies on other aspects of experienced consumers' prior knowledge. There is little doubt that experienced consumers have a great deal of information about products stored in memory (Russo and Johnson 1980). This knowledge contains more than just facts about the relations among competing brands. In addition, experienced consumers learn which attributes are most important, and should subsequently focus on these attributes in making a choice. This is a type of higher-level knowledge that goes beyond particular brands. It implies that, with experience, consumers become more selective in their search for information, and use more narrowly focused phased decision rules.

We did not collect decision protocols in this study. This means we cannot examine the decision processes directly. But we did find that differences in memory for chosen and rejected products will be greater for more experienced consumers resulting from their more selective search. This was demonstrated by the significant familiarity-by-task interaction displayed in Table 1. Also, as we have already seen, the effect of familiarity is confined to the earliest, low-level inferences. These inferences come from the initial examination of the matrix which, according to this model, experienced consumers delete from their decision processes.

Is one explanation of familiarity effects superior to the other? We cannot provide a definitive answer with these data. Our "knowledge of the attributes" explanation is very similar Bettman and Park's, and differs only for information previously unseen by the consumer. But in the real world, both mechanisms probably operate, and increase the effects of familiarity. A knowledge of important attributes, along with a knowledge of brand-specific facts, can limit the search of experienced consumers. So for many real world situations, differences in the explanations may not matter.

The unique contribution of the current explanation is limited to new information, whether from new products, or new product classes.

Discussion

What does all this mean for the study of consumer behavior? It is important to see where the present results fit in the growing literature on consumer information processing. A surprisingly consonant view of consumer information processing is emerging from a number of investigators working within different experimental paradigms, and with different products and subject populations. First, recent work calls attention to the predominance of phased strategies in choice (Bettman and Park 1980b, Wright and Barbour 1977, Payne 1976). More work is needed in display formats other than the brand-by-attribute matrix. However, we now know that phased strategies affect not only information acquisition, but also subsequent memory. Second, familiarity with product classes has an effect on decision and search behavior. At first, it leads to an increase in external search and subsequent recall. But with high levels of familiarity, both information search and memory for new information decline (Bettman and Park 1980b).

This paper attempts to make two additional contributions to this area. First, we try to extend the concepts advanced by Bettman and Park to memory for new product information. Second, we offer an alternative explanation for why experienced consumers search less information. We argue that this is not caused by the presence of relevant information stored in memory, but by their higher-level knowledge of the product class and its important attributes. They search less information and use more selective, phased decision rules that delete the preliminary exploration of external information.

These findings have implications for public and private providers of product information. They suggest that information use will be segmented by the familiarity level of the consumer. The most important implication is that providing a brand-by-attribute matrix for a low familiarity segment may do little good. Simply posting information is probably ineffectual, especially for technically sophisticated products and attributes. To help these consumers make better decisions, information about the attributes, their importance, and their relationship to quality must be provided.

References

Bettman, James R. (1979), An Information Processing Theory of Consumer Choice, Reading, Mass.: Addison-Wesley.

Bettman, James R. and Park, C. Whan (1980a), "Implications of a Constructive View of Choice for Analysis of Protocol Data: A Coding Scheme for Elements of Choice Processes," in Advances in Consumer Research, Vol. VII, Jerry C. Olson, ed., Ann Arbor: Association for Consumer Research, 148-153.

Bettman, James R. and Park, C. Whan (1980b), "Effects of Prior Knowledge, Exposure and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis," Journal of Consumer Research, in press.

Chase, William G. and Simon, Herbert A. (1973), "Perception in Chess," Cognitive Psychology, 4, 55-81.

Hansen, Flemming (1972), Consumer Choice Behavior, New York: Free Press.

Howard, John A. (1977), Consumer Behavior: Application of Theory, New York: McGraw Hill.

Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: Wiley.

Johnson, Eric J. and Russo, J. Edward (1978), "What Is Remembered After a Purchase Decision?," working paper, Graduate School of Business, University of Chicago.

Newell, Allan and Simon, Herbert A. (1972), Human Problem Solving, Englewood Cliffs, N. J.: Prentice-Hall.

Olson, Jerry C. (1978), "Theories of Information Encoding and Storage: Implications for Consumer Research," in The Effect of Information on Consumer and Market Behavior, Andrew A. Mitchell, ed., Chicago: American Marketing Association, 49-60.

Russo, J. Edward and Johnson, Eric J. (1980), "What Do Consumers Know About Familiar Products?," in Advances in Consumer Research, Vol. VII, Jerry C. Olson, ed., Ann Arbor: Association for Consumer Research, 417-423.

Wright, Peter L. and Barbour, Frederic (1977), "Phased Decision Strategies: Sequels to an Initial Screening", in North Holland/TIMS Studies in the Management Sciences, Volume 6: Multiple Criteria Decision Making, Martin K. Starr and Milan Zeleny, eds., Amsterdam: North Holland, 91-109.

PRODUCT FAMILIARITY: CRITICAL COMMENTS ON SELECTED
STUDIES AND THEORETICAL EXTENSIONS

M. Joseph Sirgy, Virginia Polytechnic Institute and State University

Abstract

This paper addresses a number of conceptual and methodological issues involving product familiarity. First, some of the general conceptualization and measurement problems were discussed. Second, the studies of Marks and Olson (1981), Tan and Dolich (1981), and Johnson and Russo (1981) were criticized for their weaknesses. Third, an eclectic cognitive view of product familiarity was introduced. And finally, recommendations for future research involving product familiarity were suggested.

Introduction

Product familiarity seems to be an emerging variable in the study of consumer behavior. Consumer researchers have used it to explain a number of consumer-related phenomena such as, message acceptance (e.g., Marks & Olson 1981), choice of decision rule (e.g. Parks 1976, Tan & Dolich 1981), product preference and purchase intention (e.g., Marks & Olson 1981), product satisfaction (e.g., Anderson, Engledow & Becker 1979), and new learning (Johnson & Russo 1981). Overall, the pattern of results utilizing product familiarity as an explanatory variable seems conflicting. The primary cause of this lack of clarity of the precise role of product familiarity as a determinant of consumer behavior may be due to the ambiguity displayed in the conceptualization and measurement of product familiarity.

This discussion will (1) focus on some of the conceptualization and measurement efforts revealed in the consumer behavior literature, (2) criticize the studies of Marks and Olson (1981), Tan and Dolich, (1981), and Johnson and Russo, (1981), which treat product familiarity as an explanatory variable of consumer behavior phenomena, (3) provide an eclectic cognitive view of product familiarity, and (4) suggest recommendations for future research in relation to product familiarity.

Conceptualization and Operationalization of Product Familiarity

One important issue concerning the conceptualization and measurement efforts of product familiarity is the unidimensional versus multidimensional treatment of product familiarity.

Product familiarity has been conceptualized and operationalized differently by different investigators. In the unidimensional tradition, Park (1976) measured product familiarity in terms of subjects' agreement with statements about the product. Woodruff (1972) used a free-recall method of purchase situations. Raju and Reilly (1979) employed self-reported "frequency of use, overall familiarity, and knowledge of how to select best brand" as measures of product familiarity. Anderson, Engledow, and Becker (1979) and Jacoby, Chestnut, and Fisher (1978) used a frequency of purchase measure as an indicator of product familiarity. Johnson and Russo (1981) used a global self-report rating measure - subjects were asked to rate their previous knowledge of automobiles compared to the rest of the population. Tan and Dolich (1981) measured product familiarity by the proportion of brands in the product class that one knew something about.

The main problem with these measures is that they are atheoretical. That is, they are not based on any one theory. In contrast, the multidimensional measurement of product familiarity, as established by Olson and his associates (e.g., Marks & Olson 1981, Kanwar, Olson, & Sims 1981, Olson & Muderrisoglu 1979), is based on the cognitive theory involving cognitive structures. The dimensions that Olson and his associates worked with are: (1) dimensionality which is defined as the number of activatable concepts associated with a particular domain, (2) articulation defined as the number of category representations or levels for each salient dimension in memory, and (3) abstraction defined as the degree to which salient, activatable dimensions are abstract versus concrete. Although more than three dimensions were identified as constituting product familiarity (reference was scantily made to the interrelatedness and centrality dimensions), Marks and Olson (1981) conceptually treated product familiarity only along the dimensionality and abstractness dimensions. High product familiarity was viewed to reflect high abstractness and low dimensionality and vice versa.

There are a number of problems which marks Olson's conceptual treatment of product familiarity. First, it is not clear how the dimensionality of product familiarity is different from its articulation. It seems to me that articulation denotes (or should I say connotes) the extent to which a psychological concept is encoded as symbolic or semantic. Symbolic psychological concepts can be construed as less articulated and semantic concepts as being more articulated. Also, how does dimensionality differ from interrelatedness? Furthermore, why wasn't the dimension of centrality incorporated in the main product familiarity schema? And finally but most importantly, why wasn't the certainty dimension considered at all? To me, product familiarity taps those beliefs (or "saliency weights" using Rosenberg-Fishbein's terminology) stored in memory. The degree of certainty of beliefs is construed to be extremely important as it affects psychological processes such as alternative evaluation and attitude change (see Fishbein and Ajzen 1975).

The measurement technique which was consistently used in most of Olson's studies (e.g., Marks & Olson 1981, Kanwar, Olson & Sims 1981, Olson & Muderrisoglu 1979) is the free-elicitation technique. Kanwar, Olson, and Sims (1981) have attempted to provide construct validation to the free-elicitation measure by relating it to product familiarity scores of a Rep test and a self-report "knowledge" measure. The results did not provide clear-cut support for the construct validity of the free-elicitation measure. However, Olson and his associates have continuously used it knowing fully that the measure's construct validity is yet to be demonstrated. Even Olson's latest study (Marks and Olson 1981), which will be discussed later in some detail, failed to provide the free-elicitation measure with the desired nomological validity. It should be noted that the free-elicitation technique is not different from the psychoanalytic technique of free-association which has been used extensively in the clinical psychology literature. It was fairly well recognized by Freud himself that the free-association technique had its share of problems. Subjects' responses are contaminated by perceptual cues, social desirability factors, memory factors, and other situational influences (The Standard Edition of the Complete Psychological Works of Sigmund Freud, 1953-1974).

Product Familiarity as a Determinant of Consumer Behavior

Marks and Olson's (1981) study was primarily concerned with the effect of product familiarity on message acceptance, product preference, and purchase intention. Tan and Dolich's (1981) study focused on the role that product familiarity plays in the selection of decision rules. Johnson and Russo's (1981) study on the other hand, examined the role of product familiarity in new learning. Each of these studies will be scrutinized in the following sections.

Effect of Product Familiarity on Message Acceptance. Marks and Olson (1981) expected that high product familiarity would induce greater cognitive responses related to support arguments, less cognitive responses related to counterarguments, and overall fewer cognitive responses than under low product familiarity conditions. The results only showed that high product familiarity generated fewer counterarguments.

There are a number of problems in Marks and Olson's (1981) message acceptance part of the study that I can see off-hand. First, even though 2/3 of their hypotheses (although not clearly stated) were not confirmed by the results, the support provided for the relationship between product familiarity and counterarguments can be due to experimental artifacts such as, experimental demand and/or nature of the sample.

Second, I was troubled by the fact that no theoretical justification was adequately provided to support the relationship between product familiarity and message acceptance in this allegedly important study in which a theory for product familiarity is advocated.

Third, one needs to realize that counterargument, among other cognitive responses such as source derogation, is a cognitive response mode designed to protect an interrelationship of existing psychological concepts. Each psychological concept has a value and a belief dimension. The belief dimension is that dimension which links one psychological concept to another (cf. Fishbein & Ajzen 1975). Product familiarity is generally represented as belief dimensions. Now, concerning the psychological dynamics involved in support and counterarguments, the following explanation is in order. A cognitive response is usually activated by a cognitive comparison between a perceptual set and an internally evoked cognitive set. Incongruence between the two sets usually leads to counterargumentation, source derogation, among other cognitive responses. This serves to protect the existing cognitive set from the threat of change inflicted by the perceptual set. Congruence between the two sets, on the other hand, generally leads to support argumentation among other responses and therefore serves to strengthen the existing cognitive set (Sirgy, in press a, in press b, Wright 1973). Based on this theoretical perspective, what would be the role of product familiarity in support or counterarguments? Product evaluation (or existing attitude toward the product) is the major determinant of support or counterargument. Product familiarity is just one component of product evaluation. Viewed from Rosenberg-Fishbein's multi-attribute attitude model, product familiarity reflects the set of saliency weights, whereas product evaluation represents the overall attitude.

Effect of Product Familiarity on Product Preference and Purchase Intention. Marks and Olson (1981) argued that since counterargument is expected to affect attitudes, subjects who are more familiar with the product (i.e., secretaries) would be more likely than those who are less familiar with the product (i.e., students) to recommend its purchase. Although the results were consistent with the hypothesis, it can be argued that the effect could have been due to an artifact of the sample and/or

experimental demands not to mention the equivocal measurement of product familiarity. Theoretically speaking, product familiarity does not independently or directly affect product preference and purchase intention but interacts with values which, in turn, interacts with perceptual elements to determine change in attitude and therefore behavior.

Effect of Product Familiarity on Decision Rule Selection. Tan and Dolich (1981) used product familiarity along with cognitive complexity to determine their effect on the use of decision rules in consumer behavior. The results indicated that product familiarity plays a negligible role in the selection of decision rules in consumer behavior, and that cognitive complexity plays a significant role. These results are extremely interesting however not totally convincing.

First, no theoretical justification was provided for the choice of the designated decision rules in the study. One may argue that product familiarity would have played a major role in the selection of other types of decision rules such as, the maximax rule compared to the minimax or maximin rules (see Edwards & Tversky 1967, Fishburn 1964, Kaufman 1968, Raiffa 1968, White 1969).

Second, the measurement of product familiarity "through the proportion of brands in the product class that one knew something about" could be a very misleading indicator of product familiarity. It is quite conceivable that a consumer could be highly familiar with one or two brands while knowing little about other competing brands.

Third, attractiveness values were used without any differentiation between its individual components (i.e., salience weights, importance weights, evaluative weights, and determinance weights). Knowing something about, for example, salience weights, can provide us with further insights concerning the dynamics involved in decision rule selection. To exemplify, it is possible that those attributes which are highly salient are included in the final decision and therefore reflecting the usage of a partial version of the compensatory model. Also importance weights (which can be construed as a dimension of product familiarity) may have a lot to do with decision rule selection. Attributes which are more important than others may lead to the adoption of a partial compensatory model which only includes those important attributes.

Effect of Product Familiarity on New Learning. Johnson and Russo (1981) examined the interaction of product familiarity and type of decision rule (attribute versus brand strategy) on learning of new information. The results provided support for the role of the "enrichment hypothesis." That is, product familiarity facilitates new learning. However, this main effect was qualified by an interaction effect between product familiarity and type of decision rule on learning of new information, and therefore providing some support for the "inverted-U hypothesis." That is, use of attribute strategy accompanied by high product familiarity contributed to new learning more than with the use of brand strategy.

Again, these results are very intriguing but not terribly convincing. This may be due to a number of problems: First, the enrichment hypothesis refers to the enhancing effect of internal existing information stored in memory on new learning (Chase & Simon 1973), whereas the inverted-U hypothesis refers to the enhancing or inhibiting effects of external information on new learning (see Jacoby, Speller & Kohn 1974, Jacoby, Speller & Berning 1974). Consequently, these two hypothesis should not be construed as competing since they operate at different levels.

Second, since the inverted-U hypothesis refers to the effect of amount of external information on learning, it

seems to me that it should have little to do with the type of decision strategy (attribute versus brand strategy) as a suitable operationalization of the principle. A more suitable operationalization of the inverted-U hypothesis would involve amount of external information as opposed to type of decision strategy.

Third, subjects were specifically instructed to make a decision by rating each automobile. They weren't specifically asked to learn the material. The surprise recall measure cannot be treated as a reliable and valid measure of new learning. New learning under these conditions can be severely affected by the nature of the task, personality and motivational factors, and other situational influences.

Fourth, subjects were asked to rate each of the seven automobiles based on the information provided, and not prior knowledge. It seems to me that it would be quite difficult for a subject to ignore previous information by simply asking him/her to do so. It is therefore very likely that previous information was indeed a confounding variable which was not adequately controlled for.

Fifth, in the judgment task, it was assumed that by judging each automobile separately, subjects are not making comparative judgments based on some specific other car or based on the preceding car, among other alternatives.

And last but not least, the measurement problem involving product familiarity has to be pointed out. Product familiarity was measured by instructing subjects to "rate your previous knowledge of automobiles compared to the rest of the population." Besides its atheoretical overtones, this measure of product familiarity is confounded by (1) individual differences in the interpretation of "the rest of the population." (2) social desirability factors, (3) as well as personality factors such as cognitive complexity (i.e., cognitively complex subjects may

have difficulty making global judgments compared to cognitively simple individuals), and (4) self-concepts related to mechanical devices (i.e., an individual who perceives himself as being mechanically inclined might rate his familiarity with automobiles as high despite his lack of knowledge in this area to avoid cognitive inconsistency).

An Eclectic Cognitive View of Product Familiarity

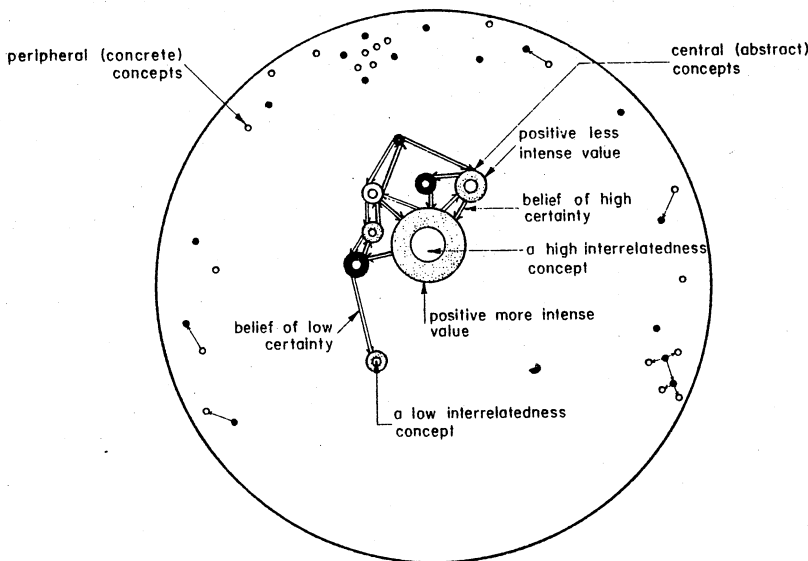
Figure 1 shows a cross-sectional and transverse view of a hypothetical physical space of cognitive structures. To fully understand product familiarity, one needs to describe how product familiarity is represented within an individual's cognitive structures. This position is quite similar to the one espoused by Olson and his associates, however, differs with respect to the delineation of cognitive structures and product familiarity dimensions.

From the cross-sectional view of cognitive structures of Figure 1, cognitions are basically represented as psychological concepts or "nodes." Each psychological concept which is content-specific is surrounded by a layer representing the value placed on that concept. The thickness of this layer reflects the positivity or negativity or direction of the value placed on that concept. Darker shades represent negative values whereas lighter shades represent positive values.

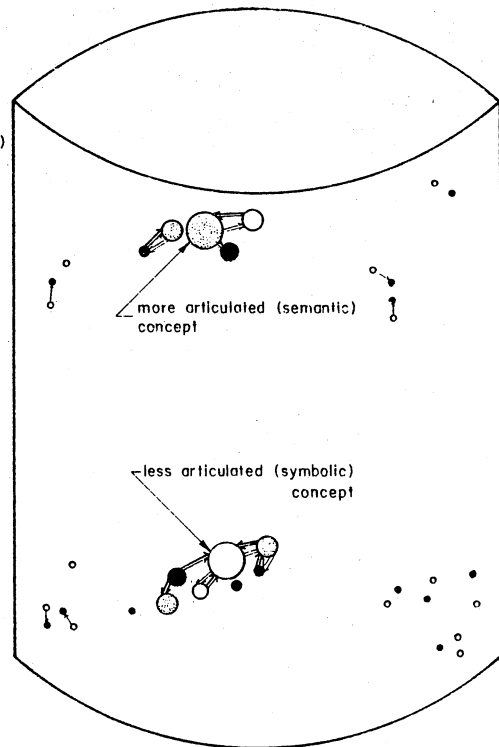
Some concepts are linked to other concepts through belief links. Each belief link is a content-specific psychological association and is directed (i.e., excitation travels in one direction). The thickness of a belief link is indicative of its degree of certainty. The thicker the belief dimension the higher the certainty projection.

Psychological concepts are also spatially arranged along a central-peripheral dimension. Central concepts are more abstract and more interrelated than peripheral concepts.

FIGURE 1



Cross Sectional View of a Hypothetical Physical Space of Cognitive Structures



Transverse View of a Hypothetical Physical Space of Cognitive Structures

Peripheral concepts can be referred to as attributes; the more central concepts as schemata and scripts.

As shown in the transverse view of cognitive structures of Figure 1, psychological concepts are also arranged along an articulation dimension. Concepts which are highly articulated are usually encoded in semantic or verbal form, whereas those which are less articulated are encoded in symbolic or nonverbal form.

This eclectic view of cognitive structures has been influenced by the writings of Kelly (1955), Rokeach (1960), and Epstein (1973), Ableson (1976), Collins and Loftus (1975), and Kosslyn and Pomerantz (1977).

Now the question arises concerning the cognitive representation of product familiarity within the space of cognitive structures. The position which is taken by this author is that product familiarity is represented by those belief dimensions connecting the psychological concept of a particular product with other psychological concepts representing the product attributes.

Those belief dimensions describing product familiarity can be characterized by at least eleven dimensions: (1) content, (2) direction, (3) certainty, (4) centrality, (5) articulation, (6) interrelatedness, (7) consistency, (8) stability, (9) accuracy, (10) verifiability, and (11) frame-of-reference.

Content of a belief dimension refers to the specific content of the association between two psychological concepts. For example, the Ford Escort automobile is an economy car. The verb "is" is a content-specific belief linking the psychological concept "Ford Escort" with "an economy car." Direction of a belief dimension refers to the flow of activation from one concept to another (see Figure 1). In the above example, the flow is from the psychological concept "Ford Escort" to "an economy car" and not necessarily vice-versa. Certainty of a belief dimension refers to the strength of the link between two concepts and is represented by the layer's thickness of a given link (see Figure 1). In the above example, the belief that Ford Escort is indeed an economy car may be of high certainty or low certainty. Centrality refers to the degree of salience, importance, or abstractness of a given concept. In this case, the product concept "Ford Escort" may be closer to the periphery than to the center (see Figure 1). Articulation refers to the extent to which a psychological concept is verbal (semantic) versus nonverbal (symbolic). The "Ford Escort" as a psychological concept may be encoded more on a verbal-semantic level than a nonverbal-symbolic level and therefore is more articulated in form. Interrelatedness refers to the number of attribute concepts which a psychological schema is linked with. A highly interrelated product concept is therefore associated with many attributes. Ford Escort may be associated with only high gas mileage in a consumer's mind, and therefore is said to have low interrelatedness. Another consumer may associate Ford Escort with high gas mileage, style, safety, reliability, roominess, and patriotism. The latter consumer's product concept is said to be highly interrelated. Consistency refers to the degree to which the attributes associated with a psychological schema are consistent with one another. For example, a consumer who knows Ford Escort is economical and stylish but also as being unsafe and unreliable is holding an inconsistent set of attributes. Whereas a consumer who thinks of Ford Escort as being economical, stylish, safe, and reliable holds a consistent set of attributes. Also, the consumer who relates the Ford Escort to being uneconomical, lacking style, being unsafe and unreliable is also holding a consistent set of attributes. Stability refers to the extent to which knowledge of a particular concept is durable across time. A person who may know something about Ford Escort today but forgets it tomorrow

lacks temporal stability. Accuracy refers to the extent to which the set of beliefs linking a psychological schema to a set of attributes are representative of the true state-of-affairs, or just simply having accurate perceptions. A consumer who thinks that the Ford Escort is a luxury uneconomical car has inaccurate product perceptions. Verifiability refers to the extent to which an individual feels that the beliefs associating a psychological schema (or product concept) with a set of attributes can be somehow verifiable. A person who believes that God controls one's destiny may feel that this kind of belief is unverifiable, however his belief that the Ford Escort is economical can be verified through actual experience. And finally, frame-of-reference refers to that psychological "script" which the individual is presently using during which certain associations between a psychological schema and its attributes become activated. For example, if one approaches the Ford Escort from a consumer's union frame-of-reference, a totally different set of attributes might be activated (e.g., safety, economy, reliability) compared to a shopper's frame-of-reference.

How does product familiarity differ from product evaluation? The only difference that I can see in these two phenomena is the value dimension characterizing the intensity and direction of the psychological concepts. Product evaluation takes into account the various dimensions of the beliefs and values pertaining to a product concept and its associated attributes whereas product familiarity deals with belief dimensions only. This formulation is consistent with traditional social-psychological attitude formulations (see Fishbein & Ajzen 1975).

Recommendations for Future Research

With this eclectic theoretical orientation, some recommendations for future research are suggested.

First, more work needs to be done in the measurement area. Atheoretical measurement of product familiarity should be discouraged. The efforts of Olson and his associates are a step in the right direction. However, since valid measures rest on adequate theory, theory development of product familiarity should have the first priority. More research is needed to test different theories of product familiarity.

Second, if product familiarity is truly multidimensional, research effort should not only be directed to identifying those dimensions but also to reveal how they are differentially combined to form overall familiarity. Can we use an additive model and sum over all the familiarity dimensions? Or would a multiplicative model be more appropriate? Should there be weights placed on those dimensions? And so on.

Third, hardly any research has been conducted treating product familiarity as the dependent variable or the phenomenon of interest. What are the differential effects of dispositional variables such as cognitive needs and cognitive complexity on product familiarity? What are the various situational factors which influence product familiarity? Examples may include time factors (e.g., repetition, schedule of reinforcement), communication factors (e.g., type of message, type of source, type of communication channel), among others.

Finally, more systematic research needs to be conducted to determine the effects of product familiarity on consumer processes such as perception, awareness, attention, need recognition, information search, alternative evaluation, actual behavior, and outcome evaluation.

References

- Ableson, R. P. (1976), "Script Processing in Attitude Formation and Decision-Making" in Cognition and Social Behavior, J. S. Carroll and J. W. Payne, eds., Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Chase, William G. and Simon, Herbert A. (1973), "Perception in Chess," Cognitive Psychology, 4, 55-81.
- Collins, A. M. and Loftus, E. F. (1975), "A Spreading-Activation Theory of Semantic Processing," Psychological Review, 82, 407-428.
- Edwards, W. and Tversky, A. (1967), Decision Making, N. Y.: Penguin Books.
- Epstein, Seymour (1973), "The Self-Concept Revisited, or a Theory of a Theory," American Psychologist, 28, 404-416.
- Fishbein, M. and Ajzen, J. (1975), Belief, Attitude, Intention and Behavior, Reading, MA.: Addison-Wesley.
- Fishburn, P. C. (1964), Decision and Value Theory, N. Y.: John Wiley.
- Jacoby, Jacob, Chestnut, Robert W., and Fisher, William A. (1978), "A Behavioral Process Approach to Information Acquisition in Nondurable Purchasing," Journal of Marketing Research, 15, 532-544.
- Jacoby, Jacob, Speller, D. E. and Berning, Carol (1974), "Brand Choice Behavior as a Function of Information Load: Replication and Extension," Journal of Consumer Research, 1, 33-42.
- Jacoby, Jacob, Speller, Donald E., and Kohn, Carol (1974), "Brand Choice Behavior as a Function of Information Load," Journal of Marketing Research, 11, 63-39.
- Kanwar, Rajesh, Olson, Jerry C., and Sims, Laura S. (1981), "Toward Conceptualizing and Measuring Cognitive Structures," in Advances in Consumer Research, vol. 8, Kent Monroe, ed., Ann Arbor: Association for Consumer Research.
- Kaufman, A. (1968), The Science of Decision Making: An Introduction to Praxeology, N. Y.: McGraw-Hill.
- Kelley, G. A. (1955), The Psychology of Personal Constructs, New York: Norton.
- Kosslyn, S. M. and Pomerantz, J. R. (1977), "Imagery, Propositions and the From of Internal Representations," Cognitive Psychology, 9, 52-76.
- Lastovicka, John L. (1979), "Questioning the Concept of Involvement Defined Product Class," in Advances in Consumer Research, vol. 6, William L. Wilkie, ed., Ann Arbor: Association for Consumer Research, 174-179.
- Marks, Larry J. and Olson, Jerry C. (1981), "Toward a Cognitive Structure Conceptualization of Product Familiarity," in Advances in Consumer Research, vol. 8, Kent Monroe, ed., Ann Arbor: Association for Consumer Research.
- Olson, Jerry C. and Muderrisoglu, Aydin (1979), "The Stability of Response Obtained by Free Elicitation: Implications for Measuring Attribute Saliency and Memory Structure," in Advances in Consumer Research, vol. 6, William L. Wilkie, ed., Ann Arbor: Association for Consumer Research, 269-275.
- Park, C. Whan (1976), "The Effect of Individual and Situation-Related Factors on Consumer Selection of Judgmental Models," Journal of Marketing Research, 13, 144-151.
- Raiffa, H. (1968), Decision Analysis, Reading, MA: Addison Wesley.
- Raju, P. S. and Reilly, Michael D. (1979), "Product Familiarity and Information Processing Strategies: An Exploratory Investigation," Journal of Business Research, 8, 187-212.
- Rokeach, Milton (1960), The Open and Closed Mind, N. Y.: Basic Books.
- Sirgy, M. Joseph (in press a), "Introducing a Self-Theory to Consumer Personality Research," Journal of Selected Abstract Service, Catalog of Selected Documents in Psychology.
- Sirgy, M. Joseph (in press b), "Towards a Psychological Model of Consumer Satisfaction/Dissatisfaction," in Proceedings of the Fifth Annual Conference on Consumer Satisfaction/Dissatisfaction and Complaining Behavior.
- The Standard Edition of the Complete Psychological Works of Sigmund Freud (1953-1974) translated by James Strachey in collaboration with Anna Freud, London: Hogarth Press.
- Tan, Chin Tiong and Dolich, Ira J. (1981), "The Moderation Effects of Cognitive Complexity and Prior Product Familiarity of the Predictive Ability of Selected Multi-Attribute Choice Models for Three Consumer Products," in Advances in Consumer Research, vol. 8, Kent Monroe, ed., Ann Arbor: Association for Consumer Research.
- White, D. J. (1969), Decision Theory, Chicago: Aldine Publishing Co.
- Woodruff, Robert B. (1972), "Measurement of Consumer's Prior Brand Information," Journal of Marketing Research, 10, 53-62.
- Wright, Peter L. (1973), "The Cognitive Process Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-62.

A MULTIVARIATE ANALYSIS OF THE PERCEPTION
OF VALUE FROM RETAIL PRICE ADVERTISEMENTS

Albert J. Della Bitta, University of Rhode Island
Kent B. Monroe, Virginia Polytechnic Institute and State University

Abstract

This study attempts to expand the scope of previous research on consumers' reactions to retail comparative price advertising. A large factorial experiment with multiple response measures examined the influence of several variables on consumers' perception of value from retail sale advertisements. The size of sale discount was found to have an interesting effect on consumers' perceptions of value. Also, the method of presenting sale information significantly influenced perceptions.

Introduction

The issue of consumers' response to price has stimulated considerable research interest in the past decade (Monroe 1973, 1977; Olson 1977). However, much of this inquiry has been directed at the price/"quality"-perception question. Except for the concept of unit pricing, one area receiving relatively little attention is the issue of presenting price information to consumers. This is unfortunate since numerous alternatives exist for presenting prices, and such information is an integral component of much retail advertising. Consequently, research into the price presentation area could have significant implications for managerial and regulatory decision-making.

Several schemes are useful for conceptualizing the problem. Jacoby and Olson (1977) have viewed price perception from the S-O-R perspective. Here stimulus variables (S) are the actual cues presented to the consumer where organismic variables (O) include those factors internal to the individual influencing the acquisition, processing, and derivation of subjective meaning from objective stimuli. These O-variables could include properties such as price awareness, subjective standards of reference and attitudes. The interaction of acquired stimuli and organismic conditions leads to responses (R) by the individual.

Clearly, presentation of an offered price qualifies as an S-variable which can significantly influence a consumer's perception of the offer. However, it should also be clear that the consumer's experiences and the context of an offering can influence perception of the offered price. Thus, price experiences over time and the prices of other products may be internalized as an O-variable to influence price perceptions. To illustrate, adaptation-level theory and empirical evidence suggest that the consumer's price experience forms a specific and quantifiable standard of reference for judging subsequent prices (Della Bitta and Monroe 1974; Nwokoye 1975).

In addition to such internalized influences, a variety of factors surrounding the price presentation (store setting, medium used to convey price information, etc.) form a context for judgment. Besides these general contextual variables, a given price presentation may be comprised of a set of more specific contextual cues also capable of influencing the meaning derived from the offered price. Walton and Berkowitz (1980) have classified these cues into comparison and semantic categories. Comparison cues are prices the advertiser references in order to provide a basis for judging the offered price. This may be a prevailing price in the market, the manufacturer's suggested selling price, or the retailer's regular selling price. Semantic cues involve phrases describing the price ("very

low"), price reduction ("significantly reduced") or amount of savings involved ("super savings"). Of course, some semantic phrases are emotionally laden while others are more neutral.

From a managerial and public policy viewpoint, a valid concern is how strongly price presentation formats influence buyers. The managerial concern is to present price information in a favorable manner while hopefully not misleading consumers about the nature of the offering. From a public policy perspective, the concern is over misleading and/or deceptive practices. In fact, in 1958 and 1964 the FTC issued a set of guides for presentation of retail price information.

Relevant Research

What is significant to note is the paucity of directly applicable research evidence to guide managers and policy makers. However, the few relevant studies available do provide some perspective. Fry and McDougall (1974) investigated buyers' perceptions of advertised sale and regular prices. Although the study probably suffered from experimental demand effects, the authors concluded that acceptance of advertised regular prices tended to vary inversely with the size of discounts. However, acceptance of the sale price as the "lowest price in town" varied directly with discount magnitude.

Barnes (1975) varied price and semantic information in a simulated newspaper advertised "sale" for unbranded products. Each presentation was ranked on a priori basis by Barnes in terms of the amount of its information content. The ranking (high to low) was disclosure of: (1) both regular and sale price, (2) sale price and the discount expressed in percentage terms and (3) only sale price. Newspaper and store type were also varied in the experiment. Dependent measures were believability of the advertisement, perceived value for the money and motivation to act. Results indicate a significant information treatment effect, with presentation format (1) above having the greatest influence on the dependent measures. Although this result is consistent with Barnes' a priori ranking of information content, one should note that such a ranking may not be valid in all circumstances. In some situations consumers may find an expression of the magnitude of the price reduction (e.g., percent off) to be more useful than knowing the specific sale price.

Keiser and Krum (1976) compared consumers' reactions to sale advertisements of different forms. In one, only sale price was shown while in a second condition both regular and sale prices were presented with the semantic phrase "one-half price sale." The authors report that the second condition influenced more subjects to perceive a "true price reduction" although the difference was not significant. Perceived truthfulness and willingness to purchase were also not significantly different between the two treatment groups. The possibility of confounding effects in this study should also be mentioned. In the presentation that disclosed both regular and sale prices, the former was set artificially high to examine consumers' reactions to overstated price savings. This element, not in the other treatment condition, may have introduced a bias into the study.

Sewall and Goldstein (1979) queried 114 catalog store

shoppers about their understanding of comparison (reference) prices used by catalog showroom retailers. They found that the vast majority clearly understood the meaning of reference price information. Also, 58 percent of the respondents understood or were explicitly skeptical (discounted the credibility) of such reference prices and regarded them as useful in their comparative shopping activities. It should also be noted that subjects can be influenced by such comparative information more than conscious responses to surveys might suggest.

Blair and Landon (1979) examined the effects of reference prices in retail advertisements for a TV and food processor using adult women in a shopping mall setting. One condition contained only the offered price while another presented the offered price, suggested list price and claimed savings. Results indicate that without reference price information subjects generally estimated savings at between 10 to 12 percent. However, advertisements with reference prices (and discounts from 16 to 36 percent) produced significantly larger perceptions of savings. Also the perceived credibility of claimed savings varied inversely with the magnitude of the claim.

Walton and Berkowitz (1980) reported a study of the effect of reference price, presentation format and several other contextual variables on consumers' price perceptions. Three products (aspirin, fan and camera) were tested at two discount levels (20% and 40%) and four information presentation formats: (1) "regular price"/"sale price", (2) "total value"/"sale price", (3) "compare at"/"our price", and (4) "x percent off"/"now only". Several dependent measures were analyzed using multivariate analysis of variance. Two particularly interesting findings are: (1) reference price effects varied across products, and (2) price information format was significant for cameras only. Here the "percent off" format (4) consistently received the lowest mean rating on all dependent measures.

Although informative, the major problem with previous studies is the lack of consistent findings. For example, Barnes' found the regular/sale price format to yield the greatest perception of value for the money. Although Walton and Berkowitz did not find this to be the case, they did observe a significant difference between the "percent off" format and all others. The two studies also show a similar contrast regarding the influence of information treatments on subjects' willingness to purchase. In addition, while Blair and Landon found the regular/sale price format to influence perceived savings, Walton and Berkowitz were unable to distinguish this influence from two other presentations. Discrepancies also appear between the findings of other studies.

Differences in test conditions may explain some of these divergent findings. Studies have varied in terms of the discount levels, information formats and dependent measures used. Further, they have tested different product classes, branding conditions and price levels. Additional research that accounts for some of these contextual influences is warranted.

The Study

This study investigated the influence of specific price presentation formats on consumers' perceptions. The focus was on a sale situation that consumers might confront in a retail setting.

Research Method

Prior to the experiment, content analysis of major metropolitan newspapers and catalog advertising revealed the frequency with which various components of sale information were employed. Four components were chosen for

investigation based on their frequency of occurrence: regular (non-sale) price, sale price, percentage price reduction, and absolute dollar amount of price reduction. Because previous research suggests that the influence of price presentation format could be effected by the price level and discount size, these factors were also incorporated.

A 2x5x8 factorial experimental design was used. The respective factors were comprised of two price levels, five discount levels (10% to 50% in 10 percentage point increments) and eight presentations of regular price (RP), sale price (SP), percent off (PO) and amount off (AO) information. The specific treatment levels are (1) SP; (2) RP, SP; (3) RP, AO; (4) RP, PO; (5) RP, SP, AO; (6) RP, SP, PO; (7) RP, AO, PO; (8) RP, SP, AO, PO.

A calculator was used as the experimental product because of its familiarity and potential interest to college students who were employed as subjects. Two specific models were used at regular suggested price levels prevailing in the market at the time (\$120 and \$50). The 400 participating subjects were randomly assigned (five per cell) to the 80 treatment conditions. All subjects were instructed to assume that they saw the provided advertisement which described the product and exposed them to the price presentation format. They were then asked to read the advertisement and respond to a series of dependent measures and personal background/debriefing questions.

The majority of price investigations have only examined a single dimension of buyers' response to price. However, since price presentations have potential for affecting a variety of buyer reactions, single variable studies forego opportunities to explore the potential richness of buyers' multidimensional responses to price presentations (Engel, Blackwell and Kollat 1978; Jacoby 1978; Wind and Denny 1974). This study focused on three response variates: perceived savings, perceived value for money at the sales price and perceived acceptability of the offer. These variables were selected because of a desire to concentrate attention on the perceived worth construct identified in other work (Szybillo and Jacoby 1974; Walton and Berkowitz 1980) and the attention given to the three variables in over nine previous investigations. Each of the three dependent variables was measured on a seven-point equal-interval scale anchored by descriptive phrases appropriate to the variable involved.

Analysis and Results

Preliminary analysis revealed that the three dependent variables exhibited substantial intercorrelations ($r \geq .58$). The three variables were also evaluated for internal consistency using Cronbach's (1951) alpha criterion. The resulting alpha value of 0.81 is quite large for this type of basic research (Nunnally 1967), suggesting high consistency among the response variates.

In cases where multiple response measures are correlated, separate univariate analysis of variance (ANOVA) runs on each dependent variate can lead to incomplete and/or inappropriate conclusions (Green 1978; Tatsuoka 1971; Wind & Denny 1974). For this reason a multivariate analysis of variance (MANOVA) procedure as summarized in Table 1 was selected. Wilks' Lambda is an appropriate statistic for examining differences among group centroids when two or more dependent measures serve as the joint set of evaluation criteria. The distribution of the F-statistic is proportional to the distribution of Lambda and is employed as the test statistic.

A first point of interest in Table 1 is that all interaction terms are nonsignificant. This allows direct assessment of main treatment effects as opposed to examining the influence of treatments at individual values of the other factors.

Table 1
Summary of MANOVA Results

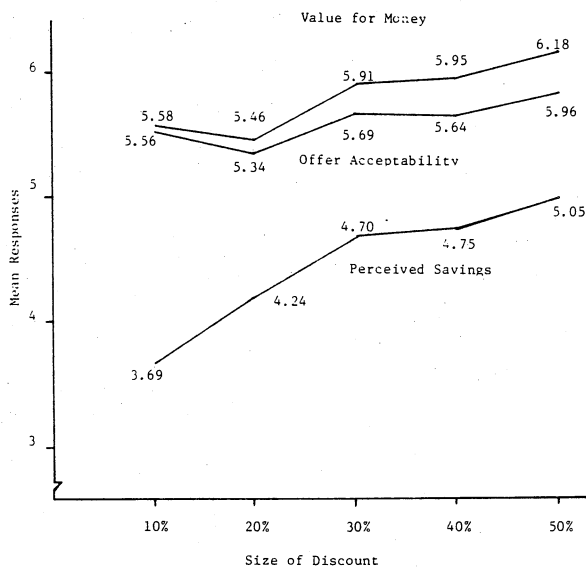
Source of Variation	Wilks' Lambda	F-Value ^a	Probability
A. Price Level	0.9875	1.34 (3,318)	0.2591
B. Size of Discount	0.7843	6.74 (12,841)	0.0001
C. Sale Information	0.8671	2.22 (21,913)	0.0014
A x B	0.9883	0.31 (12,841)	0.9872
A x C	0.9427	0.90 (21,913)	0.5864
B x C	0.7520	1.13 (84,952)	0.2023
A x B x C	0.7847	0.96 (84,952)	0.5891

^aDegrees of freedom are indicated in parenthesis under each F-value

As shown in Table 1, price level treatments did not produce a significant difference in subjects' perceptions. However, examination of the group centroids suggests some tendency for subjects to perceive less value for the money in the more expensive model of the test product.

Table 1 also reveals that discount size and information treatments generated significant differences in subjects' perceptions. For these situations it is appropriate to engage in further analysis to isolate the nature of the differences. While a variety of techniques can assist in this exploration (Green 1978; Hair *et. al.* 1979; Tatsuoka 1971) these data were further analysed using a simultaneous confidence interval approach to multiple comparison tests (Morrison 1976). This approach to follow-up testing for a MANOVA is highly conservative since it holds a maximum alpha level for all possible comparisons of a given type. Therefore, the actual alpha level on a given test can be quite small, requiring a considerably large value for the computed test statistic. For this reason, and the exploratory nature of the research, differences that approached significance at the five percent probability level are discussed along with those that actually achieved significance.

FIGURE 1
Mean Responses to Discount Treatments

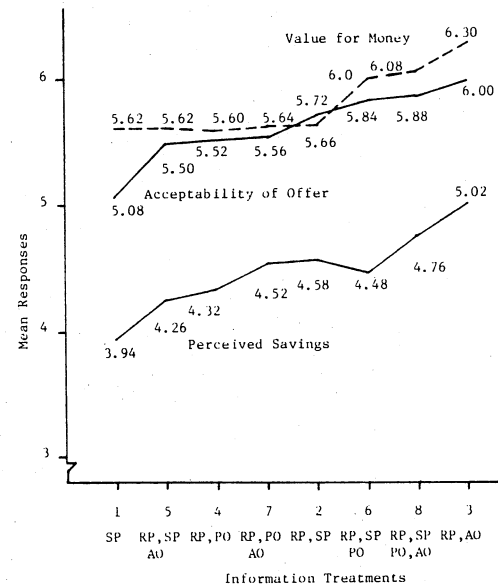


As can be seen in Figure 1, the general pattern of the three dependent variables is to increase as the magnitude of the discount increases. Significantly higher savings were perceived at the 50, 40 and 30 percent discount levels than at the 10 percent level. The 50 percent level also produced significantly higher perceptions of savings than the 20 percent treatment condition.

Larger discounts also increased the general level of perceived offer acceptability and value for money at the sales price. As Figure 1 shows, the one exception is at the 20 percent discount level where, relative to the 10 percent level, perceived offer acceptability and value for the money dropped slightly. In fact, the only group differences that approached or achieved significance on these measures were the difference in responses to the 20 percent and 50 percent-off treatments.

Since the information format treatment produced significant differences these responses were also explored through construction of simultaneous confidence intervals. Mean response patterns for the three dependent measures are presented in Figure 2. Information treatments have been arranged on the X-axis so that responses are generally increasing with movement along the axis. Also, since information treatments differed in non-quantitative aspects, their presentation on the X-axis represents categories of a factor rather than values of a continuous variable.

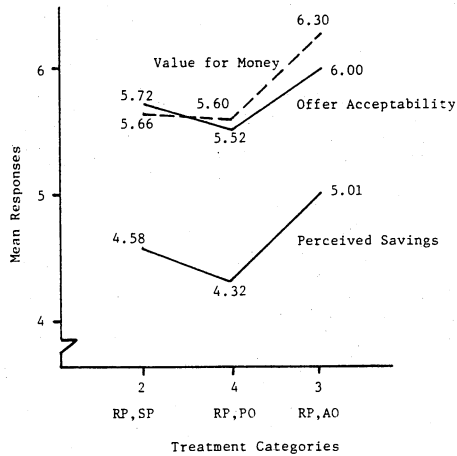
FIGURE 2
Mean Responses to Information Treatments



The difference between responses to information treatment one ("sales price" only) and three ("regular price and amount off") was accepted as significant for the perception of savings measure. As shown in Figure 2, treatment condition three accounted for a higher level of perceived savings. It also produced a nearly significantly greater perception of offer acceptability than did treatment one, and it accounted for the second largest mean difference on the perceived value for money measure. Further, the difference between mean responses to treatment four ("regular price and percent off") and three ("regular price and amount off") approached significance for perceived value for money with treatment three producing higher responses.

It is also useful to examine the pattern of responses as the character of information changes while the number of inputs in the advertisement remains constant. Figures 3 and 4 portray mean responses to two and three informational

FIGURE 3
Mean Responses to Two Information Treatments



inputs respectively. As Figure 3 shows, the pattern of responses to two inputs is quite consistent across the three dependent measures. The highest responses are under conditions of "regular price and amount off" information. This was followed by responses to the "regular price and sales price" treatments while the "regular price and percent-off" treatment generated the lowest level of perceptions.

FIGURE 4
Mean Responses to Three Information Treatments

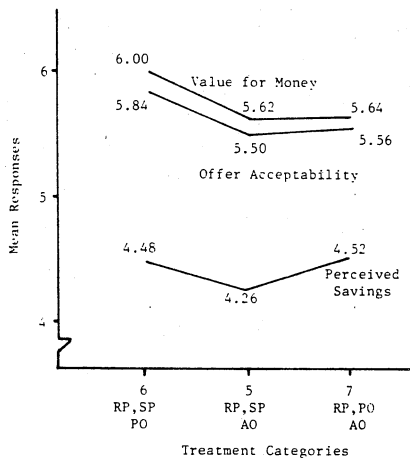


Figure 4, where responses to three informational inputs are portrayed, again shows a quite consistent pattern across the dependent measures. However, this pattern is not consistent with responses shown in Figure 3 where provision of "amount off" information as the second input led to the highest level of mean responses. That is, for three informational inputs the provisions of "amount off" information in conjunction with "regular price" and "sales price" tended to depress responses below those obtained when "percent off" information was presented with "regular price" and "sales price" information.

Discussion

This experiment was designed to investigate how individuals respond to comparative price advertisements under various conditions. The conditions involved different original

price levels, discount magnitudes and formats for presentation of price information. Response measures were perceived savings, value for the money at the sale price and offer acceptability.

Price level (\$120 vs. \$50) did not have a statistically significant effect on subjects' perceptions. However, there was a tendency to perceive more value for the money in the less expensive unit. This may suggest that the lower priced-lined unit more clearly matched the perceived needs of subjects than the higher priced model which contained more sophisticated features.

The magnitude of price discounts did produce significant differences in perceptions. Although all three dependent measures showed a very similar pattern of responses, the perceived savings variate accounted for most of the significant difference. This result closely resembles, and further confirms, the findings of Walton and Berkowitz (1980) regarding the effects of price discounts.

It is interesting to examine the character of savings perceptions across levels of price discounts. Significant differences occurred between the 10% and 30-50% levels and between the 20% level and 50% level. However, response differences between the 30%, 40% and 50% levels were not significant. This result might be attributable to subjects being suspicious of the larger discount offers. However, a separate question asked subjects to indicate the degree to which they suspected the truthfulness of the advertised prices. No significant differences in the perceived credibility of offers was detected across discount levels. Therefore, the argument that subjects' suspicions accounted for different responses to discount levels is not strongly supported. Conversely, it is interesting to observe that the threshold of significant differences in perceived savings occurs in the neighborhood of the 15% price reduction that many retailers believe must be achieved to attract consumers to a sale.

The method of presenting sale price information produced significant differences in subjects' perceptions. However, the nature of these differences appears to be complex. Presentation of only sales prices produced significantly lower perceptions of savings than did the presentation of regular price and dollar amount-off information. This result is consistent with Blair and Landon's (1979) finding that presenting only sale price led to significantly lower perceived savings than presenting reference price information with the sales price. It is also interesting to note that the average perceived savings response to all conditions employing reference price information (4.56) exceeded the average response to presenting sales price only (3.94). This shows further consistency with Blair and Landon's findings. Additionally, although the difference was not significant, format six that presented regular price, sale price and relative price reduction information led to greater perceptions of savings than when only sales price was presented. This mirrored Keiser and Krum's (1976) findings for these information presentation conditions.

Although price information formats did not significantly influence perceptions of offer acceptability and perceived value, the pattern of these responses closely reflected the perceived savings responses just described. Again, results are consistent with Barnes' (1975) findings where presentations using regular and sales prices produced significantly higher perceptions of value than expressions involving only sales price (with semantic phrase) or sales price with relative price reduction information.

When examining the effect of various price presentation formats while controlling for the amount of informational inputs, no significant differences occurred, although a rather inconsistent trend of responses was observed. For two informational inputs, presenting regular price and price reduction information in absolute terms yielded the highest

response levels across all three variates, followed by presentation of regular and sales prices. Presenting regular prices and the discount expressed in percentage terms yielded the lowest response levels. However, for presentations involving three inputs the regular price, sale price, percent-off format generally yielded the highest response levels, while the regular price, percent-off, amount-off format was only marginally higher than the regular price, sale price, amount-off format. No clear response pattern emerges from this analysis especially when comparisons between two and three informational inputs are attempted. Therefore, results suggest that while a variety of presentations involving more information than just current sales price may favorably influence consumers' perceptions of savings, offer acceptability and value for money, the relative effectiveness of these inputs is presently unclear. Further research is needed across additional subject pools, product categories and price presentation formats.

References

- Barnes, James G. (1975), "Factors Influencing Consumer Reaction to Retail Newspaper 'Sale' Advertising," in Edward Mazze (ed.) 1975 Combined Proceedings, Chicago: American Marketing Association, 471-477.
- Blair, Edward A. and Landon, E. Laird, Jr. (1979), "The Effects of Reference Prices in Retail Advertisements," Working Paper, University of Houston.
- Cronbach, L. J. (1951), "Coefficient Alpha and the Internal Structure of Tests," Psychometrika, 16, 297-334.
- Engel, James, Blackwell, Roger D. and Kollat, David T. (1978), Consumer Behavior, Hinsdale, Illinois: The Dryden Press, 579.
- Della Bitta, Albert J. and Monroe, Kent B. (1974), "The Influence of Adaptation Levels on Subjective Price Perceptions," in Scott Ward and Peter Wright (eds.) Advances in Consumer Research Vol. 1, Urbana, Illinois: Association for Consumer Research, 359-369.
- Fry, Joseph N. and McDougall, Gordon H. (1974), "Consumer Appraisal of Retail Price Advertising," Journal of Marketing, 38, 64-74.
- Green, Paul E. (1978), Analyzing Multivariate Data, Hinsdale, Illinois: The Dryden Press.
- Hair, Joseph F., Jr. et. al. (1979), Multivariate Data Analysis, Tulsa, Oklahoma: Petroleum Publishing Co.
- Jacoby, Jacob. (1978), "Consumer Research: A State of the Art Review," Journal of Marketing, 42, 87-96.
- Jacoby, Jacob and Olson, Jerry C. (1977), "Consumer Response to Price: An Attitudinal, Information Processing Perspective," in Yoram Wind and Marshall Greenberg (eds.) Moving Ahead with Attitude Research, Chicago: American Marketing Association, 73-86.
- Keiser, Stephen K. and Krum, James R. (1976), "Consumer Perceptions of Retail Advertising with Overstated Price Savings," Journal of Retailing, 52, 27-36.
- Monroe, Kent B. (1973), "Buyer's Subjective Perceptions of Price," Journal of Marketing Research, 10, 70-80.
- Monroe, Kent B. (1977), "Objective and Subjective Contextual Influences on Price Perception," in Arch G. Woodside, et al. (eds.) Consumer and Industrial Buyer Behavior, New York: North-Holland, 287-296.
- Morrison, Donald F. (1976), Multivariate Statistical Methods 2nd edition, New York: McGraw-Hill Book Co.
- Nunnally, J. (1967), Psychometric Methods, New York: McGraw-Hill Book Co., 226.
- Nwokoye, Nonyelu G. (1975), "Subjective Judgments of Price: The Effects of Price Parameters on Adaptation Levels," in Edward Mazze (ed.), 1975 Combined Proceedings, Chicago: American Marketing Association, 545-548.
- Olson, Jerry C. (1977), "Price as an Informational Cue: Effects on Product Evaluations," in Arch G. Woodside, et al. (eds.) Consumer and Industrial Buyer Behavior, New York: North-Holland, 267-286.
- Sewall, Murphy and Goldstein, Michael H. (1979), "The Comparative Price Advertising Controversy: Consumer Perceptions of Catalog Showroom Reference Prices," Journal of Marketing, 43, 85-92.
- Szybillo, George J. and Jacoby, Jacob. (1974), "Intrinsic vs. Extrinsic Cues as Determinants of Perceived Product Quality," Journal of Applied Psychology, 59, 74-78.
- Tatsouka, Maurice M. (1971), Multivariate Analysis, New York: John Wiley & Sons.
- Walton, John R. and Berkowitz, Eric N. (1980), "Contextual Influences on Consumer Price Responses: An Experimental Analysis," Journal of Marketing Research, forthcoming.
- Wind, Yoram and Denny, Joseph. (1974), "Multivariate Analysis of Variance in Research on the Effectiveness of TV Commercials," Journal of Marketing Research, 11, 136-142.

THE EFFECT OF GENERIC PRODUCTS ON CONSUMER PERCEPTIONS AND BRAND CHOICE¹

John J. Wheatley, University of Washington

Abstract

Consumer perceptions of the quality of national and private brands are apparently altered when generics are introduced in at least some product categories. National brands sales could be hurt by the improved quality image of private brands while the private brands could be adversely affected because generics offer even greater price savings to price conscious buyers.

Introduction

Generic food products were first introduced into supermarkets in the United States in 1977. They are now offered to consumers in thousands of food stores and many consumers seem to have been willing to try them. They have garnered a significant 11% market share in product categories in which they are available, suggesting that many consumers are satisfied with the quality of at least some generics (Yao 1979).

While the impact of generics on both manufacturers of "national" brands and retailers or "private" brands has been acknowledged and there has been a substantial amount of discussion and publicity given this now not-so-new development in the trade press, there has been very little attention paid to this significant change in consumer behavior by academicians. Sarel and Sewall (1980) point out what does exist "provides a mostly descriptive perspective of current position and trends." The study by Murphy and Laczniak (1979) is typical. It consisted of a telephone survey of supermarket shoppers dealing with their opinions and behavior with respect to generics. Frequency of purchase, what the respondents reported they had been buying prior to purchasing generics and future buying plans were the sorts of issues examined. Strang, Harris and Hernandez (1979) using personal interviews with shoppers in a supermarket chain determined the proportion buying generics, their income levels, the reasons given for purchase or non-purchase and shopper perceptions of the quality of generics.

No research has been reported to date dealing with consumer perceptions of specific generic products or with explaining the effect on consumers' choice behavior of introducing generics into the marketplace. The purpose of this paper is to explore these two issues.

Pragmatic and Theoretical Considerations

The generic food products that are being offered to consumers all have certain characteristics in common. They are without brand names, in very plain packages with simple labels and usually sold at prices below both the national and private brands with which they compete. The price differential is typically on the order of 10%-35% below the national brands, and 10%-20% below the private brands, but sometimes it may be as much as 65% below manufacturer's brands (Strang, et al. 1979). Occasionally, however, when national and/or private brands are on sale they may be priced below generics.

It appears that many retailers have introduced generics in order to attract price conscious shoppers on the one hand while other food retailers have added them to their product mix as a defensive measure in order to keep from losing patrons to both the low overhead "box stores" and to those merchants who initiated the introduction of generics.

Notwithstanding the trade press reports that national brands in particular have been adversely affected by the introduction of generics or the expectation that generics would take sales primarily from private brands (Strang, et al. 1979), it seems logical to suppose that both national and private brands will lose market share to generic products.

While it is plausible to believe that national brand buyers are basically quality seekers and many consumers purchasing store brands are attracted by the fact that private brands usually sell at lower prices, it seems reasonable to suggest that a significant proportion of both groups would reassess their purchase decision when confronted with an expanded choice set. A theoretical basis for expecting an altered decision making process is the fact that, as Berelson and Steiner point out (1964, p. 154):

Judgments . . . are . . . made within a frame of reference established by the total range of the relevant stimuli. For example, when the total range of light varies from x to 2x, a stimulus at 2x will be considered white; but when the total range goes from 2x to 4x, 2x will be seen as black and 4x as white.

Thus, consumers' perceptions of both national and private brands could be altered by the addition of a generic product offered at a lower price.

Helson's (1964, p. 54) adaptation level theory offers a more elaborate explanation of this phenomenon. The theory postulates that individual attitudes, judgments and behavior are the result of adaptation to environmental forces. It assumes that a new stimulus is judged on the basis of what has gone before and further that a person has adapted to those prior stimuli and internalized them. However, as Helson says:

Adaptation is . . . a two way affair. Effects of stimulation [also] initiate changes within the organism. These changes adapt the organism to prevailing [emphasis added] conditions.

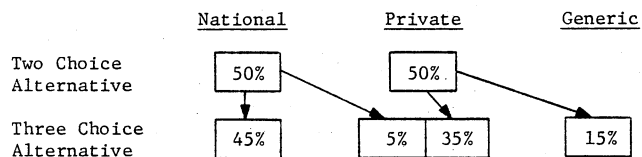
In other words, a consumer might be expected to alter his or her perceptions of an existing brand when confronted with "appropriate" information such as the existence of a new "brand" selling at a lower price. This new information will be assessed on the basis of the person's existing level of adaptation. (Adaptation level may be conceived as a point of perceived neutrality; it is a kind of "average" of all the relevant stimuli to which consumers have been exposed to in the past.) Thus, a new stimulus will be compared to this neutral anchor point to arrive at an assessment and simultaneously, *ceteris paribus*, the stimulus will push the neutral point in its own direction.

In the context of a situation in which a less expensive product is added to an existing choice set, it seems logical to expect that those individuals looking for low prices would be likely to select the generic alternative, while those seeking what they perceived to be high quality would

¹The author wishes to thank his colleagues, especially K. Pettit, D. MacLachlan, R. Yalch and J. Chiu, for their helpful comments on an earlier draft of this paper.

continue to choose the national brand. However, adaptation level theory would lead to the expectation that once the private brands were under-priced by a generic alternative, the store brands should have a tendency to lose their "bargain" image and instead be viewed as "moderately" priced by some consumers. In other words, the perception of the price of the private brand would be altered; it would no longer be "cheap" and, since it has been demonstrated that price serves as a cue to perceptions of product quality (Raju 1977, Valenzie and Eldridge 1973, Wheatley and Chiu 1977), a proportion of the national brand buyers could decide to switch to the private brand because of its enhanced quality image. In addition, some previous private brand purchasers would be satisfied with the quality-price characteristics of the store brand and would remain loyal to it. Thus, the impact of generics might look something like Figure 1.

FIGURE 1
Brand Shares



Of course, this conceptualization of the consequences of introducing generic brands does not encompass all possible outcomes. For example, some national brand users might be inclined to try a generic product just to see what it was like while others might find themselves encouraged to do so because it represents a way of coping with an inflation induced reduction in their real incomes. Some essentially random brand switching would also take place.

Following the insightful observation by Leavitt (1954) that perceived differences among brands in a product class leads to the utilization of price as a cue to product quality, it also seems plausible to suggest that consumers who believe that there is little or no difference in quality among brands would be more likely to purchase the less expensive private brand in the two-alternative choice situation and to switch to generics when they are available. Those consumers who feel that there is quite a lot, or a great deal of difference, among brands in a particular product category on the other hand should tend to do just the opposite.

This behavior would also be consistent with Rao's (1972) suggestion that consumers try to maximize perceived quality per unit price paid. Minimizing price for a given level of perceived quality, a strategy likely to emerge in a recession and in periods such as the late 1970's when the real incomes of many consumers have declined, would be a logical course of action for consumers who perceive the quality of generics or private brands to be essentially equal to that of national brands.

Research Hypothesis

This line of reasoning led to the development of the following hypotheses:

1. H_0 There will be no alteration in consumers' perceptions of private brands as a result of the introduction of generic products to a choice set.
- H_A Consumers' perceptions of private brands will be enhanced as a result of the introduction of a generic product to a choice set.

2. H_0 There will be no differences in brand choice behavior attributable to perceived differences in quality among brands in a product class.
- H_A There will be a positive correlation between the price of the brand chosen and perceived differences in quality among brands in a product class. The greater the perceived difference, the more likely the consumer is to choose a higher priced brand.

Procedure

A complete list of the generic products carried by a large food chain in a Western city was obtained and six of them were selected randomly for the purpose of this investigation. Those chosen were: catsup, canned peach halves, canned pork and beans, shortening, tea bags, and canned whole tomatoes.

The six products were displayed on a table in one of the chain's stores. The actual prices at which the products were being sold in the store were indicated on each product. Every third person entering the store was intercepted and asked to participate in a survey. A total of 75 randomly assigned respondents were presented with a display of the six products which contained only the leading national and the store's private brand of each product. A second sample of 75 subjects were shown another product display that included a national, a private, and a generic "brand" of each product.

The work was carried out between 4 and 8 p.m., on a consecutive Wednesday and Thursday. The position of the various brands in the product display was rotated for both groups of respondents to avoid a position bias effect.

Each respondent was asked to indicate which one brand they would purchase in each product category and the reason for their choice. They were also asked about their perceptions of the quality of each of the brands of one of the product categories in the display.

Results

The results of the choices made by the two groups are summarized in Table 1. Two significant observations may be made about the findings.

First, it would appear that the effect of generics on national and/or private brands varies and is product specific. This finding is consistent with Sarel and Sewall's (1980) observation that market shares for generics vary substantially, e.g., 27.3% for paper and plastic plates to 4.5% for canned tuna. It is also quite possible that the effect depends on the strength of the consumer franchise held by the national and/or private brands in a product category. The national brand of catsup was not much affected by the introduction of a generic alternative, but the private brand was hit very hard. On the other hand, both the national brand and the store brand were about equal contributors to the share-of-market gained by the generic brand of tea bags.

Second, notwithstanding the fact that generics seem to gain market share at the expense of both national and private brands, private brands were bigger losers overall in this sample.

The proportion of the subjects choosing the generic, when it was available, is larger than was anticipated on the basis of market share data. One possible explanation for this is that at least some food shoppers are still unaware of the availability of generics while all of the subjects in this experiment were made aware of their existence by virtue of the product display that they saw. It is also

true as Cagley, et al. (1980) point out generics can average a 30% market share when they are featured by a food chain.

TABLE 1
Brand Choice

		National Brand	Private Brand	Generic Brand	Number of Non Users
Catsup	n ₁ =67	77.6%	22.4%	--	8
	n ₂ =72	69.4%	4.2%	26.4%	3
Peaches	n ₁ =51	31.4%	68.6%	--	24
	n ₂ =63	34.9%	31.7%	33.4%	12
Pork and Beans	n ₁ =53	60.4%	39.6%	--	22
	n ₂ =61	52.5%	19.6%	27.9%	14
Shortening	n ₁ =58	63.8%	36.2%	--	17
	n ₂ =65	47.7%	15.4%	36.9%	10
Tea	n ₁ =58	63.8%	36.2%	--	17
	n ₂ =53	45.3%	17.0%	37.7%	22
Tomatoes	n ₁ =65	47.7%	52.3%	--	10
	n ₂ =65	36.9%	29.2%	33.9%	10
Total	n ₁ =352	58.2%	41.8%	--	
	n ₂ =379	48.3%	19.3%	32.4%	

One product, shortening, was selected from the group of six to deal with the two research hypothesis.

Table 2 reveals the reasons given by the respondents for choosing shortening.

TABLE 2
Selection Reason When the Generic Brand Is or Is Not Available - Shortening

	No Generic	Generic
National		
Price	0.0 %	--
Brand Loyalty	54.1 %	58.1 %
Quality	27.0 %	32.3 %
Name Brand	16.2 %	6.5 %
Other	2.7 %	4.5 %
Private		
Price	95.2 %	50.0 %
Brand Loyalty	0.0 %	20.0 %
Quality	4.8 %	20.0 %
Name Brand	0.0 %	--
Other	0.0 %	10.0 %
Generic		
Price	--	95.8 %
Brand Loyalty	--	--
Quality	--	--
Name Brand	--	--
Other	--	4.2 %

The null hypothesis of no alteration in consumer perceptions of private brands as a result of introducing generic products can be rejected. As anticipated, the proportion of participants claiming that the private brand was

selected for price reasons declined from 95.2% to 50% when the generic product became an element in the choice set. This was statistically significant at $p < .01$. Similarly, the proportion mentioning quality as a reason for choosing the store brand rose from 4.8% to 20%. The image of the private brand was altered; a sizeable proportion of shoppers who selected the private brand now saw them as offering quality and fewer saw them providing price savings.

Finally, the reason given for selecting the generic product was almost unanimously because of its lower price.

The subjects were asked to indicate if they felt there were any differences in quality among brands of shortening and how big these differences were. They were asked to use a 5 point scale with 1 representing no difference, 2 very little difference, 3 some difference, 4 quite a lot of difference, and 5 a great deal of difference.

It was anticipated that the beliefs held by consumers concerning quality differences among brands in a product class would affect choice behavior. If a buyer believed that the differences were small they would choose the least expensive alternative, i.e., the private brand in the two choice set and the generic in the three choice set. However, if the differences were believed to be large, such persons would be more likely to avoid the low price alternative.

The null hypothesis of no difference in brand choice behavior attributable to perceived differences in quality among brands can also be rejected, in part.

The results in Table 3 indicate, as predicted, that in the three alternative choice situations, those consumers choosing the national brand believed that the quality differences among brands of shortening was greater than the differences perceived by those choosing either the private brand or the generic. A pairwise comparison of the means revealed that the difference was large enough to be statistically significant, $p < .05$. While the perceived difference in brand quality between those choosing the private brand and those selecting the generic is in the expected direction, the difference between them is not statistically significant.

TABLE 3
Mean Perceived Differences Among Brands of Shortening*

	Consumers Choosing the National Brand	Consumers Choosing the Private Brand	Consumers Choosing the Generic
3 alternatives	3.31	2.60	2.38
2 alternatives	3.28	2.41	

* 1 represents no difference and 5 a great deal of difference

In the two alternative conditions the perceived quality difference between those choosing the national brand and those selecting the private brand is also significant at $p < .01$.

Limitations

Although the subjects in this exploratory study were allowed to indicate that they did not use the products shown to them or did not wish to indicate a purchase preference they were confronted with what some of them may have felt was a forced choice situation. They were also shown only one national brand in each product category and the experiment was conducted in a single store.

Furthermore, two different samples were used rather than a single sample of subjects.

Finally, in connection with the second alternative hypothesis, the tendency to choose the national brand was probably enhanced due to the fact that the subjects did not have to make an actual purchase.

Conclusions

The introduction of generic products into supermarkets appears to affect consumer behavior in a complex manner. Consumer perceptions of prices and the price-quality ratio or "value" of national and especially private brands can apparently be altered when generics are introduced in at least some product categories. Sales of national brands could therefore be hurt by the improved quality image of private brands while the private brands could be adversely affected because generics offer even greater price savings to price conscious buyers. In other product categories, specifically those in which consumers perceive only small quality differences among brands, it seems likely that this effect may not materialize, at least to the same extent.

Consumer's beliefs about differences in quality among brands in a product class appear to influence consumer behavior too but in a more straightforward manner. Those who perceive large differences tend to choose national brands presumably because they think that national brands are of better quality than retailer's brands or generics. (While most observers agree that there are quality differences among national, private and generic brands, there is very little objective evidence available on this point.) Those individuals perceiving smaller differences choose private brands while those who perceive the least difference among brands tend to choose generics.

Parenthetically, it would seem that the advertising strategy of some national brand manufacturers and retail food chains of stressing the quality of all their offerings as a way of blunting the inroads of generics on their share-of-market seems unlikely to succeed with those consumers who feel that all brands in a product class are very much alike.

It is to be hoped that the results reported here will interest other researchers in pursuing this subject further. Especially useful would be an experiment confirming these findings utilizing a longitudinal, within subject design. Only in this manner can an examination of individual consumer brand switching, and the possible reasons for it, be studied.

The introduction of generics is a significant development worthy of attention simply because of the magnitude of its effect on consumer behavior and because growing environmental pressures such as inflation serve to call these products to consumers' attention with increasing frequency. It is also of considerable significance to marketers of packaged goods. If sellers wish to exercise some degree of control over this development they must try to understand the process by which consumer attitudes and behavior are being affected by generics.

This exploratory study, planned to be the first of a series, also suggests that it is an intellectually interesting and challenging phenomenon as well.

References

- Berelson, Bernard and Steiner, Gary (1964), Human Behavior (New York: Harcourt, Brace and World, Inc.).
- Cagley, J.W., Neidell, L. and Boone, L. (1980), "The Wheel of Retailing Squeaks But Turns: Generic Labeling in Supermarkets," Proceedings, Chicago, Ill. American Marketing Association.
- Helson, H. (1964), Adaptation Level Theory (New York: Harper and Row).
- Leavitt, H.J. (1954), "A Note on Some Experimental Findings About the Meaning of Price," Journal of Business, 27 (July), 205-210.
- Murphy, P.E. and Laczniak, G.R. (1979), "Generic Supermarket Items: A Product and Consumer Analysis," Journal of Retailing Vol. 55 No. 2, 3-14.
- Raju, P.S. (1977), "Product Familiarity, Brand Name and Price Influences on Product Evaluations," Proceedings of the Seventh Annual Conference of the Association for Consumer Research, 64-71.
- Rao, Vithala (1972), "A Model for Brand Choice Under Price Quality Hypothesis," American Marketing Association Proceedings, No. 34, 366-371.
- Sarel, D. and Sewall, M.A. (1980), "An Exploratory Investigation of the Managerial Implications of Generic Brands," Proceedings, Chicago, Ill. American Marketing Association.
- Strang, R.A., Harris, B.F., and Hernandez, A.L. (1979), "Consumer Trial of Generic Products in Supermarkets: An Experimental Study," Educator's Conference Proceedings American Marketing Association 386-388.
- Valenzi, E. and Eldridge, L. (1973), "Effect of Price Information, Composition Differences, Expertise and Rating Scales on Product-Quality Rating," Proceedings, 81st Annual Convention, American Psychological Association, 829-30.
- Wheatley, J.J. and Chui, J. (1977), "The Effects of Price, Store Image and Product and Respondent Characteristics on Perceptions of Quality," Journal of Marketing Research, XIV (May), 181-86.
- Yao, M. (1979), "Generic Products Are Winning Noticeable Shares of Market From National Brands, Private Labels," Wall Street Journal, August 10, 6.

DETERMINANTS OF CONSUMPTION CUE UTILIZATION IN IMPRESSION FORMATION:
AN ASSOCIATION DERIVATION AND EXPERIMENTAL VERIFICATION

Russell W. Belk, The University of Utah

Abstract

This paper reports the results of two studies investigating the characteristics of product and service choices that tend to make these choices useful cues for inferences about the personality and social class of those who are observed consuming such items. The first study is a regression-based analysis of these determinant characteristics across a set of 39 types of products and services. Utilizing three characteristics that this analysis suggested are the strongest determinants of the usefulness of these product and service choices in impression formation, the second study experimentally varied the characteristic properties of a single consumption item choice in order to verify the results suggested by the regression model. The results tend to support the utility of cost and decision involvement as determinants of the item's influence on impressions about its user, but did not support a similar influence from the amount of variability in the choice set.

While a person's choice of products and services is certainly not the only source of cues for assessing that person's character and position in society, certain choices that are visible to others may be among the only cues available when observers encounter someone with whom they have had little or no prior interaction.¹ Although there has been little formal investigation of the role that consumption choices play in communicating information about the consumer to others², it seems clear that certain consumption choices such as furniture and clothing selections are generally perceived to be more expressive than other consumption choices such as insurance and menu selections. The particular choices among items within a product or service category may convey different consumer information depending upon factors such as the item's color, style, brand name, and condition. In addition some types of products and services may be better cues to the consumer's characteristics than others because, for instance, they are more familiar, costly, unique, or represent a longer commitment to the choice. This paper is concerned with the identification of the consumption choice characteristics which make that choice act as a weaker or stronger cue for inferring the states and traits of consumers observed using products and services with these characteristics.

Prior Research

Because of the lack of previous research directly investigating the consumption cue attributes that result in the utilization of certain cues in impression formation, it is necessary to begin by considering the nature of products and services that have been found to influence such perceptions of people and then attempt to infer the characteristics of these consumption choices which cause them to be treated as reliable clues to personality and social class. One category of consumption cues which has been most clearly documented to affect impression formation is clothing. While we have departed from the era when only royalty were allowed to wear certain colors and fabrics such

¹Cue visibility is in fact one of the characteristics examined in this paper as a potential determinant of cue reliance.

²For reviews of the research which has been done see Belk (1978, 1980) and Holman (1979, 1980).

as sable, there is strong evidence that clothing still imparts messages concerning social status (Douty 1963; Hoult 1954; Rosencranz 1962; Laswell and Parchall 1961; Wise 1974; Sommers 1964; Dellinger 1977; Bickman 1971; Lefkowitz, Blake, and Mouton 1955; Veblen 1899). It is also clear that clothing can serve as a strong cue for inferences about other user traits and can affect subsequent actions toward these consumers (Buckley and Roach 1974; Golden, Allison, and Clee 1978; Lefkowitz, Blake and Mouton 1955; Coursey 1973; Gibbins 1969; Rosenfeld and Plax 1977; Hamid 1968, 1969, 1972; Johnson, Nagasawa and Peters 1977; Thornton 1944; Thibaut and Riecken 1955; Holman 1980, forthcoming; Darley and Cooper 1972; Suedfeld, Bochner, and Metas 1971). Automobile ownership and selection is suggested by the literature to be another product area with definite implications for inferences about the owner's status and personality (Jacobson and Kossoff 1963; Bruce-Briggs 1977; Green and Wind 1973; Doob and Gross 1968; Ferber 1962; Porter 1966; King and King 1980a, 1980b; Martineau 1957). And the literature suggests that homes or residences also commonly allow inferences about the personality and social class of their residents (Canter, West, and Wools 1974; Felk 1978; Wedin, Avant, and Wolins 1973; Verhure, et al. 1977). Furniture and furnishings within a home, even including appliances, seem to serve similar purposes for impression formation (Sommers 1964; Chapin 1935; Ferber 1962; Laumann and House 1970). Nevertheless, Felson (1976, 1978) provides evidence that suggests that at least for inferences of social status, clothing, automobiles, housing and furnishings are no longer as reliable a set of cues as they were in earlier decades in the United States. Blumberg (1974) has suggested that such a decline in the usefulness of certain consumption cues for inferences of social status (and we may speculate, for inferences of personality traits as well) may be due to rising availability of such goods and declining constraints of income on the ability to choose freely within these product classes. This explanation also finds some support in extensions of Brock's (1968) commodity theory that support the hypothesis that scarcer goods are better able to convey messages concerning the status or uniqueness of the consumer (Fromkin, Olson, Dipboye, and Barnaby 1971; Szybillo 1975; Worchel, Lee and Adewole 1975). Thus one key feature of product and service choices which appears to affect their utility as cues for impression formation is their current uniqueness.

Among the other product and service choices that have most frequently been found to alter inferences about the personality and social class of consumers are cosmetics (Calder and Burnkrant 1977; McKeachie 1952; Belk 1978; Sommers 1964; Golden, Allison, and Clee 1978), books and magazines (Green and Wind 1973; Porter 1966), foods (Haire 1950; Belk 1978; Woodside 1972; Sommers 1964), and leisure products and activities (Porter 1966; Belk 1978; Bishop and Ikeda 1970; Sommers 1964). One factor these consumption categories seem to have in common is that they all provide a large number of substantially different choices. Without the freedom to make distinct choices which this characteristic provides, there would be little information conveyed in a consumption choice. Thus the variety of choices available may be a second determinant of the reliance placed on a particular product category in forming consumption-based impressions of others.

Several additional cue attributes appear to be related to consumption-based impression formation when the literature

cited is carefully examined. One such cue attribute, especially for inferences of social status, is the cost of the consumption item. Generally the more costly the item, the more related it seems to be to inferences about social class. Two potentially related cue attributes which may nonetheless be independent of cost in many cases, are the length of time to which the consumer is committed to a choice and the amounts of time and thought that go into the selection decision. Here it would be expected that where there is a more long-term commitment to the product choice (e.g., choice of colleges versus choice of restaurants) and where the decision is commonly more carefully thought out (e.g., choice of magazine subscriptions versus choice of television programs), the choice will be a more reliable cue to inferences about the consumer, especially that consumer's personality.

For more axiomatic reasons it may be seen that another property related to the use of the foregoing consumption cues is their visability or noticeability. A cue must be able to be detected before it can possibly have any effect on impression formation. While the consumption items found to serve as cues differ in their visability, they are all more visible than, for instance, underwear or furnaces. Two exceptions distinguishing visability from noticeability may be the cases of innovations and gifts. In both instances items that may not normally be visible may become evident through heightened interest and conspicuous display.

Two final cue properties that may elevate consumption choices to the status of consumption symbols are the complexity and the rate of stylistic change currently associated with the item. Complexity may be capable of endowing a consumption choice with surplus meaning about its consumers because of the objective ambiguity created when a product or service is beyond most observers' ability to fully understand and objectively evaluate. In these circumstances the choice not only evokes a degree of mystery, it also may give rise to more subjective inferences than would be likely for a simpler product. A product or service subject to frequent or substantial style changes may produce a similar atmosphere of mystery and subjectivity because its selection is presumably more "a matter of taste," and such tastes presumably stem from personality and social class.

In summary, a consideration of possible reasons why a scattered literature suggests that only certain product and service selections are commonly related to inferences about their owners led to hypotheses about eight consumption cue attributes. Consumption choices are predicted to be more useful to inferences about the social class and personality of their users when these choices possess or involve: uniqueness, variety, high cost, long-term commitment, thoughtful selection, noticeability, complexity, and stylistic change. All eight of these characteristics were investigated associatively in Study 1 which acted as a screen on the characteristics to be investigated experimentally in Study 2.

Study 1

Method

Sixty-eight undergraduate students were drawn from first year business students at the University of Illinois in both studies. All 68 completed ratings of the 39 products and services shown in Table 1. The products were broken into three sets to avoid fatigue and were rated by subjects on the eight characteristics derived above using a 5-point scale of agreement that a selection from the product or service category was _____.³ In addition each product or

³The attributes are shown in Table 2 as they appeared in the data collection instrument.

service was rated on the same scale as to whether it was "a good clue to personality" and whether it was "unrelated to social class".

Two types of regression analyses were conducted using the 39 products or services as observations, personality or social class ratings as dependent variables, and the eight other ratings as predictor variables. In the first type of regression the mean scores over subjects were used for both the dependent and predictor variables. In the second type of regression, 68 separate regressions were run for each dependent variable, using the scores of one subject at a time. The two sets of results were then compared.

TABLE 1
Products/Services and Examples Used in Study 1

1. Outerwear (e.g., sweatshirt, windbreaker)
2. Residence (e.g., apartment, dormitory)
3. Furniture (e.g., modern, colonial)
4. Credit Card (e.g., American Express, Visa)
5. Department Store Charge Card (e.g., Zayre, Penny's)
6. Pet (e.g., dog, cat)
7. Main Course (e.g., fish, steak)
8. Cosmetics (e.g., green eyeshadow, tan eyeshadow)
9. Automobile (e.g., sedan, station wagon)
10. Transportation (e.g., bus, train)
11. Entertainment (e.g., bar, movie)
12. Stereo (e.g., component, console)
13. Religion (e.g., Mormon, Protestant)
14. Ring (e.g., classring, turquoise stone ring)
15. Swimming Suit (Male: e.g., boxer style, brief style)
16. Watch (e.g., digital, sweep-second hand)
17. Eyeglasses (e.g., wire rimmed, horn rimmed plastic)
18. Hair Care (Male: e.g., barber, hair stylist)
19. Bicycle (e.g., ten speed, one speed)
20. Facial Tissue (e.g., 1-ply, 2-ply)
21. Camera (e.g., 35mm, pocket instamatic)
22. Laundry (e.g., coin-operated, over-the-counter)
23. Toothpaste (e.g., flouride, whitening)
24. Underwear (Male: e.g., boxer style, brief style)
25. Underwear (Female: e.g., brief style, bikini style)
26. Attache Case (e.g., brown Naugahyde, brown leather)
27. Life Insurance (e.g., term, whole life)
28. Flowers (e.g., cut, potted)
29. Book (e.g., Huckleberry Finn, Fear of Flying)
30. Newspaper (e.g., Wall Street Journal, News Gazette)
31. Magazine (e.g., Time, Readers Digest)
32. Winter Recreation (e.g., cross-country skiing, snowmobiling)
33. Cigarette (e.g., low-tar, regular)
34. Gum (e.i., bubble, sugarless)
35. Overnight Accommodations (e.i., hotel, motel)
36. Shoes (e.g., dress, athletic)
37. Telephone (e.g., desk phone, slimline phone)
38. Television (e.g., color, black and white)
39. Boat (e.g., sailboat, motor boat)

Table 2 shows a summary of the mean scores regression analysis for the two dependent variables. While there is some multicollinearity in the data, the full equations showed four partially overlapping variables to be significant predictors of both dependent variables. The three shared predictors were variety, thoughtful selection, and high cost. In addition, noticeability was strongly significant in the personality equation and uniqueness was a significant predictor in the social class equation. As expected, all but the one non-significant simple correlation coefficient (uniqueness with personality) were positive, showing that the hypotheses derived from the literature are substantially correct according to the average judgments of these subjects.

In the 68 within-subject regressions, the results are highly supportive of the mean scores regression. These results are summarized in Table 3.

TABLE 2
Summary of Regression Analyses in Study 1
Based on Mean Scores

Predictor Variable	Dependent Variable			
	Personality		Social Class ¹	
	Simple r	Beta Weight	Simple r	Beta Weight
Uniqueness	-.12	-.10 ₂	.52	.36 ₃
Variety	.49	.40 ₂	.36	.21 ₂
Stylistic Change	.52	.07	.22	.11
Long-term Commitment	.37	.01	.41	-.20
Thoughtful Selection	.59	.29 ₂	.48	.27 ₂
Noticeability	.73	.66 ₃	.31	.14
Complexity	.32	-.05 ₂	.36	-.02 ₃
High Cost	.40	.23 ₂	.52	.36 ₃
Overall F	8.8		5.2	
Significance	p < .0005		p < .0005	
Adjusted R ²	.50		.33	

¹ Reverse scoring of "unrelated to social class".

² p < .05

³ p < .025

TABLE 3

Summary of Within-Subject Regression Analyses in Study 1
(Number of Significant¹ Beta Weights by
Sign and Dependent Variable)

Predictor Variable	Dependent Variable			
	Personality - Betas + Betas		Social Class ² - Betas + Betas	
Uniqueness	12	3	0	62
Variety	16	38	14	26
Stylistic Change	1	17	2	8
Long-term Commitment	6	8	14	13
Thoughtful Selection	4	41	6	29
Noticeability	7	54	1	12
Complexity	9	5	8	7
High Cost	13	26	3	48
Average Adjusted R ²	.38		.29	

¹ p < .05 ² reverse scoring of "unrelated to social class"

In fact, the numbers of positive beta weights are in almost exactly the same rank orders over the eight predictor variables as the rank orders of beta weights in the mean score regressions summarized in Table 2. This lends strong support to the conclusions that products available in greater variety, involving more thoughtful selection, and costing more are thought to be better clues to personality and social class, and that personality inferences are also based on more noticeable consumption cues just as social class inferences favor more unique cues.

Because of the intercorrelations of the predictor variables no claim can be made that each of the significant variables in the regression equations is a better predictor than each of the non-significant variables. However we can be more confident that the significant predictor variables which are shared for the two dependent variables are judged to be strong cues to personality and social class even when the other variables are taken into account. But this conclusion is based on judgments of cue usefulness rather than actual utilization of these cues, is based on judgments of cue attributes rather than objective differences in cue attributes, and is based on associational evidence

rather than an experimental investigation. Therefore the second phase of this research took the three best shared predictor variables found in the regression study (variety, decision thought, and cost) and sought to test their ability to alter actual impression formation in an experimental context.

Study 2

Method

The three consumption cue properties identified to be predictors of social class and personality were investigated in a 2 x 2 x 2 factorial experimental design. In order to gain additional control over extraneous variables all three predictors were manipulated in the context of a single product—an attache case. Business student subjects were asked to attempt to describe the market which they felt would be attracted to a new attache case to be marketed by a luggage specialty store. The further description of the store and product contained the manipulations. The following description shows the two levels of cost, selection involvement, and variety respectively:

A luggage specialty store will be selling a new attache case made by Samsonite next month. The attache case is medium in size and comes in tan naugahyde. The suggested retail price is \$29.95 [59.95], which is an average price for an attache case today. This particular case incorporates the new slim look with hidden latches and minimal hardware. Most attache cases including the Samsonite products change their designs and materials about every five years. The salespeople of the luggage store have noticed that most customers buying attache cases spend about 10 minutes in the store and examine 2 or 3 [30 minutes...and examine most] of the dozen [three dozen] models which the store carries, prior to deciding.

Subjects were then instructed to make their best guesses about the average demographic characteristics of customers for the new product (average age, average income, percent who are female, most likely type of occupation [7 choices arrayed by status], average education [5 levels], percent who are single, and percent who have children). Of these characteristics, three were combined into an additive index of social class for further analysis: occupational status (level from 1 to 7, multiplied by 2 to increase its relative weight), income (in thousands of dollars per year) plus education level (some high school or less = 1... advanced degree = 5). In addition, subjects were asked to estimate typical customer personality traits on 13 bipolar adjectival scales with 7 scaling positions.⁴ Although a multivariate analysis of variance showed a significant effect on this set of ratings for the main effects of two of the three treatment variables (cost and decision involvement) it was felt that univariate analyses of variance would prove more enlightening. Since there were 10 subjects in each of the 8 cells of the experimental design, it was not feasible to conduct within cell factor analyses of scores on the 13 personality traits, so a single factor analysis with all 80 subjects was performed in order to reduce the number of dependent personality variables and increase the reliability of the measures through the use of factor scores. Table 4 shows the results of the final factor analysis retaining 4 factors that captured just over 76 percent of the variance in responses. It appears that the personality factors derived might be labeled "maturity", "likeability", "warmth", and "happiness". Factor scores for each of these four dimensions were derived using regression-based procedures. These factors as well as the social class index formed the final dependent variables for analysis of the experiment.

⁴ These adjectives were selected based on the work of Rosenberg, et al. (1968) and Schlenker (1975).

TABLE 4
Factor Analysis of Personality Traits (Varimax Rotation)

Trait	Factor Loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
Friendly	-.03	.52	.24	.30
Intelligent	.71	.28	.04	-.03
Dull	-.08	-.60	-.09	-.24
Likeable	.07	.73	.23	.07
Sympathetic	.08	.04	.67	-.03
Responsible	.79	.02	.08	.09
Successful	.71	.26	-.12	.08
Attractive	.25	.57	-.01	.04
Sincere	.23	.21	.33	.11
Happy	.11	.33	.02	.87
Warm	-.21	.37	.73	.28
Impulsive	-.37	.04	-.05	-.06
Creative	.30	.12	.31	.48

Results

A manipulation check was included after all other results had been collected. Based on t-tests on mean differences between treatment level means for 5-point scales to estimate the amounts of "variety available", "care and thought going into the purchase decision", and "cost" for an attached case compared to other types of purchases, all three manipulations were successful with a type I error probability of 5% or less.

Table 5 presents the results of the analyses of variance on these four personality factors and the social class index. It may be noted first that the variety manipulation was not significant as a main effect and was only marginally significant in the three-way interactions for likeability and warmth. The other treatment variables were however significant as main effects for social class and for at least two of the four personality factors. All of these effects were positive (e.g., greater cost led to judgments of higher social class). However, the interaction of cost and decision involvement for the last three personality factors shows that customers were judged to be more likeable, warmer, and happier under the combination of higher cost/less decision involvement as well as the combination of lower cost/more decision involvement.

TABLE 5
Significance Levels From Anova's on Each Dependent Variable

Source	Dependent Variable				
	Social Class Index	"Maturity"	"Likeability"	"Warmth"	"Happiness"
A Cost	p < .004	p < .023	p < .058	NS	NS
B Decision Involvement	p < .019	p < .017	NS	p < .037	p < .048
C Variety	NS	NS	NS	NS	NS
A x B	NS	NS	p < .028	p < .042	p < .038
A x C	NS	NS	NS	NS	NS
B x C	NS	NS	NS	NS	NS
A x B x C	NS	NS	p < .072	p < .080	NS

Discussion

The regression results suggesting that cost and decision thoughtfulness (involvement) are positively related to the tendency to use a consumption cue to form judgments of the consumer's personality and social class are supported by the experimental results. Such is not the case for the selection variety available for a product or service however. Contrary to the regression results the manipulated levels of variety did not affect judgments of either

social class or any of the four personality factors derived.⁵

While these results can be properly taken as strong evidence that the cost and decision involvement in selecting a product affect the likelihood that the selection will be used in forming social class and personality impressions of the product user, they should not be considered to be strong evidence that decision variety (or any of the other significant cue properties from the regression model) is unrelated to the usefulness of the consumption item in impression formation. For one thing, the untying of cue properties from the product or service category in the experimental phase of this research is artificial and, even though the manipulation checks were successful, may not have reflected the richness of differences in available variety which occur across product and service categories. Another constraint was the limited and entirely verbal information that was made available about the visual details of chosen and unchosen alternatives. Finally, the dependent variables probably do not reflect the full range of inferences which may be made about people on the basis of their consumption choices. For instance, although the personality variables were chosen from those derived by Rosenberg, Nelson, and Vivekananthan (1968) and by Schlenker (1975), they are probably not exhaustive enough to capture all of the possible nuances in the impressions formed of various product and service users.

Thus the present studies support the conclusion that products and services that are more costly and involve more carefully thought-out decisions are more likely to be utilized in forming impressions of users of these products and services. At the same time this does not rule out the possibility that some of the other cue properties investigated are also determinants of consumption cue utilization in impression formation. Specifically, the regression model suggests that the noticeability and selection variety of a product or service category enhance its usefulness for personality inferences and that uniqueness and selection variety help to determine the usefulness of product and service choices to social class inferences about consumers. In order to evaluate these predictions more fully, additional experimental testing is needed using both within- and between-product-category manipulations of the predictors.

References

- Belk, Russell W. (1978), "Assessing the Effects of Visible Consumption on Impression Formation," in H. Keith Hunt (ed.), *Advances in Consumer Research*, Vol. 5, Ann Arbor: Association for Consumer Research, 39-47.
- _____ (1980), "Effects of Consistency of Visible Consumption Patterns on Impression Formation," in Jerry Olson (ed.), *Advances in Consumer Research*, Vol. 7, Ann Arbor: Association for Consumer Research, 365-371.
- Bickman, Leonard (1971), "The Effect of Social Status on the Honesty of Others," *Journal of Social Psychology*, 85 87-92.
- Bishop, Doyle, W. and Ikeda, Masaru (1970), "Status and Role Factors in the Leisure Behavior of Different Occupations," *Sociology and Social Research*, 54, 190-208.
- Blumberg, Paul (1974), "The Decline and Fall of the Status Symbol: Some Thoughts on Status in a Post-Industrial Society," *Social Problems*, 21, 480-498.

⁵ Significant effects for a treatment on any of the personality factors are sufficient to support the hypothesis that the treatment affects inferences about personality.

- Brock, Timothy C. (1968), "Implications of Commodity Theory for Value Change," in A.G. Greenwald, T.C. Brock, and T.M. Ostram (eds.), Psychological Foundations of Attitudes, New York: Academic Press, 243-275.
- Bruce-Briggs, Barry (1977), The War Against the Automobile, New York: Dutton.
- Buckley, Hilda M. and Roach, Mary E. (1974), "Clothing as a Nonverbal Communicator of Social and Political Attitudes." Home Economics Research Journal, 3, 94-102.
- Calder, Bobby J. and Burnkrant, Robert E. (1977), "Interpersonal Influence on Consumer Behavior: An Attribution Theory Approach," Journal of Consumer Research, 4, 29-38.
- Canter, D., West, S., and Wools, R. (1974), "Judgements of People and Their Rooms," British Journal of Social and Clinical Psychology, 13, 113-118.
- Chapin, F. Stuart (1935), Contemporary American Institutions: A Sociological Analysis, New York: Harper and Brothers.
- Coursey, Robert D. (1973), "Clothes Doth Make the Man in the Eye of the Beholder," Perceptual and Motor Skills, 36, 1259-1264.
- Darley, J.M. and Cooper, Joel (1972), "The 'Clean for Gene' Phenomenon: The Effect of Students' Appearance on Political Campaigning," Journal of Applied Social Psychology, 2, 24-33.
- Dellinger, R.W. (1977), "Keeping Tabs on the Joneses," Human Behavior, 6, 22-30.
- Doob, A.N. and Gross, A.E. (1968), "Status of Frustrator as an Inhibitor of Horn-honking Responses," Journal of Social Psychology, 76, 213-218.
- Douty, Helen I. (1963), "Influence of Clothing on Perception of Persons," Journal of Home Economics, 55, 197-202.
- Felson, Marcus (1976), "The Differentiation of Material Life Styles: 1925 to 1966," Social Indicators Research, 3, 397-421.
- _____ (1978), "Invidious Distinctions Among Cars, Clothes and Suburbs," Public Opinion Quarterly, 42, 49-58.
- Ferber, Robert (1962), "Brand Choice and Social Stratification," Quarterly Review of Economics and Business, 2, 71-78.
- Fromkin, Howard L., Olson, Jerry C., Dipboye, Robert L., and Barnaby, David (1971), "A Commodity Theory Analysis of Consumer Preference for Scarce Products," Proceedings, 79th Annual Convention, American Psychological Association, 653-654.
- Gibbins, Keith (1969), "Communication Aspects of Women's Clothes and Their Relation to Fashionability," British Journal of Social and Clinical Psychology, 8, 301-312.
- Golden, Linda L., Allison, Neil, and Clee, Mona (1978), "The Role of Sex Role Self-Concept in Masculine and Feminine Product Perceptions," in William Wilkie (ed.), Advances in Consumer Research, Vol. 6, 599-605.
- Green, Paul E. and Wind, Yoram (1973), Multiattribute Decisions in Marketing: A Measurement Approach, Hinsdale, Illinois: Dryden.
- Haire, Mason (1950), "Projective Techniques in Marketing Research," Journal of Marketing, 14, 649-656.
- Hamid, Paul N. (1968), "Style of Dress as a Perceptual Cue in Impression Formation," Perceptual and Motor Skills, 26, 904-906.
- _____ (1969), "Changes in Person Perception as a Function of Dress," Perceptual and Motor Skills, 29, 191-194.
- _____ (1972), "Some Effects of Dress Cues on Observational Accuracy, a Perceptual Estimate, and Impression Formulation," Journal of Social Psychology, 86, 279-289.
- Holman, Rebecca H. (1979), "Perspectives on Product-Meaning Research," Working Paper Number 77, Pennsylvania State University, College of Business Administration.
- _____ (1980), "Clothing as Communication: An Empirical Investigation," in Jerry Olson (ed.), Advances in Consumer Research, 8, 372-377.
- _____ (forthcoming) "Apparel as Communication," in Morris Holbrook and Elizabeth Holman (eds.), Consumer Aesthetics and Symbolic Consumption, Ann Arbor: Association for Consumer Research.
- Hoult, Thomas F. (1964), "Experimental Measurement of Clothing as a Factor in Some Social Ratings of Selected American Men," American Sociological Review 19, 324-328.
- Jacobson, Eugene and Kossoff, Jerome (1963), "Self-Percept and Consumer Attitudes Toward Small Cars," Journal of Applied Psychology, 47, 242-245.
- Johnson, Barbara H., Nagasawa, Richard H., and Peters, Kathleen (1977), "Clothing Style Differences: Their Effect on the Impression of Sociability," Home Economics Research Journal, 6, 58-63.
- King, Algin B. and King, Joyce P. (1980a), "Does the Consumer Still Perceive the Automobile as a Status Symbol Under Current Market Conditions: A Behavioral Analysis," paper presented at the Midwest American Institute for Decision Sciences Conference, April 1980, Dayton, Ohio.
- _____ and _____ (1980b), "The Ranking of Automobiles as Status Symbols: A Comparative Analysis of Consumer Perceptions Based on Five Selected Socio-Economic Variables," paper presented at the Midwest American Institute for Decision Sciences Conference, April, 1980, Dayton, Ohio.
- Lasswell, Thomas E. and Parshall, Peter R. (1961), "The Perception of Social Class From Photographs," Sociology and Social Research, 46, 407-414.
- Laumann, Edward O. and House, James S. (1970), "Living Room Styles and Social Attributes: The Patterning of Material Artifacts in a Modern Urban Community," Sociology and Social Research, 45, 321-342.
- Lefkowitz, Monroe, Blake, Robert R., and Mouton, Jane S. (1955), "Status Factors in Pedestrian Violation of Traffic Signals," Journal of Applied Social Psychology, 51, 704-706.
- Martineau, Peirre (1957) Motivation Advertising, New York, McGraw-Hill, 66-80.
- McKeachie, William J. (1952), "Lipstick as a Determiner of Personality," Journal of Social Psychology, 36, 242-244.
- Porter, James N. Jr. (1966), "Consumption Patterns of Professors and Businessmen: A Pilot Study of Conspicuous Consumption and Status," Sociological Inquiry, 37, 255-265.
- Rosenberg, Seymour, Nelson, Canot, and Vivekananthan, P.S. (1968), "A Multidimensional Approach to the Structure of

Personality Impressions," Journal of Personality and Social Psychology, 8, 283-294.

Rosencranz, Mary L. (1962), "Clothing Symbolism," Journal of Home Economics, 54, 18-22.

Rosenfeld, Lawrence B., and Plax, Timothy G. (1977), "Clothing as Communication," Journal of Communication, 27, 24-31.

Schlenker, Barry R. (1975), "Self-Presentation: Managing the Impression of Consistency When Reality Interferes with Self-Enhancement," Journal of Personality and Social Psychology, 32, 1030-1037.

Sommers, Montrose S. (1964). "Product Symbolism and the Perception of Social Strata," in Stephen A. Greyser (ed.), Toward Scientific Marketing, Chicago: American Marketing Association, 200-216.

Suedfeld, Peter, Bochner, Stephan, and Metas, Carol (1971), "Petitioner's Attire and Petition Signing by Peace Demonstrators: A Field Experiment," Journal of Applied Social Psychology, 1, 278-283.

Szybillo, George (1975), "A Situational Influence on the Relationship of a Consumer Attribute to New Product Attractiveness," Journal of Applied Psychology, 60, 652-655.

Thibaut, John W. and Riecken, Henry W. (1955), "Some Determinants and Consequences of the Perception of Social Causality," Journal of Personality, 24, 113-133.

Thornton, G.R. (1944), "The Effect of Wearing Glasses Upon Judgments of Personality Traits of Persons Seen Briefly," Journal of Applied Psychology, 28, 203-207.

Veblen, Thorstein (1899), The Theory of the Leisure Class: An Economic Study of Institutions, New York: Macmillan.

Vershure, Beth, Magel, Stephen and Sadalla, Edward K (1977) "House Form and Social Identity," in Peter Suedfeld, James A. Russell, Lawrence M. Ward, Françoise Szigeti, and Gerald Davis (eds), The Behavioral Basis of Design, Book 2, Stroudsburg, Pennsylvania: Dowden, Hutchinson, Ross, 273-278.

Wedin, Carol, Avant, Lloyd L. and Wolins, Leroy (1973), "Communication of Residential Spaces by Architectural Graphics," in Richard Kuller (ed.), Architectural Psychology, Proceedings of the Lund Conference, Lund, Sweden, 301-309.

Wise, Gordon L. (1974), "Differential Pricing and Treatment by New-Car Salesmen: The Effect of the Prospects' Race, Sex, and Dress," Journal of Business, 47, 218-230.

Woodside, Arch G. (1972), "A Shopping List Experiment of Beer Brand Images," Journal of Applied Psychology, 56, 512-513.

Worchel, Stephen, Lee, Jerry, and Adweole, Akanbi (1975), "Effects of Supply and Demand on Ratings of Object Value," Journal of Personality and Social Psychology, 32, 906-914.

CONSUMER PERCEPTIONS: A DISCUSSION

Eric N. Berkowitz, University of Minnesota

Abstract

The title for this session indicates the extent of commonality between these three papers. Beyond this broad framework, little overlap exists. As a result, a discussion of macro-perceptual issues is difficult. Although the presentation may be a little more tedious, I will discuss each paper in sequence.

For each paper, three questions will be addressed. First, how adequate is the design and methodology employed? Secondly, what additional information is presented in these studies which hitherto has been unknown. And, finally, I will take some liberty to suggest areas of additional research which are suggested by these papers.

Della Bitta and Monroe's "Perception of Value from Retail Price Advertisements"

The study conducted by Della Bitta and Monroe focuses upon the presentation of price information to consumers, and the effects of these presentations on consumers' price evaluations. The authors' motivation for this study is derived some extent from the inconsistencies of prior research. Findings indicate that semantic and comparison cues do affect price perceptions, but the inconsistencies of prior research are not resolved.

Methodologically, this study has a distinct strength in its multivariate treatment of consumer response variables. Advancement in pricing research necessitates this recognition of a multiple response ["R" in the Jacoby and Olson conceptualization] variable framework. There are, however, a few aspects of this design which deserve some comment and re-examination.

Of some concern in this study is the product. The use of a single product in price perception studies must be questioned. As Della Bitta and Monroe note, this research stream while still developing has resulted in several inconsistent findings. Berkowitz and Walton have shown in a multiproduct price response study that consumer perceptions may vary by product class. As a result, findings of single product pricing studies present some difficulty in interpreting their generalizability. A second aspect of the product concern is the product itself. Calculators, as a class of products, are probably highly familiar to the subject population, students. How does this familiarity affect perceptions of discount levels? Does this possible familiarity negate the effect of more descriptive comparison cue presentations? Using the S-O-R framework, the (R)esponses may be moderated by the (O)rganismic variables. Price responses may have to be analyzed in light of "O" variables such as familiarity, reference standards, ownership, etc.

The Della Bitta and Monroe study adds several interesting aspects to the knowledge of price perceptions. Foremost, this study underscores the complexity of consumer price responses and the multidimensional nature of these judgments. Results are particularly interesting in showing the consistently poor response generated by an advertisement with only the actual price cue. And, this study finds the "percent off" semantic cue to be one version which does not favorably affect price perceptions. This finding is consistent with results observed by Berkowitz and Walton (1980). Of greatest interest may be the lack

of response differences across the three highest level discount treatments (30%, 40%, 50%). This finding should imply some reconsideration of retail mark-down policies as well as of discount operations' initial price offerings.

Finally, this study raises several interesting areas of further study. No price study has provided so many possible comparison price cues. While the most descriptive alternative (RP, SP, PO, AO) did not result in the highest positive price perceptions, this more descriptive presentation should be investigated among non-student subjects. Wheatley and Chiu (1977) and Berkowitz and Walton (1980) have observed demographic differences in price perceptions. It is possible that the amount of comparison price cue information necessary to be related to positive price perceptions is dependent upon the education level or age of the respondent.

A second important area for future investigation pertains to the discount levels. Since credibility of the price offering was not found by the authors to be an issue, other explanations for this observation should be explored. Pricing strategies for practitioners might be very dependent upon this area of research.

Finally, the "O" variables of the S-O-R model deserve greater attention. What is the effect of familiarity, usage, expected price level on consumer price responses?

As usual with pricing research, while several important and interesting findings have been added to the base of knowledge, it also has the beneficial effect of suggesting several areas worthy of further exploration.

Wheatley's "Generic Products on Consumer Perceptions and Brand Choice"

The second paper in this session examines the impact of generic brands on consumer perceptions and brand selection. The objective of the study is to assess the effect of generic brands on the share of national and private brands. Wheatley reports that generic brand market share is obtained from both the national and private label, and consumer perceptions of these existing offerings also shift with this new form of competition.

The methodology of this study is unique from the other two papers in this session by its use of non-student subjects. Yet several aspects of this study raise questions with regard to the interpretation of the results. Of primary concern is the selection and meaning of "national" brands. One group of shoppers were presented with two choices (national vs. private label) in six product categories. A second group also was presented with the same six product categories, however, a generic brand was included among the stimuli. How dependent are the observed product selections on the "national" brand being presented to subjects? The national brand was described as the "leading" offering. Is this "lead" with regard to sales at the experimental store/or in terms of national market share? If the latter definition is accurate, non-national brand selection may not be surprising.

A second concern is the demand characteristics of this experiment. Wheatley hypothesizes that the selection of the generic and private label offerings in the three-

choice condition may be understated. In the Limitation section of the paper, Wheatley states "the tendency to choose the national brand was probably enhanced due to the fact that the subjects did not have to make an actual choice". A competing explanation, however, might account for an overly high frequency of generic and private label selection. A consumer confronted with three products at three different price levels and asked which one s/he would choose, might feel compelled to give the impression of rationality. This demand artifact would result in the lowest cost alternative being selected.

A third concern with regard to this study is the previously identified strength, the subjects. Some criticism might be voiced regarding the small number of subjects, yet Wheatley should be complimented for convincing a retailer to allow an in-store study. However, the subjects who did participate do not appear to have been qualified in any way. Did the subjects use any of the existing brands? For example, what percentage of the 75 subjects presented with the three choice criteria use the national brand of catsup? Does this possibility account for the little slippage of the national brand brand choice figures (in the 2 vs. 3 condition)? Results should have been analyzed with regards to existing brand usage. Had any of the subjects in the generic group had prior experience with generic products?

A final concern with this study is the in-depth presentation of only one product category. In light of the variations observed in product selection, the choice reasons should have been gathered (or reported) for all six product categories.

Results of this study are interesting in terms of the observed brand choice shifting in the three choice condition. To some extent, these results must be disconcerting to companies in going against the trade belief that generic gains will be at the expense of store brands (Coyle 1978). Moreover, the selection reasons analysis in the shortening case reveals an interesting perceptual restructuring of brand offerings. With the existence of the generic brand, the differential advantage of private labels noticeably diminishes, but a new position on the quality dimension is revealed.

Wheatley's study suggests a couple of areas worthy of further investigation. First, some attention to the problems discussed with this study might clarify the impact of generic brands. Secondly, large variations across product classes with regard to generic brand selection are observed. What are the reasons for this occurrence? Are these variations due to perceived quality differences across product offerings? What of price differentials between alternative choices? In this study, edible products were investigated. The potential for generic brand inroad may be in non-edible commodities such as paper goods. Future investigations should consider different classes of products. Several of these issues may be resolved in the near future. The author's final statements indicate more information will soon be coming.

Belk's "Determination of Consumption Cue Utilization"

The third paper of this session is the continuation of a research stream in impression formation presented at ACR. Belk's objective in this study is to identify consumption cue determinants that result in impressions of personality and social class. The extensive literature review provides eight possible determinants. A two-stage study is conducted to assess which attributes are significant in forming impressions.

To a great extent, the design of this study appears highly appropriate. Two separate subject groups were used, one to develop the possible predictions and a second to validate the cue determinants. Following the style of this study, I'd like to raise some questions by study.

Study 1 required 68 students at the University of Illinois to assess 39 products on eight dimensions as to whether the products were related to personality and social class. Similar to the first paper in this session, the student subjects may be deserving of some criticism. These subjects, students at a State University, are quite likely to be similar in terms of social class. This homogeneity in social class might affect their own evaluations on this criterion.

A second concern with Study 1 pertains to the meaningfulness of the task. How meaningful is it to rate Pets, or Flowers (e.g., cut, potted) in terms of complexity? A table showing the 39 x 8 means and standard deviations of the student ratings could help eliminate this concern.

Study 2 investigated the significance of the predictors (variety, cost, decision involvement) identified in the first portion of this research. Again, a student subject pool was used, but the product was restricted to briefcases. Several other products such as bicycles, winter recreation, televisions, in Table 1 appear more relevant to student subject 1. Why was this particular product selected for study 2? What effect might familiarity with the product have in impression formation?

Variety is not observed in study 2 to be a significant determinant. Again, the experience/familiarity of students with this product may contribute to this lack of significance. A second possible reason, as Belk also notes, may be in the operationalization of variety. The variety manipulation was done in terms of dozen [three dozen] models. The "models" term is not explained or further defined.

How does this research extend existing knowledge? Belk has provided a good first step in identifying the determinants of cue utilization. Since products have been found to affect impression formation, the identification of significant determinants is critical. Future research in this area has an important area in which to expand. The relevancy of this research in marketplace behavior has been proposed. The need to demonstrate the results of impression formation to marketplace behavior, however, could afford further demonstration.

References

- Berkowitz, Eric N. and Walton, John R. (1980), "Contextual Influences on Consumer Price Responses: An Experimental Analysis," Journal of Marketing Research, XVII (August), 349-359.
- Coyle, Joseph S. (1978), "Generics," Progressive Grocer (February), 74-78.
- Wheatley, John J. and Chiu, J. (1977), "The Effects of Price, Store Image and Product and Respondent Characteristics on Perceptions of Quality," Journal of Marketing Research, XIV (May), 181-86.

TEMPORAL LINKS BETWEEN PREFERENCE AND PERCEPTION

James McCullough, University of Arizona
Douglas MacLachlan, University of Washington
Reza Moinpour, University of Washington

Abstract

This study examines the reliability of preference analysis using LINMAP in situations where perceptions are found to be stable over time. The results of preference analysis are shown to be less stable than perceptions but still reasonably reliable. Suggestions are made on the use of this type of analysis to understand consumer responses to information and persuasive communications.

Introduction

A major goal of promotional activity in marketing is the alteration of consumer preferences for products. In order to effectively utilize information and persuasion in promotional campaigns, it is necessary to understand how preferences are formed and what influences their change.

Marketing researchers have developed methods for examining both preference and perception. Examination of formation and change of preference would seem to necessitate examination of the link between preference and perception. Although several researchers have demonstrated the temporal reliability of measures used to measure preference and perception (Moinpour, McCullough, and MacLachlan 1976; Acito 1977; McCullough and Best 1979), attempts to link preference and perception in longitudinal studies have been limited and unsuccessful (McCullough, MacLachlan, and Moinpour 1979).

Examination of the link between perception and preference requires reliable measurement of both perception and preference, and reliable analysis of the link between the two. Individual differences scaling (INDSCAL) has been shown to be a reliable method for the analysis of consumer perceptions (Carroll and Chang 1970; Moinpour, McCullough, and MacLachlan 1976). Preference ranking is generally a reliable method for the collection of individual preferences (Acito 1977), although, as Best (1978) points out subjects need solid reference criteria to insure reliability. Linear programming techniques for the multidimensional analysis of preference judgments (LINMAP) appear to provide a suitable method for examination of the link between perception and preference (Srinivasan and Shocker 1973), but the reliability of the method has not been demonstrated.

Research Questions

This study is designed to examine the reliability of analysis of preference and perception as a tool for understanding consumer behavior. To accomplish this objective it is necessary to determine if:

1. Perceptions of unchanged stimuli can be reliably scaled at different points in time as reported in the literature.
2. Preferences are stable over time.
3. Preferences and perceptions reported at different points in time for identical stimuli can be reliably linked by joint application of preference and perception analysis.

The research reported here is an attempt to establish baseline controls for the measurement of changes in consumer perceptions and preferences resulting from informational and persuasive messages.

Experimental Method

A convenience sample of twenty-five subjects drawn from undergraduate marketing classes participated in testing sessions during which they were told they were testing the usefulness of new marketing research techniques. At each session, subjects were asked to rate all possible pairs of 10 toothpaste brands (N=45) on nine-point dissimilarity scales and to rank the brands in order of preference. The brands used in the study were chosen to match the previous study by Moinpour, McCullough, and MacLachlan (1976). Two additional brands (Aim and Aquafresh) were included to increase constraint (see Table 1). An attempt was made to replicate the previous study as closely as possible. Subjects completed this task three times at one week intervals. Twenty-four subjects completed the study, providing responses at one week intervals over a three week period.

Analysis

Pairwise dissimilarities were analyzed using the INDSCAL algorithm which produces group stimulus space coordinates and individual dimension saliences. These perceptual spaces were compared using the CMATCH algorithm (Cliff, 1966) while the preference rankings were compared using Spearman rank correlation. The group stimulus space coordinates and preferences rankings for each individual were analyzed using the LINMAP mixed mode option to produce individual dimensional weights or utilities.

Assume that each stimulus (brand) k can be represented by a vector of p coordinates

$$Y = (Y_1, Y_2, \dots, Y_p) \quad (1)$$

The INDSCAL model for measuring individual perceptions assumes the basic structure of the product space is shared by all people in the relevant market, but allows individual subjects to stretch or compress that space by a set of dimension weights or saliences. The group space coordinates

TABLE 1
Toothpaste Brands Used in Study

Aim	Gleem II
Aqua Fresh	Ipana
Close Up	Macleans
Colgate	Pepsodent [®]
Crest	Ultra Brite

by a set of dimension weights or saliences. The group space coordinates Y_{jk} for attribute j and brand k are transformed into individual coordinates X_{ijk} for individual i by the application of subject-dimension weights (saliences) W_{ij} . The algorithm developed by Carroll and Chang (1970) simultaneously derives Y_{jk} , and W_{ij} values from matrices of brand dissimilarities for all individuals using the model:

$$Y_{ijk} = (W_{ij})^{1/2} Y_{jk} \quad (2)$$

The LINMAP Model

The LINMAP is one of many "utility" generating algorithms. It has the advantage of modeling decreasing marginal

utility and ideal points for some attributes and constant marginal utility (vector model) for other attributes. It posits the following functional form for utility of brand k at time t :

$$U_{\hat{k}t}(x_k) = \sum_{j \in j_1} V_j (x_{jk} - I_j)^2 + \sum_{j \in j_2} V_j x_{jk} \quad (3)$$

where j_1 is the set of all attributes having finite ideal points and j_2 is the set of all attributes with infinite ideal points; V_j is the attribute "importance" (i.e., the value of the attribute in determining preference); and I_j is the subject's ideal point on the j th dimension.

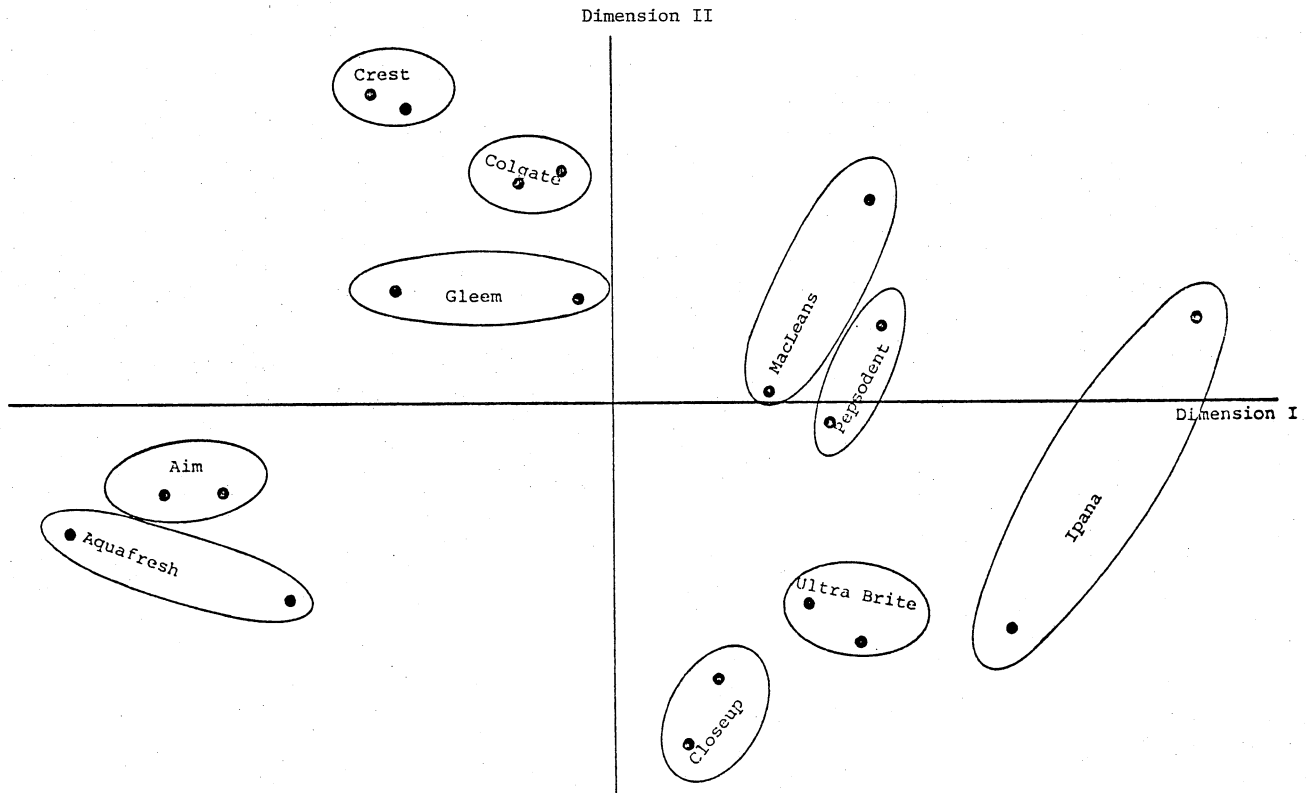
Results and Discussion

The perceptual maps derived from INDSCAL were reasonably consistent over the three week period. The group stimulus configuration is shown in Figure 1. The goodness of fit between periods 1 and 2 was .904 and between periods 2 and 3, it was .762. Distance vector correlations were .746 and .447, respectively. The slightly lower value between weeks 2 and 3 was due largely to variation in the position of Ipana, a brand not familiar to most subjects.

relative importance weights for the spatial dimensions in determining individual utilities for the attributes as they were used by the subject to determine preference. There was weak correlation between these values across the three periods. The correlation between weeks 1 and 2 was .26 and between 2 and 3 was .45.

This result is not surprising. The INDSCAL procedure develops what are in effect individual perceptual maps of the stimulus space. Although the group space may be reasonably consistent between periods there is considerable individual variation in dimensional salience as can be seen from the data in Table 2. In spite of this individual variation, the correlation between individual dimensional saliences is .75 for weeks 1 and 2 and .77 for weeks 2 and 3. In order to correct the utility values derived from LINMAP for the individual variation contained in the INDSCAL results, the utility values must be multiplied by the dimensional saliences. With their correction, the correlation between the adjusted utilities is .61 for weeks 1 and 2 and .74 for weeks 2 and 3. The values of dimensional saliences are shown in Table 2; the LINMAP-generated individual dimension utilities are shown in Table 3; and the adjusted individual dimension utilities are shown in Table 4.

FIGURE 1
Approximate Perceived Positions of Brands
Based on INDSCAL Analysis of Periods I and II



The Spearman rank correlation coefficients for the subject ranked preferences were uniformly high. Between weeks the value ranged between .84 and 1.00 with a mean value of .96. These results indicated subjects were clearly able to perform both the tasks leading to the development of perceptual maps and the preference rank of the stimuli with a high degree of reliability.

The coordinates of the stimulus space and the preference rankings were employed in the LINMAP algorithm to generate

(Tables 2, 3 and 4 follow)

TABLE 2
Dimensional Saliences from INDSICAL¹

	Week 1	Week 2	Week 3
1	.126	.097	.131
2	.212	.550	.363
3	.258	.199	.326
4	.565	.725	.751
5	.688	.667	.531
6	.271	.312	.452
7	.216	.252	.338
8	.154	.039	.244
9	.375	.108	.018
10	.027	.111	.132
11	.431	.446	.544
12	.073	.202	.255
13	.230	.423	.337
14	.335	.390	.454
15	.223	.102	.087
16	.223	.410	.295
17	.147	.300	.531
18	.696	.673	.610
19	.628	.780	.421
20	.345	.365	.086
21	.240	.415	.321
22	.171	.133	.156
23	.322	.048	.051
24	.602	.359	.251
\bar{X}	.318	.338	.320

$\rho_{12} = .75$ $\rho_{23} = .77$ $\rho_{13} = .51$

TABLE 4
Adjusted Individual Dimension Utilities¹

Subject	Week 1	Week 2	Week 3
1	.052	.050	.005
2	.126	.202	.260
3	.180	.167	.239
4	.347	.660	.651
5	.387	.485	.507
6	.154	.129	.290
7	.023	.105	.157
8	.081	.018	.129
9	.253	.017	.006
10	.005	.000	.028
11	.104	.236	.253
12	.012	.074	.165
13	.039	.224	.292
14	.259	.163	.382
15	.176	.011	.078
16	.038	.331	.266
17	.048	.219	.283
18	.229	.389	.285
19	.322	.630	.146
20	.038	.094	.030
21	.067	.001	.113
22	.041	.036	.072
23	.276	.042	.042
24	.170	.190	.217

$\rho_{12} = .61$ $\rho_{23} = .74$ $\rho_{13} = .49$

1. Table 2 x Table 3.

1. Data for Dimension I only are reported. No significant differences in reliability are found between dimensions I and II.

TABLE 3
Dimensional Utilities from LINMAP¹

	Week 1	Week 2	Week 3
1	.419	.521	.044
2	.432	.368	.717
3	.696	.841	.734
4	.615	.910	.867
5	.563	.728	.956
6	.567	.414	.643
7	.110	.418	.466
8	.530	.473	.532
9	.676	.116	.348
10	.203	.004	.216
11	.243	.529	.466
12	.170	.368	.649
13	.170	.529	.867
14	.773	.418	.842
15	.787	.109	.897
16	.170	.808	.900
17	.331	.728	.533
18	.329	.528	.468
19	.513	.808	.348
20	.110	.259	.353
21	.281	.004	.353
22	.243	.271	.466
23	.856	.860	.956
24	.283	.529	.867
\bar{X}	.415	.480	.604

$\rho_{12} = .27$ $\rho_{23} = .45$ $\rho_{13} = .37$

1. Data for Dimension I only are reported. For normalized values $Score_{II} = 1 - Score_I$.

Conclusions

The analysis of perceptions using INDSICAL appears to be reliable as was previously reported (Moinpour, McCullough, and MacLachlan 1976). The increase in the size of the stimulus set from eight to ten did not appear to alter the reliability of the scaling procedure. It was possible that inclusion of additional stimuli might have increased task difficulty and reduced reliability. There is no evidence that this occurred. It should be noted, however, that selection of stimuli can have a noticeable effect on the stability of the map. In this case, Ipana exhibited a much larger variation than Crest; for example, Ipana is not marketed in the area of the study and may not have been familiar to many subjects--it was not used by any. Crest, on the other hand, was the most commonly used product among respondents and exhibited very little temporal movement in the scaling solutions (see Figure 1).

Expressed preferences also appear to be reliable. There is no evidence of significant temporal variation in preference ranking in this study.

The link between perceptions and preference is also quite strong. Although the temporal stability is not as high as might be desired, it is significant and indicates a reasonable consistency in the results.

The major advantage of non-metric analysis is the relative ease of data collection to yield metric output. This type of analysis assumes that sufficient constraint can be placed upon the ordinal data to insure a unique solution. This assumption presents significant problems for perceptual analysis. First, sufficient stimuli must be analyzed to insure adequate constraint. In this case, ten stimuli were used, providing sufficient constraint but requiring subjects to evaluate 45 pairs of products. In spite of this there was no evidence that subjects were unreliable in responding.

The second issue concerns the relationship between the subjects' perceptions and the group space. The INDSICAL

analysis assumes the subjects utilize a common perceptual space and differ in their weighting of the dimensions. The data in Table 5 indicate this may not be true. For some subjects (subject 3, for example) the INDSICAL solution does a poor job of representing their perceptual space. When this occurs, it is not surprising that the results appear to be unreliable. In future analysis, it should be possible to identify subjects with differing perceptions and treat them separately. By doing this, the results should become more reliable.

This study has shown that INDSICAL and LINMAP can be used to jointly analyze preferences and perception. The results appear to be stable over time. Further research is needed to determine if these methods can be used to identify changes in perception and preference to provide a basis for understanding the influence of information and persuasion on the consumer.

TABLE 5
Individual Subject Correlation with the INDSICAL Solution

	Week 1	Week 2	Week 3
1	.797	.876	.855
2	.616	.589	.491
3	.395	.322	.410
4	.604	.745	.776
5	.747	.686	.635
6	.782	.744	.714
7	.732	.700	.776
8	.729	.609	.583
9	.796	.656	.768
10	.850	.800	.784
11	.509	.571	.600
12	.680	.446	.653
13	.666	.488	.483
14	.640	.467	.488
15	.776	.827	.808
16	.403	.413	.417
17	.397	.523	.632
18	.721	.718	.645
19	.653	.779	.463
20	.715	.688	.719
21	.526	.721	.712
22	.658	.803	.739
23	.468	.236	.261
24	.617	.453	.343
\bar{X}	.645	.619	.615

References

- Acito, F. (1977), "An Investigation of Some Data Collection Issues in Conjoint Measurement," Proceedings, Chicago: American Marketing Association, 82-85.
- Assal, A. (1971), "Perceptual Mapping to Reposition Brands," Journal of Advertising Research, February, 39-42.
- Best, R.J. (1978), "Validity and Reliability of Criterion-Based Preferences," Journal of Marketing Research, 15, 154-160.
- Carroll, J.D. and Chang, J.J. (1970), "Analysis of Individual Differences in Multidimensional Scaling via n-way Generalization of 'Eckart-Young' Decomposition," Psychometrika, 35, 283-319.
- Lutz, R.J. (1975), "Changing Brand Attitudes Through Modification of Cognitive Structure," Journal of Consumer Research, 11, 45-49.

MacLachlan, D.L., Moinpour, R., and McCullough, J.M. (1977), "Experimenting with Perceptual Change Strategies," Multivariate Behavioral Research, 12, 429-446.

McCullough, J., and Best R. (1979), "Conjoint Measurement: Temporal Stability and Structural Reliability," Journal of Marketing Research, 16, 26-31.

McCullough, J.M., MacLachlan, D.L. and Moinpour, R. (1979), "Linking Preference and Perception: A Longitudinal Study," Proceedings: American Psychology Association, Division 23, September.

McGuire, W.J. (1968), "The Nature of Attitudes and Attitude Change," Handbook of Social Psychology, Addison-Wesley.

Moinpour, R., McCullough, J.M. and MacLachlan, D.L. (1976), "Time Changes in Perception: A Longitudinal Application of Multidimensional Scaling," Journal of Marketing Research, 13, 245-253.

Parker, B.R. and Srinivasan, V. (1976), "A Consumer Preference Approach to the Planning of Rural Primary Healthcare Facilities," Operations Research, 24, 991-1025.

Pessemier, E.A. (1978), "Stochastic Properties of Changing Preferences," American Economic Review, 28, 380-385.

Srinivasan, V. and Shocker, A.D. (1973), "Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgments," Psychometrika, 38, 473-493.

TEMPORAL INCONGRUENCY IN CONSUMER BEHAVIOR

Philip E. Hendrix, University of Michigan
Claude R. Martin, Jr., University of Michigan

Abstract

Temporal incongruency is presented as a potentially useful concept in consumer behavior. A consumer who spends more or less time in an activity than he prefers may be labelled temporally incongruent with respect to that activity. Explanations for the presence of temporal incongruency are presented and problems with operationalizing the concept are briefly discussed. Tentative results from empirical analyses suggest the usefulness of the concept.

Introduction

Like money, time may be regarded as a scarce resource which consumers spend. The allocation of time to activities is quite analogous to the allocation of money for goods and services. One faces a (time) budget constraint--say 1,440 minutes per day--which limits the total time spent in all activities. Likewise, there exists a set of activities over which time must be allocated. The analogy is not perfect, however, as time cannot be stored and therefore must be spent as it becomes available. A basic decision facing an individual, then, is the amount of time to spend in alternative activities.

Of course, the allocations of time and money are not independent--in fact, the processes are inextricably related. Consumers spend their money for goods and services which either complement or substitute for their time spent in various activities. For instance, a wide array of products complements the time spent in activities typically regarded as leisure pursuits--witness the significant growth in demand for shoes, apparel, and information brought about by the jogging phenomenon. Other goods and services are purchased as substitutes for the time spent in such activities as home maintenance, e.g., maintenance-free siding, long-life light bulbs, etc., and meal preparation, e.g., convenience foods, microwave ovens, and meals away from home.

Some have argued that because of its scarcity, time may well be the crucial dimension of consumer behavior (Nicosia and Mayer 1976, p. 68). Certainly an increase has occurred in the number of marketing strategies in which time is implicitly reflected: express lanes and expanded hours by retailers; incentives for consumption at "non-peak" times in travel, communication, and utilities; new product offerings aimed at meeting the growing demand for convenience foods and meals away from home; and others. Our limited understanding of the temporal dimension of consumer behavior has undoubtedly impeded the development of marketing strategies consistent with consumers' actual or preferred patterns of time use. While various models of time allocation have been proposed and to varying degrees tested (Arndt and Gronmo 1977, Becker 1965, Chapin 1974, Linder 1970, Morgan et al 1966, Walker and Woods 1976), none is adequately specified. As argued elsewhere (Hendrix 1980), these models are notably deficient in their treatment of subjective needs and characteristics of the individual. The usefulness of specifying subjective as well as objective measures in models of time use was demonstrated for several activities by showing that the time which an individual spends in an activity is positively related to the extent to which the individual enjoys the activity. Such a premise is consistent with the cognitive theory of motivation.

One must question, however, whether this premise holds equally across activities and, for a given activity, whether this premise holds equally across segments of the population. In fact, we doubt that it holds unequivocally across either activities or segments. Rather, we suspect that consumers are less successful in spending the preferred amount of time in some activities than in others and that some consumers are more successful than others in spending the desired amount of time in a given activity. In view of the interdependence of the allocations of time and money, the identification of groups of consumers who spend more or less time in an activity than they prefer would facilitate the development of marketing strategies designed to alleviate this disequilibrium. We will label such consumers temporally incongruent with respect to that activity.

Our thinking was stimulated in part by Hobson and Mann (1974), who developed an overall measure of an individual's well-being by comparing his preferred time expenditures to his actual time expenditures. Consistent with their objective, they examined this relationship on an aggregate basis across activities. We feel that for purposes of marketing planning it is most useful to examine this relationship between time expenditures and preferences at the level of the activity. Analysis at this level permits one to examine the degree of temporal incongruency for each activity and to identify groups of consumers who are temporally incongruent with respect to a given activity, e.g., which consumers spend more or less time in the activity than they prefer.

Finally, we wish to emphasize that our primary objectives are to propose the concept of temporal incongruency and to illustrate its potential usefulness in marketing planning. Issues associated with operationalizing the concept are discussed but by no means resolved. For purposes of demonstration, we present an operationalization using available data and discuss the findings. In so doing, we hope to stimulate much needed research addressing the difficult task of adequately measuring temporal incongruency.

Why Temporal Incongruency Might Exist

Why would an individual spend an amount of time in an activity other than that which he preferred? Why would an individual spend any time in an activity, such as meal preparation, which he dislikes? Conversely, why doesn't an individual spend his or her time only in those activities which he or she most prefers? Various factors affect one's ability to spend time as one might prefer. One of the most significant factors is the frequent incompatibility of immediate and longer term preferences. An individual may be forced to engage in activities which he regards as less preferred in the short run to secure longer-term rewards--work, education, personal care, and home and auto maintenance may represent examples of such activities. Juster (1978) quite aptly labels such behavior "investment"--one foregoes present benefits to obtain future benefits, behavior which is, by definition, investment. We know in fact that investment motives create some temporal incongruency. In response to the question, "Would you say that you do a lot of things that you don't like to do so that your life will be better in the future?" some 50% of the respondents in the ISR Time Use Study described below replied yes or depends.

Other activities require complements which may not be present--for example, weather conducive to the activity; a suitable location; other individuals; and so on. Norms and role orientations may also compel individuals to spend more or less time in an activity than might be expected on the basis of their enjoyment of the activity.

Such factors may singly or collectively suppress the expected congruence between actual and preferred time expenditures. Examining the disparity between the actual and preferred time spent in an activity requires, perhaps obviously, valid measures of each. Less obvious, though, are the difficulties in obtaining such measures.

Data on Time Expenditures

Methods of gathering time expenditure data fall into two major categories: self report and observations by a second party. The latter is subject to its own limitations (Wells and LoSciuto 1966) and, due to cost, is practical in only a limited number of situations. There are a number of approaches to eliciting time use information directly from respondents. Perhaps the most prevalent approach has required the respondent to estimate his or her "typical" time expenditures over some limited period, e.g., a week (Hawes 1977), or frequency of participation over an extended period of time. The accuracy of such estimates has been questioned, however, by recent validity studies (Robinson 1977; forthcoming). Though it too relies on self report, the time diary method is generally regarded as the most accurate source of time use data.

The time diary is an instrument in which the respondent records (either synchronously or shortly thereafter) the activities in which he or she engages over the period of interest, usually 24 hours. Typically recorded for each event are beginning and ending times, concurrent activities (e.g., watching television while eating lunch), location-and others present. Figure 1 shows a few of the entries corresponding to more or less discrete events in a hypothetical time diary containing the elements mentioned above.

A time budget showing the allocation of time, say 24 hours, to various activities can be constructed by summing the time devoted to each activity during the day. Hence, the hypothetical respondent whose diary appears in Figure 1 spent 8 of his 24 hours in sleep. Aggregates of the time devoted to other activities may be derived similarly. These measures complement the fine-grained data found within the diary itself.

Since the diary collects data for a fairly short period of time, estimates at the individual level of time spent in activities in which individuals infrequently engage are subject to wide variation. For instance, an activity in which a respondent engages once a week only has a probability of 52/365 of showing up in his time diary for a given day. One way to reduce this variance in estimates of an individual's time expenditures is to collect a larger number of time dairies.

The analyses in this paper utilize a subset of data collected by the Survey Research Center of the Institute for Social Research, University of Michigan, in a major study of time use among Americans in 1975-1976. The national probability sample consisted of respondents, 18 years of age and older, and their spouses (1,519 and 887 respectively for a total of 2,406 individuals). The design specified that each individual provide on four separate occasions, approximately three months apart, a time diary for the previous 24 hour period. Considerable attitudinal, behavioral, and demographic data pertinent to the use of time were also obtained across the four waves of the study. In addition to controlling for seasonal variations in time use, the design avoided the bias due to differences between weekdays and weekend by collecting time diaries from each individual for two weekdays, a Saturday, and a Sunday. This complex design yielded the most reliable, representative measures of time use available for a national sample. A time budget containing estimates of the number of minutes spent in each of some 89 activities during a given week was derived for each individual from his or her time diaries.

Measures of Preferences

As suggested earlier, there exists a set of activities over which time must be allocated. Thus, one might regard activities as objects of choice capable (to varying degrees) of satisfying an individual's needs. Given an individual's hierarchy of needs, their saliences, and subjective reports of the extent to which various activities satisfy those needs, relatively preferred activities may be determined, perhaps relying on an approach analogous to multi-attribute modelling of product preferences. Though no one to our knowledge has used such an extensive approach, considerable information could be gained pertinent to preferences for alternative activities as well as underlying attributes of activities which determine their ranking.

FIGURE 1

SAMPLE OF EVENTS FROM
A TYPICAL TIME DIARY

TIME		PRIMARY ACTIVITY	SECONDARY ACTIVITY	LOCATION	OTHERS PRESENT
BEGAN	ENDED				
00:00	06:30	Slept	----	Home	Alone
06:30	06:45	Washed	----	Home	Alone
06:45	07:00	Dressed	Talked with Wife	Home	Wife
07:00	07:30	Ate Breakfast	Watched TV	Home	Wife, Child
07:30	07:55	Drove to Work	Listened to Radio	Street	Alone
o	o	o	o	o	o
o	o	o	o	o	o
o	o	o	o	o	o
22:20	22:30	Dressed for Bed	----	Home	Alone
22:30	24:00	Slept	----	Home	Alone

An alternative approach is to ask the respondent directly to articulate his or her preferences for activities, e.g., in which activities would the respondent prefer to spend his or her time. Such an approach assumes, of course, that the respondent can array his preferences. Moreover, it ignores the potential effect of situational factors on preferences--for example, the hierarchy for an evening following an 8 hour day at work is likely to differ from the one for a Saturday evening. Though simpler, this approach does not permit one to determine why one activity is preferred over another. One is forced to regard these preference measures as exogenous, which may be unacceptable for certain analytical purposes.

In using this direct approach, problems with interpretation may arise unless one employs an unambiguous term "descriptive of a person's affective state" (Jones and Pierce 1977, p. 298). For instance, a set of activities may be arrayed quite differently if one is asked to rank them from "most satisfying" to "least satisfying" versus from "most enjoyable" to "least enjoyable." Respondents may interpret the first scale to allow for intrinsic/extrinsic, concomitant/non-concomitant reinforcement corresponding to an activity while the second scale is likely to be interpreted as allowing only for intrinsic, concomitant reinforcement.

The point of this brief discussion is that alternative approaches to measuring activity preferences exist. Unfortunately, we know little about the comparative validity of measures obtained from any one of these approaches. Respondents in the ISR Time Use Study were asked to rate some 22 activities on a scale of zero (dislike a great deal) to ten (enjoy a great deal). To improve the comparability of these ratings across respondents, we standardized the scores by respondent, e.g., for each respondent, his or her ratings of the 22 activities were used to compute a mean and standard deviation which were then used to convert the 22 "raw scores" to standardized scores. Thus, negative scores reflect relatively less enjoyable activities, and vice versa, in the individual's subjective estimation.

Determining the Extent of Temporal Incongruency With Respect to an Activity

In the sections which follow, the reader should keep in mind that our data only permits us to infer the presence of temporal incongruency. That is, we do not know that a respondent would prefer to spend some amount of his time in the activity other than that which he is spending. Nevertheless, we think it reasonable to suspect that temporal incongruency exists if an individual is spending large amounts of time in an activity regarded as relatively unenjoyable, or a small amount of time in an activity regarded as relatively enjoyable.

Collapsing the time expenditure and enjoyment measures into three subsets of equal numbers and cross-tabulating the two trichotomies resulted in Table 1 for the activity of meal preparation. Approximately 58% of the sample falls in the off-diagonal cells of Table 1, tentatively suggesting the presence of temporal incongruency. Nearly 28% (above the main diagonal) appears to be spending less time than preferred, while 30% (below the main diagonal) appears to be spending more time than preferred. The latter proportions provide some insight into the nature of temporal incongruency which exists with respect to the activity.

Characterizing Temporally Incongruent Segments

Methodology

One approach to disaggregating a sample into segments in which the relationship between y and x differs is an infrequently used variant of the AID algorithm called AID

TABLE 1

Cross Classification of Time by Enjoyment for Meal Preparation

Minutes/Week	Enjoyment of Meal Preparation			
	Less Enjoyable	Neutral	More Enjoyable	
0-60	15.1%	10.5%	6.6%	32.2%
61-360	11.8%	11.4%	10.9%	34.1%
361 and up	6.5%	11.4%	15.8%	33.7%
	33.3%	33.3%	33.4%	100%

covariance analysis (Sonquist et al 1978; Rao et al 1976). Typically, AID is employed to sequentially partition a sample into mutually exclusive subgroups so as to maximize between group variation (or equivalently, minimize error sum of squares) with respect to the mean levels of some dependent variable y. In the covariance mode, AID uses mean levels of y and a covariate x to disaggregate the sample. The analysis begins with a simple regression of y on x--here, the time spent in the activity on the corresponding measure of enjoyment. The initial error sum of squares is that left unexplained by the regression. The algorithm then examines subgroup regressions of y on x within alternative dichotomous partitions of the original group--say, within groups defined by age, sex, income, or other plausible bases. The sample is split on that predictor which maximally reduces error variation and satisfies other user specified constraints, e.g., minimum group size. The partitioning is repeated for subgroups, stopping when the variance reduction and minimum groups size criteria can no longer be satisfied.

The results of the AID covariance analysis for meal preparation are presented in Figure 2. Again, the dependent variable is the time (number of minutes per week) spent in the activity while the covariate is the level of enjoyment corresponding to the activity. The predictors--bases upon which the sample may be partitioned--included both demographics and psychographics which may be expected to moderate the relationship between time and enjoyment for reasons discussed earlier.

Results

The AID covariance analysis provides some insight into the factors which may generate temporal incongruency with respect to meal preparation. Segment 6, for example, comprised of single, non-employed women, spends less time in cooking on average but enjoys the activity more than does the overall segment of women. Segment 11, on the other hand, spends a relatively large amount of time in an activity which they regard as relatively enjoyable.

Using the Results

We stress once again that our data permit us to only indirectly examine temporal incongruency. Though subject to this and other caveats noted throughout the paper, the findings from the above analyses will serve as the vehicle for suggesting the implications of temporal incongruency.

The extent and nature of temporal incongruency with respect to an activity such as meal preparation is a useful barometer of probable shifts in future demand. For example, the respondents who spend more time than preferred in preparing meals are likely to be responsive to microwave ovens, convenience foods, meals away from home, and other goods and services which allow them to reduce this time expenditure. As this segment grows, the demand for time-saving products is likely to intensify. Conversely, the respondents who are spending less time cooking than desired may continue to demand time-intensive foods and modes of preparation and ignore time saving goods and services.

Uncovering the reasons for this temporal incongruency is also useful. From Figure 2, it appears as though women who are employed full-time enjoy meal preparation nearly as much as all women on average and yet spend nearly 200 minutes per week less. Has the time devoted to employment reduced the time which these women have available to spend in the activity? Which activities, if any, provide the reinforcement which previously resulted from cooking? Can these considerations be incorporated into product development and promotional strategies?

Summary

The process by which consumers spend their time is not well understood. Neither is the interdependence between the allocation of time and the allocation of money (or, equivalently, the interdependence between the selection of activities and the selection of goods and services) fully appreciated. We hope that this discussion of the concept of temporal incongruency will stimulate research which improves our understanding of these important facets of consumer behavior.

References

Arndt, J. and Gronmo, Signumd (1977), "The Time Dimension of Shopping Behavior: Some Empirical Findings," Proceedings of the Association for Consumer Research Conference, 230-35.

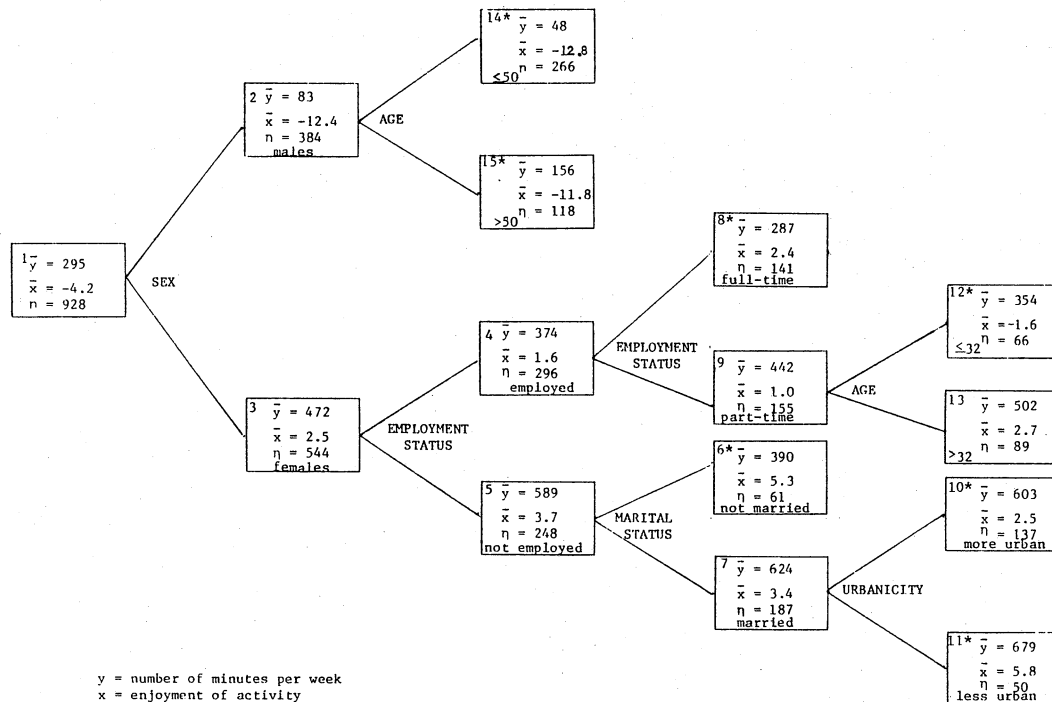
Becker, G.S. (1965), "A Theory of the Allocation of Time," Economic Journal 75 September: 493-517.

Chapin, F.S., Jr. (1974), Human Activity Patterns in the City. New York: Wiley.

Hawes, D.K. (1977), "Time Budgets and Consumer Leisure-Time Behavior," Proceedings of the Association for Consumer Research Conference, 221-229.

Hendrix, P.E. (1980), "Subjective Elements in the Examination of Time Expenditures," Proceedings of the Association for Consumer Research, 437-441.

FIGURE 2
Aid Covariance Analysis of
Time in Meal Preparation on Employment



Hobson, R. and Mann S.H. (1974), "A Social Indicator Based on Time Allocation," Social Indicators Research 1-4, 439-457.

Jones, M.B. and Pierce, J.M. (1977), "Time-Use Auditing: Approach to Validating Social Indicators," Social Indicators Research, 4, 289-315.

Juster, F. Thomas (1978), "The Investment of Time by U.S. Households: An Initial View." Paper presented to the American Psychological Association.

Linder, S.B. (1970), The Harried Leisure Class. New York: Columbia University Press.

Morgan, J.N., Sirogeldin, I.A., and Baerwaldt, N. (1966), Productive Americans. Ann Arbor: Institute for Social Research, University of Michigan.

Nicosia, F.M. and Mayer, R.N. (1976), "Toward a Sociology of Consumption," Journal of Consumer Research, 65-75.

Rao, V.R., McCann, J.M., and Craig, C.S. (1976), "Identifying Market Segments with AID3," Proceedings of the American Marketing Association Educators' Conference, 393-397.

Robinson, J.P. (forthcoming), "Methodological Studies," in F.T. Juster, ed. Studies in the Measurement of Time, Ann Arbor, Michigan: Institute for Social Research.

Sonquist, J.A., Baker, E.L., and Morgan, J.N. (1974), Search for Structure, Revised Edition, Ann Arbor, Michigan: Institute for Social Research, University of Michigan.

Walker, K. and Woods, M. (1976), Time Use: A Measure of Household Production of Family Goods and Services, Washington, D.C.: American Home Economics Association.

Wells, W.D. and LoSciuto, L.A. (1966), "Direct Observation of Purchasing Behavior," Journal of Marketing Research, 3 August, 227-233.

Rebecca H. Holman, Young and Rubicam, Inc.¹

Abstract

The "future," having no concrete basis in reality, is best conceived as part of an individual's thoughts or imagination. Many theoretical frameworks for studying consumer behavior incorporate some aspect of how consumers envision future events, yet little theoretical or empirical work has addressed this concept directly. This paper reviews several topics which implicitly include a future perception concept and suggests how these and other areas of study could benefit from models which incorporate the concept explicitly. The potential dimensionality of such a concept is reviewed in the context of a typology of futurizing styles.

Introduction

Research on time and consumer behavior has focused mainly on the time allocation process (Wilson and Holman 1980). This "economic" perspective on time is, however, only one of several worthy of attention by consumer behavior researchers (Settle 1980). Although the literature contains numerous examples of the "subjective" approaches (see Jacoby, Szybillo, and Berning 1976) relatively few works directly address how consumer decision making is affected by individuals' subjective experiences of time. (An exception is Settle, Alreck, and Glasheen 1978.) The topic of the subjective experience of time is an important one for consumer behavior researchers, however. Just as the concept of "perceived risk" has expanded the ability of researchers to explain consumer behavior toward risky products or services, so may study of the perception of time enable a more thorough explanation of time budgeting. Consumers who perceive units of time differently may plan different activities for otherwise identical time periods. Thus, perception of time may be a meaningful basis for market segmentation to the extent that consumer decision making is affected by time perception.

While it seems that consumer behavior researchers have neglected an important facet of human behavior, in fact that neglect is more apparent than real. As Hawes pointed out (1980) many of the global models of buyer behavior incorporate some references to consumers' experience or perception of time. Furthermore, one aspect of time perception, the conceptualization or imagination of the future, is embedded in several of the explanations of buyer behavior employed by past researchers. If such a concept exists, if it impacts upon consumer behavior, and if it has remained unmeasured, even "hidden" in previous studies, then there is a potential benefit to be gained by scrutinizing it and explicitly investigating its impact. The purposes of this paper are to examine the concept of the imagination of the future, to review some of the consumer behavior topics which implicitly incorporate the concept, and to suggest the dimensionality of the concept.

Imagining the Future

Before discussing the potential impact that conceptualizations of the future have on consumer behavior, it is first necessary to define the concept and to briefly trace its

history in the literature. That is the purpose of the first part of this paper.

A basic premise underlying the paper is that the future, since it can never be experienced directly, is wholly imaginary. This existential position is compatible with Lewin's ahistorical approach (see Kassarian 1973) to cognitive processes and leads to the conclusion that time itself has relevance only as the individual structures it into something meaningful. The future is no more or no less "real" to the individual than the past as both reside equally in the imagination, whether of remembered or of anticipated events (Lewin 1936).

The two major categories of events which enter into an individual's life space and impact upon that individual (within Lewin's theory) are things in the environment and things in the person's mind (Lewin 1951). As the latter may consist of current perceptions or of imaginary events (from the past or the future), it follows that the future holds a very important place in the life space of the individual.

Such an importance of the future is not found in other conceptualizations of human behavior. Freudian psychology, for example asserts that the past determines the present with imagination of the future (should it occur at all) being dominated by recollections of past events. Likewise, Marxian social theory sees the present and the future being determined rather irrevocably by past events (although it is the promise of a better future which serves as the motivating force in overturning capitalism for example).

Some researchers have attempted to deal with aspects of the imagination of the future. One group has looked at the length of one's future time perspective (how far into the future one normally thinks) and how that relates to a variety of other behaviors (e.g., Heimberg 1963; Bauer and Gillies 1972; Lamm, Schmidt, and Trommsdorff 1976). Another approach has been to study the physiopsychological bases of time perception (e.g., Michon 1975). Both of these approaches have measured time objectively with mechanical recording devices like timepieces or calendars, or in reference to such devices, which may not be how time is always perceived (especially when the individual is in a heightened emotional state). For example, there is evidence that the existence of a well-developed concept of the future is a key element in the psychological adjustment of terminally-ill patients, even when that future, when measured in objective terms, is not likely to be realizable (Susman 1980).

What is most important about one's imagination of the future is how that imagination or orientation to the future affects current behavior. (It is equally important to measure one's orientation to the past, but that is not covered here.) The concept of future imagination has appeared (albeit implicitly) in a number of past explanations given for human (and consumer) behavior. A review of several of these further illustrates the importance of the concept.

Aspects of Human Behavior Incorporating the Concept of Future Orientation

Cultural Influence

Culture incorporates time perception in the verb structure of the language of the culture. The Whorfian linguistic relativity hypothesis, that one is only able to think in

¹This paper was prepared while the author was a member of the Marketing Department, The Pennsylvania State University, University Park, PA.

the terms provided by one's language (Whorf 1956), would lead to the conclusion that if the verbs of one's language include past, present, and future tense, then one structures time into these categories. If one's language includes verb tenses which do not contain a category for events and behaviors which have not yet occurred (Arabic is one example), then one will not speak (or think) of the phenomenon of the future the same way as one who speaks and thinks in Standard American English (which does include a future tense). Furthermore, there is the possibility that one language may include verb forms which create new conceptualizations of time or span time dimensions (from the perspective of one language) making temporal translations from one language to the other virtually impossible. (See Whorf 1956, for a discussion of Hopi verb structures with these characteristics, pp. 143-5).

This is an important problem, and potential source of error in data, when one attempts research trans-culturally. For example, in the United States, approximately eighty percent of all blacks speak a language referred to as "Black English" (Baratz and Shuy 1969 quoted in Dillard 1972, p. 229), and "[i]n the system of its verbs, Black English reveals the greatest difference from white American dialects..." (Dillard 1972, p. 40). The problem is further compounded because Black English and Standard American English resemble one another superficially, sharing a great number of words. In fact the tendency in the past has been to characterize the language of most blacks as a "substandard" form of Standard American English, which it is not (Dillard 1972, p. 39-70). Thus research on blacks which does not incorporate the notions of time (and the future) inherent in the tenses of Black English risks interjecting an uncontrolled and perhaps unconscious source of bias.

The point is that by the very language used to communicate with subjects (or with members of a target audience, for example) the person forming messages imposes a concept of time (and futurity) inherent in the language used. Even if the respondent can translate from one language to another, there is a potential for misunderstanding unless the respondent is truly multi-lingual and thinks in the time categories used by the initiator of messages. (This potential for misunderstanding apparently exists even with languages which are "linguistically close" as with French and German.²)

Sociological Influence

An important component of the "sociology of consumption" is the set of cultural values held by consumer groups (Nicosia and Mayer 1976). The work by Rokeach (1973) has established the impact of values upon human behavior and their usefulness in explaining buyer behavior has also been documented (e.g., Vinson, Munson, and Nakanishi 1977; Vinson, Scott, and Lamont 1977). What has not been explicitly recognized is the implicit temporal quality of values as operationalized by Rokeach and those who use his work as a foundation.

A key element in Rokeach's concept of values is the subset referred to as "terminal values." As pointed out by Vinson, Munson, and Nakanishi, Rokeach thought of terminal values as descriptive of "the individual's desired end-state of existence..." (1977, p. 247). In order to respond to the Value Survey developed by Rokeach, a subject would need to evoke his/her ideal image of the future. Thus the concept operationally requires that a subject has formulated some concept of the future and that he/she describe it using the Value Survey.

Carman (1978) may have recognized the temporal dimension of values, although this is not discussed in his paper. Carman formulated a model relating values, life styles and consumption. One concept in his model was terminal values, another was time use activities. Carman, in effect, has stated the relationship between perception of the future and time use posited at the beginning of the current paper and has extended the relationships to life styles and other types of buyer behavior.

A subset of personal values influencing the perception of the future is the impact of religious beliefs. Religious groups in the United States have noticeably different degrees of belief in a spiritual afterlife and this apparently affects the time allocation processes of members (Azzi and Ehrenberg 1975). Thus the individual who believes that a reward in the afterlife is predicated upon laboring to "increase his talents to the greater glory of God" (Green 1972, p. 412) is likely to behave quite differently in the present than one who does not believe in an afterlife or who does not believe that efforts during one's life affect one's destiny after death. Green (1972) has done an excellent job along these lines of contrasting the temporal attitudes of Negro subcultures with those adhering to the ideals of the Protestant Ethic, a contrast which finds its basis in religious beliefs.

An even sharper contrast is found between the linear concepts of time found in most "western" religions and the non-linear concepts of some "eastern" religions, notably Buddhism and Hinduism. These differences may lead to different beliefs for example about the importance of one's death, about the significance of units of time for planning purposes, and more importantly for consumer behavior researchers may make possible an eastern acceptance of one's social status that is not feasible in western cultures (Olcott 1944; Pocock 1976).

The religious beliefs held by groups within society may predispose individuals against even speculating on the future. The argument is that as everything is under God's control, by trying to envision what will come, one is trying to "second-guess" God. Forecasting is prohibited as it presumes a knowledge that is God's alone. An example of this was reported by Eickelman (1977) who described the Muslim influence upon Moroccan tribesmen, but such an orientation toward planning and forecasting is undoubtedly present to some degree in all groups which believe in predestination.

If one were to use religious teachings (or beliefs) about the future as covariates in research, the differences in future orientation could possibly be measured and controlled in analysis. Otherwise such aspects of time and future orientation remain as random error in analyses of heterogeneous groups.

Psychological Influence

Attribution theory, the study of how individuals perceive the causes of behavior, contains a large temporal component. This relationship was pointed out by Kelley: "Implicit in the covariation principle is the important and little investigated problem of the exact temporal relations assumed to exist between a cause and its effect" (1973, p. 109). A typology of causality which defines the possible temporal relationships between cause and effect was developed by Evered (1976). Evered identified three primary types of causes: antecedent causality (behavior occurs because of some event that preceded it in time); concurrent causality (behavior occurs because of some event that occurs at the same time); and prospective causality (behavior occurs because of some event that has not yet occurred but should occur in the future).

When attributions deal with "expectations" the notion of the nature of the causality of the future is invoked.

² Anecdotal evidence supplied by Gerard M. Langneau, personal communication, October 1978.

Belief that the future is based upon events occurring in the past, a major premise of most forecasting techniques, involves a concept of antecedent causality. One expects a future that is similar, under similar conditions, to events that one has observed in the past.

Using a concept of concurrent causality would predispose an individual to expect that events occurring at a future point in time will be a function of whatever else occurs at that point in time. This is a Gestaltic notion of events, very similar to Lewin's explanation of behavior (1951). One does not expect any particular event to occur but could envision a number of alternative scenarios.

A belief that an individual is able to shape future events by the actions taken in the present is an explicit belief in prospective causality. One expects what one has planned to occur in the future. The future is, therefore, foreseeable and manipulable to some extent. (The concept of the "self-fulfilling prophesy" fits nicely within this notion of causality.)

These three types of causality produce different conceptualizations of the term "expectations," and yet the type of causality used by individuals has been empirically investigated only rarely. Mizerski and Green (1978) studied causal schemata, but ignored the temporal dimensions of that causality. However, Mizerski and Green did determine that one's causal schema was related to other cognitive processes, suggesting that a clarification of temporality in causality could produce even stronger relationships among variables.

There are also concepts of the future embedded in some of the personality traits studied in the past. Some of these are listed below along with a brief discussion of the aspect of future perception captured by the trait.

Dogmatism. The highly dogmatic individual is resistant to change preferring familiarity in events. This suggests that the high dogmatic desires a future that closely resembles the present and a present that closely resembles the past.

Need for Change. The individual with a high need for change gets bored by repetitive tasks and environments. Unlike the high dogmatic, an individual with a need for change seeks a future that is unlike what has yet been experienced.

Need for Order. The person with a high need for order can express it either through spatial ordering (keeping physical objects neat) or through temporal ordering (having things well-planned). When the ordering is expressed temporally, the individual hopes that the transitions into the future take place as planned and without discontinuities.

Locus of Control. A highly internal person believes that events occur due to his or her own intervention. This individual sees the future as both predictable and alterable, depending on what he or she decides to do. The highly external individual, by contrast, sees the future as subject to control by forces other than those of the individual.

Creativity. The person who is highly creative has an active imagination and is thus likely to be capable of producing many different scenarios of the future when asked to do so. This individual may also delight in envisioning futures that have not been conjectured before.

Deferred versus Immediate Gratification. A person who is able to defer gratifications implicitly believes in the existence of a future in which those pleasures may be enjoyed. A person who opts for immediate gratification may feel that the future is so uncertain that one should take life's pleasures as they are presented.

While it would be theoretically possible to partition the above traits into a time component and all other components, it is unlikely that such an action would materially affect the research which has attempted to relate personality to consumer behavior. As pointed out by Kassarian and Sheffett (1975), that body of research has been notably poor in isolating strong relationships and explaining much of the variance in the data. A more fruitful approach is the one taken by Settle, Alreck, and Glasheen (1978) who developed an instrument to measure time perception directly, treating it as an individual differences variable. What the above review illustrates is the pervasiveness of time (and especially future) dimensions in conceptualizations of human behavior.

The Dimensionality of Future Imagery

It is now possible to construct a theoretical statement of the dimensionality of perceptions of the future. From this statement one ought to be able to construct hypothetical types of ways of imagining the future. These dimensions are discussed next.

First, it is probably essential that any discussion of future dimensionality be limited to one cultural or linguistic group, unless a sufficient similarity in verb structures can be demonstrated. Unless the categories used to conceptualize time are essentially the same, one risks enormous errors in interpretation.³

From religious beliefs (and some personality traits) is the notion of the extent to which the future is predictable by man. Related to this, but slightly different is the degree to which an individual has control over the future, or whether that is under God's discretion alone.

Concepts of causality lead to envisioning the nature of causality itself. Next are a series of dimensions related to change: what is the rate of change between the present and the future and how does that differ from the rate of change in the past; is change valued and if not, is it disruptive (of individuals or of society in general). Finally is the concept of the number of images one has of the future (permitting this number to equal zero perhaps when the language does not possess a future tense).

An application of this theoretical dimensionality can illustrate its utility. The work of Evered (1973) is a useful vehicle, as Evered's purpose was to identify the characteristic way individuals had of thinking about the future. Using an exhaustive procedure (including content analyses of samples of writing, responses to personality tests, and performance measures), Evered, through a phenomenological analysis, derived three futurizing styles. He called these "producer," "performer," and "prospector," (although "performer" was later renamed "participant" to more clearly reflect the intra-organizational characteristics of that type of individual). Each of these three styles is presented in Table 1 and is characterized in terms of the dimensions discussed above.

As can be seen from Table 1, the producer sees a single-imagined future that is predictable because of its similarity to the present (and the past). The individual does not have control over the future since events set in motion in the past will cause those events which are to occur. Change, when it occurs is disruptive and not valued.

The participant, like the producer has one image of a predictable future, but differs from the producer in that the

³ An illustration of cultural/linguistic differences in the dimensionality of numeric measuring systems between Liberians and Americans can be found in Cole et al. (1971).

TABLE 1
A Typology of Orientations Toward the Future^a

Dimensions of Future Image	Futurizing Style		
	Producer	Participant	Prospector
Is the future predictable?	yes	yes	no
Does the individual have control over the future?	no	yes	no
What is the relationship to the future and the present?	identical or highly similar	linear but positive slope	discontinuous
Nature of causality	antecedent	prospective	concurrent
Rate of change in the future	slow or none	constant	accelerating
Is change valued?	no	yes	yes
Is change disruptive?	yes	no	no
Number of images of the future	one	one	many

^aTypes were described originally in Evered (1973).

future is different from the past. Change is valued if that change is orderly (and hence predictable). The individual has a dynamic impact upon the future, being instrumental in its design.

The prospector has traits in common with both the producer and the participant. Like the producer, the prospector does not see the individual as having control over the future; like the participant, the prospector sees a future that is different from the present. The prospector, however, is unwilling to predict the future as he/she envisions change as discontinuous with the present, and can therefore predict not one but many future scenarios. Change is not disruptive for the prospector; it is sought-out and valued.

Other hypothetical futurizing types could be generated, but these three have the advantage of being derived from observations of subjects (who were M.B.A. students at U.C.L.A.). Other work needs to identify whether these three types exhaust the universe of realized types, and to determine the distribution of these three in other populations. An instrument to operationalize the measurement of these three was developed (Evered, Reilly, Holman 1976), but that instrument has not been fully validated to date.

Theory Building Using the Concept of Future Imagery

This paper has suggested that unexplained sources of variances in buyer behavior could be decreased by incorporating the concept of consumers' images of the future as a variable in analyses. This means that the concept needs to be included explicitly in models of consumer behavior, where in the past it was subsumed in other concepts (some of which have been discussed here). The following list provides suggestions for topics which might benefit from such a practice and also raises some other issues concerning the concept itself.⁴

- To what extent are budgeting activities (whether of time, energy, or money) affected by an "inflation psychology," which is nothing more than a particular scenario of the future?

- How are investment patterns and durable goods purchases affected by one's futurizing style and by the nature of the future one envisions?
- Dissatisfaction often results when one is not able to realize the future one had envisioned. Do consumers with different ways of conceptualizing the future have different tolerance levels for this type of disappointment?
- What is the impact upon buyer-seller interactions when each has a different style of conceptualizing the future?
- How is risk perceived differently by those with different images of the future?
- Is one's style of conceptualizing the future affected by the aging process? Are age cohorts more likely to share future scenarios or is the future scenario independent of one's own age?
- Is futurizing style situationally-dependent or does it span situations?
- What are the cognitive processes underlying one's futurizing style?
- Aside from linguistic differences, is there a sense of time and the future that spans cultures? If so, what are its dimensions?
- Is the concept of the future likely to be of less importance in non-deliberative decision making than for planned purchases?

Obviously, a first step in investigating these and other questions is construction of an instrument to operationalize the concept of future imagery. The FAST instrument developed by Settle, Alreck, and Glasheen (1978) measures time orientation and has been extensively validated. Although it includes the future as one dimension, it was not designed for the work indicated here. Likewise, Heimberg's instrument (1963) measures the length of time one thinks into the future and is also not satisfactory. Constructing an instrument which does measure futurizing style is another area needing attention.

Conclusion

This paper has defined a concept of "future imagery" and identified how it has existed as an implicit concept in past research. Rarely has the concept been measured directly in the consumer behavior literature and thus any effect due to its influence remains as unexplained variance in analyses. The dimensionality of the concept was suggested and illustrated through presentation of a typology of futurizing styles developed previously. Suggestions for research questions linking the concept of future imagination to buyer behavior were provided.

References

- Azzi, Corry and Ehrenberg, Ronald (1975), "Household Allocation of Time and Church Attendance," *Journal of Political Economy*, 83, 27-56.
- Baratz, Joan C. and Shuy, R. W. (1969), *Teaching Black Children to Read*, Washington, DC: Center for Applied Linguistics
- Bauer, Rudolph and Gillies, John (1972), "Cognitive Style and Influence of Success and Failure on Future Time-Perceptive," *Perceptual and Motor Skills*, 34, 79-82.
- Carman, James M. (1978), "Values and Consumption Patterns: A Closed Loop," in *Advances in Consumer Research*, Vol. V., ed. H. Keith Hunt, Ann Arbor: Association for Consumer Research.
- Cole, Michael, Gay, John, Glick, Joseph A., and Sharp, Donald W. (1971), *The Cultural Context of Learning and Thinking*, New York: Basic Books, Inc.

⁴ Acknowledgement is given to an anonymous reviewer for his/her suggestions, some of which are included here.

- Dillard, J. L. (1972), Black English, New York: Vantage Books.
- Eickelman, Dale F. (1977), "Time in a Complex Society: A Moroccan Example," Ethnology, 16, 39-55.
- Evered, Roger (1973), "Conceptualizing the 'Future': Implications for Strategic Management in a Turbulent Environment," unpublished doctoral dissertation, Los Angeles: University of California.
- _____ (1976), "A Typology of Explicative Models," Technological Forecasting and Social Change, 9, 259-77.
- _____, Reilly, Michael, and Holman, Rebecca (1976), "An Empirical Investigation of Cognitive Styles: The Relationship Between Individual Futurizing Styles and Information-Processing Strategies in a Decision Situation," in Proceedings American Institute for Decision Sciences, ed. Howard Schneider, San Francisco: American Institute for Decision Sciences.
- Green, Helen Bagenstose (1972), "Temporal Attitudes in Four Negro Subcultures," in The Study of Time, eds. J. T. Fraser, F. C. Haber, and G. H. Müller, New York: Springer-Verlag.
- Hawes, Douglass K. (1980), "The Time Variable in Models of Consumer Behavior," in Advances in Consumer Research, Vol. VII, ed. Jerry C. Olson, Ann Arbor: Association for Consumer Research.
- Heimberg, Laura K. (1963), "The Measurement of Future Time Perspective," unpublished doctoral dissertation, Nashville, TN: Vanderbilt University.
- Jacoby, Jacob, Szybillo, George J., and Berning, Carol Kohn (1976), "Time and Consumer Behavior: An Interdisciplinary Overview," Journal of Consumer Research, 2, 320-39.
- Kassarjian, Harold H. (1973), "Consumer Behavior: A Field Theoretical Approach," in Perspectives in Consumer Behavior (rev. ed.), eds. Harold H. Kassarjian and Thomas S. Robertson, Glenview, IL: Scott, Foresman and Co.
- _____ and Sheffet, Mary Jane (1975), "Personality and Consumer Behavior: One More Time," in American Marketing Association Proceedings, Chicago: American Marketing Association.
- Kelley, Harold H. (1973), "The Process of Causal Attribution," American Psychologist, 28, 107-28.
- Lamm, Helmut, Schmidt, Rolf W., and Trommsdorff, Gisela (1976), "Sex and Social Class as Determinants of Future Orientation (Time Perspective) in Adolescents," Journal of Personality and Social Psychology, 34, 317-26.
- Lewin, Kurt (1936), Principles of Topological Psychology, New York: McGraw-Hill Book Co.
- _____ (1951), Field Theory in Social Science, ed. Dorwin Cartwright, New York: Harper and Row.
- Michon, J. A. (1975), "Time Experience and Memory Processes," in The Study of Time II, eds. J. T. Fraser and N. Lawrence, New York: Springer-Verlag.
- Mizerski, Richard and Green, Stephen (1978), "An Investigation Into the Causal Links Between Attribution Schema and Decision Making," in Advances in Consumer Research, Vol. V., ed. H. Keith Hunt, Ann Arbor: Association for Consumer Research.
- Nicosia, Francesco M. and Mayer, Robert N. (1976), "Toward a Sociology of Consumption," Journal of Consumer Research, 3, 65-75.
- Olcott, Mason (1944), "The Caste System of India," American Sociological Review, 9, 649-57.
- Pocock, D. F. (1976), "The Anthropology of Time-Reckoning," in Myths and Cosmos, ed. John Middleton, Austin: University of Texas Press.
- Rokeach, Milton (1973), The Nature of Human Values, New York: The Free Press.
- Settle, Robert B. (1980), "A Discussion of 'Time' Research," in Advances in Consumer Research, Vol. VII, ed. Jerry C. Olson, Ann Arbor: Association for Consumer Research.
- _____, Alreck, Pamela L., and Glasheen, John W. (1978), "Individual Time Orientation and Consumer Life Style," in Advances in Consumer Research: Vol. V, ed. H. Keith Hunt, Ann Arbor: Association for Consumer Research.
- Susman, Elizabeth J. (1980), "Concepts of Cancer and The Future: The Perceptions of Children, Adolescents, and Their Mothers," unpublished manuscript, University Park: The Pennsylvania State University, Department of Nursing.
- Vinson, Donald E., Munson, J. Michael, and Nakanishi, Masao (1977), "An Investigation of the Rokeach Value Survey for Consumer Research Applications," in Advances in Consumer Research, Vol. IV, ed. William D. Perreault, Jr., Atlanta: Association for Consumer Research.
- _____, Scott, Jerome E., and Lamont, Lawrence M. (1977), "The Role of Personal Values in Marketing and Consumer Behavior," Journal of Marketing, 41, 44-50.
- Whorf, Benjamin L. (1956), Language, Thought, and Reality, ed. John B. Carroll, Cambridge, MA: The M.I.T. Press.
- Wilson, R. Dale and Holman, Rebecca H. (1980), "Economic Theories of Time in Consumer Behavior: A Philosophy of Science Perspective on Their Evolution," in Theoretical Developments in Marketing, eds. Charles W. Lamb, Jr., and Patrick M. Dunne, Chicago: American Marketing Association.

TIME AND CONSUMER BEHAVIOR

Wesley J. Johnston, The Ohio State University

Abstract

The use and expenditure of time is inextricably linked to consumer behavior. As Jacoby et. al (1977) point out "Acquisition and consumption of both products and information regarding products are not cross-sectional events of short and unvarying duration." It has even been proposed that time may be the most important variable in consumer behavior (Nicosia and Mayer 1976). Our state of knowledge, however, finds us caught in an interesting anomaly. On the one side, the time dimension of consumer behavior is viewed as just beginning to emerge as a major variable of study, as three articles reviewed here indicate. On the other side, however, time has been implicitly and explicitly incorporated into consumer behavior theory and marketing strategies for quite some time. This article first reviews the history of the incorporation of time into consumer behavior theory and marketing strategies and then reviews the three articles concerned with "Perception of Time and Consumer Behavior".

Introduction

The treatment of time as a variable of importance in understanding consumer behavior has been accelerating. It is almost as if the fact that time exists and must affect consumer behavior has just been "discovered". Perhaps this is part of a larger awakening to the pervasiveness of time. As Voss and Blackwell (1979) note: "The importance of time in all aspects of American life has become a recurrent theme in the academic literature of the seventies. Interest has been cross-disciplinary."

The explicit treatment of time as a variable in consumer behavior has been sporadic. However, the implicit treatment of time has a much broader base. Many aspects of consumer behavior change over time: attitudes, perceptions, motives and values, lifestyles, and the political, cultural, technological environment. Most of the marketing literature has been directed at understanding what causes these changes and not the amount of time necessary for such changes to take place or the rate at which such changes occur. In addition to time's implicit involvement in consumer behavior, many consumer behavior concepts are explicitly dependent upon the use of a time dimension. Brand loyalty, for instance, is often operationalized as purchase patterns over time. The product life cycle examines distinct stages in the sales of a product over time. Concepts in innovation diffusion are also heavily dependent upon time. In the consumer adoption process, the consumer is seen as going through a series of stages of acceptance in adapting a new product and adopter categorization is made on the basis of relative time of adoption (i.e., innovators vs. laggards). Advertising wearout needs time to occur. Convenience shopping and specialty goods are classified according to the time and effort expended in their acquisition.

Most models of consumer behavior, especially those of an information processing type, consider a time element. Kotler (1972) in a "logical" model of buyer behavior depicts the buyer as problem solver deciding on product, brand style ... time, price, and way to pay. The Howard and Sheth (1968) theory of buyer behavior focuses on the task of explaining brand-choice behavior over

time rather than purchase frequency or purchase quantity. Bettman's (1979) information processing theory of consumer choice not only depicts consumer behavior occurring over time but explicitly acknowledges the effects time pressure can have on information processing and choice. Linear learning and Markov models of brand choice also incorporate a time factor.

It can be seen that time is integrally involved in consumer behavior and has been both implicitly and explicitly integrated in many models and concepts. Thus, "Acquisition and consumption of both products and information regarding products ... are dynamic processes that occur over time and that may involve different spans of time from one occasion and one individual to another (Jacoby, et. al. 1977)." The understanding of how the time dimension affects consumer behavior can lead to improved strategies for marketers.

Marketing managers are basically involved in decisions of two types: market segmentation and marketing/mix formulation. An understanding of the time dimension as it affects consumer behavior can improve decisions of both types. Changes in consumer segments over time have been clearly established by research in marketing. These include changes in consumer demographics, benefits sought by segments and perceptual spaces for products.

Engel et. al. (1978) provide a discussion of changes in consumer demographics and how these will affect the markets for goods and services. The two most obvious are the dramatic age changes in the population (increases in the over 65 geriatric market and 25-34 baby boom primary home buyers market and decrease in the teenage market) and the divorce, separation tendency creating many single parent households. These demographic trends translate to many new opportunities for marketers as well as the need to change current offerings.

Calontone and Sawyer (1978) discovered that while market segments defined by relative importance of product attributes (benefit segments) remained relatively stable in size, individual households changed segments frequently. This research was conducted over a two year time period. They concluded by indicating the need for more longitudinal research to monitor the benefits sought by individual consumers to help explain why and how such changes occur.

Moinpour et. al. (1976) employed multidimensional scaling techniques in a longitudinal experiment as an approach to investigate attitude processes. The study demonstrated that MDS techniques were appropriate for monitoring dynamic changes in perception caused by persuasive communications over time. Differential impact of message characteristics were also noted. The fact that perception changes over time could have perhaps been taken for granted but discovering a way to measure those changes, was an important contribution to understanding consumer behavior and aiding marketing management.

As the understanding of the influence of time on consumer behavior improves, the ability to incorporate this dimension into the modeling of market response models will increase. The purpose of a market response model is to relate sales or market share of a firm to its advertising,

price, promotion, distribution and/or other decision variables. Several studies currently exist which lead the way to the incorporation of time into marketing mix strategies.

Mickwitz (1959) discovered that advertising elasticity was highest at the growth stage of a product due to the need to create increased product awareness and lowest during maturity. Elasticities increased slightly through the saturation and decline stages of the product life cycle. Mickwitz also believed time-varying elasticities would exist for price, service, product quality, and packaging.

Voss and Blackwell (1979) hypothesize that as the value of time changes and the desire for more intensive time experiences increases, new product opportunities will develop especially among consumers with a serious leisure time scarcity. Those seeking physical fitness through recreation may switch from golf, a heavy time-investment sport, to tennis or raquetball. Thus, product or service redesign may be needed as time value changes. Voss and Blackwell also indicate the need for convenience or "time-saving" products. Appliances, such as microwave ovens and dishwashers, and services, such as lawn care, may be experiencing rapid growths in sales as a result of consumers buying increased discretionary time.

Mahajan et. al. (1980) demonstrate the need for feedback approaches to develop self-adaptive market response models. They indicate that the notion of market response varying over time was not new and actually well grounded in marketing literature. Their approach provides time-varying coefficients of advertising retention using Lydia E. Pinkham data. They conclude by noting that the use of such popular approaches of parameter estimation as least squares, Box-Jenkins, and other econometric techniques assume stability of coefficients over time and this may lead to a market response model insensitive to the reality of market conditions.

Contributions of the Articles Concerned With Perception of Time and Consumer Behavior¹

This section deals with the three articles presented in the session on the perception of time and consumer behavior. The article by McCullough, MacLachlan and Moinpour, "Temporal Links Between Preference and Perceptions" provides the only treatment of time from a longitudinal perspective of the three papers examined here. As such, I feel it examines the "main" effects of time on consumer behavior. The study shows that INDSCAL (Carroll and Chang 1970) and LINMAP (Srinivasan and Shocker 1973) can be used to jointly analyze preferences and perceptions and, as the authors argue, the results appear to be stable over time. This perception on the part of the authors is questionable, however. Do correlations in the range of .6-.7 truly show stability over time? One could argue just the opposite, especially since the interval between retesting does not permit much in the way of relevant intervening experience to occur. This position could perhaps be strengthened by citing generally acceptable levels of reliability in test-retest results at a one week interval. Another issue here is that the authors seem to feel inclusion of the brand "Ipana" reduced the stability observed in their analysis. Would it not be possible to delete this brand and rerun the experiment? The authors may then have support for their statement that selection of stimuli can affect reliability. A more general criticism of this paper deals

¹This section contains some comments made by anonymous reviewers as well as my own.

with the somewhat mechanical approach to the problem and the reporting of the results. The inability to control for the respondents' prior familiarity with the brands is a problem. The explanation given for the choice of toothpaste as the product category is only that this was the same as used in a previous study. While this may be valid because of the attempt to extend the earlier study, why not indicate the reason for choosing toothpaste as reported in the original study? Two additional brands were added to this study and in my opinion one should have been deleted. The authors report the "Ipana" is not marketed in the area where the study was conducted. Another important aspect of the paper is the use of CMATCH. This is a conservative procedure since rotation of the two spaces to maximal congruence precedes computation of correlational and proximity measures. CMATCH also uses two measures of congruence for reasons which are appropriate in calculating relationships between utilities and attribute weights.

Overall, this paper does accomplish its objectives of linking preference and perception temporally and provides direction for continued research in that direction. The authors are to be commended for this pioneering work.

The article by Hendrix and Martin, "Temporal Incongruity in Consumer Behavior" provides an interesting concept that can perhaps explain some of the variance in consumer behavior. Temporal incongruity, unfortunately, seems to be an unavoidable fact of life. While temporal incongruity can be improved for some situations such as cooking by the use of microwave ovens, instant foods, or fast food restaurants for those who want to spend less time on this activity, and by gourmet cookbooks, classes on cooking at universities, and all original ingredients for those who want to spend more time cooking, some activities still seem to resist temporal improvement; education is one example. The data analysis presented here is exploratory and indicates that further work needs to be done in modeling the relationship between time spent on and enjoyment derived from an activity. One problem that exists here is the ability of some individuals to be multi-phasic. The ability to do two or more things at one time such as drive to work and listen to the radio, watch television and write letters or balance the check book, ride the bus to school and do one's homework. Some people are capable of accomplishing three, four or even more activities at the same time. Modeling time incongruity for this type of individual presents several problems:

- A time budget will overstate the amount of time of activity for a multi-phasic person.
- How does one measure time per activity when two or more activities are carried on at the same time?
- Perhaps time incongruity - performing an activity longer than one wants to - leads to multi-phasic behavior in order to reduce mental attention to the incongruent activity.

A problem seems to exist in this article with respect to interpretation of some of the data. On page 3 the authors note that 58% of the sample falls in cells not on the main diagonal of Table 1 and this suggests the presence of temporal incongruity. "Nearly 28% (above the main diagonal) appear to be spending less time than preferred, while 30% (below the main diagonal) appears to be spending more time than preferred." It would seem that what is really the case is that only those respondents who found enjoyment of meal preparation to be "neutral" are temporally congruent, while those who found meal preparation to be less or more enjoyable regardless of the minutes/week spent in the activity are temporally incongruent. On page 4 the authors state "respondents who are spending less time cooking than desired may continue to

demand time-intensive foods and modes of preparation and ignore time saving goods and services." Fully employed, single women fall into this category per the data analysis; however, do fully employed single women demand time-intensive food or are they a prime market for microwave ovens? Why this incongruity exists is not clear. Perhaps because there is a weak relationship between time spent on cooking and enjoyment. Enjoyment does not appear to be a good surrogate for time one would like to spend on cooking. Because of this weak relationship the results could be the same for analysis using standard AID techniques, rather than AID covariates. Another possibility would be the use of standard regression analysis using dummy variables for the 8 segments identified. While the AID covariate analysis searches for intercept dummy variables, it might be interesting to look for slope dummy variables. I agree with the authors' concluding comment "... this discussion of the concept of temporal incongruity will stimulate research ..." The authors are to be commended for the early development and examination of this concept.

The final article reviewed here by Holman, "The Imagination of the Future: A Hidden Concept in the Study of Consumer Decision Making" provides a valuable review of a concept that could perhaps play a role in gaining greater understanding of consumer behavior. As the author points out, research on time has focused mainly on the allocation process. A consumer's imagination of and orientation toward the future could have significant impact on buying behavior. While the author provides no empirical results toward this end, she does raise a number of interesting questions at the end of the article. Another problem, as the author indicates, is the lack of an instrument to operationalize the concept of future imagery. Perhaps if the author's literature review had been broadened to include industrial buying behavior and marketing management this problem would have been somewhat diminished. Long range forecasting and strategic planning techniques are methodologies used by organizations to operationalize the future imagery of the market place for their goods and services. Perhaps these or similar processes operate within individuals when they consider and plan for the future. Obviously, the process within individuals is not as overt or attempted to be rational but parallels may exist. The fact that differences exist within individuals concerning belief in conceptualization of the future presents interesting problems for consumer behavior researchers with respect to measurement of and prediction from behavioral intentions. Behavioral intentions must interact with future imagery. The author could have gone further in drawing out the possible ramifications of this interaction and what might be done to control for it. The author is to be commended for her effort in bringing together literature of a highly interdisciplinary nature and demonstrating its potential to impact upon the understanding of consumer behavior. I hope that the author will follow this probe into the imagination of the future by developing an instrument to operationalize the concept of future imagery and providing some empirical results.

Conclusion

The consumer behavior and marketing literature has included both implicit and explicit treatment of the time dimension in concepts, models, and managerial decision making studies. However, this treatment has been sporadic and has lacked coordination. Studying the effects of time on consumer behavior from a cross-sectional perspective such as perception of the future and feelings of temporal incongruity provide some increases in knowledge of consumer behavior but seem to avoid the main effect of time on consumer behavior. More longitudinal studies are required for understanding and better measurement of changes in consumer behavior over time.

References

- Bettman, J.R. (1979), An Information Processing Theory of Consumer Choice, Reading, Mass.: Addison-Wesley Publishing Company.
- Calantone, R.J. and Sawyer, A.G. (1980), "The Stability of Benefit Segments," Journal of Marketing Research, August, 395-404.
- Carroll, J.D. and Chang, J.J. (1970), "Analysis of Individual Differences in Multidimensional Scaling via N-way Generalization of 'Eckart-Young' Decomposition," Psychometrika, 35, 287-319.
- Engel, James, Blackwell, R.D. and Kollat, D.T. (1978), Consumer Behavior, 3rd edition, Hinsdale, Illinois: Dryden Press.
- Howard, John A. and Sheth, J.N. (1969), The Theory of Buyer Behavior, New York: John Wiley and Sons.
- Jacoby, J., Szybillo, G.J., and Berning, C.K. (1977), "Time and Consumer Behavior: An Interdisciplinary Overview," Selected Aspects of Consumer Behavior, National Science Foundation.
- Kotler, Philip (1972), Marketing Management: Analysis, Planning, and Control, Englewood Cliffs, N.J.: Prentice Hall, Inc.
- Mahajan, V., Bretschneider, S.I. and Bradford, J.W. (1980), "Feedback Approaches to Modeling Structural Shifts in Market Response," Journal of Marketing, Winter, 71-80.
- Mickwitz, G. (1959), Marketing and Competition, Helsingfors, Finland: Centraltryckeriet.
- Moinpour, R., McCullough, J.M., and MacLachlan, D. (1976), "Time Changes in Perception: A Longitudinal Application of Multidimensional Scaling," Journal of Marketing Research, August, 245-253.
- Nicosia, F.M. and Mayer, R.N. (1976), "Toward a Sociology of Consumption," Journal of Consumer Research, 65-76.
- Srivivasan, V. and Shocker, A.D. (1973), "Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgments," Psychometrika, 28, 473-493.
- Voss, J., Blackwell, R.D. (1979), "The Role of Time Resources in Consumer Behavior," Proceedings, Second Annual Marketing Theory Conference.

CAUSAL MODELING: A GENERAL METHOD FOR DEVELOPING
AND TESTING THEORIES IN CONSUMER RESEARCH

Richard P. Bagozzi, Massachusetts Institute of Technology

Introduction

The behavioral sciences are now experiencing a revolution of sorts that is dramatically influencing the way researchers are approaching problems. The revolution entails a paradigm shift away from the many fragmented methods for inquiry currently employed toward a more holistic framework termed herein causal modeling. The purpose of this article is to outline and illustrate the use of causal modeling in consumer research. To this end, the expectancy-value model of attitude will be examined.

Before presentation of the approach, however, it will prove useful to provide some background. Causal modeling is often regarded by the unfamiliar as merely another data analytic technique to be used primarily in survey research. In reality, it is a method of enormous scope and power, where 'method' is taken here in its broadest philosophy of science sense.

In particular, we may look at causal modeling from at least four perspectives. First, causal modeling is a general philosophical orientation. It provides a way to operationalize the commonly accepted model of scientific inquiry posed by philosophers of science which stipulates that any true theory must contain, at a minimum, theoretical concepts, empirical observations, correspondence rules that link theoretical concepts to observations, and a rationale or set of laws connecting theoretical concepts (Bagozzi 1979). Further, when used as a tool for causal inference, causal modeling proceeds from a set of philosophical principles which guide the specification, test, and interpretation of any theory. Typically, the researcher will use either a neo-positivist or realist conceptualization of causation when employing causal modeling (Bagozzi 1980a).

Despite its common sense meaning, causal modeling is not limited to studies of cause-and-effect. Indeed, from a second perspective, causal modeling can be employed strictly in a measurement sense to compute internal consistency and test-retest reliabilities. Significantly, it allows the researcher the opportunity to assess a whole family of traditional measures of reliability such as Cronbach Alpha, as well as more general alternatives. It does this with less information than is required by traditional procedures in some instances, and it provides a means to take into account errors in variables and systematic error such as methods variance or other external confounds.

Third, causal modeling can be a valuable methodology for the examination of construct validity. Not only can it be used to perform a traditional multitrait-multimethod matrix analysis, but it does this in a more rigorous and less ambiguous way. Moreover, causal modeling furnishes one with a versatile means to investigate other forms of validity such as criterion related, predictive, and nomological validities. An important point to note is that causal modeling offers the advantage over traditional methods that measurement error is taken into account explicitly.

Finally, causal modeling is a general method for testing hypotheses. It can be used expediently in true experiments, quasi-experiments, cross-sectional surveys, panel studies, time series analyses, and cohort investigations, to name

a few contexts. As in measurement and construct validation usages, it offers the opportunity to take into account measurement error as well as systematic biases. For example, the causal modeling analogue to the analysis of covariance in experimental settings gives a straight forward way to accurately correct for random error, adjust for differences between experimental and control groups, and generally model the influence of non-random factors. In this sense, it avoids the major limitation of the traditional analysis of covariance which, by failing to take into account errors in variables, can sometimes fail to detect a true causal effect or else erroneously identify an effect which does not in fact exist.

As important as causal modeling portends to be, it is not without faults and shortcomings. A statement of these and some of its benefits are presented at the end of this article. For further explication of causal modeling, the reader is referred to Bagozzi (1980a), Bentler (1980), and Jöreskog and Sörbom (1979). We turn now to a prototypic development of causal modeling as a method of research.

Causal Modeling in Theory
Development and Hypothesis Testing

To illustrate causal modeling, consider the expectancy-value model of attitude and image that one desires to examine its nature, measurement, antecedents, and implications. One way to do this is to scrutinize the following issues: (1) convergent validity, (2) reliability, (3) concurrent validity, (4) discriminant validity, (5) predictive validity, and (6) nomological validity.¹ A verbal definition and mathematical expression for each of these follows.

Convergent Validity

At an intuitive level, convergent validity can be conceived as the extent to which two or more attempts to measure the same concept through maximally different methods are in agreement. The traditional standard for convergence requires that the correlations between measures of the same concept be greater than zero, statistically significant, and relatively large. However, because these criteria do not specifically take into account measurement error and can be misleading, the more rigorous causal modeling method will be used herein.

The null hypothesis for the causal model of convergence for an unidimensional attitudinal construct can be expressed as

¹Two additional standards that should be examined are the theoretical meaningfulness and the observational meaningfulness of concepts in the theory. The former refers to certain logic criteria as to well-formedness, internal consistency, and the like; while the latter concerns the adequacy of correspondence rules relating theoretical concepts to observed measures. In the interest of brevity, these criteria will not be analyzed here, and the reader is referred to other treatments of these issues by the author (Bagozzi 1979, 1980a, 1980b).

$$\underline{y} = \underline{\Lambda} \underline{x} + \underline{z} \quad (1)$$

where \underline{y} is a vector of p observations of attitude (y_1, y_2, \dots, y_p); \underline{x} is a hypothesized attitudinal construct; $\underline{\Lambda}$ is a vector of factor loadings relating \underline{y} to \underline{x} ; and \underline{z} is a vector of unique scores (i.e., errors in variables). The attitudinal construct, \underline{x} , is taken as an expectancy-value model where each y_i consists of the product of an expectancy (belief)-times-value observation. To fully specify the null hypothesis for the unidimensional attitudinal construct, the variance-covariance matrix of observations, $\underline{\Sigma}$, must be written as

$$\underline{\Sigma} = \underline{\Lambda} \underline{\Lambda}' + \underline{\Psi} \quad (2)$$

where $\underline{\Psi}$ is a diagonal matrix of error variances for attitudinal measures. In words, equations (1) and (2) hypothesize that all of the variation in responses to the attitudinal measures can be accounted for by one underlying expectancy-value construct, except for random error.

The null hypothesis for the causal model of convergence for a multidimensional attitudinal construct can be expressed as

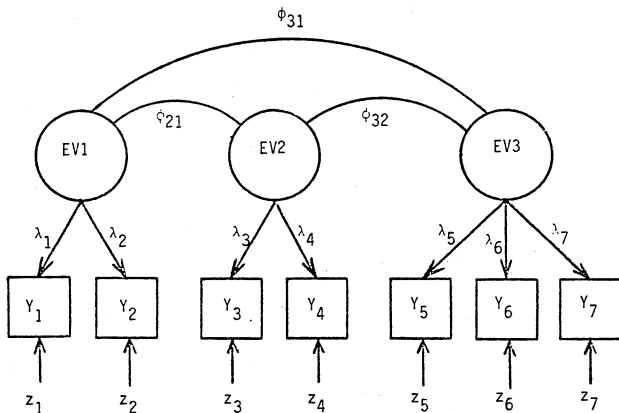
$$\underline{y} = \underline{\Lambda} \underline{x} + \underline{z} \quad (3)$$

$$\underline{\Sigma} = \underline{\Lambda} \underline{\Phi} \underline{\Lambda}' + \underline{\Psi} \quad (4)$$

where \underline{x} is now a $k < p$ vector of hypothesized attitudinal dimensions, $\underline{\Phi}$ is the intercorrelation matrix of attitudinal dimensions, and the remaining symbols are as defined earlier. As a point of interpretation, equations (3) and (4) hypothesize that all of the variation in attitudinal responses can be accounted for by k oblique dimensions, except for random error. This will occur when the responses of subjects (a) achieve a high degree of convergence among measures within attitudinal dimensions (i.e., when the within dimension measures are both highly intercorrelated and uniform in their pattern of values) and (b) exhibit uniform and significantly lower correlations among measures across dimensions. Figure 1 shows a causal diagram which is, in fact, the model achieving construct validity in the present study.

FIGURE 1

Causal Diagram for Convergent Validity



Reliability

Two internal consistency measures of reliability can be effectively applied to observations. The reliability of individual items, p_i , can be computed as

$$p_i = \frac{\lambda_i^2 \text{var } x_j}{\lambda_i^2 \text{var } x_j + \psi_i} \quad (5)$$

where λ_i is the factor loading relating the i th measure of attitude (i.e., the i th belief times evaluation product) to its respective attitudinal dimension, x_j ; and the remaining symbols are as defined earlier. Similarly, the reliability of a composite, p_c , of r observations of x_j can be calculated as

$$p_c = \frac{\left(\sum_{i=1}^r \lambda_i \right)^2 \text{var } x_j}{\left(\sum_{i=1}^r \lambda_i \right)^2 \text{var } x_j + \sum_{i=1}^r \psi_i} \quad (6)$$

As with Cronbach Alpha, p_i and p_c should only be applied to unidimensional attitude scales (or to homogeneous subdimensions) because to do otherwise would capitalize on the shared variance across dimensions and yield generally inflated values. Hence, it is recommended that one examine convergent validity first before computing reliabilities. Finally, it should be noted that equation (6) is similar to the standard Cronbach Alpha formula except that the latter assumes, a priori, that each observation of x_j contributes equally (i.e., the λ_i 's are set equal to unity). Thus, equation (6) is more general and less restrictive than Cronbach Alpha.

Concurrent Validity

The degree to which a measure of a concept correlates with a measure of a similar concept when both should naturally covary is known as concurrent validity (which itself is a special case of criterion-related validity). In the present research, concurrent validity will be examined as the degree of association between the attitude toward the act (measured with five semantic differential items) and the expectancy-value model of attitude. Figure 2 illustrates a causal diagram for concurrent validity when attitude toward the act (Aact) exists as an unidimensional construct and the expectancy-value model is multidimensional.

The null hypothesis for the concurrent validity model of Figure 2 is

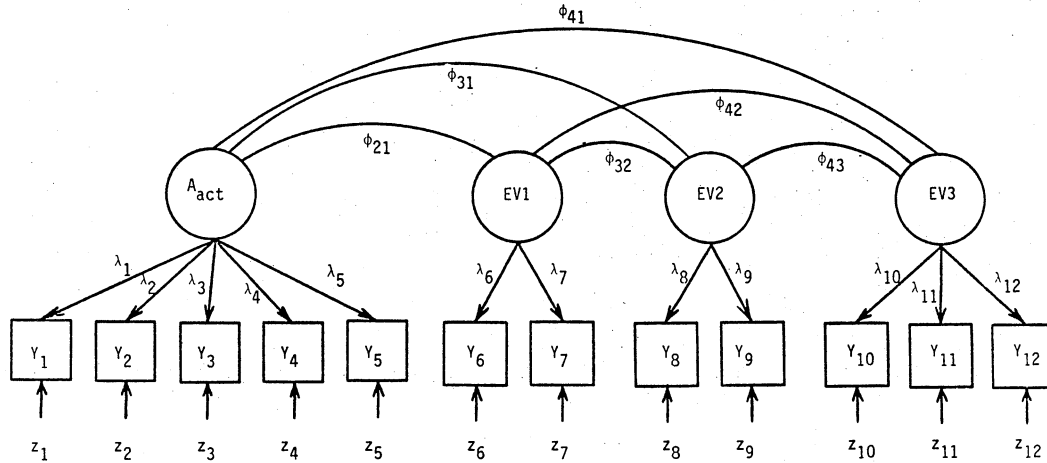
$$\underline{y} = \underline{\Lambda} \underline{x} + \underline{z} \quad (7)$$

$$\underline{\Sigma} = \underline{\Lambda} \underline{\Phi} \underline{\Lambda}' + \underline{\Psi} \quad (8)$$

where $\underline{y} = (y_1, y_2, \dots, y_{12})'$, $\underline{x} = (\text{Aact}, \text{EV1}, \text{EV2}, \text{EV2})'$, $\underline{z} = (z_1, z_2, \dots, z_{12})'$,

FIGURE 2

Causal Diagram For Concurrent and Discriminant Validity



$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \lambda_2 & 0 & 0 & 0 \\ \lambda_3 & 0 & 0 & 0 \\ \lambda_4 & 0 & 0 & 0 \\ \lambda_5 & 0 & 0 & 0 \\ 0 & \lambda_6 & 0 & 0 \\ 0 & \lambda_7 & 0 & 0 \\ 0 & 0 & \lambda_8 & 0 \\ 0 & 0 & \lambda_9 & 0 \\ 0 & 0 & 0 & \lambda_{10} \\ 0 & 0 & 0 & \lambda_{11} \\ 0 & 0 & 0 & \lambda_{12} \end{bmatrix}, \text{ and } \Phi = \begin{bmatrix} 1.000 & & & \\ \phi_{21} & 1.000 & & \\ \phi_{31} & \phi_{32} & 1.000 & \\ \phi_{41} & \phi_{42} & \phi_{43} & 1.000 \end{bmatrix}$$

This is a stringent test of concurrent validity because the hypothesis will be sustained only when (a) the within dimension correlations of measures are high, statistically significant, and uniform and (b) the across dimension correlations of measures are uniform and significantly lower than the within dimension values. Given this state of affairs, concurrent validity can be assessed by examination of ϕ_{21} , ϕ_{31} , and ϕ_{41} (i.e., the correlations between the attitude toward the act and the expectancy-value dimensions). These cross dimension correlations should be relatively high and statistically significant, yet should be lower in magnitude than ϕ_{32} , ϕ_{42} , and ϕ_{43} which, in turn, represent the intercorrelations among the three expectancy-value dimensions.

Discriminant Validity

Discriminant Validity refers to the degree to which a concept differs from other concepts. Given the establishment of convergent validity, discriminant validity can be examined through an inspection of ϕ_{21} , ϕ_{31} , and ϕ_{41} (see Figure 2). The comparison of the goodness-of-fit tests for the model of Figure 2 where ϕ_{21} , ϕ_{31} , and ϕ_{41} are left unconstrained to the same model wherein ϕ_{21} , ϕ_{31} , and

ϕ_{41} are constrained to equal unity will provide an explicit test of discriminant validity. The successful achievement of convergence and discrimination as outlined heretofore would indicate that one's construct (attitude) is valid in the sense of achieving homogeneity and uniqueness.

Predictive Validity

If a measure of a concept is related empirically as an antecedent to, or consequent of, a measure of another concept, then it is said to achieve predictive validity. The relation should not be fortuitous but rather should have its basis in the mechanism or theory connecting the two concepts.

The predictive validity of the expectancy-value model will be examined in two senses. As a predicted variable, the expectancy-value model will be observed as a function of the extent of the performance of past behavior, where the behavior relates directly to the act encompassed by the content of attitude. As a predictor, the expectancy-value model will be investigated as it forecasts three behavioral intentions that are logically entailed by a favorable attitude.

Specifically, Figure 3 and 4 show the expectancy-value model as a predicted and predictor variable, respectively. In Figure 3, past behavior (PB) is proposed to be antecedent to three expectancy-value dimensions. The structural equations for the null hypothesis of this model are

$$\begin{bmatrix} EV1 \\ EV2 \\ EV3 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} [PB] + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix}$$

Figure 4 posits that the expectancy-value model (EV) predicts three behavioral intentions (BI1, BI2, BI3). The structural equations for the null hypothesis of this model are

$$\begin{bmatrix} BI1 \\ BI2 \\ BI3 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} [EV] + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \tag{10}$$

FIGURE 3

Causal Diagram For Predictive Validity:
Past Behavior Predicting Attitude
(correlations among ζ 's omitted for simplicity)

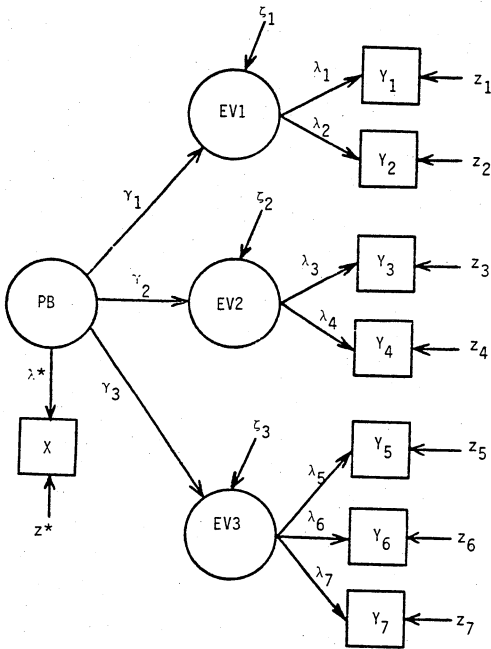
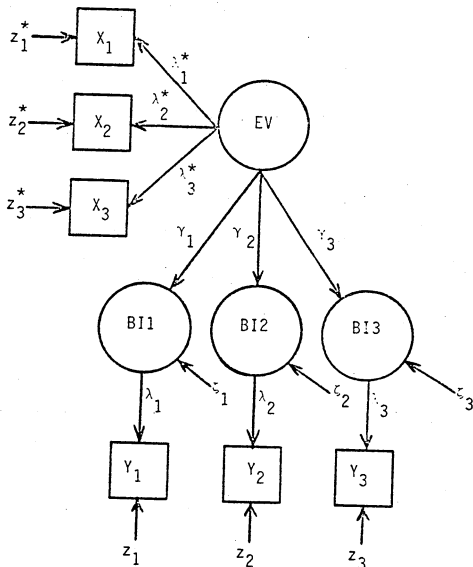


FIGURE 4

Causal Diagram For Predictive Validity:
Attitude Predicting Behavioral Intentions
(correlations among ζ 's omitted for simplicity)



A single construct is used to represent the expectancy-value model in order to avoid problems of multicollinearity which would exist had one employed the three dimensions -- EV1, EV2, EV3 -- as simultaneous predictors. The three measures of EV are, respectively:

$$x_1 = \sum_{i=1}^2 e_i v_i, \quad x_2 = \sum_{i=3}^4 e_i v_i, \quad \text{and} \quad x_3 = \sum_{i=5}^7 e_i v_i$$

where e_i and v_i are respectively, particular belief and evaluation items from a questionnaire.

Nomological Validity

Nomological validity indicates the degree to which predictions from a formal theoretical network containing a concept of interest are confirmed. From one viewpoint, the difference between predictive and nomological validities might be regarded as one of degree and not kind. Predictive validity entails the relationship of a concept to a single antecedent or consequent. Nomological validity in contrast, involves many antecedents and/or consequents in a complex theoretical system. The particular test of the nomological validity of the expectancy-value model conducted herein is based on attitude and learning theory (see Bagozzi 1980b). That is, it is hypothesized that a person's intentions to act in a particular way will be a function of (a) one's beliefs about the consequence of performing the behavior and the evaluation of those consequences (i.e., the expectancy-value attitude), (b) the extent of having performed the behavior in the past, (c) one's personal normative belief (PNB) that he or she should perform the behavior, and (d) one's social normative belief (SNB) that others whose opinions are valued feel that he or she should perform the behavior.

Figure 5 illustrates the null hypothesis for the nomological validity model. The structural equations for the key relations are

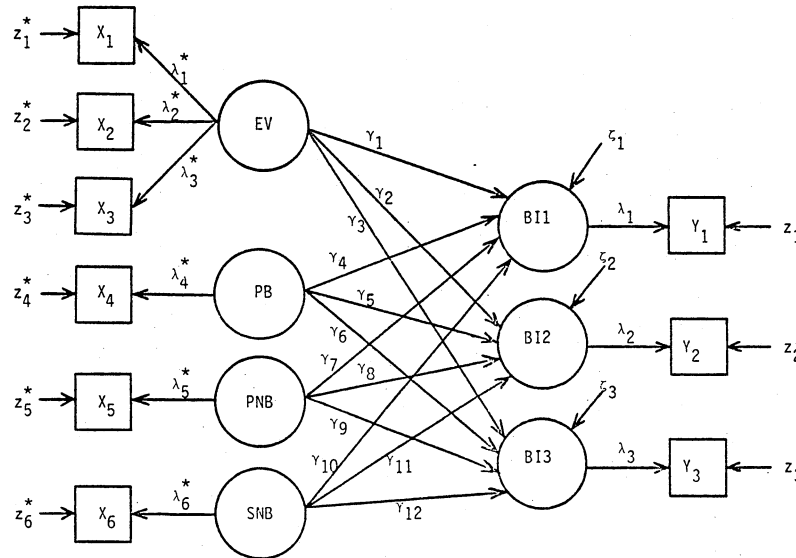
$$\begin{bmatrix} BI1 \\ BI2 \\ BI3 \end{bmatrix} = \begin{bmatrix} \gamma_1 & \gamma_4 & \gamma_7 & \gamma_{10} \\ \gamma_2 & \gamma_5 & \gamma_8 & \gamma_{11} \\ \gamma_3 & \gamma_6 & \gamma_9 & \gamma_{12} \end{bmatrix} \begin{bmatrix} EV \\ PB \\ PNB \\ SNB \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \quad (11)$$

Background of Study

The foregoing hypotheses were examined in the context of attitudes toward the act of donating blood. The results presented here constitute analyses performed on a sub-sample of 127 faculty, students, and staff which was drawn from part of a larger study. The full study was a quasi-experiment performed by the author (Bagozzi 1980b) and, for purposes of description, can be termed a post-test only design with nonequivalent multiple groups, multiple covariates, and measurement error modeled explicitly. Although a total of eight complex hypotheses were tested across and within groups on main effects and slope effects, only part of the within group analyses for one of three groups will be investigated herein. This is necessary for purposes of brevity, given space constraints. Hence, the present study is a survey analysis. The reader is urged to examine the quasi-experiment presented in Bagozzi (1980b), in order to gain a more complete picture of the scope and power of causal modeling. Also, the author presents a detailed description of the pre-tests, questionnaire, sample, methodology, and other related issues. Jöreskog and Sörbom's (1978) program, LISREL, was used for all analyses.

FIGURE 5

Causal Diagram For Nomological Validity
(correlations among ζ 's and among exogenous variables omitted for simplicity)



Results

On the hypothesis that people would form complex multi-dimensional expectancy-value attitudes rather than unidimensional ones (see Bagozzi 1980b for the rationale), convergent validity was examined. As predicted, the responses to the seven expectancy times value products did not converge to yield a single underlying attitudinal construct ($\chi(14) = 72.95, p \approx .00$). However, convergence was achieved for the multidimensional expectancy-value model, as hypothesized ($\chi^2(11) = 5.86, p \approx .88$). The second and third columns of Table 1 list the factor loadings and error variances, respectively, for the 7 expectancy times value products described in column one. Notice that each loading is relatively large in value and twice its respective standard error, and the error variances are low to moderate in magnitude.

Column four in Table 1 shows the individual item reliabilities where it can be seen that all values reach acceptable levels except the seventh which should be regarded as borderline. As presented in the final column of Table 1, the composite reliabilities are quite large in magnitude and thus indicate that the measures of the three dimensions of the expectancy-value model exhibit a high degree of internal consistency.

The goodness-of-fit test for the model testing concurrent validity shows that the hypothesis cannot be rejected (i.e., $\chi^2(48) = 56.96, p \approx .18$). The intercorrelation matrix of Aact, EV1, EV2, and EV3 is

	Aact	EV1	EV2	EV3
$\phi =$	1.000			
	.413	1.000		
	.527	.670	1.000	
	.497	.786	.785	1.000

Each of the correlations between Aact and the EVi is statistically significant at the .001 level or better and is lower than the correlations among the EVi.

Because the likelihood functions for the model represented by equations (7) and (8) could not be evaluated when ϕ_{21} , ϕ_{31} , and ϕ_{41} were constrained to unity, it was not possible to examine the difference in χ^2 -tests necessary for an explicit test of discriminant validity. However, inspection of ϕ indicates that the relevant entries are each far below 1.000, with the differences being greater than the values necessary to achieve significance at the .001 level or better. Thus, the evidence suggests that the expectancy-value model achieves uniqueness when compared to the semantic differential operationalization of attitude toward the act.²

² It should be noted that Aact, itself, attained convergent validity ($\chi^2(5) = 4.16, p \approx .53$) and demonstrated high individual ($p_i = .500$ to $.788$) and composite ($p_c = .916$) reliabilities. The expectancy-value model is largely a cognitive measure of attitude, while Aact is predominantly affective in content.

Table 2 summarizes the results for predictive validity. Looking first at column one, we can see that past behavior predicts all three behavioral intentions³, as hypothesized. All parameter values are in the proper direction and are at least twice their standard errors; and the overall goodness-of-fit test indicates a very good correspondence indeed ($\chi^2(16) = 7.67, p = .96$). The findings for the expectancy-value model as a predictor are shown in column two of Table 2. In general, the results are mixed. Although the goodness-of-fit test indicates a poor fit overall ($\chi^2(6) = 22.93, p = .00$), the parameter values are in the proper direction and are greater than twice their respective errors. Moreover, inspection of the residual matrix shows that the model captures most of the variation in responses.

Table 3 presents the results for the test of nomological validity. The overall goodness-of-fit test indicates a borderline fit ($\chi^2(18) = 28.22, p = .06$). Notice that the expectancy-value model predicts all three behavioral intentions significantly and in the predicted direction and that two of the three intentions are predicted by past behavior. Neither personal normative beliefs nor social normative beliefs function as significant predictors of behavioral intentions, however. Apparently, future volitions are only under the control of attitudes and past behavior.

But are they in reality? The results of the quasi-experiment suggest that attitudes are not validly related to intentions at all (Bagozzi 1980b). The findings in Table 3 are based on a sample of individuals who had given blood 20 minutes prior to filling-out the questionnaire or less. These individuals thus could have inferred their attitudes from their own prior behavior; and attribution and self-perception arguments could serve as rival explanations for the observed relations. However, when this sample was compared to a second sample of previous donors who had last given blood two months prior to filling-out the questionnaire or longer, only past behavior functioned as a valid determinant of intentions. It appears that attitudes do not supply any predictive power over and above learning theory arguments. Similarly, attitudes failed to predict intentions for those who had never given blood in the past. In sum, while considerable evidence exists for (a) the convergent, concurrent, discriminant, and predictive validities of the expectancy-value model of attitude, (b) reliability in measurement of attitudes, and (c) a valid relation from behavior to attitude, one must question the hypothesis that attitudes influence intentions.

As a point of comparison, it should be noted that it would not have been legitimate to use the commonly computed Fishbein model which is formed by summing the expectancy times value products to arrive at a single number for the expectancy-value attitude. The Fishbein model is justifiable only when the items form a unidimensional scale. Although few authors have ever demonstrated the reliability and validity of their measures (see Bagozzi 1980b for a review), the practice has been in consumer research to merely rely on the face validity of the expectancy-value

³Two of the behavioral intention items asked the probability one would (a) give blood "sometimes in the future" and (b) become a regular donor, respectively. These were measured on 11-point scales. The third behavioral intention item asked how frequently one might give blood if they had become a regular donor and was measured on a 7-point scale ranging from "5 times per year" to "would not become a regular donor."

model and to assume unidimensionality. For discussions and illustrations of the problems entailed by this practice, see Bagozzi (1980b,c).

Unresolved Issues

A number of shortcomings of the causal modeling method deserve mention. First, because the goodness-of-fit chi-square test is directly proportional to sample size, it is likely that virtually all models will be rejected in very large samples. There are at least three ways to mitigate this limitation. One alternative is to confine one's analysis to comparisons among nested models and to examine the differences in chi-square tests. This provides a means to assess the tenability of valid causal paths. A second procedure that might be meaningful in certain contexts is to randomly select subsamples from a larger population and then perform the causal analysis on these subsamples. One could then test the invariance of key parameters across subsamples, as well as examine the goodness-of-fit tests for each subsample. This approach might be useful when one desires to use separate fitting and validation samples. A third alternative is to scrutinize the residual matrix and ascertain whether the amount of information remaining is trivial for all practical purposes. Combinations of all three procedures might prove effective in some cases.

A second limitation of causal modeling is that it is strictly appropriate to use only large samples. Although what constitutes a large sample is open to question, this author has found that in many instances it is justifiable to employ causal modeling when the sample size minus the number of parameters to be estimated is greater than about 50. Hence, for most applications, the sample should be at least 60 to 70, or so. Unfortunately, very little is known about the small sample properties of the parameter estimates generated by LISREL, although some work is now being performed in this regard by statisticians. Hence, use of the procedures with small samples must be regarded with caution.

A third shortcoming of causal modeling is that it is designed to be used with measures that are at least interval scaled. Very little is known about the consequences of using measures that are ordinal. However, if one forms measures as the sum of independently distributed ordinal measures, say, then it is probably safe to assume that the summed variates satisfy the normality assumptions. Further, if the distribution of responses to variables is not too excessively skewed and if many scale steps are used (say seven or more), then the distributional assumptions of causal modeling may not be violated detrimentally for many applications. When in doubt, it is best to check the distributional properties of one's observations. Jöreskog and Sörbom (1978, p. 13) provide references for procedures to "robustify" one's data.

Still another limitation is the requirement that hypothesized relations be linear. This will not pose problems generally when transformations can be made to the data to reflect nonlinearities or when the underlying processes being modeled are linear to approach being linear. However, when such is not the case, causal modeling becomes less useful.

A final problem to note concerns the framework for hypotheses. Causal modeling sets-up the null hypothesis such that a nonsignificant chi-square value indicates a satisfactory fit. Ordinarily, one prefers that the hypothesis be constructed in an opposite way, but this is impossible to do so. In practice, the causal modeling approach usually gives valid results. Moreover, Bentler (1980) suggests the use of an index of the amount of information gained as a partial solution to the null hypothesis dilemma.

TABLE 1
PARAMETER ESTIMATES AND RELIABILITY MEASURES FOR CONVERGENT VALIDITY MODEL

	Factor loading, λ_i	Error Variance, z_i	Reliability, ρ	
			Individual item	Composite
EV1 I would get a sore arm, y_1 It would hurt, y_2	.822(.079) ^a .890(.077)	.325(.065) .208(.065)	.675 .792	.846
EV2 I would faint, y_3 I would be dizzy and nauseated, y_4	.819(.078) .932(.074)	.329(.061) .132(.061)	.670 .868	.869
EV3 My resistance to colds or infection would be lowered, y_5 I would be weak and have to curtail strenuous activity for a few days, y_6 I would lose time from work or study, y_7	.787(.078) .889(.073) .711(.081)	.380(.060) .209(.050) .494(.071)	.620 .791 .506	.840
Inter-component Correlation matrix, ϕ	1.000 .671 .787	1.000 1.000 .787		1.000

^aStandard errors in parentheses
n = 127
 $\chi^2(11) = 5.86, p = .88$

TABLE 2
PREDICTIVE VALIDITY FINDINGS

Parameter	Past Behavior Predicting Attitude	Attitude Predicting Behavioral Intentions
γ_1	.394(.102) ^a	.375(.093)
γ_2	.272(.097)	.371(.093)
γ_3	.338(.100)	.544(.088)
	$\chi^2(16) = 7.67$ p = .96	$\chi^2(6) = 23.93$ p = .00

^astandard errors in parentheses
n = 127

TABLE 3
NOMOLOGICAL VALIDITY FINDINGS

Predictor	BI1	BI2	BI3
EV	.344(.100) ^a	.223(.091)	.434(.089)
PB	.106(.095)	.390(.086)	.284(.083)
PNB	.023(.088)	.079(.081)	.056(.078)
SNB	-.094(.088)	-.031(.080)	.011(.077)

^astandard errors in parentheses
n = 127

Conclusion

It is important to stress that causal modeling represents only one link in the research process. Issues such as concept formation, questionnaire development, sampling, overall research design (e.g., random assignment and construction of experimental procedures), and the interpretation of findings are equally important areas for concern. Causal modeling should not be regarded as a replacement for any of these; indeed, each constitutes a necessary step in the process. Rather, causal modeling supplies the researcher with a powerful method for improving the conduct of research, given that care has been taken at all points in the process.

References

- Bagozzi, Richard P. (1979), "The Role of Measurement in Theory Construction and Hypothesis Testing: Toward a Holistic Model," in O.C. Ferrell, S.W. Brown, and C.W. Lamb, Jr., Eds., Conceptual and Theoretical Developments in Marketing, Chicago: American Marketing Association, 15-33.
- _____. (1980a), Causal Models in Marketing, New York: John Wiley & Sons.
- _____. (1980b), "On the Construct Validity of the Expectancy-Value Model of Attitude," unpublished working paper, Massachusetts Institute of Technology.
- _____. (1980c), "A Holistic Methodology for Modeling Consumer Response to Innovation," unpublished working paper, Massachusetts Institute of Technology.
- Bentler, Peter M. (1980), "Multivariate Analysis with Latent Variables: Causal Modeling," in Annual Review of Psychology, 31, 419-456.
- Jöreskog, Karl G., and Sörbom, Dag (1978), LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood, Chicago: National Education Resources.
- _____ and _____ (1979), Advances in Factor Analysis and Structural Equation Models, Cambridge, MA: Abt Books.

A COMPARISON OF CAUSAL PATH AND ECONOMETRIC MODELING APPROACHES

Leonard J. Parsons, Georgia Institute of Technology

Abstract

A tutorial is given on causal path and econometric modeling in consumer behavior. The conclusion is that conventional path-analytic techniques add little to the more substantial body of econometric theory and practice.

Introduction

Path analysis has been the primary methodology used in a spate of recent marketing papers (Lutz 1977, Moschis and Moore 1978, Bearden, Gustafson, and Mason 1978, Teas, Wacker, and Hughes 1979, Bearden, Teal, and Crocket 1980). These papers may have been influenced by an earlier piece by Christopher and Elliott (1970).

On the other hand, econometrics has had a long tradition in the study of buyer behavior. Applications range from an evaluation of the hypothesis of a hierarchy of effects (Palda 1966) to a test of Howard and Sheth's model of behavior (Farley and Ring 1970, 1972).

The purpose of this paper is to provide a brief tutorial on econometric and path analytic modeling. The goal is to assess what, if anything, path analysis adds to conventional econometric methodology. The paper begins with a review of selected aspects of econometrics, follows with an econometric reanalysis of a marketing path-analytic investigation, reviews the technique of path analysis, redoes the previous example from the perspective of path analysis, and ends with a comparison of the approaches. The discussion will be restricted to linear models and will omit the detail underlying the methodologies.

Econometric Modeling

Parsons and Schultz (1976) argued that the scope of econometric research as applied to marketing must be broadened to include theory development as well as measurement and testing processes. They were interested in causal modeling and set forth eight "standards" necessary for econometric studies to qualify as causal models. They suggested that many applications of regression analysis in marketing failed to meet their criteria.

The construction of causal models begins by developing a theory of marketing behavior and expressing this theory as a set of mathematical relations, a model. The existence of a theory which predicts a relationship among the variables within an equation causes the equation to be called structural. This theory provides the a priori information necessary to perceive the structure to be estimated, that is, to distinguish it from the other structures capable of generating the observed data. This is known as the identification problem.

A linear model can be expressed in matrix notation as

$$Y\Gamma + XB = E. \tag{1}$$

Only models containing as many equations as current endogenous variables will be considered. When this condition is met, the system is said to be complete. Suppose there are n observations on a system of L equations. Then the data on the current endogenous variables are placed in the matrix Y. The observations on the predetermined variables, either exogenous or lagged endogenous or both, are found in the matrix X.

Since the system is complete, the matrix Γ is assumed to be nonsingular. The coefficients of the predetermined variables in each equation are contained in the matrix B.

$$\Gamma = \begin{bmatrix} -1 & \gamma_{12} & \dots & \gamma_{1L} \\ \gamma_{21} & -1 & \dots & \gamma_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{L1} & \gamma_{L2} & \dots & -1 \end{bmatrix} \quad B = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1L} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \dots & \beta_{kL} \end{bmatrix} \tag{2}$$

The equations have been arranged so that the ith variable in the ith equation is that equation's dependent variable.

Finally, the disturbances are contained in the matrix E. The rows of the disturbance matrix are assumed to be stochastically independent and identically distributed with zero mean vector and an unknown but finite covariance matrix Σ .

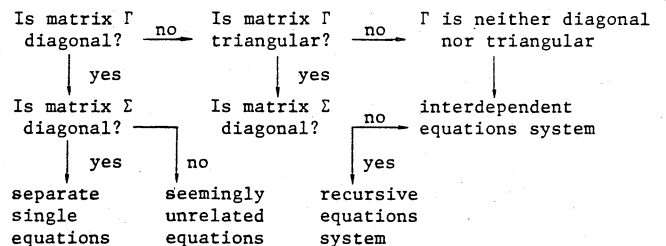
$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1L} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{L1} & \sigma_{L2} & \dots & \sigma_{LL} \end{bmatrix} \tag{3}$$

The matrix Σ , the contemporaneous covariance matrix of disturbances in different equations, is the same for all periods.

Our discussion will be facilitated by a classification scheme based upon the nature of the matrix of coefficients of the current endogenous variables Γ and the contemporaneous covariance matrix Σ . The first step is to examine the matrix Γ . If the matrix Γ is diagonal, the matrix Σ is examined. If the matrix Σ is diagonal, there are no relationships among the equations and each can be treated as a separate single equation model. If it is not diagonal, there is a seemingly unrelated equations system. If the matrix Γ is triangular and the matrix Σ diagonal, there is a recursive equations system. For the remaining categories involving the matrix Γ and the matrix Σ , there is an interdependent simultaneous equations system. Figure 1 presents a diagram of this classification process.

FIGURE 1

CLASSIFICATION OF SIMULTANEOUS EQUATION SYSTEMS



The structural equation system can be transformed into a logically equivalent system of equations in which each member equation contains only one endogenous variable:

$$Y = -XB\Gamma^{-1} + E\Gamma^{-1} \equiv X\Pi + U. \quad (4)$$

This new system is called the reduced form. Each reduced-form equation contains all the predetermined variables in an equation system. The contemporaneous covariance matrix of the reduced-form disturbances is

$$\Omega = (\Gamma')^{-1} \Sigma \Gamma^{-1} \quad (5)$$

The ability to express the structural parameters as explicit functions of the reduced-form parameters is unfortunately not automatic and, indeed is sometimes impossible. Determination of whether there is a one-to-one correspondence between the structural parameters and the reduced-form parameters is called the identification problem. Identification is logically prior to estimation. A discussion of how to test whether the relationships in system are identifiable as well as of how to apply such tests to Farley and Ring's operational version of the Howard-Sheth model of buyer behavior is given in Parsons and Schultz (1976).

If the identifiability condition is met, the parameters of the model can then be estimated. The alternative techniques include ordinary least squares, generalized least squares, seemingly unrelated equations estimation, indirect least squares, two-stage least squares, and three-stage least squares. Parsons and Schultz (1976, pp. 65-78) suggest which technique to use depending upon the structure of a model.

An Example

The heart of recent investigation of life satisfaction among elderly consumers (Bearden, Gustafson, and Mason 1978) is a two equation model. The variables can be defined in deviation form and consequently there are no intercepts shown in the equations. The first relationship says that satisfaction with level of living, X_4 , is a function of the level of alienation, X_3 , and a disturbance term, ϵ_4 or

$$X_4 = b_{43} X_3 + \epsilon_4. \quad (6)$$

The second relationship postulates that overall life satisfaction, X_6 , is a function of living level satisfaction, now an intervening variable, alienation, and health situation, X_1 , and another disturbance terms, ϵ_6 , or

$$X_6 = b_{61} X_1 + b_{63} X_3 + b_{64} X_4 + \epsilon_6. \quad (7)$$

The disturbances are assumed to be independent of one another, i.e., $\sigma_{46} = 0$. Moreover, alienation and health situation are explicitly believed to be intercorrelated. Alienation affects overall life satisfaction indirectly through living level satisfaction as well as directly. The data for this example are given in Table 1.

TABLE 1

VARIABLE CORRELATIONS, MEANS, AND STANDARD DEVIATIONS^a

Variable	X1	X3	X4	\bar{X}	s
X1: Health Situation	1.000			12.35	2.50
X3: Alienation	.374	1.000		72.08	12.16
X4: Living Level Satisfaction	.150	.259	1.000	2.59	.63
X6: Life Satisfaction	.361	.452	.346	16.06	3.37

^aSource: Bearden, Gustafson, and Mason (1977, p.387)

Using the matrix notation introduced earlier, the key matrices for this two equation system are

$$\Gamma = \begin{bmatrix} -1 & b_{64} \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & b_{61} \\ b_{43} & b_{63} \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_{44} & 0 \\ 0 & \sigma_{66} \end{bmatrix}.$$

Given our classification scheme, this simultaneous equation system is said to be recursive.

The parameters of the corresponding reduced-form equations are

$$\Pi = -B\Gamma^{-1} \quad (8)$$

or

$$\begin{bmatrix} \Pi_{41} & \Pi_{61} \\ \Pi_{43} & \Pi_{63} \end{bmatrix} = \begin{bmatrix} 0 & b_{61} \\ b_{43} & b_{63} \end{bmatrix} \begin{bmatrix} 1 & -b_{64} \\ 0 & 1 \end{bmatrix}^{-1} \\ = \begin{bmatrix} 0 & b_{61} \\ b_{43} & b_{63} \end{bmatrix} \begin{bmatrix} 1 & b_{64} \\ 0 & 1 \end{bmatrix} \\ = \begin{bmatrix} 0 & b_{61} \\ b_{43} & b_{64}b_{43} + b_{63} \end{bmatrix} \quad (9)$$

Because the model is recursive each equation can be estimated by ordinary least square regression. The least squares estimator for a given equation is

$$\hat{\beta} = (X'X)^{-1} X'y. \quad (10)$$

If we write

$$S_{XX} = (X'X)/n \quad \text{and} \quad s_{xy} = (X'y)/n$$

we can express the estimator in terms of covariances

$$\hat{\beta} = (S_{XX})^{-1} s_{xy}. \quad (11)$$

For the first equation the estimated parameter is simply

$$b_{43} = (s_3^2)^{-1}(s_{34}) = (147.866)^{-1} (1.984) = .0134.$$

The estimated parameters in the second equation are

$$\begin{bmatrix} b_{61} \\ b_{63} \\ b_{64} \end{bmatrix} = \begin{bmatrix} s_1^2 & s_{13} & s_{14} \\ s_{13} & s_3^2 & s_{34} \\ s_{14} & s_{34} & s_4^2 \end{bmatrix}^{-1} \begin{bmatrix} s_{16} \\ s_{36} \\ s_{46} \end{bmatrix} \\ = \begin{bmatrix} 3.250 & 4.252 & .236 \\ 4.252 & 147.866 & 1.984 \\ .236 & 1.984 & .397 \end{bmatrix}^{-1} \begin{bmatrix} 3.041 \\ 18.523 \\ .735 \end{bmatrix} = \begin{bmatrix} 1.249 \\ .069 \\ .281 \end{bmatrix}$$

Sometimes there is interest in the "relative contribution" of each term in the regression and the beta coefficients are calculated. The beta coefficient is

$$\text{beta}_{ij} = (s_j/s_i)b_{ij}$$

where x_i is the dependent variable and x_j is a predetermined variable. This is equivalent to running ordinary least squares regression on standardized variables. The result is that the estimator is expressed in terms of the sample correlation matrix

$$\underline{\beta} = R_{xx}^{-1} r_{xy}. \text{ Thus, } \beta_{43} = (1)^{-1} r_{34} = .259 \quad (12)$$

and for the second equation

$$\begin{bmatrix} \beta_{61} \\ \beta_{63} \\ \beta_{64} \end{bmatrix} = \begin{bmatrix} 1.0 & r_{13} & r_{14} \\ r_{13} & 1.0 & r_{34} \\ r_{14} & r_{34} & r_{46} \end{bmatrix}^{-1} \begin{bmatrix} r_{16} \\ r_{36} \\ r_{46} \end{bmatrix} \quad (13)$$

$$\begin{bmatrix} \beta_{61} \\ \beta_{63} \\ \beta_{64} \end{bmatrix} = \begin{bmatrix} 1.000 & .374 & .150 \\ .374 & 1.000 & .259 \\ .150 & .259 & 1.000 \end{bmatrix}^{-1} \begin{bmatrix} .361 \\ .452 \\ .346 \end{bmatrix}$$

$$= \begin{bmatrix} .209 \\ .313 \\ .234 \end{bmatrix}$$

A summary of these results as well as those for the reduced-form equations are given in Table 2.

TABLE 2
REGRESSION RESULTS

Variables	X1	X3	X4	R ²
<u>Structural</u>				
X4:				.067
b		.013		
beta		.259		
significance		.006		
X6:				.298
b	.281	.087	1.249	
beta	.209	.313	.234	
significance	.019	.001	.006	
<u>Reduced Form</u>				
X4:				.070
b	.016	.012		
beta	.062	.236		
significance	.538	.020		
X6:				.247
b	.301	.102		
beta	.223	.369		
significance	.015	.000		

Causal Path Analysis

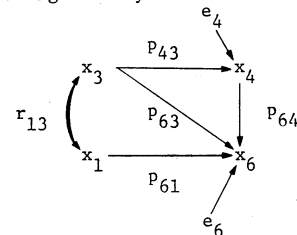
Causal path analysis was developed by a geneticist (Wright 1921, 1934) as a means of relating correlation coefficients between variables in a system to the functional relations among them. Path analysis has been used in psychology (Werts and Linn, 1970). However, its main application has been in sociology (Duncan 1966, Land 1969). The technique appears in multivariate analysis texts for the social sciences such as Van de Geer (1971, pp. 112-27).

Path analysis usually begins with the construction of a path diagram. Land (1969, pp. 6-7) states that path diagrams are drawn according to the following conventions. First, the postulated causal relations among variables of the system

are represented by unidirectional arrows extending from each variable dependent on it. Second, the postulated noncausal correlations among the exogenous variables in the system are symbolized by two-headed curvilinear arrows. Residual variables are also represented by unidirectional arrows going from the residual variable to the dependent variable. Finally, the quantities entered beside the arrows are the symbolic or numerical values of the path and correlation coefficients of the postulated causal and correlational relationships.

A path model refers to the set of structural equations representing the postulated causal and noncausal relationships among the variables under consideration. Traditionally, the variables are specified in standard unit form; that is, each variable is divided by its standard deviation as well as being centered. The symbolic form of the path coefficient is p_{ij} , where the first subscript i denotes the dependent variable and the second subscript j denotes the variable whose determining influence is under consideration.

The path diagram for the earlier example involving life satisfaction among elderly consumers is



The corresponding path model is

$$x_4 = p_{43}x_3 + e_4 \quad (14)$$

$$x_6 = p_{61}x_1 + p_{63}x_3 + p_{64}x_4 + e_6 \quad (15)$$

The next step in path analysis is to multiply through each equation in the model by a predetermined variable in the equation, take expected values, and express the result in terms of path coefficients and correlations. For instance, application of this procedure to the x_4 equation yields

$$E(x_1x_4) = p_{43} E(x_1x_3) + E(x_1e_4). \quad (16)$$

Earlier the exogenous variables were assumed to be independent of the disturbances so that $E(x_i e_j) = 0$. The key is the fact the variables are in standard form. This means that $E(x_i x_j) = \rho_{ij}$, the population correlation between x_i and x_j . Our last equation becomes

$$\rho_{14} = p_{43} \rho_{13} \quad (17)$$

Similarly, we find that

$$\rho_{34} = p_{43} \quad (18)$$

Repeating this operation for the x_6 equation and estimating p_{ij} by the sample correlation, r_{ij} , we obtain

$$r_{14} = p_{43} r_{13} \quad (19)$$

$$r_{34} = p_{43} \quad (20)$$

$$r_{16} = p_{61} + p_{63} r_{13} + p_{64} r_{14} \quad (21)$$

$$r_{36} = p_{61} r_{13} + p_{63} + p_{64} r_{34} \quad (22)$$

$$r_{46} = p_{61} r_{14} + p_{63} r_{34} + p_{64} \quad (23)$$

We now could solve for the p's in terms of the r's,

$$p_{43} = r_{43} \quad (24)$$

and

$$\begin{bmatrix} p_{61} \\ p_{63} \\ p_{64} \end{bmatrix} = \begin{bmatrix} 1.0 & r_{13} & r_{14} \\ r_{13} & 1.0 & r_{34} \\ r_{14} & r_{34} & 1.0 \end{bmatrix}^{-1} \begin{bmatrix} r_{16} \\ r_{36} \\ r_{46} \end{bmatrix} \quad (25)$$

But these are the same estimates that one obtains from ordinary least squares regression (see equation 13)!

Conversely, we could solve for the r's in terms of the p's. This would permit us, where relevant, to decompose the dependent variable. One could calculate the relative contributions of the components to the variation in the dependent variable and to determine how causes affecting the dependent variable are transmitted by means of the respective intervening variables.

More specifically, the total association between two variables is given by their zero-order correlation. The total effect of one variable on another is the part of their total association which are not due to noncausal components. These noncausal components, sometimes called spurious, are due to common causes, to correlation among their causes, or to unanalyzed correlation. The total effect is composed of both direct and indirect effects. Indirect effects are effects which are transmitted or mediated by variables specified as intervening between the cause and effect under study. The direct effect is that effect which is not transmitted by means of intervening variables (Alwin and Hauser 1975).

The zero-order correlations found from manipulation of x_6 -equation can be interpreted in the following manner.

$$r_{16} = p_{16} + p_{63}r_{13} + p_{64}p_{43}r_{13} \quad (26)$$

$$.361 > .209 + (.313)(.374) + (.234)(.259)(.374)$$

$$> .209 + .117 + .023$$

The total association is r_{16} ; the direct effect, p_{16} ; the direct correlation with another cause, $p_{63}r_{13}$; the indirect correlation with another cause, $p_{64}p_{43}r_{13}$. The remainder is unanalyzed correlation. The second relationship is

$$r_{36} = p_{61}r_{13} + p_{63} + p_{64}p_{43} \quad (27)$$

$$.452 = (.209)(.374) + .313 + (.234)(.259)$$

$$= .078 + .313 + .061$$

The total association is r_{36} ; the direct effect, p_{63} ; the indirect effect, $p_{64}p_{43}$; the correlation with another cause, $p_{61}r_{13}$. The third relationship is

$$r_{46} = p_{61}p_{43}r_{13} + p_{63}p_{43} + p_{64} \quad (28)$$

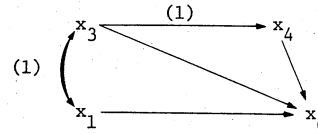
$$.346 > (.209)(.259)(.374) + (.313)(.259) + .234$$

$$> .020 + .081 + .234$$

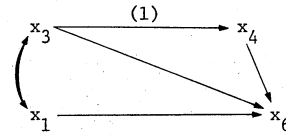
The total association is r_{46} ; the direct effect, p_{64} ; the correlation with a common cause, $p_{63}p_{43}$; the indirect correlation with another cause, $p_{61}p_{43}r_{13}$. Again, the remainder is unanalyzed correlation.

These algebraic relationships can be read directly from the path diagram according to the following rule. Read back from variable x_6 , then forward to variable x_1 , forming the product of all paths along the traverse; then sum these products for all possible traverses. In no case can one trace back having started forward. The bidirectional correlation is used in tracing either forward or back (Duncan 1966, p. 6.)

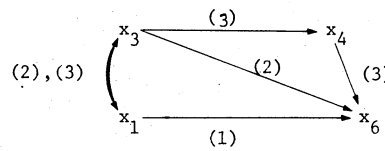
Thus, the correlation between x_3 and x_6 , r_{36} , simply equals (1) the indirect correlation with another cause (x_3), $p_{43}r_{13}$.



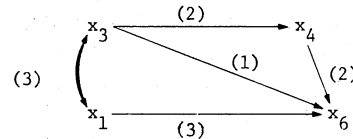
The correlation r_{34} is equal to (1) the direct effect, p_{43} .



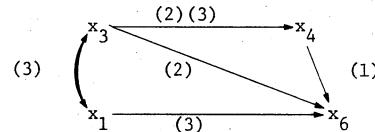
The correlation r_{16} is equal to (1) the direct effect, p_{61} ;



(2) the direct correlation with another cause $p_{63}r_{13}$; (3) the indirect correlation with another cause; $p_{64}p_{43}r_{13}$. The correlation r_{63} is equal to (1) the direct effect, p_{63} ;



(2) the indirect effect, $p_{64}p_{43}$; (3) the correlation with with another cause, $p_{61}r_{13}$. Finally, the correlation r_{64} is equal to (1) the direct effect, p_{64} ; (2) the



correlation with a common cause, $p_{63}p_{43}$; (3) the indirect correlation with another cause, $p_{61}p_{43}r_{13}$.

Reading of the path diagram must be done with care. A more prudent strategy would be to depend on algebraic manipulation.

Although path analysis provides some information about the noncausal associations among variables in a system, the focus is nonetheless on the causal influences whether direct or indirect. We have already seen that direct effects can be estimated from the structural equations using econometrics. Now we will observe that the total effects can be estimated from the reduced-form equations using econometrics (Alwin and Hauser 1975). The only indirect effects in our example occur in the path $x_3x_4x_6$ which yields the effect $p_{64}p_{43}$. The corresponding direct effect between x_3 and x_6 is p_{63} . The total effect is then $p_{63} + p_{64}p_{43}$. Equation 9 shows that this total effect is given by the appropriate reduced form equation.

Conclusions

There is considerable discussion in the literature about the conditions under which to use standardized measures such as path coefficients in contrast to unstandardized regression coefficients (Tukey 1954, Turner and Stevens 1959, Wright 1960a, Blalock 1967). One participant concluded that "Path coefficients are . . . appropriate if one wishes to measure the actual amount of impact that each variable has on the others in a given population. On the other hand, unstandardized regression coefficients seem more appropriate for describing . . . causal laws." (Blalock 1967, p. 675). In any event, both sets of coefficients may be obtained by econometric means.

Although attempts have been made to extend path analysis to interdependent systems (Wright 1960b), applications have focused on recursive systems. The danger is that the models have been formulated as recursive more because of tractability than because of theoretical soundness. The errant view that only recursive systems permit causal interpretation (Strotz and Wold 1960) was long ago debunked (Basman 1963). Moreover, even where a recursive model might be the correct causal priority, the adjustment mechanism (feedback) might be rapid relative to any practical inter-measurement period. See, for instance, Farley and Lehmann (1977). Consequently, the empirical model would still be interdependent.

Path analysis adds little that is not obtainable from econometric analysis, which is more highly developed and broader in its coverage. Indeed, the founder of path analysis (Wright 1960, p. 444) was reduced to offering the following apology: "Path analysis is an extension of the usual verbal interpretation of statistics, not the statistics themselves . . . The purpose of path analysis is to determine whether a proposed set of interpretations is consistent throughout." Even the sociologists seem to be moving away from path analysis to structural equation analysis (Duncan 1975, Heise 1975). The latter seems to be an attempt to present econometric concepts without the use of matrix algebra.

Path analysis does not seem to warrant a high priority for study by buyer behaviorists. One's time would be better spent trying to understand the strengths and weaknesses of the various approaches to making causal inferences for unobservable variables (Goldberger 1974, Griliches 1974, Wold 1975, Hui and Jagpal 1979, Jagpal and Hui 1980, Jöreskog and Sorbom 1979, Bagozzi 1980).

References

- Alwin, Duane F., and Hauser, Robert M. (1975), "The Decomposition of Effects in Path Analysis," American Sociological Review, 40, 37-47.
- Bagozzi, Richard P. (1980), Causal Models in Marketing, New York: John Wiley & Sons.
- Basman, Robert L. (1963), "The Causal Interpretation of Non-Triangular Systems of Economic Relations," Econometrica, 31, 439-48.
- Bearden, William O., Gustafson, A. William and Mason, J. Barry (1978), "A Path-Analytic Investigation of Life Satisfaction Among Elderly Consumers," in Proceedings, ed. William L. Wilkie, Ann Arbor: Association for Consumer Research, 386-91.
- Bearden, William O., Teal, Jesse E. and Crockett, Melissa (1980), "A Path Model of Consumer Complaint Behavior," in Proceedings, eds. Richard P. Bagozzi et al., Chicago: American Marketing Association, 101-4.
- Blalock, J. H. (1967), "Path Coefficients Versus Regression Coefficients," American Journal of Sociology, 72, 675-6.
- Christopher, M. G. and Elliott, C.K. (1970), "Causal Path Analysis in Market Research," Journal of the Market Research Society, 12, 112-24.
- Duncan, Otis Dudley (1966), "Path Analysis: Sociological Examples," American Journal of Sociology, 72, 1-16.
- Duncan, Otis Dudley (1975), Introduction to Structural Equation Models, New York: Academic Press.
- Farley, John U. and Lehmann, Donald. R. (1977), "An Overview of Empirical Applications of Buyer Behavior System Models," in Proceedings, ed. William D. Perreault, Jr., Atlanta: Association for Consumer Research, 337-41.
- Farley, John U. and Ring, L. Winston (1970), "An Empirical Test of the Howard-Sheth Model of Buyer Behavior," Journal of Marketing Research, 7, 427-38.
- Farley, John U. and Ring, Winston (1972), "On Land R and HAPPISIMM," Journal of Marketing Research, 9, 349-53.
- Goldberger, Arthur S. (1974), "Unobservable Variables in Econometrics," in Frontiers in Econometrics, ed. Paul Zarembka, New York: Academic Press, 193-213.
- Griliches, Zvi (1974), "Errors in Variables and Other Unobservables," Econometrica, 42, 971-98.
- Heise, David R. (1975), Causal Analysis, New York: John Wiley.
- Hui, Baldwin S. and Jagpal, Harsharanjeet S. (1979), "The Partial Least Squares Approach to Causal Marketing Models with Latent Variables," in Proceedings, eds. Neil Beckwith et. al., Chicago: American Marketing Association, 90-2.
- Jagpal, Harsharanjeet S. and Hui, Baldwin S. (1980), "Consumer Behavior Models with Unobservables: Measurement Reliability, Internal Consistency and Theory Validation," in Proceedings, eds. Richard P. Bagozzi et al., Chicago: American Marketing Association, 358-61.
- Jöreskog, Karl G. and Sörbom, Dag (1979), Advances in Factor Analysis and Structural Equation Models, Cambridge: ABE Books.
- Land, Kenneth C. (1969), "Principles of Path Analysis," in Sociological Methodology 1969, ed. Edgar F. Borgatta, San Francisco: Jossey-Bass, 3-37.
- Lutz, Richard J. (1977), "An Experimental Investigation of Causal Relations Among Cognitions, Affect, and Behavioral Intentions," Journal of Consumer Research, 3, 197-208.
- Moschis, George P. and Moore, Roy L. (1978), "Family Communication and Consumer Socialization," in Proceedings, ed. William L. Wilkie, Ann Arbor: Association for Consumer Research, 359-63.
- Palda, Kristian S. (1966), "The Hypothesis of a Hierarchy of Effects: A Partial Evaluation," Journal of Marketing Research, 3, 13-24.
- Parsons, Leonard J. and Schultz, Randall L. (1976), Marketing Models and Econometric Research. New York: North-Holland.
- Strotz, R. H. and Wold, H.O.A. (1960), "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis," Econometrica, 28, 417-27.

Teas, R. Kenneth, Wacker, John G. and Hughes, R. Eugene (1979), "A Path Analysis of Causes and Consequences of Salespeople's Perceptions of Role Clarity," Journal of Marketing Research, 16, 355-69.

Tukey, John W. (1954), "Causation, Regression, and Path Analysis," in Statistics and Mathematics in Biology, ed. Oscar Kempthorne et. al. Ames Iowa: Iowa State College Press, 35-66.

Turner, Malcolm E. and Stevens, Charles D. (1959), "The Regression Analysis of Causal Paths," Biometrics, 15, 236-58.

Van de Geer, John P. (1971), Introduction to Multivariate Analysis for the Social Sciences, San Francisco: W. H. Freeman and Co., 112-27.

Werts, Charles E. and Linn, Robert L. (1970), "Path Analysis: Psychological Examples," Psychological Bulletin, 74, 193-212.

Wold, Herman (1975), "Path Models with Latent Variables: The NIPALS Approach," in Quantitative Sociology, ed. H. M. Blalock et al., New York: Academic Press, 307-57.

Wright, Sewall (1921), "Correlation and Causation," Journal of Agricultural Research, 20, 557-85.

Wright, Sewall (1934), "The Method of Path Coefficients," Annals of Mathematical Statistics, 5, 161-215.

Wright, Sewall (1960a), "Path Coefficients and Path Regressions: Alternative or Complementary Concepts?" Biometrics, 16, 189-202.

Wright, Sewall (1960b), "The Treatment of Reciprocal Interaction, With or Without Lag, in Path Analysis," Biometrics, 16, 423-45.

INVESTIGATING CAUSAL SYSTEMS WITH QUALITATIVE VARIABLES:
GOODMAN'S WONDERFUL WORLD OF LOGITS

William R. Dillon, University of Massachusetts

Introduction

This paper is one of three directed at the important topic of causal modeling and its role in consumer research. The stimulating papers of my colleagues Professors Bagozzi and Parsons, appearing in this session, speak well to the issues that surface when causal systems are investigated in which the data have (continuous) quantitative components with at least interval scale properties; however, in many instances the available measurements are categorical in nature, having less than precise qualities, yet we may still wish to consider various causal hypotheses either of the temporal or nontemporal kind.

A point I would like to make early on is that in the case of ordinal data I favor the use of standard path analysis techniques. While I am fully aware that ordinal-based results can be altered by applying a suitable transformation, the Monte Carlo work suggests that violations of the interval assumption may not be very consequential, and this coupled with the fact that in most behavioral and consumer research settings the ordinal variables will generally have more than just two or three levels leads me to be more sanguine than some might be about the use of path analytic methods on ordinal scaled data.

This is not to say, however, that all caution should be discarded and that causal models of the Duncan (1966, 1975) or Wright (1960) variety be applied regardless of the distributional form of the data analyzed. Quite the contrary, the implicit or explicit assumptions of these models and methods that are violated when the variables are measured with only a small number of categories appear too severe to be ignored. Thus, if more than just heuristic or descriptive information is sought, the challenge is to find alternative approaches to modeling observable phenomena more compatible with the qualitative nature of the available responses. (The issue of unobservable variables and latent class models is also taken up in a later section.)

My remarks today are restricted to the case where the qualitative nature of the data is intrinsic to the attribute under study and not induced from some arbitrary collapsing of what would otherwise be a continuous variable. (I make this distinction since in the latter case the analysis may be best carried out by leaving the variables in their natural form and applying standard path analysis techniques.) There are a number of approaches to causal modeling with qualitative variables that are in the spirit of structural equation models; for example, Boudon (1965, 1967), Coleman (1964), Goodman (1972, 1974a), Lazarsfeld (1971) and Lazarsfeld and Henry (1968) are just a few of the social scientists who have made contributions to the area. Perhaps the closest analog to path analysis for qualitative variables is Goodman's loglinear/logit approach. Essentially, Goodman's system results in path diagrams, based on what usually is a series of logit specifications, and possesses the following attractive features:

- It provides the functional equivalent of structural equation models for qualitative data as well as the statistical tools necessary for testing.
- It is flexible in the sense of being able to handle temporal and nontemporal systems, and the case of unobservable variables.

- It can be easily implemented with currently available canned computer software packages suitable for analyzing multiway contingency tables.

The remainder of this paper represents an attempt, albeit an ambitious one, to satisfactorily do justice to the work of Goodman. While the primary intent of the piece is expository, I will attempt to address what I consider to be some interesting issues and problem areas. In preparation for the use of logit models as vehicles for investigating causal hypotheses, I begin the discussion with some basic ideas on fitting loglinear and logit models and, in addition, present the concept of spurious correlation in the context of cross-classified data.

Some Basic Ideas

This section provides the foundation for the subsequent material on causal modeling which immediately follows. My remarks will be clustered into three subsections focusing on loglinear modeling, logit specifications and the concept of spurious correlations as it relates to conditional independence-type arguments. The reader sufficiently educated in these areas is well advised to skip this material and turn to the next section which explicates Goodman's approach to causal modeling.

Loglinear Models

The most comprehensive work to date on the subject of fitting loglinear models to discrete multivariate data is the text by Bishop, Feinberg and Holland (1975). The interested reader might also wish to consult the recent text by Shelby Haberman (1978) for alternative treatments.

For illustrative purposes I will discuss the case of a four-dimensional table. However, the ideas to follow are easily generalized to lower- or higher-order tables. Let A, B, C, and D stand for the variables of interest with levels i, j, k, and l, respectively, and denote by f_{ijkl} and F_{ijkl} the respective observed and expected frequency in cell (ijkl). By expected frequency I mean the estimated count induced through fitting a specified model. If we further let n denote the total sample size, then the relationships hold that

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L f_{ijkl} = n$$

and

$$\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L F_{ijkl} = n. \tag{1}$$

The expected value in any cell of the $I \times J \times K \times L$ table formed by variables A, B, C, D can be parametrized in multiplicative form with parameters denoted by λ , or in additive form via logarithms, with parameters denoted by μ . My preference is to use the additive form which represents F_{ijkl} as

$$\ln F_{ijkl} = \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_{ij}^{AB} + \mu_{ik}^{AC} + \mu_{il}^{AD} + \mu_{jk}^{BC} + \mu_{jl}^{BD} + \mu_{kl}^{CD} + \mu_{ijk}^{ABC} + \mu_{ikl}^{ACD} + \mu_{jkl}^{BCD} + \mu_{ijkl}^{ABCD} \tag{2}$$

where

$$\begin{aligned} \sum_i \mu_i^A &= \sum_j \mu_j^B = \dots = \sum_k \mu_k^D = 0 \\ \sum_{ij} \mu_{ij}^{AB} &= \sum_{ik} \mu_{ik}^{AC} = \dots = \sum_{kl} \mu_{kl}^{CD} = 0 \\ \sum_{ijk} \mu_{ijk}^{ABC} &= \dots = \sum_{jkl} \mu_{jkl}^{BCD} = 0 \\ \sum_{ijkl} \mu_{ijkl}^{ABCD} &= 0. \end{aligned}$$

The right-hand side of (2) is interpretable, as in ANOVA models, in terms of individual and joint variable effects, and, consequently, has similar constraints. For completeness, it should be noted that multiplicative λ -parameters and the additive μ -terms are related by

$$\begin{aligned} \lambda &= \exp \mu; \lambda_i^A = \exp \mu_i^A; \dots; \lambda_k^D = \exp \mu_k^D; \\ \lambda_{ij}^{AB} &= \exp \mu_{ij}^{AB}; \dots; \lambda_{kl}^{CD} = \exp \mu_{kl}^{CD}; \\ \lambda_{ijk}^{ABC} &= \exp \mu_{ijk}^{ABC}; \dots; \lambda_{jkl}^{BCD} = \exp \mu_{jkl}^{BCD}; \\ \lambda_{ijkl}^{ABCD} &= \exp \mu_{ijkl}^{ABCD}, \end{aligned}$$

where "exp" denotes the exponential function. Finally, the representation given in (2) will be called a saturated model since there are as many parameters to estimate as cells in the table and, not surprisingly, the expected (or fitted) values are identical to the observed counts; that is, $F_{ijkl} = f_{ijkl}$ for all $(ijkl)$.

In almost all practical discussions the class of loglinear models is restricted to those models which satisfy the hierarchical property: If $\{\phi\}$ and $\{\phi'\}$ are any two sets of indices having the property $\{\phi\} \in \{\phi'\}$, then $\mu_{\{\phi\}} = 0$ implies $\mu_{\{\phi'\}} = 0$ and, further, if $\mu_{\{\phi'\}} \neq 0$ then all μ -term parameters containing like subscripts, that is, are a subset of $\{\phi\}$, are also not zero. Put simply, if μ_{ij}^{AB} is included in a model then the principle states that μ_i^A and μ_j^B must also be present, whereas if $\mu_{ij}^{AB} = 0$ then we must have $\mu_{ijk}^{ABC} = 0$. Given these constraints, hierarchical models can be completely defined by the minimal set of sufficient margin totals which represent the set of highest-order interactions. Adopting what I think now is standard notation, which encloses the requested model in brackets, the set of $[ABC]$, $[D]$ defines the model

$$\mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_{ij}^{AB} + \mu_{ik}^{AC} + \mu_{jk}^{BC} + \mu_{ijk}^{ABC}, \quad (3)$$

since the presence of the margin configuration $[ABC]$ implies the presence of all its lower-order relatives.

Once the set of sufficient configurations is specified, and assuming that some of their μ -term parameters are set to zero, that is, an unsaturated model is requested, maximum likelihood estimates are readily found either directly, estimates are expressible in closed form, or through some iterative proportional fitting algorithm. Most computer algorithms use proportioned fitting whether or not direct estimates exist, and rely on a result attributable to Birch (1963) which forces maximum likelihood estimates of margin totals corresponding to specified sufficient configurations to be equal to observed marginal sums. Hence, if we specify $[AB]$ and $[AC]$ as fitted margins, then the maximum likelihood estimates of the expected frequencies obey the relationship

$$\begin{aligned} F_{ij++} &= f_{ij++} \\ F_{i+k+} &= f_{i+k+}, \end{aligned} \quad (4)$$

where f_{ij++} (F_{ij++}) and f_{i+k+} (F_{i+k+}) are the observed (expected) marginal sums derived from summing over the appropriate "+" subscripts. The degree to which the quality

of the fitted model is any good can be assessed by computing -2 times the logarithm of the likelihood-ratio test statistic used for testing that the model fitted is correct versus the unrestricted alternative. Under the hypothesis that the model is correct

$$G^2 = -2 \sum_{ijkl} f_{ijkl} \ln \frac{f_{ijkl}}{F_{ijkl}} \quad (5)$$

is asymptotically χ^2 with degrees of freedom equal to # of states - # of independently fitted parameters.

Specific hypotheses can, under this framework, be easily tested by making use of the partitioning properties of the likelihood ratio statistic and the hierarchical principle of modeling. Consider, for example, the following listing of models which represent a nested hierarchy of loglinear models:

$$\begin{aligned} (M_4) & \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D \\ (M_3) & \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_{ij}^{AB} \\ (M_2) & \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_{ij}^{AB} + \mu_{jk}^{BC} \\ (M_1) & \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_{ij}^{AB} + \mu_{jk}^{BC} + \mu_{il}^{AD}. \end{aligned} \quad (6)$$

What characterizes the hierarchy is that each succeeding model differs from the model immediately preceding it by only one term. (The restriction to only one additional term is not a requirement.) Clearly, we would all agree that there is more than one set of nested hierarchical models that can be written down. For the above hierarchy denoted by $G^2(M_i)$ the likelihood ratio statistic for model M_i , then, though I omit the details, it can be shown that

$$G^2(M_4) = G^2(M_4|M_3) + G^2(M_3), \quad (7)$$

whereby $G(M_4|M_3)$ I mean the conditional likelihood ratio statistic for model M_4 given model M_3 . It follows that

$$G^2(M_4) \geq G^2(M_3) \geq G^2(M_2) \geq G^2(M_1). \quad (8)$$

If $G^2(M_4)$ and $G^2(M_3)$ are both asymptotically χ^2 random variables with degrees of freedom V_1 and V_2 respectively, then $G^2(M_4|M_3)$ is also asymptotically χ^2 with degrees of freedom $V_1 - V_2$. This property of partitioning and ordering for a nested hierarchy does not necessarily hold for the case where the Pearson statistic is used, and it is primarily for this reason that the G^2 statistic is preferred.

The conditional breakdown of G^2 into additive components is extremely useful in model building. For example, notice that the only difference between models M_4 and M_3 in the above hierarchy is the μ_{ij}^{AB} -term and, as your intuition might lead you to believe, it appears that $G^2(M_4|M_3)$ could be viewed as a test statistic for determining whether $\mu_{ij}^{AB} = 0$. (Actually, in a strict sense, we would also have to assume that variables C and D are unrelated to at least variable A or variable B in order to ensure that this test is equivalent to the test statistic used to determine whether $\mu_{ij}^{AB} = 0$ from the marginal table for variables A and B. I will have more to say about this nuance when developing the approach to causal modeling.) It is precisely in this manner that specific effects corresponding to relevant hypotheses can be tested.

Logit Specifications

In the previous discussions I made no mention of any distinctions between any of the variables. If I can assume the response variable has two levels, then it is reasonable to model the behavior of the log odds of one level of the response to the other on the basis of the explanatory variables. Upon finding a suitable log linear model, a

secondary analysis is performed and ultimately what we arrive at is a table of log odd effects from which we can better understand how changes in the combined levels of the explanatory variables affect the response measure.

To illustrate these ideas treat variable A in the $I \times J \times K \times L$ table as the response measure and the remaining variables B, C, and D as explanatory factors. In such settings we are typically interested in the relationship of the set of design or explanatory variables on the response measure. Assuming that the sample sizes for each combination of the explanatory variables are fixed by design, a little thought based on the previous discussion will show that we need to include the μ -term parameters relating the interactions between such variables in the model to insure that the estimated margin totals conform to fixed margins. Under this setup, we usually are not interested in investigating relationships between the explanatory variables but restrict ourselves to the nature of their effects on the response variables. For convenience, we can view the table as an array of rates such that

$$f_{.jkl} = f_{1jkl} / (f_{1jkl} + f_{2jkl}), \quad (9)$$

with corresponding logits given by

$$\omega_{.jkl}^{\bar{A}} = \ln(f_{1jkl} / f_{2jkl}) = \ln f_{1jkl} - \ln f_{2jkl}. \quad (10)$$

Corresponding to these observed odds, we can, in similar fashion, also define "expected odds" pertaining to variable A; that is,

$$\psi_{.jkl}^{\bar{A}} = \ln(F_{1jkl} / F_{2jkl}) = \ln F_{1jkl} - \ln F_{2jkl}. \quad (11)$$

With respect to model building and parameter estimation, we proceed essentially in the same way as was described in the previous section. The difference, as we stated, is that we must include the parameter μ_{jkl}^{BCD} in any model; in addition, suppose the best fitting parsimonious model is given by μ_{ij}^{AB} , μ_{il}^{AD} and, of course, μ_{jkl}^{BCD} . The log linear model is, therefore,

$$\ln F_{ijk\ell} = \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_\ell^D + \mu_{ij}^{AB} + \mu_{il}^{AD} + \mu_{jk}^{BC} + \mu_{j\ell}^{BD} + \mu_{k\ell}^{CD} + \mu_{jkl}^{BCD}. \quad (12)$$

The corresponding logit model is defined by

$$\begin{aligned} \psi_{.jkl}^{\bar{A}} &= \ln \frac{F_{1jkl}}{F_{2jkl}} = \\ &= [(\mu_1^A - \mu_2^A)] + [\mu_{1j}^{AB} - \mu_{2j}^{AB}] + [\mu_{1\ell}^{AD} - \mu_{2\ell}^{AD}] \\ &= 2[\mu_1^A + \mu_{1j}^{AB} + \mu_{1\ell}^{AD}] \\ &= \beta^{\bar{A}} + \beta_j^{\bar{AB}} + \beta_\ell^{\bar{AD}}, \end{aligned} \quad (13)$$

where the bar over superscript A simply is used to remind us that expected odds pertain to variable A, the dependent variable. Notice that the only interaction parameters that are included in the logit model are those terms which interact with the response variable in the corresponding loglinear specification.

We will see shortly that logit models form the basis of Goodman's approach to causal modeling. In particular, a series of logit specifications define a system of equations which are analogous to the kinds of systems of simultaneous linear equations encountered in the causal analysis of quantitative variables. Furthermore, the estimated β 's can be used to construct what are essentially path diagrams.

Spurious Correlation

Before jumping into Goodman's method for testing causal hypotheses, I would like to digress, for just a moment, in order to present a few remarks about the concept of spurious correlation when cross-classified data are analyzed. I have two reasons for doing this: first, the concept of spurious correlation is, in its own right, fundamental to the study of causal systems, and, second, a discussion of spurious correlation will, hopefully, lead to a better understanding of conditional independence type arguments used to introduce the subject of unobservable variables and latent class models.

Put simply, the association between two variables is called spurious if it vanishes in the presence of a third variable which impinges upon both. Typically, with quantitative variables, to test whether a correlation is spurious, interest centers upon the behavior of certain partial correlation coefficients. For example, a nonzero simple correlation between two variables X_2 and X_3 would be said to be spurious when upon entry of a third variable X_1 into the system, the partial correlation coefficient $r_{23.1}$ goes to zero.

In the case of qualitative variables, while we cannot couch our statements in terms of correlation coefficients, we can, nevertheless, present analogous arguments. Let us begin with the simple situation of just two variables A and B. If we now introduce a third variable C having, say, two levels, and if we find that in each of the two $I \times J$ tables formed by conditioning on the level of C, variables A and B become independent of each other, then we may say that the original relationship is spurious. In other words, we have found another variable which "explains" the relationship between variables A and B.

Similarly, if we return to our four-way table formed by variables A, B, C, D, we can say that variable D, for example, explains the relationship among the three remaining variables if the three variables A, B, C become mutually independent when the conditional relationships among the three variables in the three-way table at each level of variable D are examined. The hypothesis that variables A, B, and C are mutually independent when the level of variable C is held constant can be denoted as

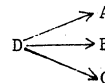
$$[A \times B \times C | D]$$

where x means independent of.

The hypothesis of conditional independence can be easily tested by assessing the goodness-of-fit of the model:

$$\ln F_{ijk\ell} = \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_\ell^D + \mu_{il}^{AD} + \mu_{j\ell}^{BD} + \mu_{k\ell}^{CD}. \quad (14)$$

Notice that no μ -term parameters pertaining to the two-way interactions between the pairs AB, AC, or BC are included and, therefore, the system could be described by the following simple diagram



where the presence of an arrow indicates a nonzero effect.

Causal Analysis: The World of Logits

I will, in this section, attempt to provide an exegesis of Goodman's approach to modeling causal systems composed solely of qualitative variables. My explication of these methods and models will, for the moment, be presented in the abstract with the following section devoted to substantive examples which highlight the issues raised in the ensuing discussion.

Goodman first introduced the modification of loglinear analysis which allowed causal hypotheses to be investigated in two 1973 papers. He argued, to varying degrees throughout both pieces, that the proposed methodology produces a framework for analysis analogous to the usual analysis of causal systems with quantitative variables. Indeed, Goodman's approach can be used to study both recursive as well as nonrecursive systems, though in the latter system the analogy to the quantitative variable case is by far the weakest, and generates path diagrams much in the spirit of those obtained with systems of simultaneous linear equations. There are, however, a number of subtleties which can introduce certain ambiguities into the analysis.

This section is largely expository, and is organized around nonrecursive and recursive systems. After each major subsection, I devote some time to commending on the nuances that are inherent in Goodman's methodology.

The Analysis of Nonrecursive Systems:
The Weakest Link

I will again make use of the four-way IxJxKxL multiway table assumed generated by variables A, B, C, and D. In addition, for simplicity, assume now that I=J=K=L=2, so that the table can be described as a 2⁴ contingency table. I begin with causal modeling in nonrecursive systems, though admittedly the situation provides the weakest analogy to the analysis of systems of simultaneous linear models with quantitative variables, because without any order conditions on the variables, the use of loglinear analysis to investigate causal hypotheses will naturally lead to systems having reciprocal causation and feedback loops.

For illustrative and discussion purposes, assume that the model described by the margin totals

[AB], [AC], [BC], [BD], [CD]

provides an adequate fit to the four-way table. Recall, these margin configurations define the model

$$\ln F_{ijkl} = \mu + \mu_i^A + \mu_j^B + \mu_k^C + \mu_l^D + \mu_{ij}^{AB} + \mu_{ik}^{AC} + \mu_{jk}^{BC} + \mu_{jl}^{BD} + \mu_{kl}^{CD} \quad (15)$$

so that we would generate nonzero estimates for five two-factor interactions and four one-factor main effects. Following the notation used in the discussion of logit specifications, we could single out each variable in turn and define the expected odds pertaining to each as

$$\psi_{ijk}^{\bar{D}} = \ln F_{ijk1} - \ln F_{ijk2} \quad (16)$$

$$\psi_{ij\cdot l}^{\bar{C}} = \ln F_{ij1l} - \ln F_{ij2l} \quad (17)$$

$$\psi_{i\cdot kl}^{\bar{B}} = \ln F_{i1kl} - \ln F_{i2kl} \quad (18)$$

$$\psi_{\cdot jkl}^{\bar{A}} = \ln F_{1jkl} - \ln F_{2jkl} \quad (19)$$

where again the bar over the superscript in the ψ 's signifies that the expected odds pertain to the superscripted variable, and that these expected odds are the dependent variable in each respective formula. In similar fashion we can express (16)-(19) in terms of the logit β -effects which yields the following system of equations.

$$\psi_{ijk}^{\bar{D}} = \beta^{\bar{D}} + \beta_j^{\bar{B}\bar{D}} + \beta_k^{\bar{C}\bar{D}} \quad (20)$$

$$\psi_{ij\cdot l}^{\bar{C}} = \beta^{\bar{C}} + \beta_i^{\bar{A}\bar{C}} + \beta_j^{\bar{B}\bar{C}} + \beta_l^{\bar{D}\bar{C}} \quad (21)$$

$$\psi_{i\cdot kl}^{\bar{B}} = \beta^{\bar{B}} + \beta_i^{\bar{A}\bar{B}} + \beta_k^{\bar{C}\bar{B}} + \beta_l^{\bar{D}\bar{B}} \quad (22)$$

$$\psi_{\cdot jkl}^{\bar{A}} = \beta^{\bar{A}} + \beta_j^{\bar{B}\bar{A}} + \beta_k^{\bar{C}\bar{A}} \quad (23)$$

where we note

$$\beta^{\bar{D}} = 2\mu_{j\cdot l}^{\bar{D}}, \beta_j^{\bar{B}\bar{D}} = 2\mu_{j\cdot l}^{\bar{B}\bar{D}}, \beta_k^{\bar{C}\bar{D}} = 2\mu_{k\cdot l}^{\bar{C}\bar{D}}, \dots \text{etc.} \quad (24)$$

and, because of the assumptions imposed,

$$\beta^{\bar{B}\bar{D}} = \beta_1^{\bar{B}\bar{D}} = -\beta_2^{\bar{B}\bar{D}}, \beta^{\bar{C}\bar{D}} = \beta_1^{\bar{C}\bar{D}} = -\beta_2^{\bar{C}\bar{D}}, \text{etc.} \quad (25)$$

The meaning of equations (20)-(23) should be clear; for example, equation (20) states that the expected odds pertaining to variable D are dependent upon the explanatory variables B and C, but not upon variable A, and, according to (10) and (11), the expected odds in favor of variable D is increased when the additive effects of variable B is introduced at level 1 and decreased when variable B is at level 2.

Equations (20)-(23) provide a system of logit equations which, in spirit, are analogous to the kinds of systems of simultaneous linear equations used in the causal analysis of quantitative variables. For those of us who might be having trouble seeing the analogy, let us introduce the dummy variables X₁, X₂, X₃, and X₄, where X₁ assumes the value 1 when variable A is at level 1, 0 otherwise, X₂ assumes the value 1 when variable B is at level 1, 0 otherwise, and so on for variables X₃ and X₄ corresponding to variables C and D. We can rewrite equations (20)-(23) as a linear function of the dummy variables; for example, in the case of equation (20) we have

$$\psi_{ijk}^{\bar{D}} = \beta^* + \beta_1^* X_1 + \beta_2^* X_2 + \beta_3^* X_3 \quad (26)$$

where

$$\beta^* = \beta^{\bar{D}} - \beta^{\bar{A}\bar{D}} - \beta^{\bar{B}\bar{D}} - \beta^{\bar{C}\bar{D}} \quad (27)$$

$$\beta_1^* = 2\beta^{\bar{A}\bar{D}}, \beta_2^* = 2\beta^{\bar{B}\bar{D}}, \beta_3^* = 2\beta^{\bar{C}\bar{D}}$$

(Note, in the above formulation, $\beta^{\bar{A}\bar{D}} = 0$.)

Finally, we can extend the analogy even further by expressing the observed odds $w_{ijk}^{\bar{D}}$ in terms of the expected odds $\psi_{ijk}^{\bar{D}}$ by introducing an error component:

$$\psi_{ijk}^{\bar{D}} = \beta^* + \beta_1^* X_1 + \beta_2^* X_2 + \beta_3^* X_3 + \epsilon \quad (28)$$

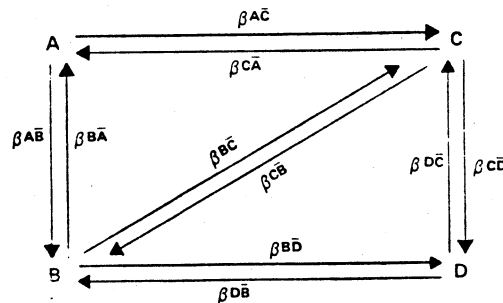
where

$$\epsilon = w_{ijk}^{\bar{D}} - \psi_{ijk}^{\bar{D}} \quad (29)$$

is the error component which, unlike in the analogous regression formulation, is not independent of the expected odds.

Systems of logit equations like those defined by (20)-(23) can also be cast in path diagrams (see Figure 1). In Figure 1, the two arrows emanating from D to C and B correspond to the effects $\beta^{\bar{D}\bar{C}}$ and $\beta^{\bar{D}\bar{B}}$ in (21) and (22) respectively; the three arrows emanating from C to D, B and A

FIGURE 1
PATH DIAGRAM FOR LOGIT EQUATIONS (20) - (23)



correspond to the effects $\beta_{\overline{CD}}$, $\beta_{\overline{CB}}$, and $\beta_{\overline{CA}}$ in (20), (22) and (23) respectively; the three arrows emanating from B to D, C, and A correspond to the effects $\beta_{\overline{BD}}$, $\beta_{\overline{BC}}$, and $\beta_{\overline{BA}}$ in (20), (21), and (23) respectively; and finally, the two arrows emanating from A to C and B correspond to the effects $\beta_{\overline{AC}}$ and $\beta_{\overline{AB}}$ in (21) and (22), respectively. The absence of any arrows linking two or more variables signifies that the respective variables are not directly related. Because $\beta_{\overline{AC}} = \beta_{\overline{CA}}$, $\beta_{\overline{BC}} = \beta_{\overline{CB}}$, $\beta_{\overline{AB}} = \beta_{\overline{BA}}$, $\beta_{\overline{BD}} = \beta_{\overline{DB}}$, and $\beta_{\overline{CD}} = \beta_{\overline{DC}}$, a single double-headed arrow could be used, as Goodman does, instead of two separate single-headed arrows in the figure. I adopt the latter approach to emphasize the presence of reciprocal causation and because it has greater potential in characterizing more complex models.

Some thought concerning the system of logit specifications defined in equations (20)-(23) will show that there is more than one path diagram which is consistent with model (15). Recall in model (15) the following parameters were set to zero:

$$\mu_{\overline{AD}}, \mu_{\overline{ABC}}, \mu_{\overline{ABD}}, \mu_{\overline{ACD}}, \mu_{\overline{BCD}}, \mu_{\overline{ABCD}}$$

Comparison with Table 1, which lists the μ -terms that are set to zero for each of the specifications (20)-(23), shows that any pair of equations (20)-(23), except for the pair (21)-(22), is equivalent to model (15). Further, any triplet of equations (20)-(23) is also equivalent. Thus, a number of path diagrams can be generated, all of which are consistent with the fitted model. For example, Figures 2(a) and (b) present path diagrams for the system of equations (20)-(21). In this figure, note that there is no reciprocal causation between variables B and D, B and C, or A and C and, therefore, only one arrow is used to describe each respective relationship; whereas, as in Figure 1, the two separate arrows from C to D and from D to C correspond to the effects $\beta_{\overline{CD}}$ and $\beta_{\overline{DC}}$ in (20) and (21), respectively.

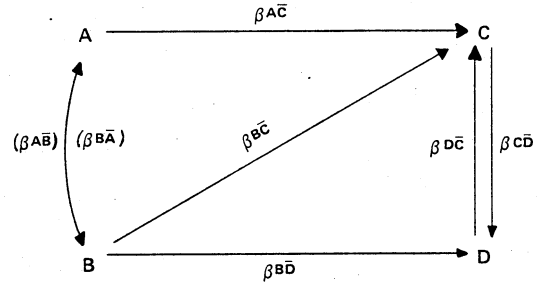
TABLE 1
THE μ -TERM PARAMETERS IN THE SATURATED MODEL SET TO ZERO IN LOGIT SPECIFICATIONS (20)-(23)

Expected Odds	Equation Number	μ -Terms Set to Zero
$\psi_{\overline{ijk}}$	(20)	$\mu_{\overline{AD}}, \mu_{\overline{ABD}}, \mu_{\overline{ACD}}, \mu_{\overline{BCD}}, \mu_{\overline{ABCD}}$
$\psi_{\overline{ij \cdot l}}$	(21)	$\mu_{\overline{ABC}}, \mu_{\overline{ACD}}, \mu_{\overline{BCD}}, \mu_{\overline{ABCD}}$
$\psi_{\overline{i \cdot k l}}$	(22)	$\mu_{\overline{ABC}}, \mu_{\overline{ABD}}, \mu_{\overline{BCD}}, \mu_{\overline{ABCD}}$
$\psi_{\overline{jk l}}$	(23)	$\mu_{\overline{AD}}, \mu_{\overline{ABC}}, \mu_{\overline{ABD}}, \mu_{\overline{ACD}}, \mu_{\overline{ABCD}}$

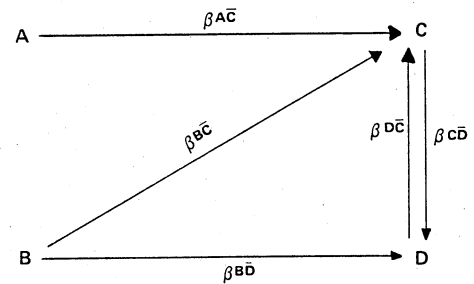
Figures 2(a) and (b) are the same except for the stated relationship between variables A and B. The distinction comes about because although the $\mu_{\overline{AB}}$ -term parameter is included in model (15), there is no logit parameter counterpart specified in either (20) or (21). This means that there are no arrows emanating from variables C or D to variables A or B and my preference is to view variables A and B as "exogenous" as opposed to "endogenous" variables. However variables A and B are viewed, their relationship can be assessed by collapsing over the other endogenous variables, C and D, and analyzing the resulting two-way table. To distinguish β -effects estimated from condensed tables as opposed to those derived from complete tables, I will use a double-headed curved arrow and enclose the coefficient in parentheses as shown in Figure 2(a). That is, in the figure the double-headed curved arrow shown in part (a) signifies that, based upon the [AB] marginal table, variables A and B are not statistically independent, as contrasted to part (b) where the absence of any arrow(s) connecting variables A and B signifies that, based upon the reduced two-way table, variables A and B are statistically independent.

FIGURE 2
PATH DIAGRAM FOR LOGIT EQUATIONS (20)-(21)

a) Mutual dependence between variables A and B



b) Mutual independence between Variables A and B



Comment. Figures 1 and 2 both describe nonrecursive causal models in the usual sense that at least one pairwise relationship in each figure is characterized by two single-headed arrows which signifies reciprocal causation. With four quantitative variables $Z_1, Z_2, Z_3,$ and Z_4 the following structural equations define a nonrecursive system:

$$Z_3 = \beta_{31}Z_1 + \beta_{34}Z_4 + \epsilon_3, \quad (30)$$

$$Z_4 = \beta_{42}Z_2 + \beta_{43}Z_3 + \epsilon_4, \quad (31)$$

where ϵ_3 and ϵ_4 are random disturbance terms with zero means. Notice that in (31) variable Z_3 , which appeared as a dependent variable in (30), appears as an independent variable used to explain variable Z_4 . In the above set up, the "causal effects" of Z_4 on Z_3 can differ (in magnitude) from the "causal effects" of Z_3 on Z_4 ; that is, there is no restriction that $\beta_{34} = 1/\beta_{43}$. On the contrary, in the Goodman approach, the reciprocal effects are by necessity equal to one another. This is particularly troublesome, and while I could suggest that one fit logit models separately (independent to the complete table), which will induce unequal estimates of the reciprocal effects, this procedure is not totally satisfactory since it seems to ignore the fundamental simultaneity of the models.

Goodman (1973b) states, in closing the section on nonrecursive systems, that while the techniques required for the analysis of nonrecursive systems when the variables are quantitative are usually more complicated than are those required for a corresponding analysis of recursive systems, when the variables are qualitative the methods and models derived from considering loglinear analysis are as easily applied to nonrecursive as well as recursive systems. Though the ease of application cannot be denied, we will see very shortly that polytomous variables and/or the presence of higher-order effects can render the analysis

complex. In addition, the multiplicity of path diagrams that are consistent with a given model means that a great deal of burden is placed on the analyst, and this, coupled with the fact that certain effects will be estimated from collapsed tables and others from the complete table, leads me to suggest that the technique be applied carefully so as to avoid misleading results. (I return to the issue of collapsing tables and the problem of path coefficients in the next section.)

Analysis of Recursive Systems

Let us return once again to our four-way table consisting of variables A, B, C, and D, each having two levels. If I impose order conditions such that variable A is causally prior to variable B, variable B causally prior to variable C, and variable C causally prior to variable D, then, under this setup, reciprocal causation and feedback loops have been ruled out. This assumed causal ordering would also seem to imply that any path diagram should be based on three (logit) models: (1) B given A, (2) C given A and B, and (3) D given A, B, and C, which when combined characterizes the conditional joint probability of B, C, and D, given A.

Suppose, on the other hand, that the true causal ordering is variables A and B are causally prior to variable C, and all three variables are prior to variable D. Under this setup Goodman would recommend that one first develop a model for the [AB] marginal table; call this model M_1 . The relationship between variables A and B would be assessed on the basis of the usual likelihood ratio χ^2 statistic based upon the condensed table. Next, attention would be focused on the three-way table [ABC] and models would be fit which included in the set of fitted margins the [AB] margin total; call this selected model M_2 . Finally, consider a model M_3 for the complete four-way table [ABCD] which contains the margin total [ABC] in the set of fitted margins.

Goodman (1973a, 1973b) shows that, under the assumption that the models M_1 , M_2 , and M_3 are all true, the test for the combined model, call it M^* , can be calculated by either estimating the expected frequencies from those obtained under M_1 , M_2 , and M_3 or by the partitioning property of the likelihood ratio χ^2 statistic. In the former case, note

$$F_{ijk\ell}^* = F_{ij}^f F_{ijk}^f / F_{ijk\ell}^f \quad (32)$$

where $F_{ijk\ell}^*$ denotes the expected frequencies under M^* , and thus the $F_{ijk\ell}^*$ can then be substituted into the formula for the likelihood ratio χ^2 statistic and the appropriate test performed. Alternatively, in the latter case, note that, because of the simple multiplicative form of the expected frequencies shown in (32), we have

$$\chi^2(M^*) = \chi^2(M_1) + \chi^2(M_2) + \chi^2(M_3), \quad (33)$$

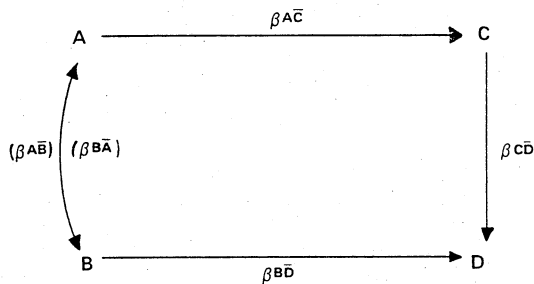
with the degrees of freedom for M^* equaling the sum of the degrees of freedom for the separate χ^2 statistics for M_1 , M_2 , and M_3 .

The path diagram for this model is presented in Figure 3. We see from the figure that: (a) variable D is posterior to the other variables since there are no arrows pointing from D to any of the other variables; (b) the arrows emanating from variables B and C to variable D reflect the situation found in equation (20), namely,

$$\psi_{ijk}^D = \beta_j^D + \beta_j^{BD} + \beta_k^{CD}, \quad (34)$$

and correspond to β_j^{BD} and β_k^{CD} , respectively; and, finally, (c) variable B is statistically independent of variable C, given the level of variable A.

FIGURE 3
PATH DIAGRAM FOR LOGIT EQUATIONS (34)-(39)



The β_i^{AB} and β_j^{BA} effect parameters are induced from fitting the saturated model to the condensed [AB] marginal table; viz.:

$$\ln F_{ij}^{AB} = \mu + \nu_i^A + \nu_j^B + \nu_{ij}^{AB}, \quad (35)$$

which generates the logits

$$\psi_i^B = \beta_i^B + \beta_i^{AB}$$

and

$$\psi_j^A = \beta_j^A + \beta_j^{BA}. \quad (36)$$

The conditional independence of variables B and C given the levels of variable A, that is, $[B \times C | A]$ can be expressed in a form which resembles equation (21). Letting F_{ijk} denote the expected frequency in cell (i,j,k) of the marginal table [ABC], the expected odds with respect to variable C at the joint level (i,j) on variable (A,B) are given by:

$$\begin{aligned} \psi_{ij}^C &= \ln F_{ij1} - \ln F_{ij2} \\ &= \beta_i^C + \beta_i^{AC}. \end{aligned} \quad (37)$$

Comments. Some reflection on Figure 3 might lead one to consider fitting the following loglinear model to the complete four-way table: [AB], [AC], [BC], [BD], [CD], where each two-factor effect corresponds to each arrow in the path diagram. However, though this loglinear model is consistent with the figure, it ignores the hypothesized structure among the variables and, in general, effect estimates derived from a single loglinear model fitted to the complete table will differ from those induced from considering a series of logit models.

On the other hand, collapsing variables may require that certain parameters be fitted that otherwise would have been set to zero if the full table had been analyzed; that is, it is not too uncommon to find that, say, in the condensed [ABC] table, the two-factor effect [BC] is needed (the model of conditional independence of variables B and C does not fit well), whereas in the next stage this effect is nil in that the model [ABC], [AD], [BD], [CD] provides an adequate fit. In fact, several of the examples used by Goodman exhibit the property. For instance, in Goodman (1973a) a four-way table relating the attitudes of school children is analyzed and various path diagrams based on the model M^* : [AB], [AC], [BC], [BD], [CD] are discussed. However, as Goodman admits, the simpler model, M^{**} : [AC], [BC], [BD], [CD], which does not fit the [AB] margin total, also fits the data quite well. If we follow Goodman's approach this term must always be included since the effect is estimated from the collapsed [AB] table.

In the preceding two paragraphs I have attempted to play the devil's advocate by arguing that, on the one hand,

fitting a single loglinear model to a complete table ignores the structure among the variables as evinced by a series of logit models, while, on the other hand, collapsing a variable may require one to fit parameters that would be unnecessary (if the complete table was analyzed). Though each argument has merit, I prefer to follow the hypothesized causal ordering of the variables and allow the reduced tables to dictate the effect parameters included, even in the face of including "additional" parameters. Further, I would argue that when causal priorities are imposed among the variables, the complete multiway table should only be analyzed in the context of the hypothesized relationship between the antecedent and posterior variable(s).

The final point that I wish to make relates to the path coefficients themselves. Though I will insert coefficients corresponding to β -effect values into the path diagrams presented in the following section, there is no formal calculus of these coefficients as exists in analysis of qualitative variables. Consequently, it is difficult to decide what values to assign to arrows not explicitly accounted for by the system of logits, and anything other than "direct" effects is difficult to assess.

Illustrative Examples

I will now devote some effort to showing two different examples of how the methods and models of Goodman can be used in modeling causal systems. Both examples deal with recursive systems. (My earlier remarks support this selection.) The examples are, however, different: the first example is rather straight-forward yet amply illustrates the basic features of the analysis of recursive systems discussed in the previous section; the second example is intended to demonstrate some of the difficulties encountered when polytomous variables and higher-order interactions are present.

Example 1

My first example will analyze data collected in a survey of 2409 individuals on their preference for a particular branch of savings institutions. The data are displayed in Table 2; the table cross-classifies each individual according to four dichotomous (Yes, No) variables: (a) whether the person is familiar with the branch, (b) the person's opinion whether the branch is conveniently located, (c) prior experience (patronage) with the branch, and (d) whether the person strongly recommends the branch.

One plausible causal ordering is that A and B precede C and D; that is, familiarity with and opinions concerning locational convenience are antecedent to one's having previously patronized the branch, and all three affect whether a person would strongly recommend the branch. I view both variables A and B as antecedent to variable C since it is difficult to argue that one precedes the other, and I prefer to treat each as simply exogenous. The relationship between variables A and B was assessed in the condensed [AB] marginal table shown in part (b) of the table. Simple computation shows the cross product ratio $\alpha = 2.7 (485 \times 1037 / 240 \times 547)$. Since an $\alpha = 1$ is consistent with the hypothesis of independence between A and B, we conclude that A and B are positively related. (We could, of course, compute a G^2 statistic and test its significance but given the magnitude of α this seems unnecessary.)

Next, prior experience, variable C, is treated as the response measure and variables A and B as explanatory. With margin total [AB] fixed (the [AB] margin proved significant) there are three unsaturated loglinear models corresponding to such a logit model:

TABLE 2
CROSSCLASSIFICATION OF FAMILIARITY (A), LOCATIONAL CONVENIENCE (B), PREVIOUS PATRONAGE (C) AND STRENGTH OF RECOMMENDATION*

(C)	(D)	(B)	1.0	A	2.0
1.0	1.0	1.0	423		187
		2.0	459		412
2.0	2.0	1.0	49		47
		2.0	68		127
	1.0	1.0	13		84
		2.0	17		407
2.0	1.0	1.0	0		22
		2.0	3		91

Part (a) [AB] Marginal Table

(B)	1.0	A	2.0
1.0	485		340
2.0	547		1037

Part (b) [ABC] Margin Table

(C)	(B)	1.0	A	2.0
1.0	1.0	472		234
	2.0	527		539
2.0	1.0	13		106
	2.0	20		498

* Variable A is whether the person is familiar with the branch; variable B is the person's opinion whether the branch is conveniently located; variable C is whether the person had previously patronized the branch; and variable D is whether the person strongly recommends the branch. The first level of each variable reflects a positive answer, the second level a negative one.

- M_1 : [AB] [AC] [BC] (1 d.f.; $G^2 = 1.01$)
 M_2 : [AB] [AC] (2 d.f.; $G^2 = 31.02$)
 M_3 : [AB] [BC] (2 d.f.; $G^2 = 533.84$)

Clearly, the only one of these models that provides an adequate fit to the data is model M_1 . Model M_1 posits no second order interactions and corresponds to the logit model

$$\psi_{ij}^{\bar{C}} = \beta_{\bar{C}} + \beta_i^{\bar{A}\bar{C}} + \beta_j^{\bar{B}\bar{C}} \quad (38)$$

The respective estimated effect parameters are 1.53 and 0.33.

The final step is to build a logit model treating variable D, whether the person strongly recommends the brand, as the dependent variable and the other three variables as explanatory. There are eight unsaturated loglinear models, each of the three two-factor effects may be absence or presence. Four of these models are listed below:

- M_4 : [ABC] [AD] [BD] [CD] (4 d.f.; $G^2 = 4.45$)
 M_5 : [ABC] [BD] [CD] (5 d.f.; $G^2 = 39.86$)
 M_6 : [ABC] [AD] [CD] (5 d.f.; $G^2 = 6.24$)
 M_7 : [ABC] [AD] [BD] (5 d.f.; $G^2 = 8.19$)

Models M_4 and M_6 provide reasonable fits to the data while models M_5 and M_7 are unacceptable. Since Model M_6 differs from model M_4 in only one term, [BD], a test for the significance of this effect can be accomplished by simply taking the difference in the G^2 statistics associated with each and then comparing this value to a χ^2 distribution with 1 d.f. Noting that $G^2(M_6) - G^2(M_4) = 1.79$, which is nonsignificant, the preference is for Model M_6 . The

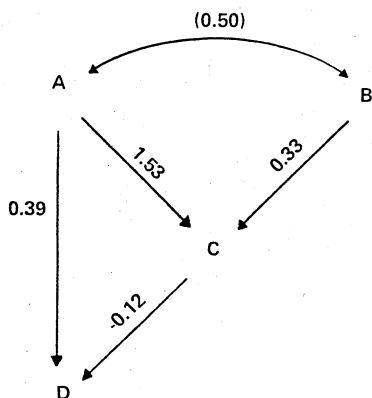
corresponding logit is

$$\psi_{ijk}^{\bar{D}} = \beta_{\bar{D}} + \beta_i^{\bar{AD}} + \beta_k^{\bar{CD}} \quad (39)$$

The effect estimates are 0.39 and -0.12, respectively.

The path diagram characterizing this system is shown in Figure 4. Notice I use the double-headed curved arrow to indicate that no decision as to whether A causes B or B causes A has been made and have enclosed the corresponding effect-estimates in () to emphasize that these variables are viewed as exogenous to the system. The causal interpretation of Figure 4 is that the exogenous variables, branch familiarity and opinions concerning its locational convenience, are related; these variables, familiarity and locational convenience, affect prior patronage; and whether the branch received a strong recommendation is affected by familiarity and prior patronage. Interestingly, the β -effect associated with prior patronage and recommendation ($\beta^{\bar{CD}}$) is negative which implies that previous experience with the branch decreases the likelihood of the branch receiving a strong recommendation. It seems, unfortunately, that the old adage: familiarity (in the form of prior contact) breeds contempt, does actually hold.

FIGURE 4
PATH DIAGRAM SHOWING CAUSAL CONNECTIONS
IMPLIED BY LOGIT MODELS (38)-(39)



Note, finally, that if we were not concerned with the hypothesized structure among the variables as expressed in terms of logit specifications, the system shown in Figure 4 might lead us to fit the loglinear model:

$$M_8: [AB] [AC] [AD] [BC] [CD]$$

to the four-way table, where each arrow in the figure is represented by a two-factor effect. The G^2 statistic for this model has a value of 7.25 with 6 d.f. Hopefully, most of us can now understand why this value is identical to the numerical value obtained from the pair of recursive logits; that is if the G^2 statistic corresponding to the combined recursive system by G_*^2 then we have:

$$G^2(M_8) = G_*^2 = G^2(M_4) + G^2(M_6)$$

$$7.25 = 7.25 = 1.01 + 6.24.$$

Example 2

The next example makes use of data, originally collected by Sewell and Shah (1968), on the relationship among five variables: (A) socioeconomic status (high, upper middle, lower middle, low), (B) intelligence (high, upper middle, lower middle, low), (C) sex (male, female), (d) parental encouragement (low, high), and (E) college plans (yes, no). The five-way table is shown in Table 3. The data present

TABLE 3
CROSSCLASSIFICATION OF SOCIAL CLASS (A), SEX (C),
PARENTAL ENCOURAGEMENT (D), IQ (B), AND EDUCATIONAL
ASPIRATIONS (E) (SEWELL AND HAN, 1968)

(C)	(B)	(E)	(D)	(A)			
				L	LM	UM	H
M	L	Yes	Low	4	2	8	4
			High	13	27	47	39
		No	Low	349	232	166	48
			High	64	84	91	57
	LM	Yes	Low	9	7	6	5
			High	33	64	74	123
		No	Low	207	201	120	47
			High	72	95	110	90
	UM	Yes	Low	12	12	17	9
			High	38	93	148	224
		No	Low	126	115	92	41
			High	54	92	100	65
H	Yes	Low	10	17	6	8	
		High	49	119	198	414	
	No	Low	67	79	42	17	
		High	43	59	73	54	
F	L	Yes	Low	5	11	7	6
			High	9	29	36	36
		No	Low	454	285	163	50
			High	44	61	72	58
	LM	Yes	Low	5	19	13	5
			High	14	47	75	110
		No	Low	312	236	193	70
			High	47	88	90	76
	UM	Yes	Low	8	12	12	12
			High	20	62	91	230
		No	Low	216	164	174	48
			High	35	85	100	81
H	Yes	Low	13	15	20	13	
		High	28	72	142	360	
	No	Low	96	113	81	49	
		High	24	50	77	98	

a number of problems not previously illustrated, namely, polytomous variables and logit models that will include second-order effects. I note that this example has been analyzed elsewhere and my discussion will follow Fienberg's (1977) analysis.

A causal model proposed by Sewell and Shah specifies

A
and
B precede D precedes E;
and
C

two logits characterize the system where in one socioeconomic status, intelligence and sex are viewed as antecedent, exogenous variables to parental encouragement; while in the second, college plans is the response measure and all four remaining variables are treated as explanatory. With respect to the exogenous variables A, B, and C, the only reasonably fitting model included all three two-factor effects: [AB], [AC], [BC] (9 d.f.; $G^2 = 11.50$). We thus conclude that socioeconomic status is associated with intelligence and with sex, and intelligence and sex are also positively related. Again we do not distinguish between any of these causal orderings.

Treating, first, variable D, parental encouragement as the response measure, some plausible logit models are:

- M₁: [ABC] [AD] [BC] [CD] (24 d.f.; G² = 55.81)
- M₂: [ABC] [ABD] [CD] (15 d.f.; G² = 34.60)
- M₃: [ABC] [BCD] [ACD] (18 d.f.; G² = 31.48)
- M₄: [ABC] [ABD] [BCD] (12 d.f.; G² = 22.44)
- M₅: [ABC] [ABD] [ACD] (12 d.f.; G² = 22.45)
- M₆: [ABC] [ABD] [ACD] [BCD] (9 d.f.; G² = 9.22)

The only model providing an acceptable fit is model M₆ which included three second-order effects. The logit counterpart to model M₆ is

$$\psi_{ijk}^{ABCD} = \beta^{\bar{D}} + \beta_i^{A\bar{D}} + \beta_j^{B\bar{D}} + \beta_k^{C\bar{D}} + \beta_{ij}^{AB\bar{D}} + \beta_{ik}^{AC\bar{D}} + \beta_{jk}^{BC\bar{D}} \quad (41)$$

The presence of second-order effects presents no real problem, except they are, of course, more complex, but we have yet to consider path diagrams induced from such effects.

Next, variable E, college plans, becomes the response measure. Some plausible logit models are

- M₇: [ABCD] [E] (63 d.f.; G² = 4497.51)
- M₈: [ABCD] [AE] [BE] [CE] [DE] (55 d.f.; G² = 73.82)
- M₉: [ABCD] [BCD] [AE] [DE] (52 d.f.; G² = 59.55)
- M₁₀: [ABCD] [BCD] [ACE] [DE] (49 d.f.; G² = 57.99)

Both models M₉ and M₁₀ provide acceptable fits. For reasons of parsimony (and since the difference in G² values of these two models is nonsignificant) model M₉ is selected; the corresponding logit is:

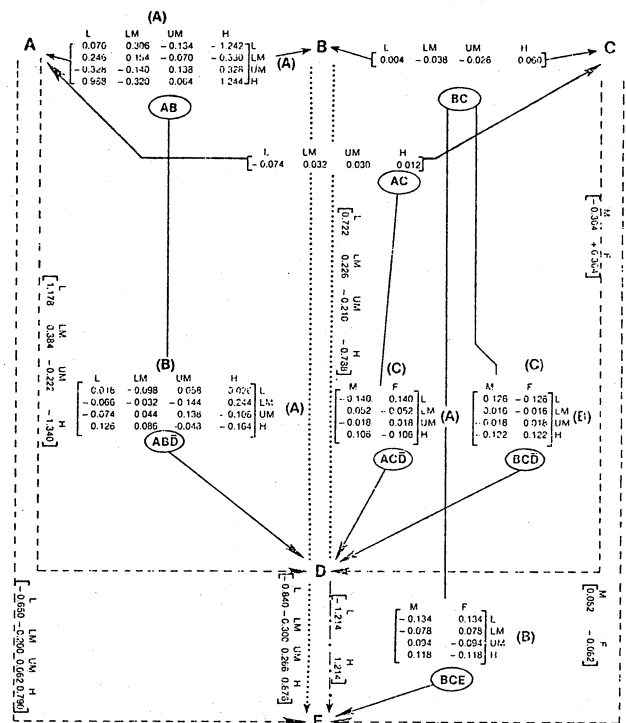
$$\psi_{ijkl}^{ABCDE} = \beta^{\bar{E}} + \beta_i^{A\bar{E}} + \beta_j^{B\bar{E}} + \beta_k^{C\bar{E}} + \beta_l^{D\bar{E}} + \beta_{ik}^{BC\bar{E}} \quad (42)$$

Note, this model includes one second-order effect corresponding to the effects of intelligence and sex on college plans, and two first-order effects which show that college plans are affected by socioeconomic status and parental encouragement.

Though a simple path diagram like those presented earlier is not possible, a diagrammatic representation, albeit more complex, of the induced logit system is possible. Consider Figure 5 which portrays the path diagram showing the causal connections implied by models M₆ and M₉. In the figure, (---) lines pertain to variable A, (····) lines to variable B, (— — —) lines to variable C, and (— · — · —) lines to variable D; association between the exogenous variables is shown by double-headed curved arrows and higher-order effects are shown by a solid single-headed arrow which emanates from the effect in question. Path coefficient effects have been included, where column and row headings have been inserted to identify particular effects.

The first order effect estimates for variables A, B, and C are in the expected direction. The likelihood of low versus high parental encouragement decreases monotonically as we move from low to high socioeconomic status and intelligence, and it seems males get more encouragement than females. The second-order effects of sex by intelligence show low intelligent males receiving relatively less encouragement than low intelligent females and high intelligent males receiving relatively more than high intelligent females; the second order effects of sex by socioeconomic status shows males receiving less parental encouragement than their female counterparts. With respect to college plans, the first order effects indicate that there are positive affects on college plans from being male, of high socioeconomic status, or of high intelligence. Notice the monotonic order to the effect estimates within the socioeconomic and intelligence categories. The second-order effects between sex and intelligence are also monotonically ordered with males having low intelligence less likely to

FIGURE 5
PATH DIAGRAM SHOWING CAUSAL CONNECTIONS
IMPLIED BY LOGIT MODELS (41)-(42)



Variable A is socioeconomic status (high, upper middle, lower middle, low), variable B is intelligence (high, upper middle, lower middle, low), variable C is sex (male, female), variable D is parental encouragement (low, high) and variable E is college plans (yess, no).

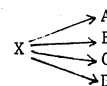
plan for college than a female with comparable intelligence, whereas the obverse is true for males and females with high intelligence.

Latent Class Models

I have already introduced the concept of spurious correlation with qualitative variables. The approach taken in that section was to argue that the association between two variables is spurious if it vanishes in the presence of a third variable which impinges upon both. In the context of our illustrative four-way table we said if three variables, say, A, B, and C, become mutually independent when the levels of the remaining variable, D, are controlled for, then their original relationship is spurious.

Consider the situation where some other variable, say, variable X, not part of the original cross-classification, can explain the relationships among the four variables when this new variable is controlled for. Once one accepts the possibility of some "outside" factor affecting the apparent relationship among the system variables, the step to considering latent class models is a rather easy, direct one since one need only assume that the explanatory variable in question is unobservable (or latent).

The situation described in the preceding paragraph can be diagrammed as



where A, B, C, and D are the observable (manifest)

variables and variable X is the explanatory latent variable. To model this relationship, let p_{ijkl} denote the observed proportion (probability) of individuals in cell (i,j,k,l) where

$$P_{ijkl} = f_{ijkl}/n. \quad (43)$$

For some specified hypotheses about the system of variables in the four-way table let π_{ijkl} be the expected proportion of individuals in cell (i,j,k,l) where the relationship holds that

$$F_{ijkl} = n\pi_{ijkl}. \quad (44)$$

Assume further that variable X has t latent classes so that π_t^X denotes the probability that an individual will be in class t and π_{ijklt}^{ABCDX} denotes the probability that an individual will be at level (i,j,k,l,t) with respect to the joint variable (A,B,C,D,X). Letting $\pi_{it}^{\bar{A}X}$, $\pi_{jt}^{\bar{B}X}$, $\pi_{kt}^{\bar{C}X}$, and $\pi_{lt}^{\bar{D}X}$ be conditional probability that an individual will be at level i with respect to variable A, level j with respect to variable B, etc., given that s/he is at level t on variable X then we can express the hypothesized system as

$$\pi_{ijklt}^{ABCDX} = \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{lt}^{\bar{D}X}, \quad (45)$$

and

$$\pi_{ijkl} = \sum_t \pi_{ijklt}^{ABCDX}. \quad (46)$$

Formula (45) simply states that the latent variable explains the relationships among the manifest variables in the sense that any relationships disappear (the manifest variables are mutually independent) when the latent variable X is held constant. It is very easy to see how path diagrams based on logit models of the form previously discussed can be used in the analysis of latent class models by simply defining $\omega_{jkl}^{\bar{A}}$ as the odds that an individual will be at level 1 rather than level 2 on variable A given his joint level on the remaining variables (B,C,D,X) is (j,k,l,t); i.e.,

$$\omega_{jkl}^{\bar{A}} = \pi_{1jkl}^{\bar{A}BCDX} / \pi_{2jkl}^{\bar{A}BCDX} \quad (47)$$

However, from (45) we have

$$\omega_{jkl}^{\bar{A}} = \pi_{1t}^{\bar{A}X} / \pi_{2t}^{\bar{A}X} \quad (48)$$

since the hypothesis under study is that the expected odds pertaining to variable A are affected by the level of variable X, but not by the level (j,k,l) of the other variables (B,C,D).

Due to space constraints I will say no more about this class of models. However, at this point, the reader should be able to see the approach that Goodman will take in the analysis of such systems: Various loglinear models are fitted which describe the relationships among the manifest variables when one or more latent variables are controlled for. Corresponding logits are then formed with the estimated effect parameters inserted as coefficients in the path diagram. The interested reader wanting further details, especially with respect to maximum likelihood estimation, should consult Goodman (1974a, 1974b).

Concluding Remarks

I have restricted my remarks to the use of loglinear model in the analysis of causal systems, and specifically to the work of Goodman. In doing so, I have no doubt slighted many. For example, Bahadur (1961), Boudon (1968), Coleman (1964), and Davis (1974, 1975) are just a few of those who have made important contributions in the area. The work of Davis (1975) is particularly noteworthy in that he

describes an interesting way of modeling cross-classified data which utilizes linear flow graphs based on differences in proportions instead of odds ratios. Even with respect to the methods and models championed by Goodman I have been remiss. Model testing, parameter estimation, and the whole class of latent models have only been superficially covered.

Critics of Goodman's system might cite me for not being more critical of his approach. After laboring through Goodman's papers several times (this itself is, I think, a major feat--reading Goodman can overwhelm one, especially his footnotes) I believe his system to be sound, if, of course, care and good judgment on the part of the user are exercised. Critics have been quick to point to situations where the analogy to the analysis of quantitative variables via systems of linear simultaneous equations is the weakest, namely: (1) the absence of a calculus for the path coefficients and (2) the problems caused by polytomous variables and higher-order effects. We have seen, however, that point (2) presents no serious problems; in particular, Example 2 demonstrated that path-like diagrams, although admittedly more complex, can be constructed in such a way as to convey the character of the causal system under study. With regard to point (1) the situation is, unfortunately, more serious. Because the Goodman approach does not impose a structure analogous to the usual normal equations and since the ψ 's do not actually become independent variables in succeeding logit specifications, the Wright multiplication theorem does not operate. Nevertheless, I see no reason why the estimated β -effects cannot be inserted in the path diagram for they serve a very important interpretive purpose. The user who, for whatever reasons, wants to employ Wright-like multiplication rules when analyzing qualitative data, especially in the case of dichotomous variables, should refer to the methods and models proposed by Boudon (1965) and Davis (1975).

The other uncomfortable feature of Goodman's approach relates to the multiplicity of loglinear models that can be fitted which are consistent with a given series of logit specifications. This, as I have indicated, places an extreme burden on the analyst. Yet this burden may well be a blessing in the sense that it can possibly provide the motivating force leading one to rely on theoretical arguments in order to generate explicit causal orderings among the variables. It is in situations where the hypothesized causal system is induced from theory that the methods and models proposed by Goodman work best; without theory to guide us, I'm afraid, the intent of most studies, no matter what the form of analysis, becomes explanatory and often subject to Procrustes-like solutions.

References

- Bahadur, R. R. (1961), "A Representation of the Joint Distribution of Responses to N Dichotomous Items," In Studies in Item Analysis and Prediction, edited by H. Solomon, Stanford, California: Stanford University Press.
- Birch, M. W. (1963), "Maximum Likelihood in the Three-Way Contingency Tables," Journal of the Royal Statistical Society (ser. B), 25, 383-400.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), Discrete Multivariate Analysis, Cambridge, Massachusetts: MIT Press.
- Boudon, R. (1965), "A Method of Linear Causal Analysis: Dependence Analysis," American Sociological Review, 22, 677-682.
- Boudon, R. (1967), L'Analyse Mathematique des Faits Sociaux, Paris: Plan.

- Boudon, R. (1968), "A New Look at Correlation Analysis," In Methodology in Social Research, eds. H. M. Blalock, Jr., and A. Blalock, New York: McGraw-Hill, 199-235.
- Coleman, J. S. (1964), Introduction to Mathematical Sociology, Glencoe, Illinois: Free Press.
- Davis, J. A. (1974), "A Survey Metric Model of Social Change," Pt. 1 (February), Lithographed, Chicago: National Opinion Research Center.
- Davis, J. A. (1975), "Analyzing Contingency Tables with Linear Flow Graphs: D Systems," In Sociological Methodology 1976, edited by David R. Heise, San Francisco: Jossey-Bass, 111-145.
- Duncan, O. D. (1966), "Path Analysis: Sociological Examples," American Journal of Sociology, 72, 1-16.
- Duncan, O. D. (1975), Structural Equations Models, New York: Academic Press.
- Fienberg, S. E. (1977), The Analysis of Cross-Classified Categorical Data, Cambridge, Massachusetts: The MIT Press.
- Goodman, L. A. (1972), "A General Model for the Analysis of Surveys," American Journal of Sociology, 77, 1035-1086.
- Goodman, L. A. (1973a), "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys," American Journal of Sociology, 78, 1135-1191.
- Goodman, L. A. (1973b), "The Analysis of Multidimensional Contingency Tables When Some Variables are Posterior to Others: A Modified Path Analysis Approach," Biometrika, 60, 179-192.
- Goodman, L. A. (1974a), "Exploratory Latent Structural Analysis Using Both Identifiable and Unidentifiable Models," Biometrika, 61, 215-231.
- Goodman, L. A. (1974b), "The Analysis of Systems of Qualitative Variables when Some of the Variables are Unobservable, Part I: A Modified Latent Structure Approach," American Journal of Sociology, 79, 1179-1259.
- Goodman, L. A. (1978), Analyzing Qualitative/Categorical Data, Cambridge, Massachusetts: ABT Books.
- Haberman, S. J. (1978), Analysis of Qualitative Data, New York: Academic Press.
- Lazarsfeld, P. F. (1970), "A Memoir in Honor of Professor Wold," In Scientists at Work, edited by T. Dalenius, G. Karlsson, and S. Malmquist, Uppsala: Almqvist & Wiksells.
- Lazarsfeld, P. F., and Henry, N. W. (1968), Latent Structure Analysis, Boston: Houghton Mifflin.
- Sewell, W. H. and Shah, V. P. (1968), "Social Class, Parental Encouragement, and Educational Aspirations," American Journal of Sociology, 73, 559-572.
- Wright, S. (1960), "Path Coefficients and Path Regressions: Alternatives in Complementary Concepts?" Biometrics, 16, 423-445.

ATTRITION BIAS IN THE ESTIMATION OF
ECONOMETRIC MODELS FROM PANEL DATA

Russell S. Winer, Columbia University

Abstract

Panel data are often used to estimate the parameters of econometric or other linear models. However, a common problem with panel data is attrition. In this paper, a model is developed that corrects structural econometric models estimated using panel data for possible attrition bias. The model is illustrated using a simultaneous equation structural model and panel data that have an attrition problem. The results indicated that the impact of attrition tended to be on the model's exogenous variables and not on the endogenous variables that were exogenous in the equations.

Introduction

Researchers in marketing and other social science disciplines often conduct studies where a sample of individuals is interviewed at two or more points in time. Such data have been referred to in various literatures as longitudinal, cross-sectional, time-series and panel. To marketing researchers, the term panel data usually connotes consumer panel data which are records of household purchasing behavior over time. However, panel studies other than those using consumer panel data have occurred frequently in the marketing literature.

Although panels have been widely used, they are not without problems (Carman 1974). One of these problems is panel attrition or mortality. Except for panels operating under tightly controlled conditions, panel attrition is unavoidable. For example, Charlton and Ehrenberg (1976) report that 88% of their initial sample completed the 25-week panel. The panel utilized by Farley, Howard, and Lehmann (1976) exhibited a 43% drop-out rate over four waves of interviewing spanning fifteen months.

There has been considerable interest in the possibility of panel bias occurring due to attrition. For example, in a panel study of economic attitude study and change covering the years 1954-1957, Sobol (1959) found that renters, people with low income, and people disinterested in the study tended to drop out. Bucklin and Carman (1967) also found the interest factor and its correlates to be highly related to attrition. Ferber (1966) discovered disproportionately high attrition rates related to older age, lower education, self-employment, and high personal asset value. Therefore, assuming an initial probability sample or an alternative sampling approach attempting to represent the population studied, the sample left at the end of the final wave may be biased with respect to the population.

Econometric or other linear models have often been used to analyze panel data. Such applications have been in the areas of evaluating advertising effectiveness (Prasad and Ring 1976, Winer 1980), market segmentation (Frank, Massy, and Boyd 1967, McCann 1974), testing general models of buyer behavior (Farley and Ring 1970, Farley, Howard, and Lehmann 1976), sales management (Futrell and Jenkins 1978), and others.

The effect of attrition on the parameters of linear models has been ignored in studies using panel data. This paper describes the conditions under which parameter bias can occur from attrition and develops a model to correct for the bias. It is demonstrated that the bias is actually specification error in the structural model being estimated.

The model correcting for attrition bias simultaneously determines the correlates of attrition and the structural model parameters. The model is estimated using panel data exhibiting attrition.

Attrition Effects On Regression Estimates

Overview

It is assumed that everyone on the panel, both attriters if they were observed and non-attriters, follow the same linear model in any given wave,

$$y = \chi \beta + \epsilon, \quad (1)$$

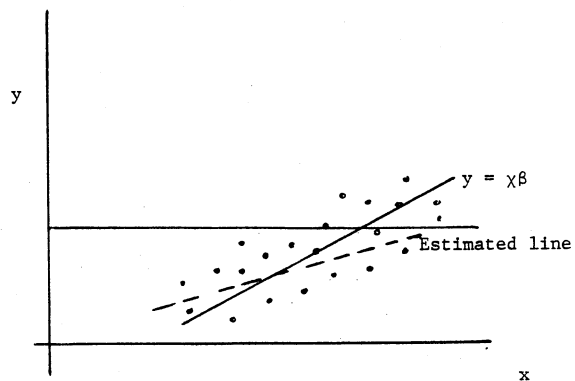
where

- y = $N \times 1$ vector of observations on the dependent variables;
- χ = $N \times k$ matrix of independent variable observations;
- β = $k \times 1$ vector of coefficients;
- ϵ = $N \times 1$ vector of normal disturbances, $\sim N(0, \sigma_\epsilon^2)$

That is, if they could be observed, the behavior of the dropouts is not different from the people remaining on the panel. Attriters could differ, of course, in the χ matrix but not in terms of β . If β differed between attriters and non-attriters, no correction in a non-attriter model could ever be made and population inferences from a sample experiencing attrition could not be done.

If attrition probability is a function solely of χ or is random, no bias of β is induced. There is, however, some loss of efficiency (i.e., the coefficient standard errors will be larger than they should be) since there is a decline in the working sample size. On the other hand, bias will exist in β if the probability of attrition is related to y and, hence, ϵ . Since it is known that interest in the study is a determinant factor in the mortality problem, it is possible that a correlate of interest is a prospective dependent variable of a structural model. For example, if behavioral measures such as awareness and attitude are being tracked over time for a product, families continuing to lack knowledge about the product may be more likely to drop out than those that are familiar with it. That the bias exists is demonstrated by the figure:

FIGURE 1
Effects of Dependent Variable-Related Attrition



An implication of this discussion is related to panel operation. For some panels such as commercial consumer panels, households that attrite are replaced by other households. It is clear that the primary criterion for using a family as a replacement should be in terms of a potential dependent variable such as purchasing rather than description variables such as family size, sex, income, etc., if linear models are to be used in analyzing the data. Replacement by household descriptors will cause the sample to resemble the population from which the original panel was drawn but does not cause the bias in β to vanish.

A Model of Attrition

To formalize the above discussion recent papers in the economics literature will be drawn upon. Hausman and Wise (1976, 1977, 1979), Heckman (1976), and Griliches, Hall, and Hausman (1977) have all developed models for dealing with the general problem of incorporating non-random missing data into econometric analysis of panel data of which attrition is a special case. Samples from which attrition has occurred are termed censored; in truncated samples, no data at all are available on a group of households as they were systematically excluded. Since data are available for panel dropouts, the probability that an observation is in the sample can be computed from censored samples. That is not the case for truncated samples. This probability is an important part of correcting the structural equation (1) for possible attrition bias.

Model Development

Again, consider equation (1), $y = \chi\beta + \epsilon$, to be the structural model for the panel members for a given wave. Let $D = 1$ if y is observed for a panel member, and let $D = 0$ if y is unobserved due to attrition. Assume that y is observed ($D = 1$) if

$$d = \alpha y + X\gamma + W\theta + \mu \geq 0, \quad (2)$$

where y and X are as before and

$W = n \times m$ matrix of variables that do not affect y but affect the probability of its being observed;

$\alpha, \gamma, \theta =$ scalar, $k \times 1$, and $m \times 1$ vectors of parameters,
 $\mu \sim N(0, \sigma_\mu^2)$

The inequality could be relative to any threshold value but is set ≥ 0 for convenience. The vector, θ , could be all zeros. If (1) is substituted into (2), the result is

$$\begin{aligned} d &= \alpha (X\beta + \epsilon) + X\gamma + W\theta + \mu \\ &= X(\alpha\beta + \gamma) + W\theta + \alpha\epsilon + \mu. \end{aligned} \quad (3)$$

A reduced form of (3) is

$$d = X\pi + W\theta + \delta. \quad (4)$$

It follows that

$$\text{prob}(d \geq 0) = \text{prob}(D=1) = F\left(\frac{X\pi + W\theta}{\sigma_\delta}\right) \quad (5)$$

$$\text{prob}(d < 0) = \text{prob}(D=0) = 1 - F\left(\frac{X\pi + W\theta}{\sigma_\delta}\right)$$

where $F(\cdot)$ is the standard normal distribution function.

As discussed in the previous section, $E(\epsilon)$ may not equal zero since some y values and, hence, errors are truncated when attrition occurs. Given the structural equation (1) and the equation describing the probability of being in the sample, equation (4), then

$$E(y|X, D=1) = \chi\beta + E(\epsilon|X, D=1). \quad (6)$$

of interest is $E(\epsilon|X, D=1)$. Since a y -observation, and, hence, ϵ , exists only if $d \geq 0$, the error can be re-written

$$E(\epsilon|X, D=1) = E(\epsilon|X, \delta > -X\pi - W\theta). \quad (7)$$

From Johnson and Kotz (1970 p. 81, 1972 p. 112),

$$E(\epsilon|X, \delta > -X\pi - W\theta) = \frac{\text{cov}(\epsilon, \delta)}{\sigma_\delta^2} E(\delta) \quad (8)$$

$$= \frac{\text{cov}(\epsilon, \delta)}{\sigma_\delta^2} \left(\frac{\sigma_\delta f(-X\pi - W\theta)}{1 - F(-X\pi - W\theta)} \right)$$

$$= \frac{\text{cov}(\epsilon, \delta)}{\sigma_\delta} \left(\frac{f(X\pi + W\theta)}{F(X\pi + W\theta)} \right).$$

Equation (8) results from the fact that $E(\epsilon)$ is affected by δ being truncated from below.

Since $\sigma_\delta > 0$, the critical quantity in (8) is the covariance between the structural error, μ , and the attrition probability error, δ . If $\text{cov}(\epsilon, \delta) \neq 0$, then the whole second term on the right side of (8) is non-zero. As noted by Heckman (1979), the general sample truncation problem can be equated with ordinary specification error arising from omitted variables in regression analysis. If $\text{cov}(\epsilon, \delta) = 0$, the term drops out from (6) and β is unbiased.

Let

$$\left(\frac{\text{cov}(\epsilon, \delta)}{\sigma_\delta} \right) = \lambda$$

and

$$\frac{f\left(\frac{X\pi + W\theta}{\sigma_\delta}\right)}{F\left(\frac{X\pi + W\theta}{\sigma_\delta}\right)} = Z. \quad (9)$$

Then, by adding an error term, equation (6) becomes

$$E(y|x, D=1) = X\beta + \lambda Z + \xi \quad (10)$$

Therefore, a convenient test for attrition bias is the significance of λ ; if it is significant, the omitted variable, Z , is an important explanatory variable and needs to be included to estimate β without bias.

Estimation

Several approaches have been developed to estimate the parameters of (10). The most complex approach is to use maximum likelihood methods (Hausman and Wise 1979, Griliches, Hall, and Hausman, 1977). However, Heckman (1976) proposed a relatively simple estimator that is useful for exploratory data evaluation.

At a given point in time (wave) in the panel, the steps of Heckman's procedure are:

1. Estimate equation (4) using probit analysis (Finney 1964) on the whole sample. That is, at that point in time, $D=1$ for a person still in the sample, $D=0$ for an attriter.
2. From the parameter estimates of step 1 ($\hat{\pi}$, $\hat{\theta}$, and $\hat{\sigma}_\delta$), estimate Z as shown in (9) for each member of the sample still left.
3. The estimate of Z , \hat{Z} , is then used in equation (10) to estimate λ and β by ordinary least squares (OLS). This procedure produces consistent estimates of $\hat{\beta}$ and $\hat{\lambda}$. However, if $\hat{\lambda} \neq 0$, σ_ϵ^2 from (10) is underestimated

as ξ is heteroscedastic (Heckman 1976, p. 480). This will produce inflated estimates of the significance of β and λ . Therefore, generalized least squares (GLS) rather than OLS can be used in step 3 of the estimation procedure (Heckman 1979). However, using OLS in step 3 makes the attrition bias correction factor straightforward to implement using commonly available computer routines.

Data

The data utilized in this study are from Farley, Howard, and Lehmann (1976). A panel was constructed from a national probability sample of respondents expressing some interest in purchasing subcompact cars within the subsequent two years. Four waves of telephone interviews were conducted over a fifteen-month period with people who would be principal drivers of the new subcompact car if purchased.

Because many questions were asked in each wave and there were eight brands in question, the panel was divided into four groups which were each questioned about only four of the eight brands. Information about the brand used in this study, the Chevrolet Vega, was requested by all four groups. Table 1 demonstrates the attrition rate for the four groups over the four waves.

TABLE 1

Group	Sample Sizes for the Groups of Respondents Over Time				Total Attrition Rate
	Wave				
	1	2	3	4	
1	305	229	213	192	37%
2	328	251	206	171	48%
3	288	219	197	169	41%
4	316	239	205	177	44%
Total	1237	938	821	709	43%

As can be seen, the attrition problem with this panel was substantial despite the fact that the panel was of ordinary design compared to other longitudinal data bases.

Models

Structural Model

The model employed is based on the Howard-Sheth model of buyer behavior (Howard and Sheth 1969) and is a variant of that utilized by Farley, Howard, and Lehmann (1976). It is a four-equation system explaining the endogenous variables Intention, Attitude, Confidence, and Brand Comprehension. The model has the form

$$By' + \Gamma x' = \mu \quad (11)$$

where

$$B = \begin{bmatrix} -1 & \beta_{f2} & \beta_{13} & 0 \\ 0 & -1 & 0 & \beta_{24} \\ 0 & 0 & -1 & \beta_{34} \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \gamma_{10} & \gamma_{11} & 0 & 0 & \gamma_{15} & \gamma_{16} & 0 & 0 & \gamma_{19} & 0 & 0 & \gamma_{112} \\ \gamma_{20} & 0 & \gamma_{22} & 0 & 0 & 0 & \gamma_{26} & \gamma_{27} & \gamma_{28} & \gamma_{29} & \gamma_{210} & 0 \\ \gamma_{30} & 0 & 0 & \gamma_{33} & 0 & \gamma_{35} & \gamma_{36} & \gamma_{37} & \gamma_{38} & 0 & \gamma_{310} & 0 \\ \gamma_{40} & 0 & 0 & 0 & \gamma_{44} & \gamma_{45} & \gamma_{46} & 0 & \gamma & 0 & \gamma_{410} & \gamma_{411} \end{bmatrix}$$

and the ij index refers to equation i and variable j . y is a 4-element vector of the endogenous variables, x is a 12-element vector of exogenous variables, and μ is a vector of disturbances. The variables can be found in Table 2.

TABLE 2

Variable List

Variable	Definition
<u>Endogenous</u>	
Intention (y_1)	Intention to purchase Vega during next three months, 1-10 scale
Attitude (y_2)	Overall liking of Vega, 1-10 scale.
Confidence (y_3)	Respondent's perceived ability to judge Vega 1-10 scale
Brand Comprehension (y_4)	The number of correct answers to a set of objective questions about Vega
<u>Exogenous</u>	
Intention ($t-1$) (x_1)	Lagged Intention
Attitude ($t-1$) (x_2)	Lagged attitude
Confidence ($t-1$) (x_3)	Lagged confidence
Comprehension ($t-1$) (x_4)	Lagged comprehension
Advertising Recall (x_5)	Summed recall of specific copy points for Vega
Propensity (x_6)	Intention to buy product class within 2 years, 1-10 scale
sex of respondent (x_7)	0=female, 1=male
age of respondent (x_8)	10 category ordinal scale
Income (x_9)	Reported, 10 category ordinal scale
Education (x_{10})	Number of years of schooling
Number of Drivers (x_{11})	In household
Child (x_{12})	If principal driver would be a child (1=yes, 0=no)

Attrition Model

As noted earlier, previous research has found that interest in the study tends to be related to the probability of a panel member dropping out. Interest in the study, when unmeasured, is likely to be a function of demographic/socioeconomic variables and other exogenous factors. Therefore, the attrition model is assumed to be composed of the reduced-form variables from the system described by equation (11). That is, by some matrix algebra, equation (11) can be represented by

$$y' = -B^{-1}\Gamma'x' + B^{-1}\mu \quad (12)$$

$$= \pi x' + v.$$

In addition, there is empirical evidence that the four groups comprising the panel differ with respect to means of endogenous variables some slopes of exogenous variables (Farley, Katz, Lehmann, and Winer 1980), and attrition rates (Table 1). To partially account for this, group dummy variables to reflect possible mean differences in attrition probabilities were added. The attrition model thus becomes

$$D = \sum_{i=0}^{12} \pi_i x_i + \sum_{i=1}^3 \alpha_i G_i + v. \quad (13)$$

where D is the 0-1 index of being in or out of the sample for a given wave, the independent variables are from Table 2, and the G_i are 0-1 dummies representing three of the four groups.

Empirical Results

Attrition Model

The results of the probit estimation of the attrition model (13) for waves 2, 3, and 4 are in Table 3. Since sex, age, income, education, number of drivers, and child are measured only once in the first wave, those variables must be repeated for each wave. However, intention through propensity (Table 3) are measured in each wave.

TABLE 3
Probit Results from Attrition

Variable	Wave 2	Wave 3	Wave 4
Constant	.7637**	.2055	.0332
Intention	.0316	.0088	.0185
Attitude	-.0040	.0043	-.0110
Confidence	-.0091	-.0195	-.0234*
Comprehension	-.0202	-.0094	.0163
Recall	-.0169	-.0193	-.0222
Propensity	-.0066	-.0016	-.0072
Sex	.0032	-.0257	-.1074
Age	.0179	.0265	-.0358**
Income	.0357	.0491	.0610**
Education	-.0035	.0043	.0009
# Drivers	-.0455	-.0230	-.0114
Child	-.2105*	-.0525	-.0455
Group 1	-.0049	.1126	.1550
Group 2	.0314	-.0608	-.1049
Group 3	-.0049	.0860	.0511
# dropouts	299	416	528
# still on panel	938	821	709
-2 log likelihood ratio	17.27	18.95	33.42
sig. log. likelihood (15d.f.)	$p < .30$	$p < .20$	$p < .05$

* $p < .10$ ** $p < .05$ two-tail tests

Since panel dropouts obviously are not measured on these variables in the wave they attrite and on, a decision had to be made on the appropriate values to use. Therefore, for all panel members, the wave one values were employed. The only difference between the data matrices for the wave 2, 3, and 4 attrition models was thus in the 0-1 dependent variable.

As can be seen from Table 3, the reduced form variables are only weakly related to the probability of attrition for waves 2 and 3. However, by wave 4, more people had dropped out providing increased information on which to base the

relative effects of the exogenous variables. The results indicate that age is negatively related to staying on the panel and income has positive influences. These results are consistent with the findings in the literature reported earlier. A weaker effect was that people with higher confidence in their ability to judge the Vega tended to drop out. Perhaps, this reflects a feeling that there was little additional knowledge to be gained from being on the panel which translates to a low interest factor.

Structural Model

As described earlier, the probit results are used to create a new independent variable to be added as an explanatory variable to the equations (11). The implications of panel attrition can be found in Tables 4-6 by comparing the results with and without the attrition bias correction factor. OLS was used to estimate each equation as the model is assumed to be recursive with independent inter-equation errors. Tables 4-6 may be found in the appendix at the conclusion of this paper.

In this instance, attrition appears to have a modest impact on the structural equation estimates. The correction factor is significant in three cases--the Confidence equation in wave 2, the Confidence equation in wave 3, and the wave 4 Attitude equation. In all three cases, the impact of attrition is selective; not all variable coefficients are affected. In fact, the effects seem to be centered on the pure exogenous variables rather than the endogenous variables that are exogenous. In the Confidence-wave 3 equation, the greatest impact is on the age variable as it doubles in size and becomes significant. In the Attitude-wave 4 equation, the income variable assumes a much larger value and becomes significant. If the model was being used for segmentation purposes, serious misallocations of resources or missed opportunities could result.

Discussion

It is difficult to determine when attrition should be modeled as part of a structural system. Hausman and Wise (1979) provide empirical evidence supporting the notion that bias is more likely to be present when the structural model is misspecified. This implies that in early stages of model development or when simplicity is sought at the expense of completeness, attrition bias can seriously distort results. The model utilized in this paper has been thoroughly tested and found to be robust to minor perturbations in specification. Therefore, it is not surprising that correcting for attrition did not have a dramatic impact on the empirical results. Even with this relatively well-specified model, however, the impact of panel attrition can be seen.

Several other issues can be raised. First, if a wave-by-wave analysis is not desired but the analyst wishes to pool, only panel members completing all waves can be used. Therefore, the implication is that only one attrition model need be estimated--the one corresponding to the wave four analysis of Table 3--since interest is focused on predicting the probability of attrition by the final wave. Second, if attrition and missing data occur, both problems could be modeled and two terms inserted in each structural equation. There may be difficulty, however, in separately identifying missing data and attrition functions.

Conclusion

This paper has developed and estimated a model that accounts for attrition bias in panel data. It was found that the main impact of attrition was on exogenous variables and not endogenous variables that were exogenous in the equations. Marketing studies employing panels should be aware that attrition can affect model parameter estimates, particularly

if the model structure is not well-known.

Appendix

TABLE 4

Structural Equation Estimates: Wave 2

Independent Variables	Dependent Variable							
	Intention		Attitude		Confidence		Comprehension	
	without correction	with correction	without correction	with correction	without correction	with correction	without correction	with correction
Intention	--	--	--	--	--	--	--	--
Attitude	.0680**	.0687**	--	--	--	--	--	--
Confidence	.0758**	.0783**	--	--	--	--	--	--
Comprehension	--	--	.4643**	.4629**	.1812**	.1805**	--	--
Intention (t-1)	.2067**	.1970	--	--	--	--	--	--
Attitude (t-1)	--	--	.1962**	.1935**	--	--	--	--
Confidence (t-1)	--	--	--	--	.1304**	.1161**	--	--
Comprehension (t-1)	--	--	--	--	--	--	.1358**	.1480**
Recall	.0136	.0156	--	--	.3687**	.3668**	.1038**	.1045**
Propensity	.1357**	.1362**	.1761**	.1763**	.0266	.0270	-.0060	-.0062
Sex	--	--	.3802**	.3717**	.9600**	.9363**	--	--
Age	--	--	-.0582**	-.0490	-.0522	-.0121	-.0245*	-.0354**
Income	-.0504	-.0658	-.1288**	-.1299**	--	--	--	--
Education	--	--	-.0355	-.0342	.0226	.0315	.0267**	.0244**
# Drivers	--	--	--	--	--	--	-.0014	.0191
Child	-.0611	.0963	--	--	--	--	--	--
Correction factor	--	-1.1112	--	.5712	--	2.4041*	--	-.7186
R ²	.172	.172	.144	.144	.281	.283	.163	.163

*p<.10 **p<.05 One-tail tests on endogenous variables, two-tail otherwise.

TABLE 5

Structural Equation Estimates: Wave 3

Independent Variables	Dependent Variable							
	Intention		Attitude		Confidence		Comprehension	
	without correction	with correction	without correction	with correction	without correction	with correction	without correction	with correction
Intention	--	--	--	--	--	--	--	--
Attitude	.1642**	.1643**	--	--	--	--	--	--
Confidence	.0350*	.0395**	--	--	--	--	--	--
Comprehension	--	--	.2281**	.2274**	.0197	.0286	--	--
Intention (t-1)	.1279**	.1290**	--	--	--	--	--	--
Attitude (t-1)	--	--	.3883**	.3855**	--	--	--	--
Confidence (t-1)	--	--	--	--	.3517**	.3423**	--	--
Comprehension (t-1)	--	--	--	--	--	--	.3793**	.3793**
Recall	.0345*	.0358*	--	--	.2848**	.2803**	.0828**	.0829**
Propensity	.1081**	.1096**	.0937**	.0918**	.0517**	.0472*	-.0085	-.0084
Sex	--	--	-.1401	-.1827	.5464**	.4718**	--	--
Age	--	--	-.0234	-.0019	.0543	.1018**	.0003	-.0003
Income	-.0428	-.0784	.0022	.0355	--	--	--	--
Education	--	--	-.0216	-.0171	-.0366	-.0167	.0414**	.0412**
# Drivers	--	--	--	--	--	--	.0204	.0207
Child	.4473**	.5585**	--	--	--	--	--	--
Correction factor	--	-1.2403	--	1.2822	--	2.6759**	--	-.0359
R ²	.178	.179	.215	.215	.340	.343	.220	.219

*p<.10 **p<.05 One-tail tests on endogenous variables, two-tail otherwise

TABLE 6

Structural Equation Estimates: Wave 4

Independent Variables

Dependent Variables

	Intention		Attitude		Confidence		Comprehension	
	without correction	with correction	without correction	with correction	without correction	with correction	without correction	with correction
Intention	--	--	--	--	--	--	--	--
Attitude	.0893**	.0917	--	--	--	--	--	--
Confidence	.0265	.0303	--	--	--	--	--	--
Comprehension	--	--	.2567**	.2625**	.1711**	.1760**	--	--
Intention (t-1)	.2095**	.2096**	--	--	--	--	--	--
Attitude (t-1)	--	--	.3662**	.3579**	--	--	--	--
Confidence (t-1)	--	--	--	--	.3392**	.3364**	--	--
Comprehension (t-1)	--	--	--	--	--	--	.3671**	.3678**
Recall	.0422**	.0439**	--	--	.2121**	.2105**	.0625**	.0635**
Propensity	.1195**	.1208**	.1177**	.1098**	.0833**	.0816**	.0149	.0155
Sex	--	--	.0932	-.1730	.6597**	.6451**	--	--
Age	--	--	-.0343	.0433	.0017	.0192	-.0024	-.0081
Income	-.0235	-.0588	.0110	.1430*	--	--	--	--
Education	--	--	-.0381	-.0291	.0012	.0067	.0188	.0173
# Drivers	--	--	--	--	--	--	.0271	.0269
Child	-.2554	-.1753	--	--	--	--	--	--
Correction factor	--	-.8208	--	3.3094**	--	.6800	--	-.2335
\bar{R}^2	.153	.153	.209	.222	.319	.318	.209	.209

* $p < .10$ ** $p < .05$ One-tail tests on endogenous variables, two-tail otherwise.

References

- Bucklin, Louis P. and Carmen, James M. (1967), The Design of Consumer Research Panels: Conception and Administration of the Berkeley Food Panel, Berkeley, California: Institute of Business and Economic Research, University of California.
- Carman, James M. (1974), "Consumer Panels," in Robert Ferber ed., Handbook of Marketing Research, New York: McGraw Hill.
- Charlton, P. and Ehrenberg, A.S.C. (1976), "An Experiment in Brand Choice," Journal of Marketing Research, 13, 152-160.
- Farley, John U., Howard, John A. and Lehmann, Donald R. (1976), "A 'Working' System-Model of Car Buyer Behavior," Management Science, 23, 235-247.
- Farley, John U., Katz, Jerrold P., Lehmann, Donald R., and Russell S. Winer (1980), "Parameter Stationarity in a Consumer Decision Model," working paper, Columbia University.
- Farley, John U. and Ring, L. Winston (1970), "An Empirical Test of the Howard-Sheth Model of Buyer Behavior," Journal of Marketing Research, 4, 427-438.
- Ferber, Robert (1966), The Reliability of Consumer Reports of Financial Assets and Debts, Urbana Illinois: Bureau of Economic and Business Research, University of Illinois.
- Finney, D.J. (1964), Probit Analysis, 2nd ed., Cambridge, U.K.: Cambridge University Press
- Frank, Ronald E., Massy, William F. and Boyd, Harper W. (1967), "Correlates of Grocery Product Consumption Rates," Journal of Marketing Research, 4, 184-190.
- Futrell, Charles M. and Jenkins, Omer C. (1978), "Pay Secrecy Versus Pay Disclosure for Salesmen: A Longitudinal Study," Journal of Marketing Research, 15, 214-219.
- Grilliches, Zvi, Hall, Bronwyn H. and Hausman, Jerry A. (1978), "Missing Data and Self-Selection in Large Panels," Annales de l'insee, 30-31, 137-176.
- Hausman, Jerry A. and Wise, David A. (1976), "The Evaluation of Results from Truncated Samples: The New Jersey Income Maintenance Experiment," Annals of Economic and Social Measurement, 5, 421-445.
- Hausman, Jerry A. and Wise, David A. (1977), "Social Experimentation, Truncated Distributions, and Efficient Estimation," Econometrica, 45, 919-938.
- Hausman, Jerry A. and Wise, David A. (1979), "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," Econometrica, 47, 455-473.
- Heckman, James J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Social and Economic Measurement, 5, 475-492.
- Heckman, James J. (1979), "Sample Selection Bias as a Specification Error," Econometrica, 47, 153-161.
- Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: John Wiley.
- Johnson, Norman and Kotz, Samuel (1970), Continuous Univariate Distributions - 1 New York: John Wiley.
- Johnson, Norman and Kotz, Samuel (1972), Continuous Multivariate Distributions, New York: John Wiley.
- McCann, John M. (1974), "Market Segment Response to the Marketing Decision Variables," Journal of Marketing Research 11, 399-412.
- Prasad, V. Kanti and Ring, L. Winston (1976), "Measuring Sales Effects of Some Marketing Mix Variables and their Interactions," Journal of Marketing Research, 13, 391-396.

Sobel, M.G. (1959) "Panel Mortality and Panel Bias," Journal of the American Statistical Association, 54, 52-68.

Winer, Russell S. (1980), "Estimation of a Longitudinal Model to Decompose the Effects of an Advertising Stimulus on Family Consumption Behavior," Management Science, 26 (in press).

SCIENTIFIC ADVANCEMENT IN CONSUMER RESEARCH: SOME PROBLEMS
ENCOUNTERED WHEN USING CONSUMER PANEL DATA¹

R. Dale Wilson, Batten, Barton, Durstine & Osborn, Inc.²

Abstract

Numerous studies have been conducted over the past twenty years which utilize diary panel data in testing particular components of consumer behavior theory. This paper addresses the use of such data and provides empirical illustrations of some of the important problems which may improperly alter the development of theoretical advances in consumer behavior.

Introduction

The widespread availability of diary panel data for use in consumer research has presented many opportunities for the creative researcher. During the past twenty years, studies in consumer research have used panel data to track consumer brand choice decisions, to investigate shopping behavior strategies, and to draw inferences about consumer response to marketing variables (such as short-term promotional programs and price differentials among national and private-label brands) among others.

Unfortunately, however, significant problems are likely to be encountered when diary panel data are used for these purposes. Some, primarily those dealing with traditional measurement error, are well documented. Other problems, however, are largely ignored in the consumer behavior literature. This latter class of difficulties include operational problems such as multiple purchases, time duration between purchases, quantity (i.e., size of package) purchased, and purchase of multiple brands at one purchase occasion.

While some researchers dismiss these problems casually, their influence on empirical findings (and thus consumer behavior theory) may be substantial. The purpose of this paper is to (1) address these problems, (2) document some of these problems with empirical data, and (3) suggest the limitations of panel data in consumer research from an advancement of science perspective. As such, the conflict between requiring near-perfect data for empirical work and accepting "less-than-perfect" panel data for testing theoretical constructs is illustrated.

Perspectives from Philosophy of Science

There seem to be several concepts from philosophy of science which may be useful in understanding the problems associated with many studies using consumer diary panel data. This section of the paper provides a brief discussion of these concepts--first as they relate to construct measurement in general and then as they relate to construct measurement when using panel data.

Empirical Indicators and Their Validity

The importance of measurement in consumer behavior research cannot be denied. Yet, at the same time, many consumer behavior studies are conducted without proper

attention to validation of the construct being measured. Dubin (1978, p. 182) uses the term "empirical indicator" to label the act of deciding the type of operation a researcher employs to secure a measurement of value on the unit under investigation. Empirical indicators simply serve as an observable representation of the theoretical construct of interest. The question remains, however, whether a particular empirical indicator is a valid representation of the theoretical construct.

Dubin (1978) addresses the issue of validity by pointing out that the term "validity" refers only to the consensus or lack thereof that a particular empirical indicator measures values on a stated unit. He continues by stating that:

"This consensus is a man-made consensus and is nothing more than a conventional agreement among a group of interested students and spectators that the empirical indicator and theoretical unit whose values it measures are homologous. We may therefore expect that what is a valid measure at some time may lose this status if the consensus upon which it is based is supplanted."

"The breakdown of a consensus usually occurs when an investigator raises questions about the empirical indicator based upon evidence that is independent of the circumstances of its employment" (p. 200).

Since construct validation is concerned with (1) what the measurement of a variable is in fact measuring and, more importantly, (2) the deductions that are being made about the theory underlying the measured variable (Churchill 1979 p. 258-9; Green and Tull 1978, p. 198-9), Dubin's emphasis on a "consensus" approach to measurement cannot be slighted. Another point worth mentioning is that the potentially large degree of measurement error in the behavioral sciences causes the issue of the validity of empirical indicators to be especially important. In the natural sciences, where there is a much closer correspondence between reality and appearance, validity is of little concern (Dubin 1978, p. 202).

Since the study of consumer behavior is a social science, the importance of measurement in our discipline cannot be denied. Yet, in many cases, we behave as if measurement issues are beyond the realm of concern. For example, Jacoby and Chestnut (1978) identify a total of fifty-three separate measures of "brand loyalty." Many of these measures, when evaluated critically, probably have little relationship to the theoretical construct of brand loyalty. Other examples from the consumer behavior literature are also available which point to the real lack of concern for construct validity. Likely candidates include many attempts to measure such variables as deal proneness, advertising effectiveness, innovative behavior, perceived risk, information processing, values, personality, brand or store preference, etc. The point here is not to criticize those researchers who have attempted to measure illusive constructs. Rather, it is to draw attention to the special problems of construct validity when measuring variables that fall outside the realm of what might be termed

¹This paper was supported by a Faculty Projects Grant from the Center for Research, College of Business Administration, The Pennsylvania State University, University Park, Pennsylvania. The author gratefully acknowledges the assistance of Manoj Hastak and Larry Newman (doctoral candidates at Penn State) in the preparation of this paper.

²The author is Director of Marketing Sciences, BBDO, Inc., 383 Madison Avenue, New York, NY 10017.

"absolute indicators."³

One insightful way to view the problems encountered when measurements are taken on a theoretical construct is to view the measured response as the sum of the true response plus possible measurement errors (see Lehmann 1979, p.106). Even more exasperating is the realization that when a statistical analysis is performed on the measured concept, the sources of potential error are many. This problem can be viewed as:

$$\text{Statistical test of a measured concept} = \text{True value} \\ + \text{Construct validation error} + \text{Measurement error} \\ + \text{Statistical error}$$

where:

"True value" represents the actual, yet unobservable, response which one is attempting to measure;

"Construct validation error" represents the unobservable difference between the true value of the response and the measurement of that response (disregarding measurement error);

"Measurement error" represents such partially unobservable factors as measurement process error, instrumentation error, and respondent error (see Lehmann 1979, p. 103-6); and

"Statistical error" represents error in the definition and selection of the sample as well as the error in the execution of the sampling plan. Statistical error also tends to be unobservable.

While many researchers have actively sought methods for limiting or controlling measurement error and statistical error, far less attention has been given to the problem of construct validation error in consumer behavior.

Construct Measurement Via Panel Data

For the purpose of this paper, discussion is limited only to measurement issues in continuous, diary panels which allow respondents to use self-administered questionnaires to record prespecified information on a regular basis (see Tull and Hawkins 1976, p. 397-401 for a concise analysis of the various types of panels, their characteristics, and their uses). Sudman and Ferber (1979), in the most comprehensive description of consumer panels and resulting data that is available in the literature, provide an analysis of the types of consumer and market research studies conducted with panel data.

It is important to note that discussions of consumer diary panels by and large ignore measurement problems associated with construct validity. Sudman and Ferber (1979) discuss measurement problems in consumer panels as including only sample representativeness, data accuracy, and panel conditioning. Similarly, others, such as Morrison, Frank, and Massy (1966) and Buck et al. (1977) mention the limitations of panel approaches to data collection but do not delve deeper into construct validity questions. Powers, Goudy, and Keith (1978), in addition to discussing the traditional limitations, compared panel data with recall data and found

³"Absolute indicators" are those that are ". . . absolute in the sense that there can be no question as to what they measure" (Dubin 1978, p. 193). Examples, according to Dubin, include all demographic characteristics. One may argue, however, that because measurement error can never be totally eliminated, the definition of absolute indicators should include some relative notion of the degree to which there can be no question as to what they measure.

major inconsistencies. In an article which discusses methods that may ultimately be useful in shedding light on validity issues, McCullough (1978) reviews four methods which are appropriate for determining causal effects in panel data studies. These methods include Lazarfeld's 16-fold table, Coleman's four-state continuous-time Markov processes, cross-legged panel correlation, and path analysis.

Thus, the literature indicates that for the most part, there is little concern for construct measurement. But as indicated by McCullough's (1978) review, it may be possible to design studies to determine cause and effect relationships and clarify measurement issues by testing alternative definitions of constructs where the causal relationships are clear. If such cause and effect relationships are thought to be known with a high degree of certainty, then several definitions of a construct could be substituted for the purpose of selecting the best definition.

Although it is not entirely clear from reading the literature on the collection and analysis of panel data, the major factor which adversely affects construct validity is the fact that researchers rarely have the opportunity to measure constructs in the way in which they desire. Thus, with rare exception in the published literature, secondary data is being used. Another factor, nearly as important as lack of control in data collection, is the tendency toward using self-reported behavioral data to attempt to measure variables that may consist either fully or partially of attitudinal or cognitive components. Jacoby and Chestnut's (1978) criticism of much of the existing brand loyalty literature is a case in point. These two problems, one associated with the inability to obtain empirical indicators in exactly the way one wishes and the other associated with the necessity of using strictly behavioral data to measure attitudinally- or cognitively-oriented constructs, causes users of consumer panel data to be less than perfect in their data requirements. Unfortunately, however, these problems probably have had an adverse effect on the growth and development of consumer behavior theory.

Empirical Illustrations

The question, "Are we really measuring what we think we're measuring?", may never be answered with any degree of certainty for many constructs. However, it may be instructive to point out how the quality of measurement is affected by what may be "oversimplistic" empirical indicators. It is hoped that the few empirical illustrations included here will stimulate more critical thinking on the hazards of using panel data in cases where the quality of the empirical indicator is in doubt.

Illustration I--Measuring Brand Purchase and the Problem of Multiple Units

A priori, one would expect that the measurement of actual brand choice would present no problem to the research given the ease with which purchase behavior could be pulled from the diary records. Unfortunately, such simplicity does not exist due to the problem of multiple purchases at a single point in time. Because of convenience, the availability of deals, or for other reasons, many consumers often buy multiple units (or even multiple brands) on a given shopping trip. For example, Wilson, Newman, and Hastak (1979) present panel data collected over a 24-month period which indicates that for the bar soap category, 68.5% of all purchases consisted of two or more units. In addition, 31.8% of the purchases consisted of three or more units and 22.5% consisted of four or more units. The most regrettable aspect of this situation is that most researchers have not revealed their methods for handling multiple-unit purchases. The available literature implies that this problem is usually handled in one of two ways--either through the truncation of all units purchased after the first (or a randomly-selected one) or the consideration of each unit

as independent purchase occasions (Wilson, Newman, and Hastak 1979). Regardless of the option chosen by a particular researcher, it is clear that the difference between the true value of brand purchase and the measured value may be great.

Illustration II--Measuring Deal Purchases

Although the trend toward purchasing in multiple quantities is interesting, there may be more to the data than is originally apparent. It stands to reason that one may be interested in determining why such behavior took place and, therefore, would want to determine how many purchases occurred in conjunction with dealing activity. In these cases, data on multiple units purchased would also contribute to an understanding of the effects of a particular type of deal. Again, however, the number of units purchased is rarely (if ever) considered in studies of consumer dealing activity even though it provides a possible explanation for the true impact of such deals.

Table 1

Type of Deal Situation	Number of Units Purchased Per Occasion					
	1	2	3	4	5 or more	Total
No Deal	10,535 (39.9%) ^a	9,315 (35.3%)	1,822 (6.9%)	3,692 (14.0%)	1,032 (3.9%)	26,396
Media-Distributed Coupons	341 (14.4%)	1,184 (50.0%)	415 (17.5%)	339 (14.3%)	90 (3.8%)	2,369
Package Coupons	254 (17.8%)	780 (54.6%)	203 (14.2%)	156 (10.9%)	35 (2.5%)	1,428
Cents-Off Marked	1,107 (14.9%)	2,702 (36.4%)	946 (12.7%)	1,656 (22.3%)	1,022 (13.7%)	7,433
Multiple Deals and Other Types	214 (9.3%)	705 (30.8%)	341 (14.9%)	798 (34.9%)	231 (10.1%)	2,289
Total	12,451 (31.2%)	14,686 (36.8%)	3,727 (9.3%)	6,641 (16.6%)	2,410 (6.0%)	39,915

Chi square with No Deal condition included = 4,527.14 with degrees of freedom, significant at $p < .0001$.

Chi square with No Deal condition excluded = 883.25 with 12 degrees of freedom, significant at $p < .0001$.

^aThe numbers in parentheses are row percentages.

Table 1 is designed to help clarify how the type of deal relates to multiple purchases. As can be seen from the table, the row percentages point to several differences in the number of units purchased across the various types of deals. Further, these trends tend to be accentuated as the number of units purchased increases. For example, four units are purchased in 22.3% of the cents-off marked coupon deals but in only 10.9% of the package coupons. The large chi-square statistic indicates that a strong relationship exists between the deal type and the number of units purchased. Thus, if a consumer researcher attempted to measure the impact of various types of deal situations without considering the purchase of multiple

units, major errors could be made due to the oversimplification of the purchase measure.

Illustration III--Types of Deals Purchased Over Time

One of the problems often associated with studies involving consumer panels is that interesting trends in the data are camouflaged by the level of aggregation of that data. Because of the high costs of processing panel data, researchers sometimes tend to draw conclusions about trends in the data without a detailed analysis. In working with panel data, it is abundantly clear that "how you cut the data" will have a large influence on conclusions that are drawn. In a philosophy of science sense, this problem is especially important whenever cause and effect relationships are being specified.

As an illustration of this problem, first consider the data in Table 2. This table presents data in the bar soap category for the total market in 1976 and 1977. Specific data are also presented for five brands, each having a relatively large market share among panel respondents over the two-year period. Table 2 points out clearly the temporal effects of sales. At first glance, the total market for the category seems to be declining over time since total units purchased declined from 4,976 to 2,959. Further,

Table 2

TOTAL UNITS PURCHASED BY BRAND							
Yr.	Qtr.	Total Units Purchased					
		All Brands	Brand A	Brand B	Brand C	Brand D	Brand E
1976	I	4,095 (10.2%) ^a	751 (18.3%)	442 (11.6%)	361 (10.2%)	235 (8.1%)	44 (1.4%)
	II	4,976 (12.3%)	974 (23.8%)	494 (13.0%)	451 (12.8%)	335 (11.6%)	59 (1.9%)
	III	5,805 (14.4%)	938 (22.9%)	612 (16.1%)	488 (13.8%)	445 (15.4%)	71 (2.2%)
	IV	6,990 (17.3%)	823 (20.1%)	620 (16.3%)	658 (18.7%)	448 (16.9%)	258 (8.1%)
1977	I	5,836 (14.5%)	262 (6.4%)	543 (14.3%)	556 (15.8%)	417 (14.5%)	834 (26.3%)
	II	5,009 (12.4%)	149 (3.6%)	424 (11.1%)	416 (11.8%)	360 (12.5%)	790 (24.9%)
	III	4,627 (11.5%)	126 (3.1%)	412 (10.8%)	353 (10.0%)	367 (12.7%)	673 (21.2%)
	IV	2,959 (7.37%)	75 (1.8%)	262 (6.9%)	245 (6.9%)	237 (8.2%)	444 (14.0%)
Totals	40,297	4,098	3,809	3,528	2,884	3,173	
Total units purchased at regular price	26,396	1,967	2,817	2,470	2,300	1,554	
Percentage of units purchased at regular price	66.1	48.0	74.0	70.0	79.8	49.0	

^aThe numbers in parentheses are column percentages.

this trend is not explained when specific brand data (at least for Brands A through E) are analyzed. One interesting feature of the inter-brand relationship is that the units purchased figures for Brand E show dramatic increases in time. Like all of the brands displayed in Table 2, however, Brand E sales seem to be declining quickly after a peak of 834 units in the first quarter of 1977.

The key to understanding the bar soap data, however, may be in the last two rows of Table 3, which represent the total number of units and the percentage of units purchased at regular price. Across all brands, 13,901 (or 34.5%) of the 40,297 units purchased were bought on deal. For some brands, such as Brands A and E, over 50% of the total units were purchased at a deal price. Thus, the simple inclusion of dealing activity may inspire us to look further into the data for explanations of the cause-effect relationships.

Table 3, 4, and 5 present the purchase data for Brands A, B, and C, respectively. These tables are designed to more clearly establish the relationship between dealing activity and units purchased across time. Except for the Brand A

Table 3
PURCHASES BY TYPE OF DEAL AND TIME PERIOD
FOR BRAND A

Type of Deal	Year	Quarter				Total
		I (Jan-March)	II (April-June)	III (July-Sept.)	IV (Oct-Dec.)	
No Deal	1976	384	463	449	399	1,695
	1977	110	66	58	38	272
Media-Distri-Buted Coupons	1976	76	115	59	45	295
	1977	22	13	10	4	49
Package Coupons	1976	24	53	35	29	141
	1977	16	10	7	4	37
Cents-Off Coupon	1976	208	279	291	240	1,018
	1977	86	47	40	15	188
Multiple Deals and Other Types	1976	59	64	104	110	337
	1977	28	13	11	14	66
Total for 1976		751	974	938	823	3,486
Total for 1977		262	149	126	75	612
Total for 1976 and 1977		1,013	1,123	1,064	898	4,098

Chi square for 1976 data = 64.97 with 12 degrees of freedom; significant at $p < .0001$.

Chi square for 1977 data = 10.78 with 12 degrees of freedom; not significant at $p < .10$.

Chi square for combined 1976 and 1977 data = 83.61 with 28 degrees of freedom; significant at $p < .0001$.

Table 4
PURCHASES BY TYPE OF DEAL AND TIME PERIOD
FOR BRAND B

Type of Deal	Year	Quarter				Total
		I (Jan-March)	II (April-June)	III (July-Sept.)	IV (Oct-Dec.)	
No Deal	1976	302	373	397	491	1,563
	1977	419	359	281	195	1,254
Media-Distri-Buted Coupons	1976	20	13	15	13	61
	1977	17	11	10	5	43
Package Coupons	1976	3	6	5	4	18
	1977	4	6	6	1	17
Cents-Off Coupons	1976	106	74	147	92	419
	1977	92	43	87	53	275
Multiple Deals and Other Types	1976	11	28	48	20	107
	1977	11	5	28	8	52
Total for 1976		442	494	612	620	2,168
Total for 1977		543	424	412	262	1,641
Total for 1976 and 1977		985	918	1,024	882	3,809

Chi square for 1976 data = 61.04 with 12 degrees of freedom; significant at $p < .0001$.

Chi square for 1977 data = 53.84 with 12 degrees of freedom; significant at $p < .0001$.

Chi square for combined 1976 and 1977 data = 126.25 with 28 degrees of freedom; significant at $p < .0001$.

data in 1977, the chi-square statistics presented in the tables indicate that the hypothesis of independence between dealing activity and time period must be rejected. Since there does seem to be an identifiable relationship here, it seems appropriate to suggest further study of the cause and effect relationships between units purchased and dealing activity. Particularly for Brand A, it appears that the lack of deal availability during calendar year 1977 may have caused, or at least accentuated, the decline in units purchased among panel members.

Illustration IV--Measuring Store Choice

Table 6 presents shopping and purchase expenditure data for a panel of 719 respondents who resided in the Chicago area. These data were collected over a four-week period in 1970 by a national consumer panel organization. The interest in these data lie in the fact that the determination of market share is quite difficult. For Store A, a large regional supermarket chain based in Chicago, the data indicate a monotonically increasing relationship between the dollar amount category and the percentage of trips made to Store A. If one considers "market share" to be equal to the number of shopping trips made to Store A divided by the total number of trips made to all stores, then the ques-

Table 5

PURCHASES BY TYPE OF DEAL AND TIME PERIOD
FOR BRAND C

Type of Deal	Year	Quarter				Total
		I (Jan.- March)	II (April- June)	III (July- Sept.)	IV (Oct.- Dec.)	
No Deal	1976	292	311	366	436	1,405
	1977	329	325	239	172	1,065
Media-Distri- buted Coupons	1976	30	33	23	64	150
	1977	95	25	11	12	143
Package Coupons	1976	13	16	14	50	93
	1977	45	18	10	12	85
Cents- Off Coupons	1976	17	40	52	57	166
	1977	40	22	68	28	158
Multiple Deals and Other Types	1976	9	51	33	51	144
	1977	47	26	25	21	119
Total for 1976		361	451	488	658	1,958
Total for 1977		556	416	353	245	1,570
Total for 1976 and 1977		917	867	841	903	3,528
Chi square for 1976 data = 65.34 with 12 degrees of freedom; significant at $p < .0001$.						
Chi square for 1977 data = 133.60 with 12 degrees of freedom; significant at $p < .0001$.						
Chi square for combined 1976 and 1977 data = 214.52 with 28 degrees of freedom; significant at $p < .0001$.						

tion becomes, "What is the relevant measure of store shopping trips?"

The data in Table 6 indicate that when all purchases are considered, Store A's market share is 25.2%. However, a more suitable measure of Store A's actual impact on the marketplace may be obtained by considering only those trips in which a particular minimum amount was spent. For example, if the relevant cutoff is at an expenditure level of \$5.00 the market share figure increases to 33.9%.

The point of this illustration is that a great deal of thinking must be done in order to establish the point at which minor, "filler" shopping trips become larger, "full-scale" trips. Using the data presented in Table 6 as justification, Wilson (1977) used the \$5.00 point as the relevant cutoff point. Similarly, Frisbie (1980) chose the \$5.00 point as one of his three criteria for establishing filler trips. In these cases, there are no firm theoretical reasons for defining filler trips in a particular way although the concept does have some support in the literature (e.g., MacKay 1973). However, as Table 6 illustrates, this problem is meaningful if a clear picture of consumer shopping behavior is to emerge from the panel data.

Discussion

One of the main points that, hopefully, will come from this paper is that because of measurement problems, the analysis of consumer panel data is no simple task. Although this point has been illustrated with only one example (i.e., the problem of multiple purchases), many other far more difficult definitional problems come to mind. For example, the early work of Kuehn and Rohloff (1967) indicates consumers surprisingly maintain a high degree of "size loyalty" and, in fact, their findings include evidence that loyalty to package size is occasionally more typical than brand loyalty. Yet, it is not surprising that researchers have avoided using panel data to explore size loyalty since variations in within-brand and across-brand package sizes would be virtually impossible to control. Not only would one have to consider the package size, but other factors such as the specific brand, deal availability, number of units purchased, and perhaps consumers' usage rate would become relevant to the analysis. Control procedures for such a study would have to be massive.

Another point to be made is that variable measurement is of utmost importance. However, the use of convergent validity, which requires the measurement of multiple dependent variables in panel studies, should be useful in limiting the extent of this problem (see Dubin 1978, p. 195-200; Jacoby 1978).

Lastly, another problem of no small significance is the fact that the huge quantities of data available on the typical magnetic tape of panel records sometimes hides relevant data. Wiggins (1973) addresses this problem when he discusses one of the three main uses of panel data--to study change on the behavior of individuals, as in studies of buying behavior. He states that:

"[this type of application]...has great significance in its own right, and its own variety of unique problems. A tremendous amount of potentially valuable findings may have been ignored by trying to eliminate what for the purpose at hand was garbage (or merely irrelevant to the problem), even though a detailed examination ...may have yielded other kinds of information of genuine value....The problem there has generally been a shortage of funds and possibly a lack of sophistication in the analysis in some instances" (p. 188).

Also related to problem discussed by Wiggins is the fact that extremely large sample sizes in most panels (see Sudman and Ferber (1979, p. 9-11) may mislead the researcher by causing immaterial deviations to be statistically significant. More care needs to be taken to avoid the problem of "failing to see the forest for the trees."

Conclusion

This paper has attempted to point out a few of the many pitfalls of using diary panel data in consumer research. More than anything else, the paper was designed to provide cautionary comments to current and future data users. While it is clear that the availability of panel data presents many opportunities for meaningful, creative research, panel data can also lead to unthinking data massaging that yields non-productive or even counter-productive conclusions. The goal of advancing the discipline of consumer research dictates that panel researchers engage in more than rote number crunching with high-powered hardware.

Table 6
SHOPPING AND PURCHASE EXPENDITURE DATA^a

Dollar Amount Category	Number of Trips	Mean Dollars Spent Per Trip	Standard Deviation	Number of Trips to Store A	Percentage of Trips to Store A
All Purchases	10,767	8.80	11.42	2,715	25.2%
All Purchases ≥ 1.00	9,419	9.97	11.76	2,548	27.1%
All Purchases ≥ 2.00	7,716	11.87	12.22	2,270	29.4%
All Purchases ≥ 2.50	7,045	12.76	12.41	2,157	30.6%
All Purchases ≥ 3.00	6,493	13.61	12.56	2,037	31.4%
All Purchases ≥ 4.00	5,602	15.23	12.80	1,845	32.9%
All Purchases ≥ 5.00	4,950	16.64	12.97	1,679	33.9%
All Purchases ≥ 10.00	3,029	22.68	13.41	1,151	38.0%
All Purchases ≥ 15.00	2,048	27.69	13.69	854	41.7%
All Purchases ≥ 20.00	1,443	32.05	14.18	647	44.8%
All Purchases ≥ 25.00	979	36.63	15.17	440	44.9%
All Purchases ≥ 30.00	637	41.59	16.80	292	45.8%
All Purchases ≥ 35.00	400	47.08	19.16	182	45.5%
All Purchases ≥ 40.00	241	53.70	22.32	115	47.7%

^aSource: adapted from Wilson (1977, p. 59)

References

- Buck, S. F., Fairclough, E. H., Jephcott, J. St. G., and Ringer, D. W. C. (1977), "Conditioning and Bias in Consumer Panels--Some New Results," Journal of the Market Research Society, 19, 59-75.
- Chrchill, Gilbert A., Jr. (1979), Marketing Research: Methodological Foundations, Second edition, Hinsdale, Illinois: Dryden Press.
- Dubin, Robert (1978), Theory Building, Revised edition, New York: Free Press.
- Frisbie, Gil A., Jr. (1980), "Ehrenberg's Negative Binomial Model Applied to Grocery Store Trips," Journal of Marketing Research, 17, 385-90.
- Green, Paul E. and Tull, Donald S. (1978), Research for Marketing Decisions, Fourth edition, Englewood Cliffs, New Jersey: Prentice-Hall.
- Jacoby, Jacob (1978), "Consumer Research: A State of the Art Review," Journal of Marketing, 42, 87-96.
- _____ and Chestnut, Robert W. (1978), Brand Loyalty: Measurement and Management, New York: John Wiley and Sons.
- Kuehn, Alfred A. and Rohloff, Albert C. (1967), "Consumer Response to Promotion," in Promotional Decisions Using Mathematical Models, ed. Patrick J. Robinson, Boston: Allyn and Bacon, 43-145.
- Lehmann, Donald R. (1979), Market Research and Analysis, Homewood, Illinois: Richard D. Irwin.
- MacKay, David B. (1973), "A Spectral Analysis of the Frequency of Supermarket Visits," Journal of Marketing Research, 10, 84-90.
- McCullough, B. Claire (1978), "Effects of Variables Using Panel Data: A Review of Techniques," Public Opinion Quarterly, 42, 199-220.
- Morrison, Donald G., Frank, Ronald E., and Massy, William F. (1966), "A Note on Panel Bias," Journal of Marketing Research, 3, 85-8.
- Powers, Edward A., Goudy, Willis J., and Keith, Pat M. (1978), "Congruence Between Panel and Recall Data in Longitudinal Research," Public Opinion Quarterly, 42, 380-9.
- Sudman, Seymour and Ferber, Robert (1979), Consumer Panels, Chicago: American Marketing Association.
- Tull, Donald S. and Hawkins, Del I. (1976), Marketing Research: Meaning, Measurement, and Method, New York: Macmillan Publishing Company.
- Wiggins, Lee M. (1973), Panel Analysis: Latent Probability Models for Attitude and Behavior Processes, San Francisco: Jossey-Bass Inc., Publishers.
- Wilson, R. Dale (1977), "Generalized and Embedded Versions of Heterogeneous Stochastic Models of Consumer Choice Behavior: An Empirical Test and Statistical Evaluation in a Dynamic Store Selection Context," unpublished Ph.D. thesis, The University of Iowa.
- _____, Newman, Larry M., and Hastak, Manoj (1979), "On the Validity of Research Methods in Consumer Dealing Activities: An Analysis of Timing Issues," in 1979 Educators' Conference Proceedings, eds. Neil Beckwith et al., Chicago: American Marketing Association, 41-6.

DECOMPOSING THE CORRELATION MATRIX IN PANEL DATA

Donald R. Lehmann, Columbia University
John U. Farley, Columbia University

Abstract

ANOVA is suggested as an approach to decomposition of a correlation matrix as an aid to model building. The decomposition proceeds sequentially from a simple "direct effects" to a "significant effects" model relating four constructs measured in two waves of a panel study tracking the introduction of a new small automobile.

The correlation matrix serves as the basis for many types of analysis, including two-variable association, multiple-relationship structural models and reliability and validity studies. Methods of analysis include coefficient alpha, multi-trait and multi-method approaches to the analysis of within-construct and among-construct correlations, factor analysis for assessment of variable structure and dimensionality, and regression and path analysis for estimation of structural parameters linking different constructs. When multi-wave and/or multi-brand data are available, the correlation matrix has regularities that may offer special information useful for systematic model building. This paper describes a direct approach to analysis of patterns in a correlation matrix. The approach is designed to help gain insight into structure in a set of variables in situations when repeat measurement is available.

The Approach

The basic approach involves decomposing the set of zero-order correlations with analysis of variance, attempting to explain the correlations on the basis of combinations of variable, construct and remeasurement patterns. The approach, which has similar goals as approaches for deducing structural relationships among a set of variables (Baggozi 1979, Joreskog and Sorbom 1977), can be summarized as follows:

1. Underlying constructs are assumed to be linked causally.
2. Constructs are assumed to be measured with error.
3. Errors are assumed to have a correlated error structure.
4. Multiple measures are available on constructs over time, over multiple items or both.
5. An ANOVA incorporating direct effects (e.g., wave, brand, construct or variable) is used to assess "base-level" correlation in the data which may represent response style (Gruber 1979) or some other "common factor" characteristic of the measurements.
6. Interactions (combinations of waves, constructs and variables) are specified to model particularly strong or weak relations among variables and hence to suggest intertemporal or intervariable links for a causal model.

This ANOVA approach is similar in spirit to the log-linear method for analysis of individual traits, in that it sequentially tests various effects to assess the impact of that particular variable on patterns of intercorrelations.

Different types of data configuration make different types of ANOVA models appropriate. Some possibilities are shown for illustration in a two-wave correlation matrix in Figure 1.

a) for any single-wave study, a triangular correlation matrix such as I is available. In this case, it is possible

to develop a model including both multiple measurements within construct and relationships between constructs.

b) for a two-wave study using different respondents, two such triangles (I and II) are available. Insofar as measurement is on parallel constructs (e.g., $K = M$ and each set of measurements is identical) a parallel model can be estimated viewing the two triangles as replications. A similar approach can be used for two stimuli measured identically on either the same or different samples. Both will yield two triangular correlation matrices which can be viewed as replications.

c) remeasurement on the same respondents either over stimuli in a cross section or over time as in a panel (or both) produces a series of rectangular correlation matrices like III, in addition to triangles I and II. When identical measurements are used on each wave or stimulus, III is square and test-retest reliabilities of the measurements appear on the main diagonal.

Direct decomposition of the correlation matrix uses various features of construct, repeated within-construct measurement, time and stimuli to develop design patterns to be used in analysis of variance models for assessing patterns of systematic differences in arrays of correlations of the type shown in Figure 1. The design matrix for the ANOVA will be ideosyncratic to a particular application, depending on the number of waves, the extent to which parallel measurements are used, and the extent to which measurements on different stimuli (usually brands) can reasonably be viewed as replications. Further, the number of available square cross-wave and cross-stimuli matrices determines the number of replications or semi-replications available, which in turn determines the extent to which various second- and higher-order interactions can be incorporated into the ANOVA design.

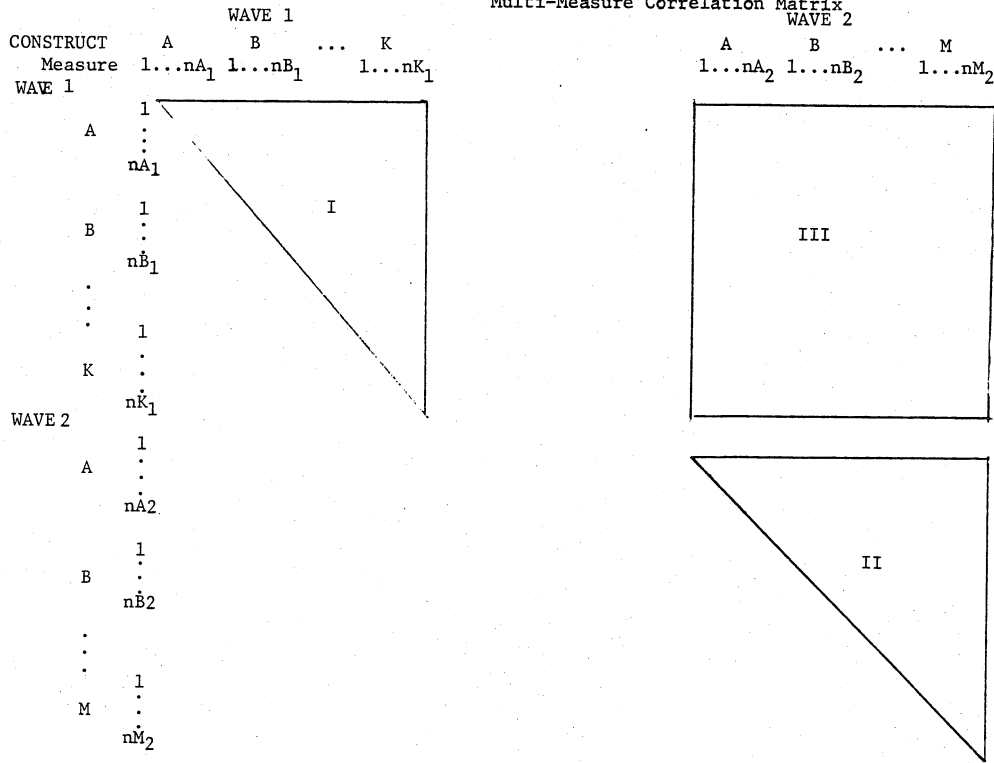
An Application

The approach is illustrated in an analysis of four single-measure constructs which constitute the core endogenous variables in a consumer choice model (Farley, Howard and Lehmann 1976):

Intention to buy brand i
Attitude toward brand i
Confidence in ability to judge brand i
Perceived knowledge of brand i

There is theoretical reason to expect systematic contemporaneous inter-construct relationships, and there may be important inter-temporal relationships as well. The data consist of measurements on two waves of a national panel on small cars, with one wave taken immediately following the introduction of Vega and the other three months later. All constructs were measured by self-report on a 10 point bi-polar adjective scale. The analyses uses the correlation matrices for Volkswagen (an existing brand) and Vega (the new brand) shown in Table 1.

FIGURE 1
 Illustrative Two-Wave Multi-Construct
 Multi-Measure Correlation Matrix



The "full decomposition" ANOVA model is of the form:

$$r_{ijklmnpqs} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \phi_m + \xi_{nn} + \psi_{pq} + \epsilon_{ijklmnpqs} \quad (1)$$

- i = 1,2 brand
- j,k = 1,2 wave of row and column position in Table 1 respectively
- l,m = 1,2,3,4 variables in row and column in Table 1 respectively
- nn = 1-1,2-2,3-3,4-4 indicating remeasurement on the same variable
- p,q = 1,2,3,4 (p≠q) interaction terms pairings of different variables
- s = 1, ..., 56 the observation

TABLE 1
 Simple Inter-Variable Correlations for Volkswagen
 (Above Major Diagonal) and Vega (Below Major Diagonal)

VOLKSWAGEN									
Construct	WAVE A				WAVE B				
	Intention	Attitude	Confidence	Knowledge	Intention	Attitude	Confidence	Knowledge	
WAVE A									
V Intention	*	.261	.164	-.097	.266	.184	.138	.015	
E Attitude	.218	*	.346	.133	.254	.712	.312	.134	
G Confidence	.211	.444	*	.302	.133	.291	.488	.215	
A Knowledge	.050	.338	.270	*	-.057	.132	.232	.435	
WAVE B									
V Intention	.144	.125	.091	.012	*	.239	.184	.002	
E Attitude	.166	.419	.261	.215	.305	*	.373	.053	
G Confidence	.133	.279	.398	.293	.220	.431	*	.267	
A Knowledge	.060	.185	.227	.495	.002	.154	.267	*	

Regression procedures were used for estimation. The design variables are binary and were coded +1, -1 so coefficients of a given effect sum to zero. Pair-wise interaction terms are formed by direct multiplication of the three direct effects (brand, wave and variable). The two brands are viewed as replications of the same data-generating process. Two different specifications of ψ_{pq} are used—one symmetric (e.g., $\psi_{pq} = \psi_{qp}$) and one in which the coefficients are not constrained to be equal.

Results

The results of the decomposition are discussed in terms of:

1. an assessment of residuals from a baseline "direct effects" model which leads to further inferences about model structure in the context of this potentially incomplete model. This model excludes ξ and ψ from (1).
2. tests on direct effects and pair-wise interactions in the full version of (1).
3. interpreting the parameters of an ANOVA model containing effects identified as significant in steps (1) and (2).

The Use of Residuals. A "direct efforts" model provides the baseline for assessment of possible interactions. Residuals from this ANOVA model, often disregarded in analysis of variance, provides guidance for improved specification. Patterns in the signs of the residuals from this parsimonious direct effects only model (Table 2) show clear underprediction of reliabilities and over-

prediction of cross-wave combinations of non-paired variables. The direct effects model does a reasonably good job predicting correlations of contemporaneous pairings. Further, it appears that a relatively small number of large underpredictions (particularly involving the reliabilities) offsets a larger number of smaller overpredictions (Table 3) in the cross-wave combinations.

These patterns motivate the use of a model incorporating interactions combining various pairings of variables measured across waves. The residuals thus provide guidance in structuring tests for specific interactions—a step necessary because the data base lacks the degrees of freedom for a full set of factorial interactions of all orders.

ANOVA Results. Conventional analysis of variance for the full model containing both direct effects and pair-wise interactions are shown in Table 4. Since the design is not orthogonal because of the configuration of the correlation matrices, only the partial sum of squares are attributed to the individual effects. Common variance components are omitted from the numerators of all F statistics. The results:

- 1) no significant difference was found over time in the correlations, even though the Vega was newly-introduced. It appears that intervariable structure becomes stable quite quickly, probably in part because the product class is well-known and heavily advertised.

TABLE 2
Residuals From Direct Linear Effect
Model Including Wave, Car and Variable

Brand	Vega				Volkswagen			
Variable- Combination	Cross-Variable Combination Residuals							
	Wave 1	Wave 2	Wave 1- Wave 2	Wave 2- Wave 1	Wave 1	Wave 2	Wave 1- Wave 2	Wave 2- Wave 1
Intention-Attitude	-.043	.093	-.056	-.063	.052	.041	-.024	.079
Intention-Confidence	.007	.027	-.070	-.091	-.026	.005	-.051	-.360
Intention-Knowledge	-.028	-.065	-.171	-.105	-.161	-.152	-.049	-.161
Attitude-Confidence	.076	.072	-.089	-.121	-.010	.028	-.090	-.121
Attitude-Knowledge	.094	-.071	-.158	-.101	-.117	-.167	-.058	-.107
Confidence-Knowledge	.041	.039	-.010	-.005	.077	.052	-.010	-.005
	Residuals for Reliabilities							
	Vega				Volkswagen			
Intention	.122				.257			
Attitude	.032				.338			
Confidence	.035				.098			
Knowledge	.322				.275			

TABLE 3
Sign Patterns of Residuals From
Direct Effects Model

	Residual sign positive (regression underpredicts)	Residual sign negative (regression overpredicts)	Number of correlations
Reliabilities (variable paired with itself over time)	8	0	8
Different variables paired in same wave	14	10	24
Different variables paired in different wave	1	23	24
	<u>23</u>	<u>33</u>	<u>56</u>

2) no significant difference was found between Vega and Volkswagen correlations despite significant shifts in means of three of these variables for Vega over the introductory period and absence of such shifts for Volkswagen (Farley, Katz, Lehmann and Winer, 1979). Stable inter-variable relationships can thus exist among key constructs describing consumer behavior even when the means of those constructs are shifting over time.

3) significant differences occur in correlations involving different individual variables. This indicates that the variables are not simply one "common factor" but measure different constructs.

4) a significant test-retest effect was found, indicating reliability and/or the existence of significant carry-over effect for individual variables.

5) an interaction effect also exists for individual variable pairs, indicating the presence of some type of model linking these variables. A test for symmetry of interactions (e.g., that the correlations of attitude and knowledge are different for different wave pairings) shows that the interactions are not asymmetric. This implies that the interactions are stable over time.

The Regression Coefficients. Additional insights into the nature of the data structure is provided by the regression coefficients from an ANOVA because the analysis of the sums of squares indicates existence but not direction of an effect. A reduced model consisting only of those variables identified by the ANOVA just discussed is used--the equivalent of pooling the sums of squares of the insignificant effects in Table 4 with those of the residual error term. The resulting "significant effects" model contains direct effects of variables, plus reliability and symmetric interaction terms. The coefficients (Table 5) are expressed in units of increment or decrements of correlation around the constant term which represents the grand mean. Among the individual variables, knowledge and attitude have incrementally positive effects while intention is incrementally negative and significant. The reliabilities add significant increments of correlation, while the interactions have modest negative effects.

TABLE 4
Results of ANOVA on Correlations

Source of Variation	Degrees of Freedom	Sum of Squares	F ratio
Brand	1	.00252	.579
Wave in row and column	2	.01555	1.787
Variable in row and column	6	.39256	15.040
Reliabilities	3	.24010	18.398
Interactions			
Symmetric effect	3	.26742	20.492
Increment provided by asymmetry	3	.00236	.181
Error	37	.16104	

TABLE 5
Regression Coefficients for Variables
in "Significant Effects" Model

Direct Variable Effects	Row	Column
Intention	-.088*	-.120*
Knowledge	.081*	.077*
Attitude	.029	.068*
Reliability		
Intention		.172*
Knowledge		.171*
Attitude		.083*
Symmetric interactions		
Intention-Knowledge		.000
Intention-Attitude		-.045*
Knowledge-Attitude		-.037

*significant t statistic

R² .848
R²_{adj} .806
constant .242

Reports of ANOVA results often omit measures of goodness of fit which can be useful both for comparing alternative models for a given body of data and for comparing results from different but related studies (Farley, Lehmann and Ryan 1979) in terms of the ability of a model to grasp essential characteristics of a data generating process. The explanatory power of the "significant effects" model compares favorably with those of many semi-aggregate cross sectional models (Farley and Howard 1977), and with those containing various combinations of direct effects and interactions (Table 6) in this specific case.

TABLE 6
Coefficients of Determination
for Alternative ANOVA Models
Coefficients of
Determination

	Adjusted	Unadjusted
ANOVA model with direct effects only	.341	.449
ANOVA model with direct effects and reliabilities	.782	.830
ANOVA model containing only significant effects	.804	.848
ANOVA model with direct effects, reliabilities and constrained interactions	.816	.866
ANOVA model with direct effects, reliabilities, and unconstrained interactions	.804	.868

Discussion

This paper describes an ANOVA approach to direct decomposition of a correlation matrix on the basis of various

direct effects and specified interactions. The analysis can be developed sequentially from a parsimonious "direct effects" specification and can lead to a "significant effects" model involving direct effects and interactions. The approach can be used to:

- 1) identify factors causing systematic variation in the correlations;
- 2) identify patterns in appropriately classified residuals from an ANOVA for extension of the decomposition process;
- 3) describe the process generating the correlations using the coefficients of individual variables and the coefficient of determination for an ANOVA.

The approach can be especially useful for constructing a design for analysis of panel measurements when a theory that is essentially contemporaneous is being used. The ANOVA models are likely to be parsimonious and can provide integrated information on reliability of individual measurements over time, on within-construct validity and on general patterns in among-construct relationships. While such a correlation decomposition is not an end in its own right, it may be useful for integrating inter-temporal features of panel data into a general contemporaneous specifications of a structural model (Farley, Katz, Lehmann and Winer 1979).

References

- Bagozzi, R.P. (1980), Causal Models in Marketing, New York: John Wiley and Sons.
- Farley, J.U. and Howard, J.A. (1975), Control of "Error" in Market Research Data. Lexington, Mass.: D.C. Heath.
- _____, Katz, J.R. and Lehmann, D.R. (1977), "Patterns in Repeated Attitude Measurements on New and Established Brands of Subcompact Car," working paper, Columbia University.
- _____, _____ and _____ (1978), "Impact of Different Sets of Comparison Brands on Evaluation of a New Subcompact Car Brand," Journal of Consumer Research, (September, 5, 2, 138-143).
- _____, _____ and Winer, R.S. (March 1979 forthcoming), "Two Approaches to Enriching Specifications of Consumer Decision Process Models," Proceeding, First Annual TIMS/ORSA Conference on Market Measurement, Stanford University.
- _____, Lehmann, D.R. and Ryan, M.J. (1979), "Generalizing From Imperfect Replication," working paper, Columbia University.
- Gruber, Robert E. (1978), "The Impact of Response Style and Response Set on Large Scale Survey Research Involving Activity, Interest, and Opinion (AIO) Variables," unpublished Ph.D. dissertation, Columbia University.
- Jöreskog, K.G. and Sörborm, D. (1977), "Statistical Models for Analysis of Longitudinal Data," in D.J. Aigner and A.S. Goldberger, Latent Variables in Socioeconomic Models, Amsterdam: North Holland Publishing.

RESEARCH PANELS IN CONSUMER BEHAVIOR

Robert Ferber, University of Illinois
Linda B. Lannom, University of Illinois

Introduction

Most panel or longitudinal studies are undertaken to study some aspect of consumer behavior, such as purchase patterns over time or people's experiences moving into and out of the labor market. Such uses of panels are becoming more commonplace, and deservedly so, because they provide information about consumer behavior that only can be obtained by the panel technique.

At the same time, the growing popularity of the panel technique has led to a few attempts to set up panels to study the technique itself. This type of panel is much less glamorous, since no substantive question or policy issue is under investigation, and hence there is no expectation of dramatic results that, say, 46% of milk drinkers have seen the light and switched to scotch. The fact remains, however, that research panels serve a vital function in investigating the reliability of this technique and showing how the efficiency of these panel operations can be improved. Accordingly, the purpose of this paper is to discuss the objectives and the functions of research panels and to indicate the types of uses to which they can be put. In addition, we shall present some results obtained with one type of research panel that has been operated by the Survey Research Laboratory for more than ten years.

Objectives of a Research Panel

Before discussing the objectives of this type of panel, it might be a good idea to explain what we mean by a research panel. First, we should make clear that by a panel or longitudinal study we are referring to a series of reinterviews of the same sample on the same subject. We classify a study as a panel only if at least two reinterviews are involved, since the one-reinterview case is more likely to be an experiment than a panel study. The reinterviews are usually made at regular intervals, and the same people and the same basic information are sought in each interview, though this does not preclude asking for additional information as well.

By a research panel we mean one where the primary focus is on some methodological aspect of a panel operation. This may include a panel that is to be used solely for experimentation with panel techniques or to investigate the cost or the feasibility of a large-scale panel by means of a pilot study. Research panels that we know of are invariably of the latter type. They may have objectives such as:

- a. To ascertain the types of problems likely to be encountered in that type of panel operation.
- b. To explore means of dealing with these problems.
- c. To obtain operating experience with such a panel over a period of time, especially, information on panel mortality, sampling biases over time, conditioning effects, and reactions to sensitive information.

Even though the focus may be on methodology, a research panel has to have a substantive focus as well. This is necessary to make the panel members happy, as well as the funding agency. Thus, it is hardly feasible to inform a sample of households that we plan to study how their answers or behavior may change over time as a result of membership in a panel. It is a lot easier, and more feasible, to tell them that we want to study how their saving and spending patterns change over time as their little monsters grow up and take over the household.

Moreover, this provides the opportunity to obtain useful information at the same time as valuable data are obtained on the more important methodological aspects. Also, it is easier for a funding agency to justify a methodological study when substantive results are to be obtained as well--policy makers do not easily comprehend spending tens of thousands of dollars to improve methodology with which they have little contact, whereas they are much more quick to see the value of a study that will give them more information on the feeding and control of the two-legged little monsters whom they run into all the time.

Some Experimental Considerations

The planning of a research panel involves all of the usual considerations of a sample survey plus a few additional wrinkles. Thus, we still have to consider what population will be studied, how to select a proper sample in light of the given objectives, how to design the questionnaire, how to deal with possible response and nonresponse biases, and so on. In addition, however, some factors have to be kept in mind that do not usually arise in the ordinary pilot, or experimental, study.

For one thing, any experimentation should not be destructive in the sense of losing an appreciable part of the sample, at least not till the end of the study. For example, in a long-term panel operation, one cannot start out by asking, say, half of the sample for a complete listing of their assets and debts on the very first interview, to measure the effect of asking for highly sensitive information. While the results of such a procedure would certainly provide a good measure of the effects of asking for highly sensitive information at the beginning, this procedure will also very likely lead to the loss of an appreciable part of the sample, so that the sample size available for later observation and experimentation will be sharply reduced.

For these reasons, any experiments built into a research panel have to be planned keeping in mind their effect on the amount of information to be collected and the type of experiments to be carried out at a later time. One other consideration is that any methods used with the panel to improve response have to be planned in terms of their possible effect on the conditioning of the panel members and the types of information that will be sought in the future. Thus, any incentives or premiums have to be selected very carefully. For example, if the focus of a panel study is on the financial behavior of households, one can hardly give as an incentive a book on how to save and invest money. While it would be nice for some purposes to obtain an exact measurement of the bias due to this type of incentive, it is much the wiser approach from the point of view of study design to simply assume there will be some appreciable effect and to focus on other types of incentives for the purposes of the study. To be sure, there is no reason why some experimentation with incentives, or with other variables, cannot be built into a research panel, but it becomes virtually crucial that each of the experimental treatments be independent of the attitudes or behaviors to be studied as part of the substantive purpose of the study.

A third consideration is that if the research panel is to be operated over a number of years, every effort has to be made to maintain rapport with the panel members and to keep in contact with them. This means that the interviews should not be so long as to irritate the panel members to

the point of dropping out, that procedures have to be developed for keeping in contact with the panel members and finding out if they moved, and that where possible, other measures be used to maintain their interest, such as periodic reports or even a newsletter. Incentives can also be very effective.

Of course, ideally a research panel, like any survey study, should be based on a carefully thought-out set of objectives accompanied by a detailed set of data collection plans and analytical methods for meeting these objectives. In practice, this is hardly feasible with a research panel, and less so as the length of the panel study is increased. This is not to say that a long-term panel study may not have a set of specific objectives, and also general plans for meeting these objectives. However, conditions do change over time, resources may change, new ideas may arise from previous waves (if one is not careful...), and the priorities may be changed. Hence, while a general set of objectives is useful, it is also wise to stay flexible and be able to accommodate changes in objectives and procedures as they seem warranted.

Such changes are especially likely in a long-term research panel of the type that we shall describe very shortly. In such a case, there was the additional complication, that is not unusual for long-term panel studies, of a change in the funding sources, which can lead to changes in both the objectives of the study and the procedures.

In general, there seems to be somewhat of a tendency in the case of a research panel to place less emphasis on pure experimentation and to obtain more information on the techniques and methodology by means of observation and ex-post data analysis. This is not the place to indicate the various ways in which this may be done, except that we might note that this type of information can be obtained both internally--by studying the characteristics of the data--as well as by comparing characteristics and behavior of the panel with that of the population from which the panel was drawn.

The Illinois-Berkeley Panel

Background and Objectives

Three features make our consumer panels unique among studies of family economic behavior. The panel aspect is, of course, the primary of these. To date, 17 waves of data have been collected from the Decatur-Peoria panel, and 11 waves of data have been collected from the Chicago panel, creating two data sets which are rich sources of information about changes in family economic behavior and purchase decisions. Second, in these consumer panels, the family is not viewed as a single homogenous unit. Rather, it is regarded as being composed of individuals with separate backgrounds and personality characteristics which all influence family economic behavior to some extent. Finally, all couples in the panel are in the early stages of family formation and growth. This fact, together with the panel design, allows the collection of data from families while their economic and consumption habits are being formed and while many major life style events, such as the birth of children, migration, divorce, and remarriage are taking place.

Outline of the Panel Design

The consumer panels are two samples of young married couples who have been interviewed regularly since the time of their marriages. One of the panels consists of couples married in the summer of 1968 in Peoria and Decatur, Illinois. These sites were selected to avoid the complexity of a large city and to obtain respondents in sites which were reasonably self-contained economically and socially. At the time the panel was formed, the population

of these two cities was between 100,000 and 250,000, their economies were primarily industrial, and they were relatively isolated from other major urban centers, as is still the case.

The Peoria-Decatur sample was selected from a list of 1,300 couples who married in Peoria or Decatur between June 1, 1968, and September 30, 1968. Three sources were used to develop this list--county marriage license data from the three counties forming the Decatur-Peoria SMSA, newspaper accounts of new marriages, and records of clergy. From this list a sample of 400 newly married couples was selected, 150 in Decatur and 250 in Peoria. Addresses for these newly married couples were obtained by matching the last names with telephone directory listings or by contacting secondary sources, such as parents. Only couples who lived either in the Peoria or Decatur SMSA after their marriage were used. In addition, the husband had to be 30 years of age or less at the time of the marriage. The first wave of interviewing was completed in the fall of 1968, with 311 eligible couples completing the questionnaire.

When the Peoria-Decatur panel was four years old, the Chicago panel was formed, composed of young couples married in Chicago between June 1, 1972, and August 31, 1972. This sample was drawn from microfilm marriage records obtained from the Illinois Bureau of Vital Records. A sample of 1,003 couples married in Cook County was selected from this list. Location of the newly married couples' addresses was accomplished as in Decatur-Peoria, using telephone directory listings and information supplied by secondary sources. Only couples who lived within the Chicago SMSA after their marriage were eligible.

Four other eligibility criteria were imposed in Chicago:

1. The husband should be 30 years of age or less at the time of the marriage;
2. The marriage should be the first one for both members of the couple;
3. The couple could not own a home at the time of the marriage;
4. The couple's income should exceed \$5,000 per year, unless one of them was a student.

Since it was difficult to find couples who met all of these criteria, the first criterion was relaxed to include couples with husbands who were 34 years of age or less at the time of their marriage. The Chicago panel, first interviewed in the late fall of 1972, contained 409 couples.

Topics Included

Economic Behavior

The major purpose of these panels is to collect data on the durable goods ownership and financial assets of these young couples. For this reason, on each wave couples are asked about their purchases of durable goods during the time between interviews and their intentions to purchase durable goods during some specified period in the future. At regular intervals, information on their stock of goods is updated. For many of these goods, data are available on such characteristics as the price paid, the brand name, on whose advice the purchase was made, and who made the decision to purchase. Extensive information is also collected on the couples' financial portfolios, saving and budgetary practices, and use of credit. Many of the questions are geared toward determining the role of each marital partner in the decision making process.

Data relevant to the economic personalities of the spouses are also collected periodically. Both husbands and wives respond to questions by which they can be characterized as economically minded, price conscious, or bargain seeking.

Subjective assessments of the country's current and future economic situation, including inflation, have also been gathered from the respondents. Intermittently, the subjects have assessed their own economic situation relative to that of their parents, their friends, their own past, and their anticipated futures. Data on household income have normally been collected once a year.

Labor Force Participation

Data on additional economic characteristics relevant to labor force participation, such as occupation, industry, part- or full-time employment, and involvement in a second job have been collected at each interview. The surveys have also included an extensive number of questions relevant to female labor force participation (most of these after Wave 9 in Peoria-Decatur and Wave 3 in Chicago). These included asking nonworking wives about their reasons for not working as well as their own assessments of their occupational prospects if they were to seek a job, and the cost of child care arrangements they would incur by working. Working women reported on the cost that they actually incurred while working as well as their reasons for working. All women responded to questions on their work plans after having children and in the more general future.

Background Characteristics

A series of questions on the respondents' background characteristics was included in the first wave for each of the panels. These included birthplace, occupation of father, number of siblings, parents' marital status, and whether parents were currently living. Current religion and religious background were sought on a later wave.

Aspirations and Goals

Periodic attempts have been made to measure respondents' aspirations and goals. These have included educational, occupational, and familial aspirations, as well as financial aspirations. One wave also included questions on respondents' aspirations for their children.

Children

At each wave, respondents indicated whether they had a birth during the interval between waves. On the tenth wave in Peoria-Decatur and the fourth wave in Chicago, the exact dates of all births were obtained. This provided a check on earlier answers and filled in missing data for a few cases. Respondents also reported at each wave whether they were expecting a child or in the process of adoption. After Wave 6 in Peoria-Decatur and beginning in the first wave in Chicago, respondents answered questions periodically on their desired and expected family sizes. Respondents have also been asked about birth control as well as about abortion. On various waves, an attempt was made to review the decision-making process surrounding fertility.

General Attitudes and Personality Characteristics

Throughout, attempts have been made to measure general attitudes toward life and the personality characteristics of the respondents. Questions regarding attitudes toward life in general and satisfaction with one's own life and marriage have been asked repeatedly. Among other characteristics, measures of experiment-proneness, timidity, conservatism, and self-assuredness were constructed from responses to individual questions.

Sex Role Orientation

Both husbands and wives have responded periodically to questions concerning at what age and under what circumstances a mother/wife should work, the importance of job advancement, factors important in job selection, decision-

making concerning employment of other family members, and work histories of respondents' mothers. Husbands and wives have also been asked about the division of labor within their homes as well as those of their families of origin.

Time Use

Use of time for household tasks and leisure activities was the focus of a number of questions on Waves 8, 9, 12 and 15 in the Peoria-Decatur study and Waves 2, 3, 6 and 9 in the Chicago study. A series of questions about vacation plans and habits was asked of the Peoria-Decatur respondents on the eighth wave.

Social and Political Attitudes

Some topics reflect the political and social climate of the times in which they were asked. For instance, respondents' attitudes toward electric cars, reusable containers, and organically grown food have been assessed. Questions on transportation were asked during the energy crisis of 1973-74 and again on the current 1980 wave of the panels. Attitudes toward protests by blacks, students, and truck drivers were measured at the time truckers were blocking the highways. At the same time, questions concerning violence on television and police violence were asked. During earlier waves, questions on media preferences (newspapers, magazine, television) and exposure were asked.

This list represents a fraction of measures taken during the study and should not be regarded as exhaustive. Many of the topics just mentioned were covered on the spouses' separate questionnaires, so that answers are available from both members of a couple. In addition, many of the topics were repeated, so answers are available to the same questions at different points in time and after different durations of marriage. The table provided as an appendix summarizes selected data currently available for the consumer panel.

Response Experience

By June 1980, the Decatur-Peoria panel had been interviewed 17 times, and the Chicago panel 11 times. From 1968 until 1977, the aim was to interview both panels every six months. Beginning in 1978, the aim was to interview each of the panels once a year. Because of the vagaries of funding, it was not always possible to achieve these aims; thus, the actual intervals between waves of interviews varies between six and fifteen months.

The date of each wave of interviewing and the number of households completing interviews at each of the waves is shown in Table 1. From the first through the seventeenth wave, the total number of households in Decatur and Peoria dropped from 311 to 206. The reduction in the number of cases was even more severe in the Chicago panel, which lost over half of its members from the first to the tenth wave, dropping from 406 to 194 cases. As expected, attrition was most serious during the earlier waves of the panel, due largely to couples refusing to participate or inability to locate couples who had moved.

We were concerned that such attrition represented an important source of bias in the consumer panel. To measure this bias, we conducted a follow-up study of nonrespondents in 1975, making intensive efforts to locate and obtain completed interviews from a sample of people who had left both panels. We found the dropouts of these consumer panels differed only slightly from the current panel members. The relatively high rate of attrition, particularly for the Chicago panel, appeared to have had very limited effects on the demographic structure of the panel. Significant differences between panel members and dropouts did appear in relation to expenditure characteristics, in particular to the ownership of various durable goods. However, even

here, no discernible pattern of bias emerged.

Experience with Sensitive Information

The focus of these panels is family financial decisions. Extensive information is collected on the couple's financial portfolios, saving and budgetary practices, use of credit, and ownership of life insurance. Many of these questions are highly sensitive, asking for the value in dollars of specific assets and debts. Our experience with item nonresponse for such items parallels our experience with attrition; that is, the rate of item nonresponse has dropped over the life of the panel. For example, item nonresponse to household income was .3 percent on the seventh wave of the Peoria-Decatur panel and .9 percent on the first wave of the Chicago panel. On the most recent wave of each of these panels, nonresponse to this item had dropped to .2 percent.

We believe that item nonresponse has dropped for several reasons. First, those most likely to refuse sensitive items have already left the panel. Second, over the years, the panel respondents have come to trust our guarantee of confidentiality. Periodically, we issue special reports to respondents, summarizing some of the findings of the study. In reading through these reports, respondents discover that we do, in fact, report their responses as aggregate estimates, and that it is not possible to identify individuals on the basis of what we report.

Third, in 1977, the method of data collection was changed from face-to-face interviewing to telephone interviewing. When we made this switch, we were concerned that it would depress response rates and possibly increase item nonresponse. In fact, it did not. We found that respondents were less reluctant to report dollar amounts when interviewed on the telephone, possibly because this form of data collection had the effect of making interviews more objective by insulating respondents from the interviewer.

Experience with the Use of Incentives

Because the panel members are asked to do a substantial amount of record checking during the course of an interview, we give them a small gift each time they are interviewed. When respondents were interviewed face-to-face, the interviewers presented the gift at the conclusion of the interview. Now that respondents are interviewed by telephone, gifts are mailed in advance to each panel member. The monetary value of these gifts is small, ranging from \$2 to \$5 per couple. However, we feel the psychological value of the gift is important both for minimizing attrition and item nonresponse.

Some Uses of the Data

Because the consumer panels are ongoing projects, analysis to date has been somewhat limited. Most of it has been done using data from the Peoria-Decatur sample, since the Chicago marriages are still of relatively short duration. The analysis which has been done has concentrated on two major areas--various facets of family financial behavior, and fertility and female labor force participation.

Very briefly, some of the principal findings from these studies may be summarized as follows:

Family Financial Behavior

1. The studies support the idea that family financial behavior can be better understood if the family is treated as interrelated individuals rather than as a homogeneous unit. In their analysis of asset accumulation by the Peoria-Decatur couples, Ferber and Nicosia (1972) found that while

education of both the husband and the wife influenced asset ownership, the effects of education for the wife tended to be in the opposite direction from that of the husband. In addition, the effects on asset accumulation of some personality variables and individual's priority on savings also tended to work in different directions for husbands and wives.

2. Ferber and Lee (1974a) found that who played the role of the so-called Family Financial Officer (FFO) did not appear to be determined by socioeconomic characteristics of the couple, but rather by savings attitudes and economic personalities. Most frequently, the couple acted jointly as the FFO; however, over time there was a pronounced shift toward the wife's assuming this role. The identity of the FFO did appear to influence economic behavior. For instance, when the husband was the FFO, the proportion of income saved was greater than when the wife alone or the couple acting jointly played this role.

3. Lee and Ferber (1977) found tendencies for assets to be greater when the husband and wife acted jointly as the FFO. They also found that when wives worked the family had less debt. However, the couples best off financially after five years of marriage in terms of net assets were those who were venturesome enough to acquire substantial amounts of debt. In this sample, both debts and assets primarily related to the purchase of a home.

4. Ferber and Lee (1974b) have attempted to integrate lifestyle into their model, believing that lifestyle may provide an additional dimension beyond the socioeconomic role and personality variables studied to date. Operationalizing lifestyle according to time use, they classify couples into lifestyles on the dimensions of career orientation, pleasure orientation, and familial orientation. Findings indicate that lifestyle helps to explain why women work and whether couples purchase automobiles.

5. Several additional papers by Ferber and his associates (Ferber and Lee 1973; Liebman and Ferber 1977; Nicosia and Ferber 1976) have been concerned with the ability to predict purchases of certain durable goods from responses to questions on past purchases, satisfaction with current goods, and the likelihood of purchasing of goods in the future. The most recent paper of the three (Liebman and Ferber 1977) found that later actual auto purchases were related to earlier predicted and current auto purchases.

6. Anderson and Nevin (1975) found that demographic, socioeconomic, psychographic and other variables were relatively weak predictors of household life insurance purchasing behavior. They reasoned that this was due in part to the exclusion of life cycle change as an explanatory variable, since they used data for the first one or two years of marriage only. They did find that six variables were significant in explaining the amount of life insurance purchases.

7. Lee and Ferber (1980) in a further study of life insurance acquisition found that three variables were significant in explaining the type of life insurance purchased. The purchase of term insurance was found to be much more likely in households where the net worth was large; the wife purchased term insurance before marriage; and the insurance agent did not influence the couple's decision.

Fertility and Female Labor Force Participation

8. Several papers using data from the consumer panels have concentrated on the determinants of fertility behavior. In "Relative Status and Fertility" Crimmins-Gardner and Ewer (1978) tested the relative income hypothesis, that fertility is affected by the relationship between desired and actual living levels. Relative status was measured as the difference between the husband's occupation and the occupations of both the husbands' and the wives' fathers.

Also, each spouse compared his (her) family financial situation at the age of 16 to the couple's financial situation at marriage, in the seventh year of marriage and with the expected situation in the seventeenth year of marriage. When these measures of relative status were related to fertility, only the wife's subjective assessment of relative status showed a consistent and significant pattern. Wives who felt they were or would be better off than their parents, want more children than wives who compare their own situations unfavorably to their teenage years. hypothesis that people with certain backgrounds which emphasized children over material possessions have children earlier relative to the time they acquire material goods (when income is controlled) than people without these backgrounds. Occupation and religion were the two background characteristics examined. However, the results showed that occupation and religion have little effect on the timing of the first child, the purchase of durable goods, or the relative timing of these two acquisitions. Income does affect when people acquire goods, both absolutely and relative to children. Couples with higher incomes acquire goods absolutely earlier, and earlier relative to the birth of the first child. This last relationship is the result of both earlier goods acquisition and later births of first children in couples where the wife works.

10. A panel design is particularly appropriate for investigating the causal dynamics underlying the negative relationship between working and fertility. Data from the first seven years of the Peoria-Decatur panel were used in a path analysis to disentangle the effects of husband's income, wife's education, wife's age at marriage, family size, and wife's employment and subsequent time intervals (Ewer, et al., 1979). The results indicated that during the early stages of marriage and family formation the presence of young children exerted a strong negative effect on wife's employment. In contrast to this effect of fertility on employment, the effects from employment to fertility are less consistent.

11. In an additional study, only wife's labor force participation was used as the dependent variable (Crimmins-Gardner and Ewer 1978), to explain factors affecting employment among young mothers. In addition to the usual determinants of working, expenditures connected with working and the emotional cost of child care were important determinants of the labor force behavior of young mothers.

12. Bagozzi and VanLoo (1978a, 1978b, 1979), have begun to develop a theory of fertility as consumption, examining two classes of socioeconomic determinants--economic variables and social structure variables. The socioeconomic determinants are hypothesized to act as exogenous variables, either facilitating or constraining psychological and social decision-making processes within the family. These latter processes are posited to be the primary endogenous variables affecting fertility.

Tests of the overall theory were performed on three separate samples using a structural equation methodology. In general, fertility decisions were shown to be affected directly by the social psychological processes within the family and only indirectly by socioeconomic constraints, as predicted.

If We Had To Do It Over

Recounting these experiences and results leads us to raise a final question that is likely to be of general interest, namely, if we had the opportunity to do this panel study over again, would we, and how would we change things?

First, would we do it over again? Certainly--the topic is fascinating, the study design is full of challenges, and the results seem to be both very useful and personally rewarding. On the other hand, we would also certainly want

to make various changes in the study design. Some of these changes are pretty obvious, such as having a larger sample size and interviewing panel members every six months on a regular basis, something that we could not always afford. Indeed, interviewing the panel members in a study of this type every six months initially, say for the first five years, and every year thereafter would seem just about optimal, based on our experience.

Some of the other changes are perhaps less obvious, such as the following:

1. Even with separate interviewing of both spouses (a highly desirable approach for data collection purposes), we would have started using the telephone approach much earlier than we did. Our experience, as well as that, for example, of the Health Interview Survey of the U.S. National Center for Health Statistics, indicates that such simultaneous interviews can be obtained very efficiently by telephone. Personal interviews would still be used the first couple of times, and would also be used where families are not reachable by telephone, but the principal reliance would be on the telephone approach.
2. We would seek more powerful methods of maintaining panel member cooperation. This could range anywhere from sending more letters and reports between interviews to introducing an incentive system similar to that of the commercial panels where the rewards for remaining in the panel increase geometrically with the length of time. However, there would certainly be problems in doing so, given the University sponsorship and likely limitations on funds.
3. Insofar as possible, we would develop at the outset a clear set of objectives for the types of decision problems that we wished to study, or hypotheses to test, and develop a long-run data collection plan based on such objectives. Unfortunately, this is more easily said than done. Because funding sources change, there is always uncertainty regarding future funding, and even with a fixed set of objectives, one's ideas how to best obtain the necessary data change with time and with experience.
4. Last but not least, we would take our programmers and systems analysts, lock them in a room, and feed them only mashed potato sandwiches and warm water (or, worse yet, milk), until they came up with a simple and economical means of putting these data in a computer file that is amenable to analyses both over time and over space. Getting a manageable file that would permit the many uses to which panel data could be put is one of the most frustrating aspects of a panel operation. In fact, we are not sure that we would put the mashed potatoes into the sandwiches until the programmers were close to a reasonable solution of this problem.

Thank you.

TABLE I

Interview Date, Collection Methodology, and Number of Households Completing Interviews at Each Wave:
Consumer Panels of Young Married Couples

Date	Method of administering the main questionnaire ¹	Peoria-Decatur				Chicago			
		Wave	Total	Originals households	Households formed after Wave 1	Wave	Total	Original households	Households formed after Wave 1
Fall, 1968	Face - to - Face	1	311	311	0				
Spring, 1969	Face - to - Face	2	286	286	0				
Fall, 1969	Face - to - Face	3	259	258	1				
Spring, 1970	Face - to - Face	4	253	248	5				
Fall, 1970	Face - to - Face	5	237	231	6				
Spring, 1972	Mail with telephone	6	224	220	4				
Fall, 1972	Face - to - Face	7	224	212	12	1	409	409	0
Summer, 1973	Face - to - Face	8	227	213	14	2	334	325	9
Winter, 1974	Face - to - Face	9	225	203	22	3	281	268	13
Spring, 1975	Face - to - Face	10	225	191	33	4	257	238	19
Fall, 1975	Face - to - Face	11	220	185	35	5	253	219	24
Spring, 1976	Face - to - Face	12	221	184	37	6	235	215	20
Fall, 1976	Face - to - Face	13	215	180	35	7	212	194	18
Spring, 1977	Face - to - Face	14	207	165	42	8	199	176	23
Fall, 1977	Telephone	15	201	160	41	9	205	173	32
Winter, 1979	Telephone	16	210	146	64	10	203	165	38
Spring, 1980	Telephone	17	206	N.A.	N.A.	11	194	N.A.	N.A.

N.A. = Not Available

¹Panel members who have moved outside of the Peoria, Decatur, or Chicago SMSA since the initial interview are sent mail questionnaires. In addition for Waves 1 - 14 of the Peoria/Decatur Panel and for Waves 1 - 8 of the Chicago Panel, the male and female heads of the household each received separate self-administered questionnaires.

APPENDIX

SELECTED DATA CURRENTLY AVAILABLE--CONSUMER PANELS OF YOUNG MARRIED COUPLES

TOPIC	Wave: Peoria/Dectaur																		
	Chicago		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ECONOMIC BEHAVIOR																			
Durable goods purchases	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Durable purchase decisions	x	x																	
Relevance of factors in purchase decisions	x																		
Durable goods purchase plans	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Durables purchase priority rank	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Home ownership			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Income data																			
Income from job	x																		
Income relative to expenses				x															
Awareness of price increases																			
Financial assistance and debts	x																		
Finances and money management	x		x																
Financial situation expectations, attitudes, etc.	x																		
Investments in variable dollar form (Ownership of, satisfaction with, financial services)																			
Factors affecting investment decisions																			
Shopping attitudes	x	x																	
Credit cards	x																		
Family balance sheet																			
Attitude toward, decisions concerning savings	x																		
Saving rate																			
Use and ownership of savings accounts																			
Awareness of ads on savings																			
Ownership of life insurance	x																		
Insurance other than life insurance	x																		
LABOR FORCE PARTICIPATION AND SEX ROLE ORIENTATION																			
Employment (unemployment)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Women and work	x																		
Comparison of couples roles with parents																			
Discussion of decisions																			
BACKGROUND CHARACTERISTICS																			
Miscellaneous	x																		
Age	x																		
Race																			
Draft status																			
Religion																			
Education	x																		
GOALS AND ASPIRATIONS																			
Education	x																		
Employment	x																		
Income																			
Homeownership	x																		
CHILDREN																			
Fertility plans																			
Changes in family circumstances																			
Parents and children																			
GENERAL ATTITUDES AND PERSONALITY CHARACTERISTICS																			
Self-determination																			
Self-assessment																			
TIME USE																			
Use of leisure time																			
Vacation plans and habits																			
Media preferences and exposure	x																		
SOCIAL AND POLITICAL ATTITUDES																			

OVERVIEW OF THE CONSUMER EXPENDITURE SURVEYS ^{1/}

N. Gail Hoff, Bureau of the Census

General

The United States Government has recently begun a statistical program to provide current data on consumer expenditures of the population. Until now, there have been periodic surveys but no continuing program to provide such data on a regular basis. This gap in the government's statistical program has hindered many important analyses of the interrelationships of income and expenditures for various population groups, has created a lack of means to keep the Nation's Consumer Price Index more current, and has caused a void in studying and measuring the effects of a broad range of important policies and programs on the population such as inflation, taxes, energy, housing, medical care, and so on. After many unsuccessful attempts in the past, the Bureau of Labor Statistics (BLS) received appropriations in fiscal year 1979 to develop and conduct an ongoing Current Consumer Expenditures Surveys Program. The project has been given high priority by the Secretary of Labor. The program consists of three interrelated surveys; the two expenditure surveys are designed to provide the expenditure data for revising the expenditure item weights and for studying family buying habits; the Point of Purchase Survey provides data on the outlets at which people shop, to be used in updating the outlet samples for the BLS consumer pricing surveys.

In addition to producing the data necessary to update the CPI market baskets of goods and services and to revise the CPI weights, there are many other uses of the data. Some of these uses are to (1) provide a continuous flow of data on income and consumption patterns for use in economic analysis and policy formulation, (2) provide one of the basic data inputs required to revise and update the BLS Family budget program as required by the Comprehensive Employment and Training Act of 1973 (CETA), and (3) provide a flexible consumer survey vehicle that will be available for use by other Federal Government agencies.

The Census Bureau is responsible for the survey design, data collection and processing activities for the program.

The surveys were last conducted during the period of 1972 through 1974.

This paper will present an overview of the operational aspects of each of the surveys and point out any significant differences with the earlier surveys.

Sample Design

Data in the two expenditure surveys are collected from a national probability sample of households designed to be representative of the total civilian noninstitutional population. A sample of 102 primary sampling units (PSU's) was selected, 86 of which have been previously defined and selected by the BLS for the CPI revision. These 86 PSU's represent the urban part of the United States with the additional 16 PSU's added to represent the rural population. Only the 86 urban PSU's are used in the Point of Purchase Survey.

The unit for analysis for all three surveys is the set of eligible individuals comprising a consumer unit. A consumer unit consists of all members of a housing unit or other type of living quarters who are related by blood, marriage, adoption, or some other legal arrangement. The consumer unit determination for unrelated persons is based upon financial independence.

The sample design for the Quarterly Interview Survey will be a rotating panel scheme. Interviews with each sample unit will be conducted quarterly over a period of five consecutive quarters (15 months) and then dropped from the survey. The first interview for a household will serve to "bound" data for expenditures and to obtain information on the household composition and characteristics of its members. Data collected in the initial survey quarter will not be included in developing the regular series of expenditure estimates. Data collected during the second through fifth interviews will form the basis of the estimates derived from the quarterly survey. After the fifth quarter in sample, the panel will be dropped from the survey and replaced by a new incoming panel. The quarterly sample will be divided into three equal parts with each part designated for interviewing in a particular month of the quarter and every 3 months thereafter while in sample. The monthly sample consists of approximately 2,360 assigned units.

This scheme differs markedly from the 1972-73 survey, partly because it is a continuing survey rather than a one-time survey. In 1972-73, approximately 23,000 addresses were selected in 216 PSU's. Half the sample was introduced in January 1972 and half in January 1973. Each yearly sample was in the survey for five consecutive quarters.

The survey methodology for the Diary Survey requires each selected sample unit to keep two 1-week diaries of expenditures over two consecutive weeks. The earliest possible day for placing a diary with a household will be pre-designated so that each day of the week has an equal chance to be the start of the reference week. The annual target sample size at the United States level for the Diary Survey is 5,470 completed interviews with a total estimated workload of 7,700 sample units. During the last 6 weeks of each year, the Diary sample will be supplemented to twice its normal size (about 850 of the above designated units) to increase the reportings of types of expenditures unique to the holiday season. The basic difference between the current Diary survey and the one conducted in 1973 and 1974 is again a one-time survey versus a continuing one. One additional difference is in the assignment of a first possible placement day. In the previous survey, each sample unit was assigned a week for placement, with the effort made to place the Diary as early as possible in the sample week. Results of the 1973-74 survey showed that the amount of expenditures reported was highest during the first day of the recording period regardless of what day of the week the first day was and tended to taper off towards the end of the 7-day period. By attempting to control the first day of the recording period, each day of the week has an equal chance of being selected as the first recording day. This should spread any reporting bias resulting from calendar day placement among all days of the week.

^{1/} This paper cannot be copyrighted. It may be reproduced.

Preliminary results of the relationship between the earliest placement day assigned for a diary case and the actual day the diary was placed show that 62 percent of the diaries are placed by the third day and 72 percent are placed by the fourth day. This indicates some difficulty in placing a diary on a specific day.

The sample for the Point of Purchase Survey is a clustered systematic national sample of housing units in 86 SMSA's or ZIP code areas across the country. These areas will be treated essentially as PSU's. The program will be designed so that the entire 86 areas in the sample will be updated every 5 years, approximately one-fifth of the PSU's will be surveyed each year. In the first year, 18 of the PSU's will be surveyed. Each year for the succeeding 4 years, a different fifth of the areas will be sampled. In the sixth year, the cycle will begin again. Through the 5-year cycle, 29,000 units will be surveyed. The only difference between the 1974 survey and the current survey is the spread of the PSU's across 5 years. This was done mostly for financial reasons.

Quarterly Interview Survey

Data Collection

The Quarterly Interview Survey is designed to collect data on certain types of expenditures which respondents can be expected to recall fairly accurately for periods of 3 months or longer. In general, expenses reported in the survey are relatively large, such as purchases of property, electrical appliances, and vehicles, or are expenses which occur on a fairly regular basis, such as rent, utility bills, or insurance premiums. The purpose of the Diary Survey is to collect information on the small, inexpensive items which respondents cannot be expected to recall even for relatively short periods of time. Most of the items for which estimates will be derived from the Diary Survey are not sufficiently covered, or are not included at all in the Quarterly Interview Survey. Data collection began in October 1979.

Each occupied Quarterly Interview Survey sample unit will be interviewed once per quarter for five consecutive quarters. In the first interview, information will be collected on socioeconomic characteristics of the household, an inventory of major durable items, and expenditures for selected goods and services during the reference period. Since the initial interview is meant to serve as a bounding interview, the reference period for expenditures is kept to 1 month in order to reduce the amount of interview time. Since duplicate reporting is most likely to occur in the period closest to the interview, the 1-month reference period should provide an adequate period for bounding purposes. A uniform questionnaire with a 3-month reference period has been designed for interviews 2 through 5. Replacement households in interviews 2 through 5 will be unbounded and have a 3-month reference period.

The exact reference period which will be used in all interviews after the first will be from the first of the month, 3 months ago, up to the time of interview. Although this allows the reporting of expenditures for slightly longer than a 3-month period and creates overlaps in reference periods, it is preferable to collect purchases made during the month of interview at this time since recall will be the shortest and most accurate of these purchases. Because of the intentional overlap in reference periods, it was necessary to develop a technique for bounding the data in order to control for duplicate reporting of

expenditures in subsequent interviews. Bounding also prevents the reporting of expenses in a current reference period which actually occurred during an earlier period. As stated previously, the first interview's data will be used primarily to bound data reported in subsequent interviews. To prevent duplicate reporting from one quarter to the next, a method was developed for the pretest which has been implemented for the ongoing survey. For most purchases reported, the date of purchase and a description of the item is obtained. At the completion of the interview, the field office will transcribe the information reported for selected items onto designated areas of the next quarter's questionnaire for the sample unit. The interviewer then has the previous quarter's responses available for bounding purposes at the next interview.

The inventory technique is also being used in the Quarterly Interview Survey to improve the accuracy of reporting. An inventory of certain items, such as owned property, major household appliances, automobiles, and insurance policies possessed by the household, is obtained at the initial interview. The inventory serves as a launching point in subsequent interviews, reminding the respondent of any additional or replacement purchases of such items made since the time of the initial interview.

In addition to the initial interview questionnaire and the uniform quarterly questionnaire for the second through fifth interviews, there will also be annual supplemental questions which will be asked at the fifth interview. These questions will obtain more detailed income information for the past 12 months--cash receipts, taxes, contributions, occupational expenses, and detailed mortgage payment data. Another of the differences in methodology between the 1972-73 survey and the current one is the way the data is collected. In the 1972-73 survey, the reference period was not constant as it is now, but varied for different kinds of expenditures.

In the 1972-73 survey, the first interview collected information on socioeconomic characteristics of the household, an inventory of major durable items, and expenditures for selected goods and services purchased since the first of the year. The second, third, and fourth interviews focused primarily on relatively inexpensive items, such as utilities, clothing, small durable goods, household furnishings, vehicle operating expenses, and the like. Data on income received during the preceding year also is collected during the second quarter's interview. The fifth and final interview obtained data on income, assets, and liabilities for the survey period, as well as expenditure data for most items covered in the survey.

The recall period for reporting expenditures varied according to the size and frequency of the expenditure class. Frequent and relatively inexpensive items, such as utilities, clothing, and vehicle operating expenses, were collected at quarterly interviews. Slightly more expensive items, such as small kitchen appliances and furniture, were collected using a 6-month recall period. For very expensive and infrequent purchases, such as real estate, automobiles, and major appliances, the recall period was 12 months.

One of the reasons for such a change in the collection methodology is the estimation procedure to be used. In 1972-73 only annual estimates were produced, with all data collected referring either to 1972 or 1973. In the current survey, "quarter of occurrence" estimates will be produced. This will require that data collected in a 5 consecutive-month period be used for

such estimates. For example, expenditure data for the October through December quarter will be collected in the November, December, January, February, and March interview months. For this reason, data must be tabulated by the particular month of the reference period in which the expenditure was reported to have occurred. For some characteristics, this will not be possible as expenditures are not reported by month but are provided either in aggregate figures for a 3-month period or by an average expenditure pattern over time such as monthly or weekly costs. The BLS is exploring different methods of allocating the reported 3-month aggregate expenditures or the expenditure pattern data to individual months within the reference period.

One of the primary concerns in the Quarterly Interview Survey is the accuracy of the expenditure data reported. It is unlikely that any respondent could accurately report all of the detailed information requested without referring to records of some type; however, pretest interviewers estimated that only 50 percent of all respondents used any records at all during the interview. Encouraging respondents to keep records for the quarterly visits in the ongoing survey will continue to be an important part of the program. Attractive calendars and files will be offered to respondents; interviewers will be trained to persuade respondents to keep records; and a letter will be sent before each quarterly visit to remind respondents of the type of expenditures they will be asked about.

Data collection is under the direction of the Census Bureau's permanent professional field staff, operated through 12 regional offices throughout the United States. Each staff member is thoroughly trained prior to beginning his/her work on the survey. Initial interviewer training consists of 3-1/2 days of classroom training plus self-study training. A 2-day classroom training session was held prior to the start of the second quarter with subsequent self studies and classroom sessions being provided as needed. In addition, a 1-1/2 day supervisor's training session was held prior to interviewer training.

Data Processing

In 1972-73, all clerical and computer processing was performed at the completion of the five quarters of interviewing at a centralized location. Since quarterly estimates are a necessary part of the ongoing survey, a different method of processing had to be developed. Although a number of alternatives were discussed, the most feasible was decentralized processing in our 12 regional offices.

When the questionnaires are returned to the offices, they are clerically coded for household and consumer unit relationship, make and model of automobile, trip destination, and detailed codes for type of housing repairs. In addition, all outlets (insurance companies, schools, etc.) are coded uniquely by name. Upon completion of the clerical processing, the data are keyed using a Nixdorf data entry system and transmitted to Washington via communication lines where it passes through a very detailed computer preedit. Inconsistencies, errors, and identification of missing questionnaires are transmitted back to the regional offices for reconciliation by the field staff through office review or interviewer followup. Corrections are then keyed and transmitted to Washington and cycled through the computer preedit. This continues until errors identified by the preedit no longer appear. Once a panel month's preedit is complete, data necessary for bounding is transcribed to the next quarter's questionnaire. The current quarter's

questionnaire is sent to our processing office in Jeffersonville, Indiana, for microfilming and storage.

Industry and occupation coding is completed via computer listings sent to Jeffersonville; the codes are keyed and transmitted to Washington to be merged into the main data file.

The data then goes through a rather complex computer edit. In addition, demographic and work experience items are imputed when missing or invalid. The file is then run through an adjustment to net out business and reimbursed expenses; adjusted to always include sales tax and then weighted.

For interviews 2 through 5, an inventory management system is being developed in order to show expenses for items identified using the inventory technique for which payments are still being made at the same level.

The BLS will be provided with both monthly and quarterly weighted tapes beginning in June 1981, for the first calendar quarter of 1980.

Diary Survey

Data Collection

The purpose of the Diary Survey is to obtain expenditure information for small, relatively inexpensive items which are subject to severe recall loss and for which the Quarterly Interview Survey collects only cumulative expenses or does not cover at all. Expenditure classes not sufficiently covered in the Quarterly Interview Survey include food and beverages, personal care products and services, small household supplies, and a few other relatively inexpensive items.

The survey methodology for the Diary Survey calls for each selected sample unit to keep 1-week diaries of expenditures over each of 2 consecutive weeks. The earliest possible day for placing a diary with a household is predesignated so that each day of the week has an equal chance to be the starting day of the reference week.

Respondents are asked to report all expenditures made during the recording period, not just those which will be used for CPI revision purposes. The main reason for this is to take away from the respondent the decisionmaking about which items to record in order to reduce the chance of error due to the respondent's confusion over whether to record an item. It was felt that this additional recording would not increase the respondent's burden for reporting, since the vast majority of the items reported are those of interest to the survey.

The interviewer makes three visits to the Diary household over the 2-week period. At the first visit, the interviewer conducts a brief interview with the household respondent which includes obtaining the household composition and selected demographic characteristics of each member. The interviewer then explains to the respondent how to keep a diary, the type of detail that is needed, and where to record the information. The interviewer then makes an appointment for 1 week later. At that time, s/he returns to the household and reviews the first diary for completeness and legibility, updates the household composition, and asks a few selected questions regarding common types of expenditures. The third appointment is made, again for 1 week later. When the interviewer returns for the

third visit, the second diary is reviewed and selected questions on income and work experience are asked.

If the interviewer arrives to pick up the diary on the appointed day, and the diary has not been completed, the expenditures for the week are obtained through recall. The interviewer would mark a box on the front of the questionnaire to indicate the Diary was completed by the recall procedure. However, if the Diary is not picked up on the appointed day, the expenditures cannot be obtained through recall.

Because the weekly Diary workload is so small, interviewers in the Quarterly Interview Survey also work on the Diary Survey. A 3-day training session was held prior to the start of the Diary Survey and included concentrated instructions on concepts common to both the Quarterly and Diary Surveys as well as training for the Diary Survey. This was done to shorten the Quarterly classroom training which has traditionally been 5 to 7 days.

A 1-day supervisors' training session was held prior to interviewer training. Any additional supervisor or interviewer training will probably be combined with that for the Quarterly Interview Survey.

Data collection for the Diary Survey began in September 1979, using the product-structured diary which is composed of a similar set of adjoining pages for each day of a 7-day period. The pages are structured by product category headings to aid the respondent in recalling and recording purchases made during the week.

Data Processing

When Diary questionnaires are returned to the offices, they are checked in and reviewed for completeness and consistency. Any problems are immediately referred back to the interviewer. The questionnaires and diaries are then shipped to a centralized processing location (Jeffersonville) for coding and data keying. The data is then transmitted to Washington where a computer preedit is performed to identify errors resulting from coding and keying. Any remaining errors are corrected and rekeyed.

A computer edit for consistency is then performed along with the imputation of missing or invalid demographic or work experience data. As in the Quarterly Interview Survey, adjustments are made to include sales tax. However, since the respondent is asked only to report personal expenditures, no adjustments have to be made for reimbursed expenses. The data is then weighted. Monthly data tapes will be provided to the BLS beginning in February 1981.

Point of Purchase Survey

Data Collection

Data collection procedures are similar to those used in the 1974 survey, with modifications to the item checklists and to the recall periods for certain expenditure items. The survey is conducted over a period of 4-6 weeks beginning in April of each year. Each consumer unit is interviewed once. The field work takes place under the direction of the Bureau of the Census' permanent field staff operated through the 12 regional offices and the Washington field staff.

Each staff member is thoroughly trained prior to beginning work on the survey. Interviewers will receive 2 days of classroom training plus 8 hours of self-study training materials. Formal training is

supplemented by on-the-job training as necessary to ensure job performance at a level of established standards.

The interview itself consists of two major parts. In the first part, the interviewer elicits information on the demographic and socioeconomic characteristics of the household. Then the interviewer conducts the second part of the interview, which consists of the checklist and questionnaire administered together. The checklist, a list of consumer goods and services, is sectioned into a number of different recall periods. The respondent identifies the items on the checklist which the household has purchased in the stated recall period. Then, the interviewer requests information about all outlets at which a particular type of purchase was made. The name and location (including the closest intersecting street) of each store at which items were purchased and the costs of the items are collected. The interviewer then proceeds to the next items with the same checklist/questionnaire procedure, until the interview is completed.

There are two separate checklists, but only one is administered to a given consumer unit. The checklists are split to reduce the reporting burden for each consumer unit. Unlike for the 1974 survey, there is considerable overlap between the two checklists with respect to the items included. The recall period on each of the checklists varies from 1 week to 5 years depending on the relative frequency and expense of the purchase. A shorter recall period is used for the more frequently purchased items, such as foods and gasoline, whereas a longer period is employed for less frequently purchased and more expensive items, such as major equipment items or vehicles.

At the end of the day's interviewing, the interviewer codes as many of the store names obtained during the day as possible. An alphabetized, precoded computer list of retail, wholesale, and service outlets in the area is provided for this purpose. When possible, the interviewer matches the exact store name and address obtained in the interview to one on the list and codes the outlet accordingly. Those outlets which cannot be coded from the list are coded in a centralized operation at the Bureau's Jeffersonville operation's office.

Data Processing

In the precomputer processing, all store names not coded in the field are coded. A card file is created to code stores which do not appear on the coding list. The processing of the final Store Code List comprises a major portion of the data processing. BLS has requested that outlets retain the same code from one year to the next. This involves matching codes for two lists compiled 5 years apart, since each PSU is to be interviewed only once every 5 years.

The remaining precomputer and computer processing is similar to that performed in the 1974 survey. After check-in, general and specific coding and verification, data keying, precomputer editing, and reject processing, the computer processing will take place. This involves reformatting and editing, matching coded stores back to the original list to obtain additional geographic detail, and weighting the final records. A microdata tape is delivered to BLS by September 30 of each year.

Preliminary Results

Currently, the only results available from the Quarterly Interview Survey or the Diary Survey are

field response rates, inasmuch as the computer processing system is still in the developmental stages. The Point of Purchase Survey is conducted only for internal use by the BLS, so again only data on the collection activities are available.

Quarterly Interview Survey

Response rates in the first 9 months of the survey are quite pleasing and compare well with the 1972 survey. Tables 1 and 2 show response rates by month and by quarter for the current survey and the 1972 survey respectively.

Although the initial response rate of 90.2 percent for the first quarter is somewhat lower than 1972, by the end of the third quarter, the difference is less than one-half percent. In the third quarter, the response rate of 88.6 percent is only 0.4 percentage points lower than 1972. However, it should be noted that the current survey has rotation groups at various interviews, i.e., first through third interviews, whereas in 1972, all were in their third interview. But given the general decline in response rates in the past few years, the current survey would appear to be doing well.

Diary Survey

Response rates during these first several months have been considerably lower than expected. Although they are comparable to those obtained in the 1972 survey, they are considerably lower than those obtained in the 1973 survey. (See Table 3)

During the first 4 months (September-December), the cumulative response was 78.2 percent, with the rates ranging from 93.4 percent to a low of 59.0 percent among the 12 regional offices. The highest response rate for any week during this period was 84.9 percent for Week 10, and the lowest rate was 58.2 percent for week 17.

Since the sample is an annual sample, the weeks start again with the beginning of each year and response rates are cumulated on a calendar basis. As of Week 21 for 1980, the cumulative response rate is 82.8 percent with the highest rate reported being 90.9 percent in Week 12 and the lowest rate reported being 69.4 percent in Week 21. There is an overall improvement of about 4.6 percentage points in the 1980 work to date over the rate obtained in the 1979 period, with most of this improvement being noted during the last several weeks.

One of the main reasons for the improvement in the response rate is the increased emphasis and stressing the importance of the Diary Survey to the regional office staff. In addition, interviewers are requested to contact the offices earlier in the placement week when they encounter refusals, that the supervisors must make a better effort to convert refusals, and that we keep constant watch over the rates, getting back to problem offices on a timely basis. Probably the most important effect of our efforts with the field staff has been their placing more effort on contacting households within the allotted time.

When the supervisors were brought together in February, special emphasis was put on the noninterview problem. Supervisors from offices with low Type A rates spoke on methods they use to deal with the problem.

Point of Purchase Survey

The Point of Purchase Survey has just completed its fourth year.

A comparison of response rates for 1974 and 1977-80 are shown in Table 4. The 1980 response rate of 94.9 percent is only 1.6 percentage points lower than when first conducted in 1974. Almost all of the change can be observed in a corresponding increase in the refusal rate.

TABLE 1
Response Rates by Month and Quarter
Quarterly Interview Survey

Month/Quarter	Response Rate	Noninterview Rate	
		Total	Refusal
October-December 1979	90.2	9.8	7.5
October	89.8	10.2	7.8
November	90.9	9.1	7.1
December	89.8	10.2	7.7
January-February 1980	88.5	11.5	9.7
January	87.1	12.9	10.6
February	88.5	11.5	9.9
March	90.0	10.0	8.7
April-June 1980	88.6	11.4	10.4
April	87.3	12.7	11.7
May	89.2	10.8	9.8
June	89.1	10.9	9.6

TABLE 2
Response Rates by Month and Quarter
1972 Consumer Expenditure Survey (Quarterly)

Month/Quarter	Response Rate	Noninterview Rate	
		Total	Refusal
January-March 1972	94.9	5.1	3.3
January	94.3	5.6	3.3
February	94.8	5.2	3.3
March	93.6	6.4	4.6
April-June 1972	91.3	8.7	6.7
April	91.7	8.3	6.7
May	91.8	8.2	6.5
June	90.4	9.6	7.1
July-September 1972	89.0	11.0	8.3
July	88.9	11.1	8.0
August	89.4	10.6	8.4
September	88.6	11.4	8.6

TABLE 3
 Diary Response Rates by Quarter
 Comparison of 1972 with Current Survey
 Diary Survey (Week 1 Pickup)

Survey Quarter	Response Rate			Noninterview Rate					
	1972	1973	Current Survey	Total			Refusal		
				1972	1973	Current Survey	1972	1973	Current Survey
1	74.2	90.0	78.2	25.8	10.0	21.8	10.5	5.8	9.9
2	77.7	89.7	83.1	22.3	10.3	16.9	11.8	7.4	9.3
3	93.9	91.0	<u>1/</u>	16.1	9.0	<u>1/</u>	8.8	6.0	<u>1/</u>

1/ Through Week 21, full quarter results not available.

TABLE 4
 Response Rates by Year
 Point of Purchase Survey

Year	Response Rate	Noninterview Rate	
		Total	Refusal
1974 <u>1/</u>	96.4	3.6	2.4
1977	96.5	3.5	2.6
1978	94.4	5.6	3.8
1979	93.4	6.6	4.4
1980	94.9	5.1	4.3

CONSUMER EXPENDITURE SURVEY DATA: USES BY BLS

Eva E. Jacobs, U.S. Bureau of Labor Statistics

Abstract

Consumer Expenditure Surveys conducted by the Bureau of Labor Statistics provide essential data for BLS statistical series as well as for general economic analysis and making purposes. Use of these data by BLS in the CPI, expenditure tables, and Family Budgets as well as the differences in concept among the series are described.

The long history of consumer expenditure surveys in the United States reflects the variety of needs which they were designed to serve, which have generally been associated with the need for cost of living information. The inquiries over time have been directed to the welfare of immigrants in 1870, the influence of the costs of living on the cost of production in connection with setting tariffs in 1888, the incorporation of the data as weights in the Consumer Price Index and the use of the data in developing Bureau of Labor Statistics (BLS) Family Budget cost estimates. BLS has always published tables, and more recently tapes, of expenditure, income and asset data to permit users to track changes in consumption patterns and to examine the relationship among economic variables for different types of families. The words "different type of families" cannot be overemphasized because it is the provision of demographic characteristics associated with the economic data that is the major contribution of this survey. This is in contrast to the other widely used measure of consumption, the personal consumption expenditure component of the national accounts, which is an aggregate estimate of consumption of all families in the U.S. and does not tell us who is buying what.

The survey data are essential to several major BLS programs and I will describe how these uses affect the design of the survey.

For the Consumer Price Index (CPI), the data are used for the derivation of the expenditure weights of the components and for the selection of items to be priced. Since the 1977 revision, the CPI has been constructed for two populations. One is the total urban population, about 80% of the total non-institutional population. Derivation of weights for this group requires expenditure data and geographical identification. The other is for the traditional concept of urban wage earner and clerical workers, about 40 percent of the population. The determination of this group requires data on geography, occupation, employment status, and income because the group is defined as limited to certain occupations based on the proportion of income derived from these occupations.

The particular definitions of the CPI components determine to some extent the kinds and number of questions that are asked in the survey. It is not enough to ask the respondent for his monthly mortgage payment although the response may well be reliable and might be sufficient for certain types of analyses of outlays. Homeownership cost in the CPI includes the net cost of houses purchased in the base period and the interest cost contracted for over half the term of the mortgage. In order to estimate these weights it is necessary to know the price of houses purchased and sold, the interest rate contracted for, number of years, down payment, etc.

Some components require additional detail because of experience with price movements. Telephone charges provide an example. Since intra-state and interstate prices may move differently, expenditures for long distance and local service are treated as separate items. It is not sufficient to know what the total telephone bill was for CPI purposes.

The requirements of the CPI change over time and the change in the 1978 procedure for selecting the detailed item to be priced for the CPI has allowed us to reduce some of the detail collected. Prior to the revision, BLS pricing agents were given detailed descriptions of items to be priced. These items were selected from the detail collected in the survey. In the new disaggregation process the agents have more general descriptions to choose from. For example, previously the pricing agent was given a specific type of toaster to price - all agents attempted to price the same specification. Now the agent is provided with the item "small kitchen electrical appliances." Through the disaggregation process, which involves the relative sales of items in the store, the pricing agent selects the specific item, a toaster, in one outlet, a can opener in another. Since this procedure was introduced after the 1972-73 survey, the question was changed for the continuing survey from asking about a long list of items, all of which are frequently responded to with "no", to a question that asks about small kitchen appliances.

This may be satisfactory for the CPI but may prove to be a great disappointment for marketers of small electrical appliances. This is part of the constant conflict between the desire for more and more data and the desire to reduce respondent burden. Other requirements for the CPI have added new kinds of questions to the continuing survey. With the 1977 revision of the CPI, a separate point of purchase survey was introduced from which the sample of outlets which agents visit to collect prices for most components is drawn. For certain components, such as telephone and gas and electric utilities, the name of the company supplying the service is now obtained in the expenditure survey. The responses will form the basis for selecting outlets for these components in the future. However, point of purchase information may not be made available for individual outlets because of CPI confidentiality requirements.

The second major activity in the Bureau concerns use of the data for general economic analysis of consumption. The Bureau has published a series of tables which shows selected characteristics, expenditures, income and assets and liabilities for all families and families classified by income, age, family type, region, etc. From these for example, one can compare the consumption patterns of the elderly with the general population, or the prime working age group. Or one can compare one earner husband-wife families with two earner families. For this use even more data are required than for the CPI since the CPI does not cover income taxes, social security taxes, life insurance, cash gifts and contributions. All of these are included in a family's outlay and clearly influence the amount that will be spent for consumption items. Assets and liabilities come into play since borrowings or asset depletion are used to finance expenditures and loan repayments and additions to assets constitute additional outlays.

The methodology and format of the 1972-73 survey were markedly different from past surveys and these new features had a considerable impact on the process of developing the published tables. In earlier surveys, the household's income, total expenditure and assets data were collected through an annual recall procedure. The 1972-73 survey provided for two surveys -- one, consisting of a quarterly interview of one sample covering major expenditures and another, consisting of a separate sample keeping a two week diary from which frequently purchased components such as food detail and household supplies could be obtained. The new form had some inherent problems. One complication was that the time periods covered in the two surveys were not the same, although there was an 18 month overlap. That was finally resolved through a deflation procedure for the diary portion to provide an estimated two year period to match the interview. There was an additional difference in the time period covered by the income question. Since income is one of the most important variables in the tabulations, it was necessary to make an adjustment to make the income variable in the diary comparable with the interview. These adjustments having been made, the two surveys were integrated by groups of consumer units. For example, the food component of diary families of four persons were combined with interview families of four persons to form the four person family column in the integrated table showing expenditures by family size.

What was learned about changes in consumption shares from 1960-61 to 1972-73? The major movement was away from food and towards transportation for all segments of the population. The exception was that the food share increased at the lowest income level, probably as a result of the food stamp program. These were the expected results, since food shares have been declining since the 1930's, but the extent of the change in transportation was greater than expected. For all consumer units, total food declined as a share from 24 percent to 20 percent even though expenditures for food away from home were increasing. While for the total population this was partly the result of the decrease in family size, the decline was significant for all family sizes. The increase in the transportation share from 15 to 21 percent is associated with more cars per family as well as increased use. This trend held across the board for all types of families. The fact that the share for transportation now equals or exceeds that for food, is a graphic confirmation of the importance of the automobile to the American lifestyle.

In addition to published tables, we make available micro data tapes. However for the tapes the problem of the two surveys has not been solved -- separate tapes are produced for each survey. For many users who wish to analyze a particular component such as gasoline, there is no difficulty -- the interview tape is self-contained with all its associated characteristics. For researchers wishing to have a micro record of all expenditures, however, there is a problem, for matching individual diaries to individual interviews is much more complicated than matching by group. I don't know of anyone who has undertaken this activity as of now.

In BLS, expenditure data also have been used in the construction of the 4-person family and retired couple's budgets. These are hypothetical constructs that estimate the amount of income that is required to live at three specified levels of living -- lower, intermediate and higher. The 1960-61 survey was used as a basis for determining the long list of goods and services which constitute a significant portion of the budget estimates.

There has been a good deal of criticism of the methodo-

logy of the current budgets and at the very least, the market basket should be updated to reflect 1972-73 data.

A committee of experts was appointed to advise BLS on revising the budget program. The committee's report is now being reviewed. The major recommendation was to establish the level of adequacy as the median expenditure of the 4 person family as being a reflection of the prevailing American standard of living. If the recommendations in the report are adopted, the Bureau's new Continuing Consumer Expenditure Survey will provide the basis for establishing this standard in the future.

I have briefly described three data series published by BLS which measure levels of, differences in or changes in what is popularly known as living costs, all of which use Consumer Expenditure Survey data. Furthermore, because all three series provide data for selected areas as well as the U.S., they are widely used to satisfy the great demand for information on differences in living costs among areas. Since the differences in what each is designed to measure are not always clear to users the data are frequently misinterpreted. I will briefly describe the differences in concept among the series.

First, the family budgets, because it is the series most widely used for interarea comparisons of living costs. These costs represent differences in cost of living among areas at a point in time. The budget market baskets are not the same in all areas -- there are variations in many of the components. These have not been changed since 1967, but the consumption costs are updated through time by the components of the CPI for the particular area. Therefore, only the change in prices since the base period affects an area's relative position in subsequent periods.

The CPI measures the change in prices over time in each area and is often considered to be a measure of the change in the cost of living over time. A more rapid increase in the CPI between any two periods of time in one area versus another does not mean a higher cost level in the second period but simply a more rapid escalation of prices in one area. However, if one area's CPI increases more rapidly than another over a long enough period, its living costs will eventually reach a higher level even if those costs started at a lower level.

The published tables from the Consumer Expenditure Survey for selected metropolitan areas describe what people actually spend in those areas. The differences in expenditures for an area result from different characteristics of the populations as to income, age, ratio of homeowners and renters, price levels and taste. Thus, we are measuring what people do in these areas rather than how living costs vary.

In addition to the basic difference in what these series are designed to measure, there are differences in definition of particular components, most notably in the treatment of homeownership cost. This component has recently been receiving a good deal of attention in connection with the CPI. While homeowner costs are a very large component of living costs, the concepts of measurement and comparability are difficult and there does not appear to be a consensus as to the "correct" way to measure such costs.

Since the family budgets are supposed to represent an established family, they include the cost of owning a home purchased six years earlier, so that mortgage payments and interest costs are calculated for a mortgage obtained six years earlier than the budget year. This clearly is not relevant to someone just moving into the

area or a young family purchasing a home for the first time. The CPI, on the other hand, includes as a weight in its market basket the net purchase of home (in the base period) and the interest costs for half the term of the mortgage for those consumers entering into mortgage agreements in the reference period. These are then moved by changes in house prices and interest rates. These changes in the housing component clearly are not relevant to homeowners who purchased home years before. In our publication of data from the expenditure survey, we present still another concept of homeownership costs, this time in terms of the total household current account and balance sheet. The current costs for all homeowners for interest, taxes, insurance and maintenance, are included in current consumption but principal payments are considered repayment of a loan and are shown as a reduction in liabilities with other loan repayments. This concept of consumption costs might not suit someone who wanted to know how much they needed to earn to cover current outlays.

Users need also to be aware of the concepts and coverage of the expenditure survey data in making comparisons with other data sources. The comparison most frequently made is between the survey data (CEX) and Personal Consumption Expenditures (PCE) in the national accounts. Such a comparison should not be made without adjusting for differences in concept which affect the shares, the aggregates and the trends over time. Significant differences exist in the housing, automobile, health care and financial components. For example, rent in PCE is space rent, in CEX contract rent; homeowner costs in PCE are estimated from rental equivalent values, in CEX an outlay approach is shown; new automobiles are gross in PCE, net of trade-ins in CEX.

These cautionary words are intended not to discourage but to enhance understanding of the data and thereby help make analytical results more reliable. While the surveys provided essential information for government policy making and statistical activities from the start, marketing interest in the data has been more recent -- from the 1960's, when Life Magazine funded an extensive publication produced by the Conference Board based on the 1960-61 data. For the 1972-73 survey we have sold tapes to over 300 universities, individual companies and other research workers in addition to thousands of copies of the published bulletins. We anticipate even wider use of the results of the continuing survey than of the past sporadic surveys.

The Continuing Consumer Expenditure Survey is not in the field and we are looking forward to getting the first tapes from the Census Bureau. Our review and publication plans are being formulated now. We plan to release regular quarterly and annual publications based on the integrated diary and interview data. However, because of the infrequency with which individual items are purchased and the smaller sample size, the quarterly expenditures will be highly aggregated and the number of characteristics quite limited. On an annual basis, both the detail and the number of classes of classifying characteristics can be increased. Deriving reliable data may require aggregating two or more years of data. In addition to providing timely data on consumption patterns, the survey data will eventually be very useful in evaluating the need for and timing of revisions of the CPI market basket.

We also plan to make public use tapes but specifications of content or frequency are not yet available.

It is not until we review the data that we will be in a position to recommend changes in the questions or the

methodology. We hope to introduce a research panel on which proposed major changes will be tested but that is for the future. The goal is to provide useful, reliable and timely information.

AN EVALUATION OF THE METHODOLOGY OF THE
1972-73 U.S. CONSUMER EXPENDITURE SURVEY ^{1/}

Robert B. Pearl, Survey Research Laboratory
University of Illinois

Abstract

This paper evaluates the accuracy/quality of various types of expenditure data collected in the 1972-73 Consumer Expenditure Survey. Recommendation for changes in the methodology - some of which already have been implemented by the Census Bureau, are made.

Introduction

Collection of data on consumer expenditures is probably the most widespread of all statistical endeavors in this country, engaging both the public and private sectors. Such information is, of course, virtually the lifeblood of the market research industry. The Government is also heavily involved in this field for a variety of purposes. Expenditures by consumers constitute one of the principal components of the Gross National Product accounts. Data on outlays for food, housing, and other living costs are an essential ingredient in establishing benefit levels for social programs. Perhaps most important, this kind of information provides the technical framework for construction of consumer price indexes.

Although the objectives and type of detail may differ a great deal among these various public and private undertakings, many of the technical problems are common to all. In particular, nearly every survey organization has experienced serious reporting errors and biases in attempting to collect information from consumers about their purchasing behavior. This is one reason the findings from the large-scale Government expenditure survey conducted in 1972-73 should be of general interest to practitioners in this field and to others who use or could use similar techniques.

The 1972-73 Survey comprised two major components:

(1) An interview panel consisting of about 10,000 households each year which was visited on a quarterly basis primarily to obtain the larger items of expenditure and certain repetitive items (rent, utilities, etc.) Particular categories were covered either quarterly or on a semi-annual or annual basis, depending primarily on expenditure size.

(2) A diary operation consisting of about 200-250 households per week asked to keep a diary or record of all expenditures for the subsequent, two-week period. Although the main focus of the diary was the smaller items of expenditure, the fact that all categories were covered provided various options in compiling estimates as well as many research opportunities. For further information on the methodology used for this survey, see other papers presented in this session.

^{1/} This work was undertaken under a Joint Statistical Agreement between the Survey Research Laboratory and the Research Center for Measurement Methods of the Bureau of the Census.

Purpose

Because of the critical uses which are being made or which could be made of the data and the marked changes in methodology, considerable attention has been devoted to evaluation of the survey results. The objectives have been to assess the adequacy of the survey procedures in general, the relative merits of using interviews as compared to diary keeping for particular categories of expenditures, and the implications of the findings with respect to improvements in methodology for future use in either a continuing survey program or other similar undertakings. Since there was little prospect that all of the methodological issues would be resolved, a determination of further research needs was also an important goal.

This paper presents a summary of the findings and conclusions deriving from this evaluation effort. Detailed results appear in publications in the Census Bureau's Technical Paper Series (Nos. 45 and 46).

Summary of Findings for Expenditure Categories

The following is an attempt to assess the adequacy of the expenditures data obtained in the 1972-73 survey for the various categories of goods and services. The general approach used in this appraisal has been to compare the estimates from the quarterly panel with those from the diary operation, where the same subject was covered in both, and to relate either or both to various independent sources of expenditure data. The principal objectives are to assess which of the survey procedures appeared to be more effective for particular categories of expenditures and to determine what types of improvements and modifications may be suggested by the results. The conclusions can only be tentative because of major uncertainties about the validity and comparability of the independent data used as a standard and because adequate detail was often unavailable to explore the subject in sufficient depth. Nevertheless, in a substantial number of cases, persistent patterns emerged across category lines which pointed in rather specific directions.

The most frequently used of the independent data sources are the Personal Consumption Expenditure (PCE) estimates prepared by the Department of Commerce in conjunction with the Gross National Product Accounts. The other independent data derive mainly from Government administrative, census, or survey sources although some private sources are also used. In some sectors, such as education and health, the conceptual differences between the PCE and the 1972-73 survey data were such that dependence had to be placed entirely on the other independent sources for comparative purposes.

It should be noted that the survey results used in this evaluation are derived from special tabulations of re-weighted original data tapes. They do not reflect editing charges which may have been made at later stages of processing by BLS. As a result, the figures may differ somewhat from those already published or to be published by that Agency or which may be compiled from the public-use data tapes recently issued. Certain differences in time reference and conceptual approach would also contribute to the disparities.

Table 1 presents a summary of the findings for the various expenditure categories. For purposes of summarization, a number of the detailed categories have been combined and averaged. The table designates the "best" survey source, that is, the one generally closest to the independent data, in cases where the two survey estimates are significantly different. The ratios of the "best" survey estimates to the independent estimates are indicated in terms of broad class intervals, allowing insofar as possible for conceptual differences between the sources, but the actual computed values are also provided. Although the magnitude of these ratios may provide some indication of the adequacy of the survey estimates, the margins can be regarded only as rough approximations, for reasons stated above, and have been used primarily for purposes of detecting any consistent and meaningful patterns which have a bearing on the methodology.

1. Food and beverage expenditures--After allowance insofar as possible for various conceptual incomparabilities, there appeared to be a reasonably close correspondence between the diary estimates of food purchases for home use and the independent sources. The fact that the homemaker--the usual diary keeper for the family--is ordinarily responsible for most of the purchases was undoubtedly a positive factor. The allocation of maximum space on the diary record to this expenditure class probably contributed as well.

There were considerable disparities, however, in the precision with which various food categories were reported. The reporting was apparently most complete for relatively costly items, such as meat and poultry, and for those used promptly and on a daily basis, such as milk, eggs, fruit juices, and bread and fresh-baked items. The coverage seemed to be considerably less complete for food staples which are bought less frequently, with each purchase used over a considerable period of time. One of various possible explanations for these differences is that many respondents may not start keeping their diaries promptly--or do not make entries, as requested, on a daily basis--but later attempt to reconstruct the omitted periods by memory. In doing so, items which represent the main course in a meal or which are purchased and used relatively frequently might be more readily recalled.

A less anticipated finding was the close correspondence between the survey and independent estimates for meals in restaurants or other eating places, where a substantial proportion of the outlays would be made by individual family members other than the homemaker. The prominent positioning of the section for reporting purchased meals on the diary record and some emphasis to this subject at the time of the diary checking procedure might have contributed to this outcome. At the same time, the marked deficiency for alcoholic beverages confirms the continued failure of household surveys to measure a sector where there is considerable sensitivity about reporting.^{2/}

2. Small expenditures other than food--For various small expenditure items other than food, for which the diary was the principal if not only source, a predominant factor appeared to be the role of the various family members in making purchases. Where the responsibility was principally that of the homemaker, such as for laundry or cleaning products or household services, the reporting appeared to be considerably more complete than in cases

^{2/} Some special check questions, asked at the time diaries were collected in an effort to overcome this deficiency, provided data that would have increased the survey estimates by about 15 percent (no actual correction was made, however). Even on that basis, the survey aggregates would have amounted to only about half of the independent levels.

where other members were substantially involved, as for toiletries or hair care. Even for those expenditures where the homemaker predominated, however, the reporting appeared to be generally less adequate than for food purchases, probably partly a reflection of the much smaller amount of space and attention given to non-food items on the diary record. Some limited efforts to measure small non-food expenditures through summary questions in the quarterly panel did not appear to be especially productive, although in one or two cases the figures appeared to be more complete than the weakest of the diary estimates.

3. Clothing expenditures--As was anticipated to some extent, this expenditure category represented one of more troublesome sectors, with neither survey source exhibiting any clear cut overall advantage and neither corresponding very closely with the independent data.^{3/} Following the pattern observed throughout the analysis--and expected from previous experience--the larger items (suits, coats, etc.) were apparently more adequately reported and the quarterly panel emerged as the superior source in this case. Also not surprisingly, the diary procedure represented the "best" source for a diversified category such as accessories, where it was probably difficult to communicate the full range of items in an interview procedure. For no apparent reason, the diary estimates also provided the closer correspondence with the independent data for footwear, although this subject was probed in much greater detail in the quarterly panel.

For the broad range of middle and lower priced clothing products, the advantage seemed to alternate between the two survey sources, without any consistent relationship to the importance of the item. One problem which complicated the appraisal--and which extended to most other expenditure classes as well--was the existence of a large residual clothing group in the diary estimates, consisting mainly of incomplete or inadequate entries which could not be assigned to specific categories.

4. Household appliances--Reporting of expenditures for major appliances was apparently one of the more successful survey outcomes. The high cost associated with these items and the more probing nature of the "inventory" technique, described earlier, probably contributed to this favorable showing.^{4/} There was some evidence in favor of pooling the two estimates obtainable for a given year under this approach, the one based directly on the survey for that year and the other derived indirectly from the initial inventory in the survey for the following year. Comparison between those two sets of estimates indicated no significant differences for several of the major appliance groups and no evident superiority of one over the other in relation to the independent data. Pooling of the estimates would have roughly the effect of doubling the sample size at virtually no additional cost, a considerable gain for items with especially large variances.

The picture was considerably less favorable for small appliances, which would be consistent with the findings for other product categories. The inventory approach is

^{3/} Clothing expenditures were covered each quarter in the interview panel. Some 16 broad categories were specifically probed, using detailed item checklists for each. A small section was provided for clothing expenditures on the diary record and it was, inadvertently, located in an especially inconspicuous position on the form.

^{4/} The inventory was updated only once, at the final quarterly visit, for large appliances. The updating occurred twice, at 6-month intervals, for small appliances.

apparently less successful where a multiplicity of small products is involved (such as for small kitchen appliances) or where the family may possess several articles of the same kind (e.g., cameras or radios). An interesting by-product of the inventory approach in these instances, however, was that estimates which included the value of gifts received from outside the household--which derived directly from the inventory on items on hand--invariably exceeded those which included the cost of items bought by the family as gifts to be given to others, which was based on their recollection of purchases. This finding suggests that the inventory method, based on items in the family's possession, has certain advantages over the more conventional approach which requires recall of previous actions.

5. Household furnishings--This broad category provided a rather clear-cut demonstration of the relationship between the size of an expenditure and the likelihood of its being reported.^{5/} The closest correspondence with the independent estimates was found for furniture, the most costly class, followed by the next most significant group--floor, window, and furniture coverings. In both cases, the quarterly panel appeared to be the superior source, partly on the basis of sampling variances. The survey estimates fell considerably short for household linens and especially for smaller products such as dinnerware and cookware, luggage, and decorative items. In these latter instances, the diary estimates, although themselves deficient, appeared to be at least equivalent in coverage to those from the interview panel.^{6/}

6. Automobile and vehicle expenses--In most categories, the survey data corresponded closely with the independent sources. The main disparities were for tires and other accessories. Although the homemaker would normally have less responsibility for vehicle expenses than most others, the diary estimates held up surprisingly well for some of the smaller categories, especially gasoline. The quarterly panel, as expected, was clearly the superior source for vehicle purchases and also appeared to provide the more reliable data for vehicle registration, insurance, and similar items.^{7/}

7. Housing expenditures--Once again, with the exception of one rather diffuse category (fuel purchases), the survey and independent estimates corresponded rather closely.^{8/} Also, in the main, the diary-based data, again

^{5/} Most items in this class were covered at 6-month intervals in the quarterly panel. The exception was household linens which were covered each quarter. A small section was provided on the diary form for "housewares, furnishings, hardware, and garden supplies."

^{6/} In fact, for the dinnerware-cookware category, the diary estimates were appreciably higher and, thereby, closer to the independent levels.

^{7/} The "inventory" approach was used for vehicle purchases in the interview panel, with the inventory updated only once at the final quarterly visit. Questions were asked each quarter, however, for gasoline purchases and vehicle maintenance and repair. A section was provided for "gas, oil, tools, parking fees, and other vehicle expenses" on the diary form.

^{8/} A rather complex set of procedures, including use of the "inventory" approach in some cases, was used to determine housing expenses in the quarterly panel. For home repairs and maintenance, detailed questions (adapted from the Census Bureau's special survey on this subject) were asked each quarter. Utilities were also covered each quarter using the "last payment" approach (asking for the amount of the last bill and the period it covered). A section for "rent, utilities, fuel, phone, insurance" was provided on the diary record.

unexpectedly, matched those from the interview panel. The exception, in the case of mortgage payments, may have resulted from the fact that some such disbursements are made directly through bank accounts, which could easily escape the attention of the diary keeper.

8. Health expenditures--One of the more pleasant surprises was the relatively close correspondence between the survey and independent estimates for most health expenditures, usually considered to be one of the more treacherous areas in this kind of undertaking.^{9/} A less optimistic reading might be that the survey results were at least as good as most previous endeavors of a similar nature, without attempting to categorize their accuracy in an absolute sense. In any event, the diary procedure once more provided the most unexpected outcome, in that not only for small items such as drugs and medicines, but also for most professional health services, the data seemed to hold up surprisingly well. An exception was for hospital services but even the quarterly panel data in that instance seemed somewhat deficient, possibly because of complications introduced by the pervasive role of health insurance and other third-party payors.^{10/} Also, for health insurance premiums themselves, where the interview panel data appeared to be reasonably adequate, the diary procedure was understandably ineffective, in view of the fact that payments are often made through payroll deductions which would rarely come to mind in completing a household diary.

9. Education, travel, and miscellaneous expenses--Because of the diversity of items, it is more difficult to summarize for this remaining expenditure group.^{11/} Although the evidence was limited, there appeared to be some deficiency in reporting of education expenditures, possibly because of the difficulty of adequately covering college students in a household survey. The quarterly panel still appeared to be the better source for the larger payments (tuition, etc.), but it seemed desirable to alter the survey procedures for students attending college away from home, whereby they would be interviewed directly at their college quarters instead of obtaining proxy information, in some instances, from their parents at home. No direct comparisons were possible for other school expenses, but it appeared likely, based on other findings, that the smaller expenses--books, school supplies, meals outside of school, boarding arrangements, etc.--as well as related items, such as recreational lessons, might better be approached through the diary procedure.

^{9/} Detailed inquiries were made in the interview panel at 6-month intervals for health expenditures. A section for "personal care, drugs, and medical supplies" was provided on the diary form but other medical expenses would presumably be entered in the catch-all section.

^{10/} The survey objective was to measure "out of pocket" costs, net of insurance and other reimbursements.

^{11/} Education expenses were covered at 6-month intervals in the interview panel. Detailed inquiries were made each quarter on trips and vacations. Other items were covered at various intervals depending on expenditure size (but with a quarterly inquiry for watches and jewelry). Most of these items were not specifically mentioned on the diary record.

Comparisons with independent sources for travel and vacations could only be made in terms of the reported number of trips taken rather than directly for expenditures. On this basis, the quarterly panel data appeared to be consistent with other findings. Since the validity of expenditures data may depend more on whether or not an expenditure is reported than on the precise amount of the expenditure, even this limited finding with regard to trips taken could be meaningful.

For the remaining subjects--transportation costs and miscellaneous products and services--a few generalizations may suffice since the patterns appeared consistent with those already cited for items of a similar nature. Public transportation expenses, which were available in detail only from the diary, were drastically underreported in relation to the independent sources, probably because such outlays are usually made on an individual basis. Among the miscellaneous categories, the larger items, such as pianos, organs, and funeral expenses, were much more adequately reported in the quarterly panel and the survey estimates agreed, in large measure, with the independent data. A large disparity for watches and jewelry could have resulted, in part, from sensitivity in reporting such expenditures, whereas a similar deficiency for moving expenses could reflect some undercoverage of recent movers in the survey.

Conclusions and Recommendations.

The survey findings, although predictable in a number of respects, carry some implications for further efforts in this field, not only on behalf of Government programs but possibly for market research and similar private endeavors. A rather unexpected result was the relatively effective performance of the diary procedure in some sectors, such as housing and health, where the quarterly panel has been assumed to be the only realistic option. The diary results also appeared to be about as effective as those from the quarterly panel in many cases where both sets were evidently deficient, including some situations where a great deal of attention had been devoted to expenditure reporting in the interview procedure.

A general rule of thumb suggested by these findings might be that, unless a clear-cut reason existed for using an interview procedure such as the quarterly panel, dependence might better be placed on diary-keeping for purposes of expenditure data collection. The diary might be a dubious source for items with exceptionally large variances (vehicles, appliances, furniture, etc.) or where unusual payment arrangements might require special questioning, such as insurance paid through payroll deductions, mortgage payments made automatically through bank accounts, hospital bills paid largely but belatedly by insurance, etc. The diary might also be inferior to an interview process in cases where a composite set of questions may be necessary in obtaining complete responses as may be the case, for example, in cataloging trips and vacations. In other instances, the diary procedure appears to be at least as good a risk as the interview approach, and probably a less costly one as well.

A number of modifications and improvements in the diary procedure are clearly necessary, however, in order to overcome some rather evident deficiencies. Some possible steps in this direction are the following:

1. Limit the range of items any one family would be asked to report--The use of a diary covering all items of expenditure, as was done in the 1972-73 Survey, may have certain theoretical benefits, but considerably proscribes the ability to improve the overall process

sufficiently to satisfy the expanded requirements just cited. Evidently, as stated earlier, one of the reasons for the more successful coverage of food purchases than other small items in the 1972-73 diary was the much greater amount of space and attention accorded the former. Moreover, for all categories including food, there was a considerable undifferentiated residual group, resulting mainly from incomplete or inadequate entries which could not be classified in detail, which detracted materially from the usefulness of the results. The general lack of space and the inability of interviewers to focus on so wide a range of items in reviewing the diaries probably largely accounted for this latter deficiency.

Most market research endeavors and similar undertakings are already restricted in terms of item coverage. However, since exhaustive detail is often requested about specific items (type of packaging, package sizes and weights, brand names, ingredients, marketing features, etc.), it is possible that better results would be obtained if the samples were split and particular respondents were asked to report on only a subset of products. For more comprehensive projects, such as the 1972-73 Survey, where the amount of detail is more limited, a possibility would be to split the sample along category lines. One subsample, for example, might be asked to report only on food and the supermarket products, a second on clothing and household linens, a third on health expenditures, etc. There is obviously some practical limit to the number of subsamples that could be simultaneously operated and a good deal of thought and some experimentation would be necessary to derive a workable plan.

Probably even more important than space considerations, the use of this kind of specialized approach would make it feasible to provide for a more focused set of check questions and procedures at the time of diary pickup to overcome some of the disparities noted in the present survey (such as underreporting of certain food items relative to others). In fact, a modified procedure would likely entail much more of a combination of interviewing and record keeping than is now the case. A good deal of developmental work would be required in devising effective checking procedures.

2. Vary length of record-keeping periods--As previously noted, the 1972-73 Survey provided for two weeks of record keeping for each sample family, covering all items of expenditure. In most private undertakings also, the practice is to use a standard reporting period for those products which are covered. If specialized subsamples are developed as proposed above, it is obvious that either a larger overall sample would be needed or much higher sampling variances would have to be accepted. One way out of this dilemma would be to vary the length of the record-keeping period depending on the variances of the subjects covered for a given subsample. For example, for a low variance category, such as frequently purchased food items, a period as short as a week might be adequate. For most categories besides food, however, such as clothing or health expenditures, record keeping periods of up to 3 months or longer might be considered. The fact that only a limited set of items is covered might reduce the reporting burden sufficiently to secure extended cooperation of this kind without an unacceptable attrition in the sample.

In such a system, the use of less costly collection methods already employed in many private endeavors, such as having respondents mail in completed diaries on a periodic basis (monthly, semi-monthly, etc.) would be more practical. Provision would, of course, have to be made for follow-up (by mail or telephone, where possible) for non-respondents or to carry out special checking for returns that did not meet prescribed standards.

3. Provide separate diaries, where indicated, for individual members--As noted, only one person, usually the homemaker, probably maintained the diary for the entire family in the 1972-73 Survey. Not surprisingly, the results were clearly more favorable for the kinds of expenditures for which the homemaker was mainly responsible than for those likely to be made by other members. One possible way of obtaining more consistent results, where the expenditures to be reported are of a more dispersed nature, would be to provide separate diaries for all family members above a certain age (perhaps 12 and over) on which to record their individual disbursements. For this purpose, the diaries could be briefer and less formal than the main record for the family. Experimentation with various versions would obviously be important in developing a procedure of this kind.

4. Providing monetary or other incentives for cooperation--Most market research endeavors provide various cash or non-cash incentives to promote greater cooperation. There is considerable reluctance to follow this course in Government surveys, although the approach is not unknown. In fact, an experiment with cash incentives was conducted in the early stages of the 1972-73 Survey to determine whether cooperation could thereby be improved. The results were inconclusive in this regard and the incentives were dropped from the procedure. As noted earlier, cooperation was high in spite of the absence of material rewards.

Since most studies have shown that the quality of reporting as well as the degree of cooperation appear to be benefitted by some type of material inducement, the matter of providing incentives should be seriously considered in any operation where there is primary dependence on diary keeping. The rewards should, of course, be attuned as closely as possible to achieving the main objectives. If, for example, it is important for individual family members to keep separate records, the incentives should be offered only if all such members agree to cooperate, which would presumably promote intra-family pressure for individual compliance. Similarly, if cooperation over extended periods is requested, a useful approach is to offer a small reward for each sub-period with a sizeable bonus for completing the entire cycle. Special consideration for prompt and complete returns by mail (where used) might be still another element. Payment of incentives might also be made contingent on retention by respondents of cash register tapes, bills, and other evidences of expenditures which could be consulted in reviewing the diaries.

5. Continue exploration of timing biases--Although not mentioned up to this point, perhaps the most conclusive survey finding was affirmation of the traditional bias found in diary operations, whereby a higher level of expenditures is reported in the earlier as opposed to the latter stages of the record keeping period. In a 2-week diary procedure, for example, the estimates for the first week are almost invariably higher than those for the second week. Differences of this nature were found for virtually every expenditure category in the 1972-73 survey. The margins of difference varied a good deal, however, and seemed to be almost random in nature. There was no evident relationship of the weekly differences for a category to other survey measures, such as the completeness of reporting relative to the independent estimates.

Since separate pages were provided in the diary for each reporting day, it was possible to examine the day by day recordings in this context. It was found that the higher levels for the first week diaries were primarily the result of exceptional amounts of expenditures recorded for the first reporting day (about 40-50 percent above the daily average). Levels were also above average

for the first day of the second diary week (8th reporting day) but not nearly to the extent found in the first week. There was a tendency for a gradual decline in expenditure levels in the course of each week. There were indications that this pattern persisted for most categories of expenditures.

Although various theories have offered for this phenomenon, the one that appears most likely on the basis of the evidence is that there is a certain amount of "telescoping" of expenditures on the first day. According to this theory, many respondents may not begin their diaries immediately, but skip the first day or two and then attempt to reconstruct the information from memory. In doing so, they may inadvertently include some expenditures made prior to the reporting period. Another suggestion that the patterns can also be attributed to some element of fatigue as the period lengthens can also be supported by the findings. Still another hypothesis to the effect that respondents may alter their buying habits temporarily because a diary is being kept does not appear to be supportable as a major element because the differences are so widespread among categories.

Among the suggestions made to control for this tendency is to extend the period of reporting somewhat (such as to 17 days for two-week diaries) and then to exclude the entries for the early and final days from the estimation. Another is to conduct a brief retrospective interview covering a day or two prior to the reporting period and to use that information as a "bounding" or control device in examining the early diary entries. The problem with such procedures is that, by reducing the impact of the exceptional diary levels at the outset, they would accentuate the tendency for understatement in the overall diary levels in general. The better approach might be first to attempt improvements in diary reporting through various of the possibilities cited earlier in order to be able to afford the use of any devices for controlling telescoping.

6. Explore use of universal product codes--A promising technological development that could affect future survey work in this field is the inclusion of "universal product codes" on most canned and packaged supermarket and drug store items (and likely to extend to many others). A useful experiment would be to ask respondents to record these codes in their diaries where available, as well as brief product descriptions, to assess how accurately this information is reported. If the effort is sufficiently productive, respondents might be relieved of the necessity of describing products in any detail where code numbers are available. More importantly, this step could result in far more accurate and consistent classification of products reported in surveys and a major reduction in the coding effort required at the data processing stage.

A related effort would be to make more effective use of the much more descriptive cash register tapes now being provided in automated checkout systems. If retained by respondents, they could certainly augment and, in some cases, take the place of diary entries.

In a comprehensive expenditure data program, such as required for purposes of the Consumer Price Index, there would be need for an interview survey capability, especially for high variance items, even if greater dependence were placed on a diary procedure. An interview approach would also be the indicated procedure in market research surveys related to durable goods and similar products. The main lesson from the 1972-73 Survey was that exhaustive interview surveys are likely to be self-defeating in this sector, and that the inquires at any one time should be limited to those products and services not amenable to the diary technique. The use and possible extension of the so-called inventory approach for

durable goods was also generally supported by the findings. Also confirmed was the notion of using variable time references in interviews depending on the size and importance of particular items.

One final matter might be mentioned which has a bearing primarily on comprehensive surveys such as those conducted by the Government, but which would be of interest to analysts of consumer behavior in general. The major recommendation made--curtailing the scope of the interview phase and introducing specialized diaries--would mean that only very partial expenditure data would be available for individual families, either for annual or shorter periods. Unless some counter measures were taken, there would be some evident limitation in the ability to analyze individual consumer behavior.

Actually, some of the largest items, which distinguish one family's expenditures from another's, would still be obtained by interview under these proposals for periods up to a year, but most categories would not be covered. A first step to close the gap might be to add a few questions, perhaps on an annual basis, to identify exceptional expenditures in other sectors, such as large orthodontic bills, major house additions, purchase of expensive paintings, and the like.

Beyond that point, a further although admittedly experimental step would be to obtain certain limited information in the interview procedure on behavioral patterns which would help in imputing data from the diary operation or another source to the individual family record. For example, in the health sector, questions could be asked on the frequency of doctor or dental visits, consistent usage of costly drugs, etc., which, together with demographic and socio-economic characteristics, could provide a basis for imputing a value for health expenditures to the family from another source. (The other source would, of course, also require the inclusion of such behavioral items.) Obviously, considerable research and testing would be needed before embarking on a large-scale endeavor of this kind.

TABLE 1

Summary of Findings for Expenditure
Categories: 1972-73 Expenditure Survey
Estimates Compared to Independent Sources

Category	"Best" Survey Source ^{1/} QP=Quarterly Panel D=Diary Operation N=No Significant Difference	Ratio of "Best" Survey Estimates to Independent Sources			Independent Sources Used ^{2/}	
		Best Judgment as to Range of Ratios (Allowing for conceptual and other differences between sources) ^{3/}	Survey Results			
			Actual Ratio of Survey to Independent Estimates ^{3/}	Estimated Standard error of Ratios		
Food Purchases		A= .9 or higher B=.8 to .89 C=.7 to .79 D=.6 to .69 E=Less than .6	Re-vised	Preli-minary		
<u>For home use</u>	D**	A, B ^{5/}	.86	.85	.01	ERS, PCE
Meat or poultry		A	1.04	1.02	.03	
Eggs		A	.98	.87	.03	
Fresh milk		A	.91	.93	.02	
Bread and other fresh baked items		A	.92	.98	.02	
Food staples		D, E	.60	.63	.02	
Fruits		B	.88	.85	.03	
Vegetables		C	.78	.74	.02	
<u>Purchased Meals</u>	D	A, B	1.05	1.11	.03	Cen-Bus PCE
<u>Alcoholic Beverages</u>	D	E	.34	.39	.04	Cen-Bus PCE
Small Non-Food Exp's	D**					PCE
<u>Homemaker's Purchases</u> (laundry, cleaning products)		B, C	.82	.72	.02	
<u>Dispersed Responsibility</u> (Toiletries, Film, Books)		E	.55	.57	.02	
Services						Cen-Serv PCE
<u>Homemaker's Purchases</u> (laundry services, household help)		A	1.04	1.04	.02	
<u>Dispersed Responsibility</u> (hair care, shoe repairs)		E	.54	.68	.02	
Clothing Expenditures	N	C	.73	.72	.015	Cen-Bus PCE
<u>Larger items</u>	QP	A	.98	1.08	.03	
<u>Medium and Smaller Articles</u>	N	D ^{6/}	.66	.76	.015	
<u>Accessories</u>	D	D	.72	.72	.03	
<u>Footwear</u>	D	C	.77	.79	.015	
Household Appliances	QP**					
<u>Major appliances</u>		A ^{7/}	.96	1.00	.02	
<u>Small appliances</u>		C ^{2/}	.75	1.01	.03	
Household Furnishings						
<u>Furniture</u>	QP	A	.91	.91	.04	Cen-Bus PCE
<u>Other large items</u>	QP	B	.81	.77	.04	Cen-Bus MRCA PCE

Category	"Best" Survey Source ^{1/} QP=Quarterly Panel D=Diary Operation N=No Significant Difference	Ratio of "Best" Survey Estimates to Independent Sources			Independent Sources Used ^{2/}	
		Best Judgment as to Range of Ratios (Allowing for conceptual and other differences between sources) ^{3/}	Survey Results			
			Actual Ratio of Survey to Independent Estimates ^{3/}	Estimated Standard error of Ratios		
<u>Household linens</u>	N	C	.77	.69	.02	Cen-Bus MRCA PCE
<u>Dinnerware, Luggage, etc.</u>	N	D, E	.59 ^{2/}	.50 ^{8/}	.02	PCE
Automobile and Other Vehicle Expenses						
<u>Vehicle Purchase</u>	QP**	A	1.01	1.01	.04	AHS, PCE
<u>Gasoline and Oil</u>	N	A	.98	.98	.015	PCE
<u>Tires and Accessories</u>	QP	C	.75	.88	.03	PCE
<u>Maintenance</u>	N	A, B	.91	.88	.03	PCE
<u>Insurance</u>	QP	A	.93	.93	.02	A.M. Best
Housing Expenses						
<u>Rent</u>	N	A	1.06	1.03	.03	PCE
<u>Mortgage and taxes*</u>	QP	A	1.02	1.02	.04	AHS
<u>Repairs and Alterations*</u>	N	A	1.04	1.04	.06	SURAR
<u>Utility Costs</u>	N	A	1.00	1.01	.03	PCE
<u>Other Fuel Costs</u>	N	B, C ^{9/}	.76	.61	.03	PCE
Health Expenditures						
<u>Hospital Services*</u>	QP	C	.76	.76	.04	CHAS, SSA
<u>Professional Services*</u>	N	A	.98	.98	.02	CHAS, SSA
<u>Drugs and Medicines</u>	D	A	.96	1.02	.07	PCE, SSA
<u>Medical Supplies</u>	D	C	.72	NA	.10	PCE
<u>Health Insurance</u>	QP	A	1.10	1.10	.03	HIS
Other Expenditures						
<u>Education Tuition and Fees</u>	QP	B ^{10/}	.82	.82	.02	NCES
<u>Trips and vacations</u>	QP**	A ^{11/}	1.02	1.02	.06	NTS
Public Transportation						
<u>Local (bus, taxi)</u>	D**	E	.44	.58	.06	
<u>Other (airplane, rail)</u>		D, E ^{12/}	.53	.85	.11	PCE
Miscellaneous						
<u>Large items (pianos, organs, funeral costs)</u>	QP	A	.96	.93	.08	
<u>Moderate Items</u>	QP	B, C	.81	.86	.08	
<u>Watches and Jewelry</u>	N	E	.48	.56	.10	
<u>Moving and Storage Costs</u>	QP	E	.51	.88	.06	

*Signifies that comparisons are based on 2 years of data. For other categories, comparisons were based on only 1 year, usually 1972.

1/ The "best" survey estimate is defined as the one--quarterly panel or diary estimate--which was closest to the independent figures. Where a double asterisk (**) is appended to the code, this indicates the specified source was the only one for which a comparison with the independent estimates could be made in the required detail. In most of these latter cases, the specified source was also the only realistic source for the data in question.

2/ The ranges in this column are not always entirely consistent with the computed ratios in the next column, but attempt to make allowances for conceptual differences between the survey and independent sources, disparities among the independent sources themselves, and other factors for which numerical adjustments could not be made.

3/ The ratios are based on weighted averages of the survey data (numerator of ratio) and of the independent data (denominator of ratio) for the various individual categories combined on a given line of the table. Where the two survey estimates were not significantly different (Code "N" in the "best" survey source column), the quarterly panel and the diary data were averaged. Where more than one independent source is specified, those data were averaged for this purpose. Revised ratios reflect PCE data adjusted to 1972 benchmarks. Preliminary ratios based on PCE data projected from 1967 benchmarks (and used in the preliminary evaluation report--Census Technical Report No. 45).

4/ The independent sources, in alphabetical order, are as follows:

AHS--Annual Housing Survey, Bureau of the Census; A. M. Best--Insurance company premium data compiled by A. M. Best and Co.; Cen-Bus--1972 Census of Business Merchandise Line Data; Cen-Serv--1972 Census of Selected Service Industries; CHAS--Center for Health Administration Studies, University of Chicago; ERS--Economic Research Service, U.S. Dept. of Agriculture; HIS--Health Interview Survey, National Center for Health Statistics, U.S. Dept. of HEW; MRCA--Panel data from the Market Research Corp. of America; NCES--National Center for Education Statistics, U.S. Dept. of HEW; NTS--National Travel Survey, Bureau of the Census; PCE--Personal Consumption Expenditures, CNP Accounts, U.S. Dept. of Commerce; SSA--Social Security Administration, U.S. Dept. of HEW; SORAR--Survey of Residential Alterations and Repairs, Bureau of the Census.

5/ Uncertainty about the magnitude of a conceptual difference precludes a more precise judgment for this borderline value. The ratio of the survey to the independent estimate is understated because the PCE figure used as a base relates to "off premise" consumption and includes (in addition to food purchased for home use) snacks purchased in stores (and other establishments which are not eating places) but consumed outside the home. The survey estimate, however, relates only to food purchased for home use. The reverse situation holds for the PCE and survey estimates for "purchased meals and snacks" where the ratio is overstated.

6/ This is an average value for the eight major categories included in this bracket. Of these, three were significantly more completely reported than indicated by the coded mean value, but there was no clear-cut distinction among these by expenditure size.

7/ Using optimum estimate (including value of gifts received and averaging the two alternative estimates for the year), see text. Under the standard estimation, the rating for major appliances would be "A, B" but that for minor appliances would be "D".

8/ Survey estimates used in the computation were diary estimates for dinnerware-cookware and hand tools and average of the two estimates for decorative items and luggage.

9/ Computed ratio is understated (magnitude uncertain) because the PCE base figure (but not the survey estimate) includes fuel used in rental quarters which is supplied and directly paid for by the landlord.

10/ Comparison based on college tuition only.

11/ Comparison based on number of trips taken rather than expenditure.

12/ Uncertainty about allocation of a large unclassifiable residual (about 15 percent of transportation expenditures) precludes a more precise judgment for these categories.

THE CONSUMER EXPENDITURE SURVEY: PROSPECTS FOR CONSUMER RESEARCH

Gregory D. Upah, Virginia Polytechnic Institute and State University
Seymour Sudman, University of Illinois

Abstract

This paper provides a brief overview of methodological changes in, and particular benefits of, the Consumer Expenditure Survey as they relate to prospects for consumer research. The major focus of the paper is to point out some of the major issues/research questions which can be addressed through the use of the C.E.S. data.

As early as 1965, there were major attempts to stimulate increased interest in longitudinal and other types of over-time studies of consumer behavior (Andreasen 1965; Granbois and Engel 1965; Nicosia 1965). However, to date there have been few such efforts undertaken and reported in the literature by consumer researchers. While such research is widely done by industry sponsored research groups (e.g., The Conference Board), major marketing research organizations and major retail and manufacturing firms, seldom are these results published and/or available to consumer researchers not associated with these organizations (for exceptions, see Linden 1965 and Linden and Axel 1978).

The Consumer Expenditure Survey sponsored by the Bureau of Labor Statistics and conducted by the Bureau of the Census provides in many ways the most comprehensive over-time data on consumer behavior of any such "panel" study in existence. This data base has been frequently utilized by home economists and consumption economists. However, this data and similar data collected in other countries have received only scant use by, and attention from, consumer researchers.

As Sheth (1980) points out, much existing government and private data on consumer behavior goes unanalyzed, while more data is collected; more theories are left untested while new ones are developed. Furthermore, much of this government data is collected from large, national probability samples--just the sort of data base which is generally not economically feasible for most academic research. Sheth suggests that the analysis of existing data become a priority for future consumer research efforts. The existence of the Consumer Expenditure Survey data provides one major opportunity for those interested in any number of aspects of consumer behavior to do just this.

Purpose

The purpose of this paper is to: (1) provide background information on the methodological changes in the Consumer Expenditure Survey, (2) compare and contrast the current C.E.S. with other well-known panel studies, (3) outline the unique advantages and potential drawbacks of the C.E.S. for consumer research and (4) suggest several sets of illustrative research questions for which C.E.S. data should be particularly appropriate for answering.

Background

As has been discussed in other papers in this session, the current methodology used for the current C.E.S. is markedly different from that which was used in the previous surveys done in 1950 and 1960. In those years, respondents were asked, in basically one or two long interviews, to recall all of their household expenditures throughout the previous year. There is substantial evidence that this

task was too difficult for many respondents, particularly for non-durable goods. The recall method has, even for shorter recall periods, been shown to produce seriously inaccurate reporting (Sudman and Ferber 1974).

Beginning with 1972-1974 survey, and continuing through the current survey, the diary method has been used to collect expenditures on inexpensive non-durables. The recall method is used to collect expenditure data for more expensive durable goods and regular expenses such as rents, utility bills and insurance premiums. The recall period for durables is 3 months. Diary data is collected only over a 2 week period.

BLS is now engaged in an ongoing or continuous Consumer Expenditure Survey Program. Whereas the 1950, 1960 and 1972-74 surveys were one-time surveys, the current survey collects data on at least a monthly basis. The survey's dynamic character, along with its improved data quality, allows for analyses not possible with previous surveys. For instance, analyses of the short-term impacts of major social events could not be made with the 1950 and 1960 surveys since the data collected was for a single one-year period. While such analyses--over a 15 month period were possible with the 1972-74 data, measures of expenditure changes in the years after 1961 and before 1972 and in the years after 1972 and prior to the current survey were not.

The data quality especially for non-durable goods purchases has been markedly improved (since the 1960 survey) through the use of the diary method. The data on these expenditures may be more comprehensive than any such data previously collected. As Pearl (1981) points out, the Consumer Expenditure Survey obtains the most extensive travel expenditure data ever collected.

Some Major Commonalities and Distinctions Between the C.E.S. and Commercial Panel Studies

There has been somewhat substantial use of and reference to consumer panel data in the marketing literature (e.g., Ahl 1970; Cherington 1943; Frank and Strain 1972; Massy and Frank 1965). For a review of the use of consumer panels in marketing see Sudman and Ferber (1979). Some of the same objectives illustrated in these studies can be satisfied through use of the C.E.S. data. A comparison of the C.E.S. to a major consumer panel such as Market Research Corporation of America (MRCA) national consumer panel, does however, reveal major similarities and differences. Both "panels" provide over-time data on the quantity and type of item purchased. Both panels provide data on prices paid.

The C.E.S. does not, however, provide information on the particular brand purchased or whether or not the item was purchased with a coupon or some other promotional "deal." The latter types of data are usually obtained in commercial panel studies. The C.E.S. does not obtain data on outlets shopped. However, a separate study sponsored by BLS, the Point-of-Purchase Survey, does provide a means for some types of outlet studies.

Some commercial panels also have measures of a wider variety of household characteristics to relate to expenditures--e.g., life styles, than are obtained in the C.E.S., as well as attitudinal measures regarding products for

which purchase histories are obtained. The Consumer Expenditure Survey obtains no data on attitudes, life styles or behavioral intention. Thus, while desirable, over-time analyses of the relationships between these other household or individual characteristics and behavior are not possible with existing C.E.S. data. However, as Jacoby (1978) points out, the analysis of consumption itself and its consequences for subsequent acquisition decisions deserves more attention in the consumer behavior literature.

The C.E.S. and MRCA panels differ in terms of the length of time participants remain in the panel. The maximum length of time for participation in the C.E.S. is 15 months. MRCA's panel members may, for instance, remain in the panel indefinitely. This continual replacement effectively removes the C.E.S. from categorization as a longitudinal study. Longitudinal analysis requires over-time observations to be made on the same sampling units. The C.E.S. is longitudinal for a maximum of 15 months--the five quarters for which data is collected. However, no longitudinal analysis is possible for non-durables since this data is collected for only one two-week period.

Perhaps the major difference between the C.E.S. and other major panels is the breadth of the consumption expenditures measured for each sampling unit. MRCA has a panel for grocery items and another for clothing. General Electric has a panel which is concerned with major appliances. The C.E.S. measures virtually every conceivable type of expenditure using the same panel. The existence of this almost totally comprehensive expenditure data provide the unique opportunity for assessing expenditures on certain goods and services relative to expenditures on other products. This extensive detail is necessary to allow for the continual updating of weights and market basket items used in calculating the Consumer Price Index (CPI). Within some limits to amount of detail in, and specificity of, product breakdowns researchers interested in consumer expenditures on virtually any type(s) of product(s) can obtain high quality data through use of the C.E.S.

Some commercial research firms have been amenable to releasing panel data to academic researchers a specified number of years after the data has been collected. Current panel data is however, for most academic researchers, too costly. In contrast, the data tapes sold by BLS are relatively inexpensive.

Sales and distribution audits also provide data on consumer expenditures for a variety of products. Information on expenditures across types of retail outlets also is available. However, these data offer no opportunity for analyses of differences in expenditures by various subgroups of the population. An additional drawback is the relatively high cost of current data.

Other Benefits of The Consumer Expenditure Survey Data

As mentioned above, the C.E.S. provides interested researchers with the opportunity to analyze over-time data collected from a large national probability sample. There are some methodological concerns regarding the validity and reliability of some of the data (see other papers in this session--e.g., Pearl 1981). Nevertheless, the data quality, overall, is likely to be superior (i.e., in terms of producing lower total survey error on any given item of data) to most data which has been utilized in academic research (especially those data derived from student or local household convenience samples).

The C.E.S. data is particularly well-suited to macro or group level, as opposed to micro, analyses of consumer behavior. The former approach has been classified as a major need for consumer research in the 1980's (Sheth 1980).

The sheer comprehensiveness of expenditure data present in the C.E.S. allows researchers to overcome the weaknesses in analyzing consumption of a single product and making generalizations about consumer behavior based on that single category of product. The C.E.S. data allows for an analysis of the total consumption structure of a household.

The unit of analysis--the sampling unit for the C.E.S. data, called the consumer unit, is almost always the household. The view that the household, as opposed to the individual, is the appropriate unit of analysis for consumer research has gained growing support (Ferber 1980). Only in certain circumstances (i.e., based on the financial independence of one cohabitant from another cohabitant) will a household--i.e., place of residence, contain more than a single consumer unit. BLS estimates that 95% of all households will contain a single consumer unit.

Although the C.E.S. and commercial panel data are based on self-report data as opposed to direct observation or record checks, the data deals with actual purchases--i.e., expenditures. There are no attitudinal measures taken.

Kassarjian, for one, has argued for more research into consumer behavior which relies on unobtrusive and nonattitudinal measures--(e.g., analysis of current periodicals, advertisements, etc.) insofar as they reflect the changing values and attitudes of a society. Actual expenditure data deals with the end-result or activity that an analysis of these factors intends to predict. Researchers who wish to make inferences as to what value/attitude changes are taking place can look to the current production of visible tangible symbols of a society (its art, literature, etc.). However, it also is important to analyze the actual consumption patterns of various segments of the society and society in general. There are, of course, relative disadvantages to any single approach--e.g., determining whether or not changes in art, or literature are lagging behind or preceding the value or attitude changes that help to produce changes in consumer expenditures. Furthermore, consumer expenditures are constrained by what sellers choose to offer in the marketplace. Thus, consumer behavior may be reflective not of what consumers are thinking, feeling or desiring, but rather what compromises from what is wanted or needed they are willing to make given the limits of the marketplace.

The complex and interesting questions surrounding the macro aspects of consumer behavior render any one means of assessing long term shifts in value and expenditure somewhat inadequate. Multiple means/indicators are required. Nevertheless, the comprehensive expenditure data available through the C.E.S. is one such means of assessment that has not been fully utilized by consumer researchers.

Possible Research Programs Utilizing the C.E.S. Data

The following are illustrative examples of the research questions and issues for which the use of the C.E.S. data would be most appropriate and useful.

1. Analysis of Long Term Predictions for Changes in Consumer Behavior:

How do changes in consumer expenditures compare to the long-term changes in consumer behavior predicted by various pundits and forecasters?

For instance, the Yankelovich research organization regularly tracks changes in a number of major cultural values. Based on this organization's surveys, predictions for major changes in consumer values in the U.S. are made. These changes in values will ultimately influence attitudes and purchase behavior toward a wide variety of goods and services (Kotler 1980: Loudon and Della Bitta 1979). The

Consumer Expenditure Survey data can be used as one means of testing specific propositions about changing values and consumer behavior through an analysis of changes in consumer expenditures.

2. How are consumer expenditure patterns influenced by the major changes in characteristics of households?

(a) How do consumption expenditures of married households in which both spouses work differ from those married households with wives who are not employed outside the home? This is a popular issue and source of conjecture among home economists, consumer researchers and marketers. Still, there have been few empirical studies examining the full range of implications of the working wife on household buying behavior.

One study dealing with this issue utilized data obtained from a 1967-70 Survey Research Center study of consumer finances. Holding income, education, family life cycle and other socioeconomic and demographic measures constant, the labor-force participation of wives was related to overall household consumption expenditures, and expenditures on durable goods (both as a percent of total income) (Strober 1977). Strober found that holding other factors constant, households with working wives had a significantly higher consumption expenditure to income ratio than did households without working wives. The labor-force participation of the wife, however, had no significant effect on total expenditures or durable goods (as a percent of total household income). This study tests several hypotheses developed in economics regarding the expenditures of households with working wives (see Strober 1977 for a brief overview of these hypotheses). There seems to be a great opportunity to merge the consumer behavior theory developed in sociology, marketing and psychology with these economic theories to further explore these issues. At the very least additional tests of the hypotheses developed in economics could be made. The Consumer Expenditure Survey data again provides an excellent data base for such research.

(b) How do these consumption expenditure patterns change as a function of the age of the wife/husband, number and ages of children or overall family life cycle of the household?

There have been two recent studies of the impact of family life cycle on various consumer expenditures. The first of these studies was, in fact, based on Consumer Expenditure Survey data collected by the Government of Norway (Arndt 1979). The other study was based on a local commercial survey in the U.S. (Landon and Locander 1979). Both of these studies have, however, been criticized for failing to allow for, or to hold constant, other factors (aside from family life cycle) likely to influence the expenditures under consideration--e.g., income, education and occupation (Ferber 1979). As such they provided no opportunity to assess the efficacy of family life cycle relative other household characteristics in explaining consumer expenditure.

The Consumer Expenditure Survey provides ample data for constructing various formulations of family life cycle. Those interested in conducting such research might well consider the criticisms, and suggestions for modification, of the family life cycle concept itself (Murphy and Staples 1979). In addition, the Consumer Expenditure Survey data contains measures of the household characteristics needed for the multivariate analyses required to establish the usefulness of the family life cycle concept--in whatever form, for explaining consumer behavior. The continuous nature of the survey allows for an analysis of the short-term impact (i.e., impacts on durable goods expenditures) of changes in household composition or life cycle--e.g., a new baby.

(c) Assessing the impact of other changes in household composition.

The dynamic character of the current C.E.S. allows for an analyses of the short and long-term effects of the aforementioned changes in household composition. The effects of other changes such as a loss of job or death in the family also can be assessed.

3. How do consumption expenditure patterns of Blacks, Whites and Hispanics differ? Do these expenditures differ when age, education, income and other relevant demographic characteristics are held constant or controlled for across the two groups? If so, which types of expenditures differ?

Recent research indicates that in similar occupational, income, educational age and life style categories there are few Black-White differences in consumption (Edmonds 1979). The Consumer Expenditure Survey offers an extensive data base for making these cross-sectional comparisons over time.

4. What changes in consumption (e.g., percent of income spent on the good or service) occur as income increases?

These issues have concerned economists (e.g., with respect to Engel's Law) and other researchers for a number of years (see Prais and Houthakker 1971). It should be interesting to note where these shifts have been occurring in recent years.

BLS has, in past years, calculated income sensitivity ratios. These ratios represent the percent increase in expenditure on some item (i.e., usually a broad category of items) resulting from a one percent increase in real disposable income. One drawback in the use of these ratios has been their timeliness (Reynolds and Wells 1978). However, with the switch by BLS to an ongoing C.E.S., these ratios can be kept much more current. They provide a means for marketers to calculate likely changes in consumption of their products with changes in disposable income levels. They also provide a means of assessing the effects of sheer increases in disposable income on relative increases in the consumption of various products (i.e., tests of income--related consumption hypotheses.)

5. Profiling the Heavy User:

Comparisons of heavy users of a particular category of good or service with light users are easily facilitated with the Consumer Expenditure Survey. These comparisons can be made in terms of the various household characteristics which are measured or can be derived (e.g., family life cycle) from the C.E.S. survey. Comparisons in terms of use of related goods or services across user categories also could be made. One possible research focus would be to compare the results to results of past "heavy half" studies.

6. The Growth of Services in the Economy

Services have, over the past ten years, grown as a percent of GNP. Potential research questions regarding the growth of services include: Which particular service expenditures have exhibited the greatest increase? How do these rates of increase compare across categories of households? Are these expenditure changes reflective of recent predictions for consumer shifts to the service sector of the economy? These changes in the over-time expenditures on services are best measured by the continuous survey now in progress rather than by separate surveys conducted at yearly or 5 or 10 year intervals.

7. Applications of Life Style or Consumption Style Approaches to the Study of Consumer Behavior

While the Consumer Expenditure Survey does not obtain life style data based on activity interest and opinion questions, it does allow for a means of inferring life styles from actual consumption behavior. The use of actual consumption data is one of several means of assessing life styles (Wind and Green 1974). The use of particular products, which might be considered to be part of the activity component of life styles, has been correlated with the purchase of other products of interest (see, for example, Hustad and Pessemier 1974--for a review of such studies, and Bass, Pessemier and Tigert 1969 and Reynolds and Wells 1978).

Sheth (1980) discusses the need for more research dealing essentially with a typology of consumption life styles. This research would involve an analysis of "the time dependent covariances of preselected and representative goods and services." The Consumer Expenditure Survey data is ideally suited to just such a research approach. Life styles can thus be inferred from actual consumption patterns and not merely activity interest and opinion data. Analysis of trends in life styles based on changes in consumption of various types of products would presumably be of major interest to marketers, consumer researchers interested in the overall consumption trends in the market place, and sociologists and others who wish to discover evidence for changes in values and life styles, over time.

There has been some recent research in these areas. One study analyzed the correlates and consumer behavior implications of changes in consumption style (Uusitalo 1980). The author determined a variety of consumption styles by forming eight combinations of three consumption indices. The consumption indices were based on factor scores produced in a factor analysis of consumer expenditures on 220 items. The expenditure data was derived from Consumer Expenditure Survey data collected by the Statistical Central Office of Finland. The consumption styles derived from these analyses were then related to total consumption expenditures and individual expenditures on specific categories of products.

One possible caution with respect to such analyses is in relating predictor variables derived from particular expenditures to essentially the same expenditures. A better method would be to derive the consumption style measures from one set of expenditures, and then using those measures to predict another set of expenditures.

8. Cross-Cultural Analyses of Consumer Expenditures

Already in this review of possible research programs involving the use of the C.E.S. data, three different national consumer expenditure surveys have been discussed. In fact, there are nine European countries with which cross-cultural comparisons are possible. Such studies are now more easily facilitated due to the use of much more comparable expenditure survey methods (see, Statistical Office of the European Communities 1980). One additional interesting aspect of such research is the possibility for comparisons of expenditures with Iron Curtain Countries.

9. Examining Trends in Consumption Using New Categorizations of Goods and Services

What do sports equipment, vacations, and tickets to a play have in common in terms of satisfying consumer needs? How are they representative of different types of consumer needs? What are the underlying consumption trends in the vast array of changes in consumer expenditures? Clearly, better characteristics of products (goods and services) theories--i.e., categorizations of products according to what value expression or sought need they represent are

needed to answer these types of questions. The convenience goods--shopping goods--specialty goods breakdown clearly is not sufficient for making projections in needs-based categories of expenditures, or in making inferences from expenditure data to value shifts. Work in analysis of consumption as it pertains to the acquisition of particular benefits has been done by Lancaster (1975) and Ratchford (1975).

Some trend studies of consumer expenditures published by BLS break all consumer expenditures down to expenditures on food, housing, clothing, medical care, transportation, recreation (which includes among other things expenditures on education, reading, alcoholic beverages and tobacco). There are certain to be subtle differences within each of these expenditure categories which when guided by a theory-based approach could yield new insights into consumer behavior. Furthermore, as Nicosia and Mayer (1976) point out, differences in the nature of expenditures made within a particular expenditure category may have important public policy implications. For example, the satisfaction of entertainment needs via attendance at spectator sports versus purchase of records or books will influence energy utilization, use of roads and mass transit, consumption of food and drinks away from home etc.

10. Impact of Major Social Events on Consumer Behavior

As mentioned above, the continuous collection of expenditure data allows for an assessment of the short-term impacts of major social events on consumer behavior. For instance, the impact of such events as oil embargoes and resulting gasoline shortages, the eruption of Mt. St. Helens, major new product introductions such as K-Body cars or video disc players, shortages of certain products, new laws, major economic changes (e.g., rises or falls in inflation) or political changes (e.g., the election of a new president) can be measured by the continuing survey.

Conclusion

The Consumer Expenditure Survey has provided, and continues to provide, interested consumer researchers a rich data base which can be used to satisfy a variety of research interests. There is certainly room for creative, theory-based approaches to the analysis of the C.E.S. data. Fundamental tabulations and cross-tabulations of expenditure data--of the type published by BLS and the Conference Board are some of the most important analyses performed. However, there is much room to go beyond these analyses to answer some of the more complex questions consumer researchers and public policy makers have, especially in recent years, been posing. The C.E.S. data holds out the promise of not only enabling consumer researchers to see what is happening and what is likely to happen in terms of actual consumer behavior in the market place, but also providing the insights and tests of hypotheses which will assist these researchers in determining the reasons for these behaviors.

References

- Ahl, David H. (1970), "New Product Forecasting Using Consumer Panels," Journal of Marketing Research, 7 (May), 160-67.
- Andreasen, Alan (1965), "Potential Marketing Applications of Longitudinal Methods," in Marketing and Economic Development, Peter D. Bennett (ed.), Chicago: American Marketing Association, 261-278.
- Arndt, Johann (1979), "Family Life Cycle as a Determinant of Size and Composition of Household Expenditures," in Advances in Consumer Research, 6, William L. Wilkie (ed.), 128-132.

Bass, Frank M., Edgar A. Pessemier and Douglas J. Tigert, (1969), "Complementary and Substitute Patterns of Purchasing and Use," Journal of Advertising Research, 9, p. 19.

Cherington, Paul T. (1943), "New Economic Patterns Found by Consumer Panels: Introduction," 8 Journal of Marketing (July) 41.

Edmonds, Linda L. (1979), Clothing Buying Practices and Life Style Differentials Between Employed Black and White Women, unpublished Ph.D. Dissertation, Virginia Polytechnic Institute and State University (April).

Ferber, Robert (1979), "Comments on Life Cycle Analysis," in Advances in Consumer Research, 6, William L. Wilkie (ed.), 146-148.

_____ (1980), "What Do We Know About Consumer Behavior," in Marketing Management and Strategy, Philip Kotler and Keith Cox (eds.), Englewood Cliffs: Prentice Hall, 97-108.

Frank, Ronald E. and Charles Z. Strain (1972), "A Segmentation Research Design Using Consumer Panel Data," Journal of Marketing Research, 9 (November), 385-90.

Granbois, Donald H. and James F. Engel (1965), "The Longitudinal Approach to Studying Marketing Behavior," in Marketing and Economic Development, Peter D. Bennett (ed.), Chicago: American Marketing Association, 208-221.

Hustad, Thomas P. and Edgar A. Pessemier (1974), "The Development and Application of Psychographic Life Style and Associated Activity and Attitude Measures," in Life Styles and Psychographic, William D. Wells (ed.), Chicago: American Marketing Association, 97-126.

Jacoby, Jacob (1978), "Consumer Research: A State of the Art Review," Journal of Marketing, (April), 87-96.

Kotler, Philip (1980), Marketing Management: Analysis, Planning and Control, 4th ed., Englewood Cliffs: Prentice Hall, 122-128.

Lancaster, Kelvin (1975), "Hierarchies in Goods Characteristics Analysis," in Advances in Consumer Research, Vol. 3, Beverlee B. Anderson (ed.), Ann Arbor: Association for Consumer Research, 348-352.

Landon, E. Laird Jr. and William B. Locander (1979), "Family Life Cycle and Leisure Behavior Research," in Advances in Consumer Research, 6, William L. Wilkie, (ed.), 133-138.

Linden, Fabien (1965), Expenditure Patterns of the American Family, New York: The Conference Board.

_____ and Helen Axel (1978), Consumer Expenditure Patterns, Vol. I., New York: The Conference Board.

Loudon David J. and Albert J. Della Bitta (1979), Consumer Behavior: Concepts and Applications, New York: McGraw-Hill, 120-132.

Massy, William F. and Ronald E. Frank (1965), "Short Term Price and Dealing Effects in Selected Market Segments," Journal of Marketing Research 2 (May), 171-85.

Murphy, Patrick E. and William A. Staples (1979), "A Modernized Family Life Cycle," Journal of Consumer Research, 6, (June), 12-22.

Nicosia, Francesco M. (1965), "Panel Designs and Analyses in Marketing," in Marketing and Economic Development, Peter D. Bennett (ed.), Chicago: American Marketing Association, 222-243.

_____ and Robert N. Mayer (1976), "Toward a Sociology of Consumption," Journal of Consumer Research, 3, (September), 65-75.

Pearl, Robert (1981), "An Evaluation of the 1972-73 Consumer Expenditure Survey Data," in Advances in Consumer Research, Vol. 8, Kent B. Monroe (ed.).

Prais, S.J. and H.S. Houthakker (1971), The Analysis of Family Budgets, Cambridge: The University Press.

Ratchford, Brian T. (1975), "The New Economic Theory of Consumer Behavior: An Interpretive Essay," Journal of Consumer Research, 2, 65-75.

Reynolds, Fred D. and William D. Wells (1977), Consumer Behavior, New York: McGraw-Hill, 385-412.

Sheth, Jagdish N. (1979), "The Surpluses and Shortages in Consumer Behavior Theory and Research," Journal of the Academy of Marketing Science, 7, (Fall), 414-427.

Strober, Myra (1977), "Wives' Labor Force Behavior and Family Consumption Patterns," The American Economic Review, 7, (February), 410-417.

Sudman, Seymour and Robert Ferber (1979), Consumer Panels, Chicago: American Marketing Association.

_____ and Robert Ferber (1974), "A Comparison of Alternative Procedures for Collecting Consumer Expenditure Survey Data for Frequently Purchased Products," Journal of Marketing Research, 11 (May), 128-135.

Uusitalo, Liisa (1980), "Identification of Consumption Style Segments on the Basis of Household Budget Allocations," in Advances in Consumer Research, 7, Jerry C. Olson (ed.), 451-459.

Wind, Yoram and Paul E. Green (1974), "Some Conceptual Measurement and Analytical Problems in Life Style Research," in Life Styles and Psychographics, William D. Wells (ed.), Chicago: American Marketing Association, 97-126.

Womer, Stanley (1944), "Some Applications of the Continuous Consumer Panel," Journal of Marketing, 9, (October), 132-136.

NONRESPONSE IN CONSUMER SURVEYS

Fred Wiseman, Northeastern University

Abstract

Response rates in consumer surveys have declined in the 1970's due to a number of controllable and uncontrollable factors. This decline has been a source of concern to many professionals in government and private industry. As a consequence, consumer researchers now find their results and inferences being challenged when they are based upon survey data which are susceptible to substantial non-response bias. This paper reviews the findings of recent nonresponse research investigations and discusses their importance and implications for consumer researchers.

Introduction

During the 1970's there was a growing interest among many social scientists and statisticians within the Federal government, private industry and academia about the negative impact that certain uncontrollable, environmental factors were having on the survey research process.

Changing life styles, increased female participation in the labor force, flexible working hours and more leisure time made it increasingly difficult for a potential respondent to be contacted by an interviewer. Further, if and when contact was made, numerous other factors such as privacy related concerns, fear and suspicion of strangers, questions regarding survey legitimacy and excessive interviewing resulted in a substantial percentage of contacted respondents deciding to refuse participation.

In 1973, the American Statistical Association brought together a diverse group of social scientists and survey methodologists in order to discuss the problems of conducting present-day surveys of human populations. From the discussions that took place, a number of conclusions were reached. Among these was that "Survey research was in some difficulty, and, to an undetermined scale, that difficulty was increasing." For example, some participants noted that completion rates on general population surveys averaged 60-65% compared to 80-85% in the decade of the 1960's (The American Statistician, 1974).

What are the consequences of increased nonresponse on data quality? This is a question that cannot easily be answered in a survey because nonresponse bias depends not only on the magnitude of nonresponse, but also on the degree to which nonrespondents differ from respondents on the variables of interest. However, in discussing the effects of nonresponse, Platek (1977) points out that since the sampling variance of the estimates is inversely proportional to the response rate, estimates based on a simple random sample with 80% response rate will have a sampling variance that is 12½% higher than the variance of corresponding estimates with 90% response rate.

This paper will briefly review some of the key findings of recent studies and investigations that have focused on nonresponse and data quality related issues. In addition, it will discuss efforts that are now underway that will hopefully result in an improved understanding of the nature and extent of nonresponse in consumer surveys. Finally, implications for the academic research community will be explored.

Concerns About Nonresponse/Data Quality in Federal Agencies

As an outgrowth of the 1973 ASA conference, the National Science Foundation agreed to support a pilot study in which 36 surveys were examined in detail in order to determine the extent to which stated survey objectives were achieved (Bailar and Lanphier 1978). Of the 36 surveys, 26 were conducted by or for the Federal government and the remaining ten by state governments, academic institutions or professional associations. The major finding of this investigation was that:

Twenty-two of the 36 surveys did not meet their objectives because of technical flaws such as a low response rate, the inclusion of inferences in the final report that could not have been substantiated by the survey results, no validation of survey interviews and no data "cleaning."

With respect to response rates, Bailar and Lanphier stated that they were difficult to collect and compare as they were found to have different names and different definitions in different places and circumstances. Further, when response rates were reported, they were often inflated.

A second major investigation was recently completed for the Comptroller General of the United States upon request of three members of the U.S. House of Representatives (Comptroller General 1978). The study examined the potential for incorrect or unreliable information being generated by public opinion polls and attitude surveys sponsored or conducted by Federal agencies. The Congressmen were concerned that poor quality data from surveys were being used as the basis for making Federal management decisions that affected national programs and policies.

In carrying out the study, the General Accounting Office (GAO) identified a number of recently conducted surveys at various agencies and decided to review five in detail. Their overall conclusion was:

Although there were no indications that survey results were intentionally misused, use of the results of all five should have been limited because each contained serious technical flaws.

Numerous examples of these technical flaws were cited by the GAO in their report including those in a survey designed to provide an information base for use in a model that compared potential national carpool incentive policies. The GAO concluded that findings could not be used for national projections because of the type of sampling used and the extremely low response rate.

As a result of that study, it was recommended that Federal agencies be given more explicit guidance as to what constitutes a good attitude survey or opinion poll and that the use of surveys which contain extensive technical flaws should be discouraged.

Given the concern within the government about the widespread use of faulty surveys, it was not surprising to see that the Federal Trade Commission, in a precedent-setting case, recently ruled that Litton Industries advertisements for its microwave oven violated Federal law because the survey research data used to justify the ad claims were

defective (Marketing News 1980). Irving Roshwalb, of Audits & Surveys, Inc. was asked to testify in that case and he'll later describe in more detail the specific nature of the issues raised.

- Overall, response rates were low, with a median rate of 30% for surveys in the data base.

Concerns About Nonresponse/Data
Quality in Private Industry

In general, it appears that both users and producers of survey research in private industry have not paid as much attention to nonresponse related issues as have their counterparts within the Federal government. Among the reasons for this are (1) time pressures, (2) budget limitations, (3) client specifications, and (4) less stringent precision requirements. However, there is a growing awareness of the problem within the industry and a realization that improvements must be made in terms of research methodology in order to insure that accurate and reliable data are obtained.

During the past four years, Prof. Philip McDonald and myself have conducted a major nonresponse research investigation. This investigation carried out under the sponsorship of the Marketing Science Institute, has had the cooperation of their member firms as well as those in CASRO, the Council of American Survey Research Organizations. In 1978, 32 MSI and CASRO firms supplied data from 182 consumer telephone surveys that were conducted over a six week period. Collectively, for these surveys, there were over one million unique sample members selected to be interviewed. We found (Wiseman and McDonald 1978):

- A relatively large percentage of potential respondents/households were never contacted. The median non-contact rate was 40%.
- Of those individuals contacted, slightly more than one in four refused participation. The median refusal rate was 28%.

The analysis of the data received from these commercial companies suggested that low response rates were due more to controllable, than to uncontrollable factors. For example, in almost 40% of the surveys only one attempt was made to contact a potential respondent and rarely did a research firm make a concerted effort to convert reluctant respondents.

Potential Consequences of Nonresponse

The seriousness of low response rates depends upon the extent to which respondents differ from nonrespondents on the variables of interest within a survey. It may be that those individuals who are difficult to reach or who are unwilling to be interviewed share the same general attitudes, opinions, preferences, etc., as do individuals who are readily accessible and who are willing to be interviewed. If this be the case, then the potential consequences of a low response rate are substantially reduced.

If, however, significant differences do exist between respondents and nonrespondents then survey results, no matter how large the sample size, are likely to be of little value to decision or policy-makers.

Little is known about the characteristics of nonrespondents. However, in a number of recent studies, differences on a number of dimensions have been found to exist among those who readily respond in a survey, those who initially refused, and those hard-to-reach people who responded only after a large number of callback attempts. For example, Table 1 presents the results of four such studies in which various categories of nonrespondents were found to differ from their readily accessible and cooperative counterparts.

TABLE 1
Dimensions Upon Which Refusers and Hard to Reach Individuals Differed From Accessible and Cooperative Respondents

Study	Population	Topic of Survey	Compared to readily accessible and cooperative respondents	
			Refusers	Hard-to-Reach Individuals
Ankelberg and Day [1973]	National Survey - H-O-H conducted in 1967 by Survey Research center, University of Michigan.	Consumer finances	----	<ul style="list-style-type: none"> ● Young ● Better educated ● Higher income ● More likely to live in large cities
The Data Group, Inc. [1967]	National Survey - female H-O-H	Attitude, opinion and consumption habits	<ul style="list-style-type: none"> ● Older ● Less likely to be employed full time ● Shop less often ● Go out less often ● Less aware of new products 	<ul style="list-style-type: none"> ● Younger ● Higher income ● Socially active ● Shop more often ● Spend more on entertainment
van Westervhoven [1978]	National Survey (The Netherlands) - female H-O-H	Attitude, opinion and consumption habits	<ul style="list-style-type: none"> ● Older ● Lower social class ● Urban ● Make fewer purchases 	<ul style="list-style-type: none"> ● Working woman ● Age: 35-49 ● 1-2 person hh ● Few small children
O'Neil [1979]	General population survey of Chicago telephone households	Attitudes toward crime in Chicago and how police are handling it	<ul style="list-style-type: none"> ● Blue collar ● Service workers ● Less educated ● Lower income ● Older ● Less social participation 	

As can be seen, the results of the four studies with diverse populations show a measure of similarity with respect to the characteristics of individuals who are difficult to reach and those who are reluctant to grant an interview. However, it appears that on a number of dimensions, the characteristics of refusers are opposite those of hard-to-reach individuals. Thus, a strategy that involves a large number of callbacks without including any extra effort to convert initial refusers is one that is likely not to improve the representativeness of the sample. In addition, to completely assess the nature of nonresponse bias, we also need information on how the various response and nonresponse segments vary with respect to the key variables of interest in a survey and not just on their socioeconomic and demographic compositions.

Nonresponse Task Forces

As a result of the research that has taken place during the last few years, two Nonresponse Task Forces have been recently established involving members of MSI and CASRO.

The first Task Force, chaired by Lester Frankel (former President of the American Statistical Association) has been given the charge "to develop a uniform formula for measuring completion rates in the survey research industry for all modes of data collection -- mail, telephone and personal interview." This Task Force not only includes MSI and CASRO representatives, but also includes representatives from the Bureau of the Census, and Office of Federal Statistical Association.

It is anticipated that the Task Force will make its recommendations within the next year and, hopefully, after discussion within the survey research community, a set of standardized definitions and reporting procedures will be adopted. Such standards are long overdue.

A second Task Force is working on the development of a research design for a large scale study that will examine, in detail, the characteristics of nonrespondents and the impact that their exclusion has on survey results. At the present time we envision a national sample of 4,000 heads-of-household with up to 25 callback attempts made to locate as many hard-to-reach and "nonresponders" as possible.

Implications For Academic Researchers

The current thrust and interest in nonresponse related issues has important implications for academic researchers in three different areas: (1) teaching, (2) research, and (3) consulting.

Those in academia have a responsibility to stress, more than has been done in the past, the importance of data collection techniques and the need for quality data. We now have sophisticated sampling and experimental designs, along with methods of analysis. The element that links these two areas together is data collection and, too little attention has been given to this subject. Researchers should be concerned about low response rates and the potential impact that non-response and other forms of nonsampling errors can have on data quality. In a 1977 panel discussion sponsored by Advertising Age, leading research professionals indicated that one of the major problems facing the industry was the lack of competent research professionals (Advertising Age 1976).

In consulting activities and in testimony, as Keith Hunt emphasized in his presidential address last year to ACR members in San Francisco, it is incumbent on the research professional to stress the need and importance of sound methodological procedures especially if the results of consumer research will be used externally as well as internally. The Litton case should serve as a warning that the FTC and other government agencies will be looking very closely at claims and inferences that are made as a result of

attitude surveys and opinion polls.

Finally, in conducting our own research, attention again must be given to data quality and improved response rates even if this means working with smaller samples. Lipstein (1975) made an eloquent plea for this in his excellent article entitled, "In Defense of Small Samples." It appears now that government sponsored studies will be more closely scrutinized than they have in the past. This is also likely to be true for articles submitted to scholarly journals. For example, in the current Journal of Marketing and Journal of Marketing Research style sheet, there is the following statement: "Do not ignore the nonrespondents. They might have different characteristics than the respondents." Bob Middlestaedt and then Bob Ferber can tell you a little more about the quality of our own data reporting and implications for journal acceptance.

References

- Bailar, Barbara A. and Lanphier, C. Michael (1978), Development of Survey Methods to Assess Survey Practices, Washington: American Statistical Association.
- Comptroller General (1978), "Better Guidance and Controls Are Needed to Improve Federal Surveys of Attitudes and Opinions," Washington: General Accounting Office.
- Dunkelberg, William C. and Day, George S. (1973), "Nonresponse Bias and Callbacks in Sample Surveys," Journal of Marketing Research, 10, 160-168.
- "FTC: Litton Used Defective Research for Microwave Ads," (1980), Marketing News, (July) 23, 12.
- "Is Business Hurting Research? Users Differ," (1977), Advertising Age, 1.
- O'Neil, Michael J. (1979), "Estimating the Nonresponse Bias Due to Refusals in Telephone Surveys," Public Opinion Quarterly, 43, 218-232.
- Platek, Richard (1977), "Some Factors Affecting Non-Response," presented at the International Statistical Institute Meetings, New Delhi.
- "Report on the ASA Conference on Surveys of Human Populations" (1974), The American Statistician.
- The Data Group (1977), "A Study of Nonrespondents," Unpublished paper, Philadelphia: The Data Group, Inc.
- van Westerhoven, Emile (1978), "Covering Nonresponse. Does It Pay?" presented at the Congress of the European Society of Opinion and Market Research.
- Wiseman, Frederick and McDonald, Philip (1978), The Nonresponse Problem in Consumer Telephone Surveys, Cambridge: Marketing Science Institute.
- (1980), Toward the Development of Industry Standards for Response and Nonresponse Rates, Cambridge: Marketing Science Institute.

RESPONSE RATES AND INTERNAL VALIDITY:
SOME IMPLICATIONS FOR OUR LITERATURE

Robert Mittelstaedt, University of Nebraska, Lincoln

Introduction

One of the dangers of exposing one's students to some of the classic writings in the Philosophy of Science is that they begin to ask serious questions about the adequacy of the empirical base of our generalizations and our explanatory models. Three years ago, having read of the role of replication in the advancement of a discipline, members of the graduate seminar, which I was privileged to lead, began to consider replication in the literature of our field. The semester ended, the seminar was over, but their - and my - interest continued and led to a wider consideration of the issue and a small-scale study of the replicability of studies reported in the marketing and consumer behavior literature (Madden, Franz and Mittelstaedt 1979).

As that study progressed, several points became evident. First, many are willing to extoll the virtues of replication but, when pressed, oppose those moves which would facilitate or support studies which attempt to replicate. Second, much of this opposition seems to proceed from a distrust of the motives of would-be replicators. Third, this distrust, in turn, largely follows from a misunderstanding of the nature of replication. Many seem to believe that, necessarily, the process requires one to provide raw data to anyone who asks to see it and/or answer innumerable foolish questions from those who are too inert or inept to devise their own research projects.

It is the purpose of this paper to describe the several kinds of and purposes for replication and to relate these to the nature of the sampling process and the reporting thereof. Following that, the results of a modest scanning of a sample of ACR papers will be reported which will lead to a few conclusions and recommendations.

Types and Purposes of Replication

A useful classification of the types of replication is given by Lykken (1968) who differentiates among literal, operational and constructive varieties.

Literal replication is the exact duplication of the original study's sampling procedures, measurement techniques, experimental conditions and methods of analysis. In consumer behavior, literal replication is almost impossible, although testing for split-half reliability or the use of a holdout sample in discriminant analysis comes close to Lykken's concept. If this sort is possible at all, it seems to be the kind of replication which could be conducted only by the original study's investigator or his/her close associates.

Operational replication begins with the methodological recipe of the original study and attempts to follow it as closely as possible. For one investigator to replicate the work of another, it is obvious that it would be necessary to know a great deal about the methods of the original study. Realistically, no one expects researchers to report this kind of detail; conference papers and journal articles would become impossibly long and exceedingly dull.

Constructive replication occurs when one deliberately avoids the methodology of the first study to verify its

findings by other means, thereby enhancing the generalizability of those relationships. The researcher attempting the constructive replication of another study needs enough information about the original to determine the probability that the relationship exists, its possible strength and the general circumstances in which it is likely to be found. Thus, a description of methodology which allows meaningful interpretation of results, is sufficient to permit the constructive replicator to decide whether or not a study's findings are worth replicating, to choose another methodological recipe and to compare the findings of the replication with the original.

The purposes of replication (as suggested by Rose 1954, pp. 262-8; Selltitz, Wrightsman and Cook 1976, pp. 63-70) are to advance the knowledge base of a discipline by:

Testing Methodologies for Soundness. Obviously very tight operational replication, with prescribed variations, is needed to investigate the soundness of a methodology. Sawyer (1975) has provided some excellent examples of this type of replication in his investigations of demand artifacts.

Testing Predictions Derived From a Model, in a Context Different From That of the Original Study(ies). Since so many of our models are borrowed from disciplines which have no direct interest in consumer behavior, their predictions are most often untested in any context resembling those of interest to researchers in our field. Thus, while some findings may be well established in, say, social psychology, the first step in demonstrating their relevance to consumer behavior is to show that they hold in circumstances similar to those of the market place. To avoid total confusion if they do not work out, it is a necessity to hold methodology as constant as possible implying, again, some type of operational replication.

Testing the Generalizability of a Relationship by Extending the Findings of a Particular Study or Set of Studies to New Time, Place and Situational Contexts. Obviously, confidence in the existence of any relationship between two variables is enhanced by finding that it holds across a wide variety of circumstances. When Rogers and Shoemaker (1971) report that 203 of 275 studies have found education to be positively related to innovativeness, we accept this generalization, partly because of the sheer number of supportive findings but, more importantly, because those 203 supporting studies represent the efforts of different investigators working in various times, places, and cultural settings, using different measures of innovativeness. Extending findings involves constructive replication and, in terms of building the knowledge base of our field, may be more useful than many operational or literal replications.

Response Rates and Internal Validity

When considering the research paradigm most of us have been trained to use, and especially those with a background in the non-experimental tradition of marketing research, all questions about sampling seem to be related to the issue of external validity. In the apt phrasing of Petrinovich(1979, p. 376):

...great care is extended to obtain a representative sample of subjects from the reference population, and concern is with the reliability (and hence the standard error) of the subject population. The focus is on subject sampling, and the statistical procedures used assign individual differences between subjects to the error term against which the significance of mean differences is assessed. Great care is used to sample subjects, and pains are taken to assure that generalizations can be made to a relevant population of subjects. There is little concern, however, regarding the legitimacy of the generalizations across situations.

By contrast the central concern of constructive replication is exactly with the "legitimacy of generalizations across situations." Thus, the whole question of response rates might be thought to be nearly irrelevant to the issue of replication. Of course, knowledge of response rates would be helpful to the designer of an operational replication but, in the end, not really crucial to such an effort.

However, the informational requirements of constructive replication are essentially those needed to synthesize a study's results into the relevant literature. Thus, to the would-be constructive replicator, as to the knowledge synthesizer, the whole question of response rates - and the reporting of such rates - is important to the extent it bears on the issue of internal validity.

One obvious question arises from the likelihood that most non-response, for whatever reasons, includes the more "extreme" cases, regardless of the variable. To the extent this is true, the statistical effect of non-response is that the sample variance becomes a biased understatement of the population variance and, therefore, the calculated value of alpha understates the true probability of Type I error in any hypothesis test. While proper reporting would not lessen this problem in real terms, it would alert the interested reader to the necessity of making some subjective revision of the reported alpha level and, therefore, of effect size.

However, the issue of non-response cuts deeper. It seems unlikely that the factors which produce non-response are "random". To the extent they are interactively related with the independent variable and/or related to the dependent variables of a given study, response rate is an important internal validity issue.

To take the most obvious example, the probability of a householder being interviewed in a given study is a function of that person's allocation of time to those activities which he or she does at home. The literature of "time budgeting" supports the intuitive notion that this allocation of time between "away" and "at home" is systematically related to many other behaviors of interest to consumer researchers. However, because the nature of the probability function is partially, perhaps even substantially, controllable by callback procedures, it is essential that information about such efforts be appropriately reported, along with contact and response rates, to allow one to make an adequate interpretation of any given study.

A less obvious example involves refusals. Anyone who has done any interviewing probably has some explanations for the reason, or reasons, behind most refusals. A reasonable surmise is that many result from some sort of fear and, while we may not know the roots of those fears, it seems likely that they are systematically related to other variables of interest. But, whatever the reasons, refusals are not a random phenomenon and the reasons for their occurrence are likely to be related to other important

variables.

In summary, there are serious and substantial questions of internal validity associated with sampling procedures and, ultimately, replication is a matter of internal validity. In Kaplan's (1964, p. 128) words:

...The methodological importance of what is called repeatability is, I think, made more plain by its restatement as intersubjectivity. A scientific observation could have been made by any other observer so situated; nature plays no favorites, but exposes herself promiscuously . . . The methodological question is always limited to whether what is reported is an observation that can be used in subsequent inquiry, even if the particular observer is no longer part of the context. I ask "Do you see what I see?" to help decide whether what I see is to be explained by self-knowledge or by knowledge of the presumed object.

Reported Response Rates In A.C. R. Conference Papers

Described in this section are the results of reading of 30 A.C.R. papers randomly selected from the "Competitive Papers" sections of Volumes V, VI and VII of Advances in Consumer Research. Conference papers were chosen because they more closely reflect our collective reporting habits than do journal articles which have been revised before publication according to the standards of the journal.

Table 1 describes the general nature of the 30 selected papers. Twenty reported surveys or, at least, non-experiments. Of the remaining 10, 6 were field experiments and 4 were conducted in a classroom or laboratory. Cut along a different dimension, 15 of the studies used an ad hoc sampling procedure designed for the particular study. In 11 studies, respondents were contacted through, and because of, their membership in an extant grouping such as a panel, classroom or volunteer organization. In the remaining 4 studies, one could not tell how respondents qualified for inclusion in the study.

Regardless of the type, every paper reported a value for "n." Sometimes it was the only aspect of the sample which was reported. Most papers contained a very brief description of the pool from which the sample was drawn; these ranged from the terse (e.g., "a sample of students") to a two or three sentence description of a panel's composition.

Beyond these basic items, what one might expect depends mostly on the type of study involved. At the risk of oversimplification, but in the hope of adding some structure to what follows, it is useful to think of the potential for non-response occurring at two levels. First, there is that level or stage which contains the processes by which the sample is framed and, second, there is the level containing the processes by which usable responses are obtained from a chosen sample.

To begin with the first stage, the choice of a sample frame, in effect, defines the actual universe of which the sample may (or may not) be representative. In all studies involving an ad hoc sampling procedure, it may be presumed that some set of screening questions or interviewer instructions are used to "qualify" respondents. Of the 15 studies using ad hoc sampling procedures, one field experiment reported the use of screening variables but gave no indication of the number rejected because of them. Another field experiment reported the number of persons contacted and the number lost and described some of the

TABLE 1

Selected Characteristics of Sampled Conference Papers

<u>Data Collection Method</u>	<u>Ad Hoc Sampling Procedure</u>	<u>Extant Groupings Including Panels</u>	<u>Indeterminate</u>
Surveys			
Personal Interview	3	2	
Telephone	1		
Mail	5	2	
Drop-off	2		
Group administration		2	
Indeterminate		1	2
Experiments			
Field	4	1	1
Classroom or Laboratory		3	1
Totals	15	11	4

screeners used. Two of the surveys stated the qualifications for respondent participation but did not give any indication of the number of contactees lost as a result. Eleven of the 15 studies using ad hoc sampling procedures simply did not mention any conditions or actions which, in effect, defined the universe being sampled.

When sampling from extant groupings, the issue is a bit more complex. First, there are the qualifications for participation in the group itself. As noted, most of the panels were briefly described but only one paper made any serious attempt to describe the means by which the panel had been formed or maintained. It is apparent that those who have access to commercial panels believe that mentioning such a panel by name is sufficient to describe its "representativeness." In a similar fashion, those who use classroom or volunteer organizations as samples seldom feel constrained to describe the "profile" of such groupings; only one paper described the general nature of the volunteer groups from which its respondents were drawn. Second, some further qualifications may or may not be imposed within an extant grouping. Two of the classroom experiments and one survey which drew its sample from a panel described the screening devices employed and the respondent losses associated with each.

At the second stage, it is necessary to divide the studies according to the method of data collection employed and the degree of researcher control over data collection. When a study is conducted by mail, for example, there is a considerable loss of respondents between the presentation of the instrument and the recovery of usable returns while, in the case of an interviewer-conducted survey, the major potential for non-response occurs before the instrument can be presented to the respondent. Even in experiments with extant groups, or in surveys that are administered in a "group" setting, a certain amount of subject loss can be expected, although it was mentioned in only 2 of the 6 such studies examined. In all 7 mail surveys, 2 "drop-off questionnaire" surveys and 2 of the field experiments, respondents completed instruments while out of the presence of the researcher. In these instances it is difficult to sort out the various forms of non-response. However, all but one of the field experiments reported some sort of "response rate" which, in effect, aggregated all losses between the distribution of the instrument and the counting of usable responses and expressed it as a percent of the number of instruments distributed. For example, each of the 7 mail surveys

reported the number of questionnaires sent out and the number of usable returns, although none gave any indication of any follow-up procedures not did any attempt any analysis based on quickness of return.

The surveys conducted by personal or telephone interview and the two field experiments which involved the recruitment of volunteers share a somewhat different set of problems. None of these 9 studies gave any indication of the number of potential respondents contacted, none mentioned refusals, nor did any suggest that there were any less-than-complete responses to presented instruments.

Summary and Conclusions

In the end, what a given researcher does or does not do is not really as important as whether or not what was done is openly reported in enough detail to allow the interested reader to make an evaluation of a study's internal validity. By this criterion, our collective reporting habits appear, to put it generously, rather loose. From the previous section, it can be concluded that the interested reader of A.C.R. papers would find it difficult, if not impossible, to answer either of two questions about the sample of the typical study: (1) What universe was being sampled? (2) How representative is the sample of that universe?

No one expects sampling procedures to be reported in such detail as to permit operational replication. Few would claim that such was desirable. But, what should concern everyone is that sampling procedures appear to be reported in such a way as to prohibit the extension of findings through constructive replication and the useful literature synthesis that both precedes and follows such efforts.

Some cynics claim that the reporting of methodology is so casual because many wish to obscure their sloppy procedures. Others acknowledge that the problems of reporting largely follow from a lack of understanding and propose that young scholars receive better training so that, over the next 10 or 20 years, reporting habits will be improved by attrition.

What is apparent to me is that the problem of reporting standard deserves attention now. If there is to be a body of literature, in any meaningful sense of that word, some standards must be imposed. Widely diffused and understood

definitions of response rates would be a useful first step; the quality of research might not be improved as a result, but the quality of interpretation would be enhanced considerably. Until all researchers are willing to ask, in Kaplan's phrase, "Do you see what I see?" we are not going to get very far.

References

- Kaplan, Abraham (1964), The Conduct of Inquiry, San Francisco: Chandler Publishing Company.
- Lykken, David T. (1968), "Statistical Significance in Psychological Research," Psychological Bulletin, 70 (February), 151-159.
- Madden, Charles S., Franz, Lori Sharp, and Mittelstaedt, Robert (1979), "The Replicability of Research in Marketing: Reported Content and Author Cooperation," in O.C. Ferrell, Stephen W. Brown and Charles W. Lamb, Jr., eds., Conceptual and Theoretical Developments in Marketing, Chicago: American Marketing Association, 76-85.
- Petrinovich, Lewis (1979), "Probabilistic Functionalism: A Conception of Research Method," American Psychologist, 34 (May), 373-390.
- Rogers, Everett M. and Shoemaker, Floyd (1971), Communication of Innovations, 2nd ed., New York: The Free Press.
- Rose, Arnold M. (1954), Theory and Method in the Social Sciences, Minneapolis: The University of Minnesota Press.
- Sawyer, Alan G. (1975), "Demand Artifacts in Laboratory Experiments in Consumer Research," Journal of Consumer Research, 1 (March), 20-30.
- Selltiz, Claire, Wrightsman, Lawrence S. and Cook, Stuart W. (1976), Research Methods in Social Relations, New York: Holt, Rinehart and Winston.

THE ROLE OF RESPONSE RATES IN EVALUATING MANUSCRIPTS FOR PUBLICATION

Robert Ferber, University of Illinois

The sad news I have to report is that in the manuscripts I have seen that involve surveys, a very large portion, at least half, are either rejected or sent back to the author for revision on account of inadequate attention to response rates and their effects. Let me cite some recent examples, mostly based on composites of past studies we have seen.

1. A mail survey is conducted of the patrons of a particular service to ascertain people's preferences for different attributes of this service. The response rate after one follow-up mailing is 40%. No further attempts are made to find out about the characteristics of the nonrespondents. However, the data are subjected to an elegant multivariate analysis, the results of which are used to recommend to management how their future services should be structured, and also to suggest to this, and other worlds, how all such services should be structured. Management was very happy with the recommendations; the JCR reviewers did not go into similar ecstasies.
2. A survey of the members of a consumer mail panel yields a 70% rate of response. A multivariate model is then developed to differentiate between the users and the nonusers of particular products, projecting to the population based on a finding of nonsignificance between the responders and the nonresponders to the mail questionnaire by various demographic characteristics. In this instance, two things were overlooked. First, virtually none of the demographic characteristics included in these comparisons for nonresponse bias were significant in the later analysis. Second, the mail panel by definition is already self-selected, a good estimate for that panel being that approximately 20% of those originally contacted were in the panel. Hence, the effective response rate was not 70%, but 14%.
3. A telephone survey is made of a random sample of users of a hot line assistance service provided by a state agency. No information is provided on the type of response obtained, though the later analysis proceeded to generalize on the characteristics of the users of this service.

In giving these examples, I do not necessarily mean to imply that a high response rate necessarily means a good survey, or even that special attention should be given to nonresponse in every survey. For example, if the purpose of a mail survey is to ascertain how many and what type of people are likely to be interested in a new service, the nonrespondents do not need special study, since by not replying these people have indicated they are not interested. (In a few instances, nonresponse may not indicate lack of interest, but these are so few as to be inconsequential.) The fact remains, however, that in the large majority of cases, one of the best ways how to assure rejection of a manuscript is to have a high rate of nonresponse, pay no attention to it, and then generalize to this and other populations.

Needed Information

For manuscript evaluation, what do we need to know about response rates? The answer is essentially the same as what we need to know in evaluating a particular study. First and foremost, we need to know the magnitudes of the re-

sponse rates both for the population as a whole and for key strata. If, for example, the purpose of a study is to compare certain differences in behavior between blacks and whites, it is pertinent to give the response rates for each group.

Important also is a clear definition of response rates, since without such a definition it is difficult to evaluate the significance of particular rates. Thus, a response rate could refer to the total sample list as a base, only eligible sample members, or only those who were contacted. Since the difference between these quantities can sometimes be very great, the term must be defined before it can be interpreted.

In addition to information on the size of the response rates, it is useful to have some interpretive comments from the author on the reasons why particular response rates were encountered, especially if they are very low. Such insights can be very useful for interpreting the results of the later analysis--and also indicate whether the author is cognizant of the sort of analytical problems that may be engendered by low response.

The other key item about response rates, something that is stressed everywhere but is frequently ignored, is the effect of the response rates on the results. Some people feel this is only likely to be a problem if the response rate is low, by some suitable definition, but this is not a very useful rule and can be very misleading. Even if the response rate in a mail survey is, say, as high as 80%, if the survey deals with something like people's willingness to undertake certain energy conservation programs, the chances are that the nonresponding 20% will have very different (probably much more negative) attitudes on the subject than the responding 80%. To leave out this group of nonresponders and to generalize from the 80% could lead to very biased results.

In fact, what is meant by a low response rate? Analytically, there does not seem to be any meaningful definition. In fact, most authors seem to try not to recognize that a response rate is low, and often try to minimize the size of the response rate by referring to lower response rates obtained in similar studies, or to special circumstances that mitigated against getting a better rate of response. If they do mention that the response rate is low, they tend to point out that it is still within the wide range of response rates obtained in this type of study, and is therefore acceptable--this was the exact argument used in a manuscript that was recently processed by us. Of course, by this argument it is hard to imagine what sort of response rate would be unacceptable.

In the final analysis, it is really immaterial whether a response rate is "good" or "bad". The key question is whether the response rate is such that the results of the study are likely to be affected by some sort of nonresponse bias. Clearly, so much depends on the subject of the study and the structure of the questionnaire. If willingness to undertake conservation programs is imbedded in a questionnaire covering many other topics, the effect of nonresponse on this question is likely to be much lower than if that topic is the sole subject of the questionnaire. In either event, some consideration is needed in virtually all instances of the possible effect of nonresponse on the results obtained in a survey study. One could generalize and say that such an evaluation is even more essential in the case of mail surveys, since nonresponse rates there can

be very low, but the same basic questions have to be posed in almost any type of survey.

What Response Rates Tell Us

The fact of the matter is that the treatment of response rates can tell a journal editor a great deal about the quality of a particular survey and the soundness of the data on which the analysis rests. If the author does not even report any response rates, he or she has a probability of just about zero of getting that manuscript accepted by JCR. This is true even in the case of a purposive sample, where some authors argue that since the idea was to obtain only a certain number of interviews with people having particular characteristics, and since they did so, response rates are of no consequence. (This is only true if the author could show that the nonrespondents are no different from the respondents on all relevant characteristics for that study; if that were possible, the study would be unnecessary.)

More important to editors is not only the reporting of response rates, but how the topic is handled. If the author gives short shrift to response rates in a situation where nonresponse bias is highly likely (such as a one-topic mail survey with no follow-ups), there is a strong indication that the author is not at all sophisticated in the analysis of survey data, and the manuscript will be returned. The one exception would be where the objective of the study is to illustrate some new analytical technique, and the survey data are used only for illustrative purposes. Even then, however, any sort of substantive inferences would have to be hedged on account of the possible bias due to nonresponse, and are actually best not made at all.

Personally, I find it rather curious that there seems to be little correlation between sophistication in the use of analytical techniques, such as multivariate analysis, and understanding of the problems that stem from nonresponse and other types of biases in survey data. With the increasing use of causal modeling techniques in consumer research, one would expect an increasing positive correlation in these two skills. So far, however, the correlation seems to be pretty close to zero--those who are sophisticated in survey techniques do not seem equally sophisticated in data analysis, and vice versa.

Recommendations to Authors

On the basis of these comments, let me conclude this presentation with a few suggestions on how authors of articles making use of surveys as the primary data source should deal with the subject of response rates:

1. Cite the overall response rate as well as the response rates for the subgroups.
2. Offer explanations for the reasons why the response rates are what they are, including comparison with the response rates on similar studies.
3. Discuss the possibility of nonresponse bias and indicate, hopefully with some justification, why you feel such bias is or is not present.
4. If nonresponse bias is expected, evaluate the nature of the bias as best you can. Moreover, keep this bias in mind when making generalizations about population attitudes or behavior, as well as in evaluating the significance of your findings (even though the significance will be diminished somewhat as a result).

In making these suggestions, I am not suggesting that you lengthen the manuscript appreciably. As many of you know, space is at a premium in both JCR and most other journals, and authors invariably write too much anyway. The fact remains, however, that this information can be presented very

briefly, often in footnotes, and the results will be a much more useful and informative manuscript. The result will also be a lesser probability that the manuscript will be rejected or returned for major revision (though even with this change I am afraid that the chances of any manuscript being returned at least for revision will remain quite high).

RECENT CONTROVERSY IN WASHINGTON: AN FTC CASE

Irving Roshwalb, Audits & Surveys, Inc., New York, N.Y.

Abstract

A recent FTC case provided the opportunity to review three important aspects of survey research employed in legal proceedings -- the role of confidentiality of survey responses in determining the admissibility of surveys as evidence, a definition of the independence of surveys, and the definition of survey response rates.

Introduction and Background

The "recent controversy" of the title involved the FTC's complaint that the survey findings used by a division of Litton Industries in its advertising were based on surveys which "...did not provide a reasonable basis for or prove the claims of the advertisements" (USA-FTC, p.5).

Later, the complaint goes on to say, "The Litton surveys had a very high rate of non-response. However, Litton failed to determine whether there was a bias of non-response, that is, whether the answers of non-respondents would have differed significantly from those of respondents" (USA-FTC, p.6).

This is, for those interested in the use of survey findings in legal proceedings, a very lively case. It presents several interesting aspects of the place of survey research in such proceedings. One of these is the problem of non-response and how it should be reported. Another deals with the terms under which a study should be deemed admissible in legal and/or regulatory proceedings. Finally, a third deals with the definition of an "independent" survey. The issues of survey "admissibility" and "independence" are discussed briefly before turning to the non-response problem.

Survey Admissibility

At one point in the hearings, the judge was asked to deny the admissibility of an FTC-sponsored survey on the grounds that there was no way for Litton counsel to identify the individual respondents and connect them with their questionnaires. The ruling of Judge John Mathias was quite to the point:

"...I find that the admissibility of surveys depends on their relevancy and trustworthiness and not upon whether respondents gain access to the codes identifying survey-respondents with their respective questionnaires" (USA-FTC Order, p.1).¹

This statement parallels the "Federal Rules of Evidence", which state, in part, "Attention is directed to the validity of the techniques employed rather than the relatively fruitless inquiries whether hearsay is involved" (West Publ. Co. 1977).

¹To keep the language straight, "respondent" in this context refers to those answering the FTC charges; "survey-respondent" refers to those interviewed as part of the studies.

What is of greater interest is that the order then proceeds to give the reasons for this finding, stressing the confidentiality of the relationship between a survey agency and its respondents. It is worth quoting in full:

"... Both (...survey agencies...) claim not only that they have given pledges of confidentiality to the interviewees, but that to break such confidentiality would be against the ethics of their profession and would be detrimental to the efficacy of future surveys to be conducted by themselves and others of their profession...there is some legal support for the surveyors' claims in this regard. Although there is certainly no pollster-client privilege, it is recognized that such claims of confidentiality cannot be lightly brushed aside. Where they can be recognized without depriving a litigant of discovery adequate to fairly meet his opponent's case, due deference should be given to legitimate claims of confidentiality" (USA-FTC, pp.6-7).

The concluding argument was:

"Since we are so dependent on surveys for much of the data upon which daily decisions are made, every attempt should be made to protect their accuracy. Yet the pollsters depend, to a great extent, on assurances of confidentiality to promote the efficacy of their product. If the survey-respondents believe they are going to be routinely re-interviewed and cross-examined in connection with any poll in which they participate, they will either refuse to participate or be guarded in their answers. In either event the survey method of gathering information is impaired. If enough people become so disenchanted with surveys that they refuse to participate, the ability to obtain a meaningful universe is seriously affected. If their answers are not free and open, the results of the poll are severely twisted" (USA-FTC, p.8).

Survey Independence

A second charge levelled against the advertising was that some of the ads carried the misleading notation that the surveys were independently conducted. The studies were actually designed by Litton and then turned over to an outside agency to execute the field work and prepare the tabulations. The judge's comments are again worth quoting:

"In either case (whether or not the ad carried the 'independent' notation) I find that the reader was not likely to believe that the Litton surveys were totally independent. It is difficult to perceive how any reader of the advertisements in question could possibly believe that the surveys were conceived, designed and conducted without any input by Litton, in view of their narrow focus. Further, the contact part of the surveys -- which might be thought of as the 'conduct' of the surveys -- was, in fact, conducted independently by (... the survey agency ...)" (USA-FTC, p.6).

This represents a judicial attempt to deal with the concept of an "independent survey" by pointing out that independence is not an absolute judgment but one that must be dealt with on some relative scale and reasonably.

Non-Response

The discussion of non-response in the affected studies was part of a much broader discussion of whether the studies were conducted according to the standards of the industry. The use of surveys in legal proceedings is generally governed by this concern. The U.S. Judicial Conference has recommended that the party offering the survey in evidence has the burden of establishing "...that the survey was conducted in accordance with accepted principles of survey research" (McCarthy 1973, p.508). One of the areas on which evidence would be required is that, "the sample, the questionnaire, and the interviewing were in accordance with generally accepted standards of objective procedure and statistics in the field of such surveys" (McCarthy 1973, p.509). It would appear, however, that defining "generally accepted standards" or "accepted principles" of survey research can be a difficult task.

Before turning to the discussion of the non-response rates in the studies directly involved in this case, a brief description of the two would be useful. One called for interviewing technicians working for independent companies that service consumer microwave ovens; the other called for interviewing technicians working for independent firms that specialize in servicing commercial microwave ovens. Lists of the service agencies were provided by Litton Industries to the survey agency that conducted the field work. The survey agency's role in the study was to execute the field assignment by telephone interview and to provide tabulations of the findings.

In the "Statement of Issues" section of the complaint filed by the FTC, the following charge was made:

- "II. The surveys do not provide a reasonable basis for or prove the claims of the advertisements.
- 2. The Litton surveys suffer from basic deficiencies in survey execution.
 - B. The Litton surveys had a very high rate of non-response. However, Litton failed to determine whether there was a bias of non-response, that is, whether the answers of non-respondents would have differed significantly from those of respondents" (USA-FTC, p.6).

The basis for this charge was obtained from the report summaries prepared by Litton Industries. In one instance, the report stated that the response rate for the Consumer Microwave Oven Technician Survey was approximately 47% (234 divided by 500) of the sample. In the second survey, the finding was that among commercial microwave oven technicians the response rate was 42.2% (211 out of 500).

The available data on sample disposition for the two surveys are contained in the following table. The first column of data refers to the Consumer Microwave Oven Technician Study while the second refers to the Commercial Microwave Oven Technician Study.

	CONSUMER MICROWAVE OVEN TECHNICIANS	COMMERCIAL MICROWAVE OVEN TECHNICIANS
Total Names Called	500	500
Not Qualified	80	95
Refusals	15	25
Could Not Be Reached	171	139
Disconnected, Out Of Business	--	30
Completed Interviews	234	211

It is clear from a review of this table and from the charges made by the FTC that the calculation of the response rate contained in the charges was more like that of a "hit rate", as much a measure of the quality of the list, perhaps, as of the quality of the effort expended to complete the interviews.

There are at least two entries in the table that should not be ignored in the calculation of response rates. The first is the listing of 'not qualified' persons. As far as evaluating the effort put into completing the interviewing assignment, these 'not qualified' respondents represent successful conclusions of interview attempts. That is so, at least from the point of view of the interviewer. If we consider these as successful attempts in the Consumer Technicians Survey, then the calculation of the response rate for this group, R(C1), is

$$R(C1) = (234+80)/500 = 314/500 = 62.8\%$$

Before we attempt a similar calculation for the commercial technicians, we have to consider what to do with the 30 telephone numbers which were found to be 'disconnected, out of business'. These 30 numbers are not part of the universe. Eliminating them from the base yields a response rate for the commercial technicians, R(CM1), of

$$R(CM1) = (211+95)/(500-30) = 306/470 = 65.1\%$$

A second approach is to measure response only among those who qualify. The rationale here is, of course, that it is only those who qualify who are the object of the research and only a measure of success among this group is reasonable and appropriate. Using this definition of non-response, a second response rate among the consumer technicians, R(C2), may be calculated --

$$R(C2) = 234/(234+15+171) = 234/420 = 55.7\%$$

The comparable calculation among the commercial technicians, R(CM2), is

$$R(CM2) = 211/(211+25+139) = 211/375 = 56.3\%$$

A third definition of response rate would now argue that the 139 consumer technician telephone numbers that were classified as 'nonreachable' and the 171 commercial technicians who 'could not be reached' should not all be judged as non-respondents. Some portion of them, could they be reached, would end up as unqualified. One estimate of how many should be so classified is to take the ratio of non-qualified among those that were reached and then apply this ratio to those who could not be reached. First, among consumer technicians:

The proportion who are estimated to be unqualified among the 171 who could not be reached is

$$80/(80+15+234) = 80/329 = 24.3\%$$

Therefore, the estimated number of the 171 who could not be reached who should be counted as non-respondents is $(1.000-.243)171 = (.757)(171) = 129$. The calculation of the response rate among consumer technicians, $R(C3)$, then becomes

$$R(C3) = (234+80)/(234+80+15+129) = 314/458 = 68.6\%$$

Among the commercial technicians, in a similar manner, the 139 who could not be reached are reduced to an estimated 99 non-respondents. The calculation of the non-response rates, incorporating this estimate is

$$R(CM3) = 211+95/(211+95+25+123) = 305/451 = 67.6\%$$

The following table summarizes these calculations.

	CONSUMER MICROWAVE OVEN TECHNICIANS	COMMERCIAL MICROWAVE OVEN TECHNICIANS
Hit Rate	46.8%	42.2%
R 1 (RC1 Or RCM1)	62.8	65.1
R 2 (RC2 Or RCM2)	55.7	56.3
R 3 (RC3 Or RCM3)	68.6	67.6

Clearly, the lowest value, that corresponding to the 'hit rate', is the least appropriate measure of response rate. In this study, either R1 or R3 is more appropriate. In both of these instances the number of respondents who are contacted and who prove to be unqualified for the purposes of the study are given fair representation in the calculation of the response rate. R1, the estimate that counts as successful interviews those contacts that establish the respondent as 'not qualified', is probably the better of the two, since it involves fewer assumptions. R2 simply assumes that all who are not reached are non-respondents, even though some of them would -- if ultimately contacted -- count among those who qualify.

The need to find a qualified respondent via some form of preliminary screening doesn't appear in all telephone surveys. In many studies, any adult member of the family is eligible for interview. As the screening procedure becomes finer, the incidence of unqualified (i.e., ineligible) respondents increases, and the problem of coping with them in the calculation of non-response becomes more acute.

Setting Standards

One of the difficulties facing an expert witness in a case such as this is to be able to testify on industry standards. Wiseman and MacDonald comment in their study of response rates that "there is a lack of agreement among industry leaders regarding terminology and reporting procedures" (1978). The exercise we have gone through is a reasonable display of the variety of response rate calculations that are available and all are based on reasonable definitions of what it is we are trying to measure. In many instances, the major problem is whether response rates are reported at all, and, if reported, whether they are clearly defined.

There is one more example of response rate reporting that should be considered. Suppose the interviewing assignment requires that a given number of interviews

be completed in each of several strata. Suppose further that the response rate varies among strata. In a simple example, the following table summarizes response rates and the required initial sample sizes needed to achieve the fixed number of completed interviews (in this case 10) in each stratum. Assume that the strata sizes are equal.

STRATUM I	RESPONSE RATE = R(I)	INITIAL SAMPLE = N(I)	COMPLETED INTERVIEWS
1	.80	12.5	10
2	.50	20.0	10
3	.20	50.0	10
4	.10	100.0	10
TOTAL		182.5	40

The simplest reading is that the response rate for this survey is $40/182.5 = 21.9\%$. If the assignment were stated a bit differently - as, for example, to estimate the response rate in the universe from which the sample was drawn, the solution would be to assign equal numbers of telephone numbers to call in each stratum, observe the response rates in each, and then construct the appropriate estimate of the universe parameter, as in the following table.

STRATUM I	R(I)	NO. OF CALLS	COMPLETED INTERVIEWS
1	.80	25	20
2	.50	25	12.5
3	.20	25	5.0
4	.10	25	2.5
TOTAL			40.0

The response rate calculation for this set of data is $40/100 = 40.0\%$. In fact, the first procedure produces the harmonic mean of the strata response rates, while the latter calculation yields the arithmetic mean of the strata response rates. The harmonic mean is generally less than the arithmetic mean and would always produce an underestimate of the response rate.

This example bares another source of confusion in the calculation of response rates. Is the purpose of the calculation to come up with a descriptive, mechanical exercise to run the observed data through, or is the purpose to estimate the response rates in the universe from which the sample data have been drawn? The first is characterized by a (perhaps, apocryphal) researcher's comment made in the wake of the 1948 election poll experience that "we have no non-response problem. If we promise 100 interviews, we keep on going until we have 100 interviews." That response rates are governed by some probability mechanism has long been recognized, particularly in procedures proposed by Hartley and Politz and Simmons to weight for not-at-homes by the inverse of the probability of finding respondents at home (Hartley 1946, Politz & Simmons, 1949).

In this simple example, which calls for a fixed number of interviews per stratum where response rates may vary among strata, the mechanical calculation of response rates results in an underestimate of the response rate for the universe of the study. However, the simple arithmetic mean of the strata response rates yields an unbiased estimate of the universe response rate. So, not only must the elements of the response rate be properly designated, but the estimation procedure itself must be properly selected.

Conclusion

A number of things are quite clear as one considers the experience in this FTC case. This is an example of the ever-growing use of survey data by the parties to legal proceedings. As survey evidence finds wider usage in such matters as trademark cases, advertising substantiation and the support of product claims, the reviews of these documents can be expected to become more critical.

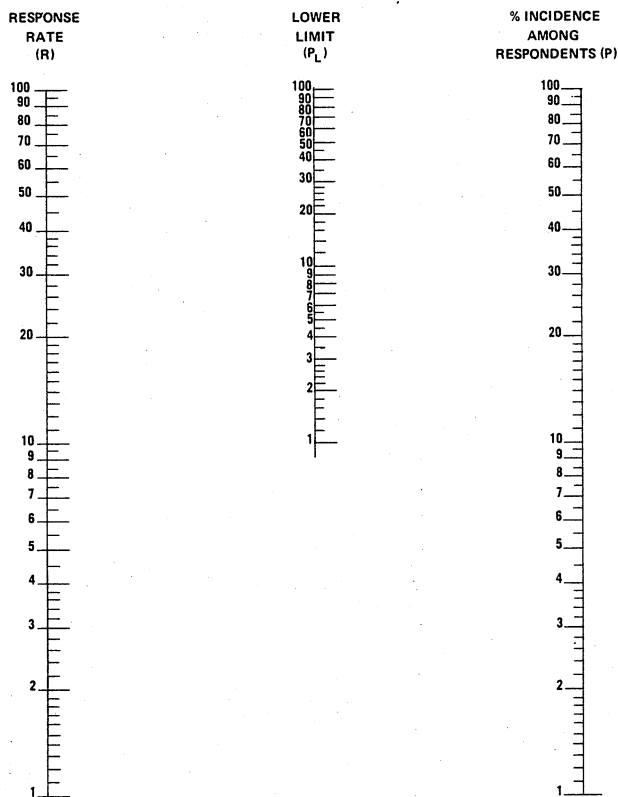
Creating and adhering to standards of survey research are going to become increasingly important to the users of survey research and, therefore, to the practitioners of the research art. And I use the word 'art' advisedly. It was not without careful thought that Stanley Payne named his book, "The Art Of Asking Questions" (1951).

The difficulty in setting standards for response rates is the temptation to be dogmatic, that is, to establish by fiat that response rates below some arbitrary level are inadequate or unacceptable. That is the easy way out. The FTC charge in this case states explicitly what concerns us when we discuss response rates. To repeat the relevant sentences from the charge, "The Litton surveys had a very high rate of non-response. (Forget for the moment that the calculation actually used was inappropriate. The next sentence is what counts.) However, Litton failed to determine whether there was a bias of non-response, that is, whether the answers of the non-respondents would have differed significantly from those of respondents" (USA-FTC, p.6).

Of course, that's where the answer to the problem lies, regardless of the rate of non-response. In most surveys, while this solution is implicitly recognized, it rarely is investigated thoroughly. For one thing, the nature of the non-response beast is that it doesn't make itself readily available for measurement. In any case, time and budget constraints usually preclude much effort in this direction. But there are other things that can and should be done. The first of these is to report the non-response rate, along with a precise definition of how it was calculated. The second is to state what the effect of the non-response could be, given some reasonable assumptions about the range of possible differences between respondents and non-respondents. One such solution is contained in

the nomogram reproduced below (Roshwalb 1970, p.17).

BIASES DUE TO NON-RESPONSE IN SURVEYS Binomial Variate



$$\text{LOWER LIMIT: } P_L = RP$$

$$\text{UPPER LIMIT: } P_U = RP + (1-R) \\ = P_L + (1-R)$$

For a stated response rate and the observed binomial proportion in the respondent sample, the nomogram yields directly the estimate of the lower limit of the total sample (i.e., respondents plus non-respondents) estimate if the binomial estimate for the non-respondents is taken as 0.00. It also indicates the value for the total sample if the value among the non-respondents is taken as 1.0. The possible effect of non-response on the estimate based on the total sample is thus available under very extreme conditions. With more information about the universe and subject matter under study it is possible to place more reasonable limits on the values for the non-respondents and thus obtain more reasonable estimates for the total sample.

References

Hartley, H.O. (1946), "Discussion of paper by F. Yates," Journal of the Royal Statistical Society, 109, 37.

Manual for Complex Litigation, St. Paul, Minn.: West Publishing Company (1977).

McCarthy, J. Thomas (1973), Trademark and Unfair Competition, Rochester, N.Y.: The Lawyers Co-operative Publishing Company.

Payne, Stanley L. (1951), The Art of Asking Questions, Princeton, N.J.: Princeton University Press.

Politz, Alfred N. and Simmons, Willard R. (1949), "An Attempt To Get The 'Not-at-Homes' into the Sample Without Callbacks," Journal of the American Statistical Association, 44, 9-31.

Roshwalb, Irving (1970), Nomograms For Marketing Research, New York: Audits & Surveys, Inc.

United States of America, before Federal Trade Commission. In the matter of Litton Industries, Inc., a corporation, and Litton Systems, Inc., a corporation. Docket No. 9123.

United States of America, before Federal Trade Commission. In the matter of Litton Industries, Inc. Docket No. 9123. Order concerning the identification of individual survey respondents with their questionnaires. 6-19-79.

Wiseman, F. and McDonald, P. (1978), The Non-Response Problem in Consumer Telephone Surveys, Cambridge, Mass.: Marketing Science Institute.

LACK OF AGREEMENT ON THE STANDARDIZATION
OF RESPONSE RATE TERMINOLOGY IN THE SURVEY RESEARCH INDUSTRY

Joy Williams-Jones, Marketing Science Institute

Abstract

A survey of commercial research suppliers and users revealed widespread inconsistency in definition and calculation of four commonly used response rate terms. This lack of standardization may be a function of lack of formalized instruction at the academic level, and/or the fact that many response rate terms can be used for distinctly different purposes. This paper reviews results of that survey, illustrating the wide range of variation in terminology definitions; and stresses, in light of recent developments, the need for standardization of terminology across all survey research practitioners.

Introduction

Fred Wiseman earlier reported on the combined results of 182 consumer surveys shared with us by the member companies of the Marketing Science Institute (MSI) and Council of American Survey Research Organizations (CASRO) which encompassed over one million interview attempts. The main purpose of the synthesization of those results was to determine current response rates in U.S. survey research, in an effort to either corroborate or dispel the suspicions that readily reachable and cooperative respondents were becoming more and more difficult to contact.

In Wiseman's and McDonald's (1978) efforts to compare response rate data from one survey to the next, however, they encountered two problems: (1) varying terminology used to describe survey outcome "rates"; and (2) lack of comparable calculation methods used to develop these rates.

A Steering Committee composed of eight MSI member company representatives suggested surveying MSI and CASRO members in order to determine current practice in the research industry regarding the use and calculation of selected response measures used. This paper will report some of the results compiled from that survey.

Procedure

A short questionnaire was developed which asked respondents (field directors and appropriate marketing research personnel who were involved in the computation or use of response rates) to indicate:

- the relative frequency with which the following four response terms were calculated in attempting to describe the outcome of particular consumer data collection efforts;
 - Response rate
 - Completion rate
 - Refusal rate
 - Contact rate
- and next, using the results of three actual telephone survey outcomes as data sets, they were asked how they would calculate the four rates, for each survey.

The data sets included three types of sampling: (1) telephone directory, (2) random digit dialing, and (3) pre-determined respondent lists (Exhibits 1-3):

Exhibit 1

Source of Sample: Telephone Directory

Category	Frequency
Disconnected/non-working number	426
Household refusal by someone other than designated respondent	153
No answer, busy signal, not at home	1757
Interviewer reject - language barrier	150
Interviewer reject - other	37
Respondent refusal	711
Ineligible respondent	74
Completed interviews	<u>501</u>
Total	4175

Exhibit 2

Source of Sample: Random Digit Dialing

Category	Frequency
Disconnected/non working number	702
Business/non-household number	237
No answer, busy signal, not at home	1467
Interviewer reject - language barrier	15
Interviewer reject - other	8
Respondent refusal	189
Ineligible respondent	531
Termination by respondent	1
Completed interviews	<u>99</u>
Total	3249

Exhibit 3

Source of Sample: List

Category	Frequency
No phone number available for designated respondent	175
No answer, busy, not at home	13
Interviewer reject - language barrier	10
Interviewer reject - other	4
Respondent refusal	169
Ineligible respondent	10
Over quota	8
Termination by respondent	5
Completed interviews	<u>211</u>
Total	605

Results: Frequency of Terminology Usage

In total, 55 professional field directors (on the "supplier" side) and marketing research personnel (on the "user" side) responded to our questionnaire. Table 1 illustrates that completion rate is the term most commonly calculated by these researchers with 40% of the sample indicating "always"

using that figure. Contact, refusal, and response rates were not as frequently used.

Table 1

Reported Frequency of Calculation

Rate	Frequency of Calculation						
	Always 1	2	3	4	5	6	Never 7
Response	16%	22%	4%	9%	7%	13%	13%
Completion	40	29	7	2	0	9	4
Refusal	22	25	15	5	4	13	7
Contact	24	13	11	11	2	18	13

(N = 55)

It is not surprising that "completion rate" was the term most commonly used, due to the standard practice of cost bidding based on estimates of eligible respondent incidence in the population.

When a survey research supplier is requested to "bid" on a proposed project, the client is commonly asked to provide an incidence estimate for his required sample. For example, if one wished to conduct a telephone survey of 200 female heads-of-household who are regular users of a product category they might refer to TGI (Target Group Index) or other syndicated sources, to learn that 40% of females between the ages of 21 and 65, are regular users of that category. Assuming females compose about one-half the population, the incidence would be estimated at .4 x .5 or 20%. The research supplier can then assume that 2 out of every 10 contacts made should yield an eligible respondent. Based on this estimate, the research firm can then determine the number of interviewer hours that should be required to produce 200 completed interviews. (It should be noted that length of interview, sensitivity of the topic area, and "day-parts" in which the interviewing is conducted are additional factors which can be expected to affect ultimate response rates.)

Upon completion of the interviewing, the research supplier will commonly base the final study cost on the actual number of staff hours required to complete the field portion of the survey. If time estimates based on the assumed incidence figures are substantially "off," the actual incidence encountered during the field work is often used to proportionately increase or decrease field work charges.

While "completion rates" are useful in costing and planning research surveys, the common usage of this rate does not necessarily indicate that commercial researchers are utilizing that calculation as an aid to assessing data quality.

Results: Rate Calculations

Among those respondents attempting the 12 calculations we requested (Ns ranged from 37 to 50 out of 55 possible), we found substantial variability in the interpretation of the rates. Table 2 indicates the minimum and maximum reported rates for each data set.

Using the telephone directory sample data set as an example, a brief discussion of each rate illustrates the broad variation in definition, calculation and potential use revealed in the survey.

Completion Rate

Two basic interpretations, with vastly different meanings, were found for the most often calculated "completion rate"

Table 2

Summary Statistics

	Minimum	Median	Maximum	Range	N
Response Rate					
Telephone directory	12%	30%	90%	78	40
Random digit	3	20	97	94	39
List	35	54	97	62	37
Completion Rate					
Telephone directory	12	25	61	49	50
Random digit	3	12	34	31	47
List	35	51	96	61	46
Refusal Rate					
Telephone directory	17	47	65	48	49
Random digit	6	23	68	63	48
List	28	42	47	19	47
Contact Rate					
Telephone directory	23	48	53	31	48
Random digit	10	26	55	45	46
List	39	69	97	58	45

(Table 3). These were: (1) the percentage of completed interviews out of all respondents/households selected and (2) the percentage of completed interviews out of all respondents/households contacted. The latter completion rate will always have a value greater than or equal to the former rate.

A key issue with respect to the denominator of the term is whether or not it should include all potential respondents/households contacted or all originally selected.

As can be seen in Table 2, while completion rate is the most commonly used term among our sample, 13 different definitions were obtained from the 50 respondents reporting their calculation methods. Based on the specific definition used, this rate could be set anywhere from a minimum of 12% to a maximum of 61%. This range of variability may in part be due to differences in the purposes for which individual researchers calculate this rate (to be discussed later).

Response Rate

The greatest amount of confusion and uncertainty existed for "response rate," with 29 different definitions specified out of the 40 responses obtained (Table 4). Fifteen respondents did not attempt this computation, either because they never compute its value or because they were not sure how it should be done. According to our professional practitioners' calculations, the telephone directory survey achieved a response rate "somewhere" between 12% and 90%. The definitions suggested most frequently occurred only three times.

Definition 1 suggests that, for many researchers, this term measures the extent to which selected respondents/households were accounted for (i.e., were contacted and could be classified into any category at all). This interpretation differs dramatically from the traditional meaning of the term and its subsequent use as a measure of data reliability.

The second definition differs from the first only by eliminating household refusals from the numerator, suggesting that the proponents of this formula feel that the intended respondent must be reached by the interviewer if he is to be counted as a response -- even if he later refuses, terminates the interview, or is found to be ineligible.

The third method of calculation seems to be more appropriate as a "completion rate" than a "response rate."

Table 3

Completion Rate Calculations for Telephone Directory Sample

		Computed Value	No. Using	Total No.
Most Frequency Used Definitions				
1.	$\frac{\text{Completed Interviews}}{\text{All}}$	= 12%	13	50
2.	$\frac{\text{Completed Interviews}}{\text{HH Refusals + Rejects + Refusals + Ineligibles + Terminations + Completed Interviews}}$	= 25%	11	
3.	$\frac{\text{Completed Interviews}}{\text{Rejects + Refusals + Ineligibles + Terminations + Completed Interviews}}$	= 27%	7	
(Number of Different Definitions Reported = 13)				

Minimum and Maximum Reported Definitions

$\frac{\text{Completed Interviews}}{\text{All}}$	= 12%
$\frac{\text{Completed Interviews + Terminations}}{\text{Completed Interviews + Terminations + Ineligibles}}$	= 61%

Table 4

Response Rate Calculations for Telephone Directory Sample

		Computed Value	No. Using	Total No.
Most Frequency Used Definitions				
1.	$\frac{\text{HH Refusals + Rejects + Refusals + Ineligible + Terminations + Completed Interviews}}{\text{All}}$	= 48%	3	40
2.	$\frac{\text{Rejects + Refusals + Ineligibles + Terminations + Completed Interviews}}{\text{All}}$	= 44%	3	
3.	$\frac{\text{Completed Interviews}}{\text{All}}$	= 12%	3	
(Number of different definitions reported = 29)				

Minimum and Maximum Reported Definitions

$\frac{\text{Completed Interviews}}{\text{All}}$	= 12%
$\frac{\text{Refusals + Ineligibles + Terminations + Completed Interviews}}{\text{Rejects + Refusals + Ineligibles + Terminations + Completed Interviews}}$	= 90%

Looking at the difference between the definitions which produced the minimum and maximum calculations, it is obvious that researchers are confused regarding the appropriate populations to consider as the "universe" and the "responses." The definition yielding a 90% response rate, could be interpreted to mean response rate is the proportion of all respondents who interacted with the interviewer in any way that were not rejected by the interviewer as inappropriate (i.e., "We count everyone we talked to as a response except for those we were responsible for eliminating").

Contact Rate

The greatest amount of consistency existed for "contact rate," with 38 out of 49 respondents calculating rates within a range of 11 points (44% to 53%, Table 5). It's rather ironic in this discussion of precise rate calculations to consider an 11-point spread as an example of consistency, but when we realize just how great the variation is in the existing definitions of terminology, 11 points begin to sound reasonably small.

It is interesting to note that the two most frequently specified definitions for contact rate were identical to those specified for response rate (definitions 1 & 2 in Tables 4 and 5).

To simply percentage the number of refusals on all attempts -- the formula yielding the lowest rate (17%) also seems a bit unrealistic. How could non-working numbers, or not-at-homes be considered possible refusers?

Table 5

Contact Rate Calculations for Telephone Directory Sample

		Computed Value	No. Using	Total No.
Most Frequency Used Definitions				
1.	$\frac{\text{HH Refusals} + \text{Rejects} + \text{Refusals} + \text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}{\text{All}}$	= 48%	20	49
2.	$\frac{\text{Rejects} + \text{Refusals} + \text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}{\text{All}}$	= 44%	11	
3.	$\frac{\text{HH Refusals} + \text{Rejects} + \text{Refusals} + \text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}{\text{All} - (\text{Disconnected/non-working})}$	= 53%	7	

(Number of different definitions reported = 12)

Minimum and Maximum Reported Definitions

$\frac{\text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}{\text{All}}$	= 23%
$\frac{\text{HH Refusals} + \text{Rejects} + \text{Refusals} + \text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}{\text{All} - (\text{Disconnected/non-working})}$	= 15%

Definition 1 assumes that if a household is reached at all, it is considered a contact. The second limits this a little further, by requiring the specified respondent to have interacted with the interviewer.

Both of these definitions essentially imply that contacts represent the sheer proportion of households or respondents reached, out of the number of all attempts made.

The third definition does not include non-working numbers in its denominator (i.e., as a counted "attempt") thus raising the contact rate proportionately.

The method of calculating which produced the lowest (23%) rate, does not inherently seem to be well thought out. If respondents who terminate the interview are included in the numerator, why aren't refusals? Similarly, if ineligibles are included, why not those rejected by the interviewer?

Refusal Rate

Most researchers agreed that once a household contact is made, any way in which cooperation is refused or terminated (by the specified respondent or another household member) would be defined as a refusal (in the numerator, Table 6).

The denominator definition (i.e., the universe on which to percentage the refusal rate), however, accounted for most of the variation in responses with 23 different definitions supplied by 49 respondents.

It would seem rational to assume that only those eligible for inclusion in the study would be asked to participate and thus, would be the only ones able to "refuse." Definitions 1 and 2 (Table 6), however, both include ineligible respondents in the denominator. And definition 1 also leaves in interviewer "rejects."

Overall, it seems that the 65% rate makes the most sense: of those who could have been interviewed (i.e., eligible, reached, and not rejected) what proportion refused to complete or participate in the interview. However, this definition did not receive many supporters among the sample.

Conclusions About Definitions

Based on even this relatively small sample of those involved in survey research on a day-to-day basis we can draw the following conclusions:

- There is little or no consistency among either users or suppliers of survey research (within or across organizations) as to how various response and non-response terms are or should be defined.
- The traditional meaning of response rate appears to have been lost and, as a result, the term can no longer be considered as an indicator of data quality.
- The most commonly used "completion rate," when defined as "the percentage of completed interviews obtained from the total number of contacts" is of some value. It is most useful for determining the number of attempts that will be required for future surveys with the same respondent requirements as an aid for planning, scheduling and budgeting. However, unless there are no eligibility requirements posed in a survey, completion rate cannot be considered an indicator of data quality.
- New measures are needed (or old terms should be redefined) which will enable data users to assess the representativeness of their samples and the reliability of the obtained results.

Table 6

Refusal Rate Calculations for Telephone Directory Sample

		Computed Value	No. Using	Total No.
Most Frequency Used Definitions				
1.	$\frac{\text{Household Refusals} + \text{Refusals} + \text{Terminations}}{\text{HH Refusals} + \text{Rejects} + \text{Refusals} + \text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}$	= 47%	11	49
2.	$\frac{\text{Household Refusals} + \text{Refusals} + \text{Terminations}}{\text{Household Refusals} + \text{Refusals} + \text{Ineligibles} + \text{Terminations} + \text{Completed Interviews}}$	= 52%	4	

Minimum and Maximum Reported Definitions

	$\frac{\text{Refusals}}{\text{All}}$	= 17%
	$\frac{\text{Household Refusals} + \text{Refusals} + \text{Terminations}}{\text{Household Refusals} + \text{Refusals} + \text{Terminations} + \text{Completed Interviews}}$	= 65%

Sources of Confusion

In subsequent discussion of these findings with MSI and CASRO researchers we began to wonder how the problem of different interpretations materialized. First of all, none could recall the definitions of these terms being formally taught in statistics or research courses. The general experience was learning to use the definitions accepted within the companies one had worked for, or with, as a user or supplier.

Secondly, it was noted that one of the major sources of definition discrepancy was that terms such as "response, contact, completion and refusal rates" can be used for two distinctly different purposes: (1) to assess the reliability of the data obtained, and (2) to assess the quality of the data collection effort of the research supplier. Consider the following example:

A research company is asked to conduct interviews with 100 individuals whose names were obtained from a warranty card list. This list turns out to be of poor quality in that for 30 names, no phone numbers or addresses are available. With the remaining 70 names, the research supplier completes 60 usable interviews.

What response rate should be reported: 83% (60/70) or 60% (60/100)? If we are trying to assess the quality of the data collection effort, then clearly the higher rate is more relevant. If we are concerned with reliability, then the lower figure is more appropriate indicating the larger proportion of respondents who were unaccounted for. It is here that dilemma exists.

Implications for Future Survey Research

So far this discussion has centered around the issue that researchers do not agree on standard definitions for response rate terminology. Some might be inclined to conclude, "So what? As long as we know how we calculated our rates and can tell how we did it, then we can each call that percentage by any name we choose." Actually, that "so what" attitude could be supported; if all researchers agree to report their calculations, and if we as researchers were the exclusive users of the terms. That, however, is not the case on either count.

Government agencies, in an effort to improve the reliability of their survey data, which is often used as bases for policy information and program development, have increasingly begun to require researchers to guarantee a minimum response rate. In a recent national probability sample survey MSI conducted for the U.S. Department of Agriculture, we were asked to guarantee an 85% response rate. Other of MSI member companies are also reporting requirements. The Litton case illustrates that the FTC, and, we must assume, other agencies are increasing scrutiny of research already completed and using as a basis for some of their judgements the "adequacy" of the achieved response rate.

This raises the obvious question: How can a researcher guarantee to achieve a specific "rate" that can be defined (at least in the case of 39 of our researchers, Table 2), to range anywhere from 3% to 97% for the same data base?

In addition, do we really know what an acceptable response rate should be (i.e., is 85% really better than 80% in terms of the reliability of the data obtained)? The requirement that individuals and commercial researchers who use survey results in testimony, advertising claim substantiation, rate change hearings, grant research, etc., will be held responsible to meet an as yet undefined and possibly unrealistic criterion, is unsettling.

In addition, arbitrary response rate requirements can force the user of surveys into increased time and financial expenditures, which may be unnecessary for certain categories (product usage, attitudinal studies, etc.) in which non-response bias may not be a serious problem. Without the results of a definitive study to determine what topics may be more, or less, susceptible to such errors, research users leave themselves open to potentially increasing outside control of survey efforts.

Toward Resolution of the Issues

As Fred reported earlier, MSI and CASRO have recently joined forces along with representatives from the Bureau of the Census, Office of Federal Statistical Policy and Standards, Advertising Research Foundation, and the American Statistical Association, in coordinating work in response rate terminology standardization and designing a large scale national study to investigate the impact of nonresponse on the quality of data collected in surveys.

The academic community too, should play a supporting role in helping to resolve the issues at hand. Once standardized definitions have been established, it is incumbent on the academic researchers and educators to adopt these definitions, report the rates in all published research, and disseminate this terminology to future researchers currently in the universities.

References

Mathias, John J. (1980), Initial Decision: In the Matter of Litton Industries, Inc. a corporation, and Litton Systems, Inc., a corporation, Docket No 9123, Federal Trade Commission.

Wiseman, Frederick and McDonald, Philip (1978), The Non-response Problem in Consumer Telephone Surveys, Cambridge, Marketing Science Institute.

INVESTIGATING "INCOME REFUSALS" IN A
TELEPHONE SURVEY BY MEANS OF LOGIT ANALYSIS

Robert A. Peterson, University of Texas at Austin
Robert P. Leone, University of Texas at Austin
Mohammad H. Sabertehrani, University of Texas at Austin

Abstract

A logit analysis was employed to investigate the relationship between refusal to answer an income question and four demographic variables. Data were drawn from a telephone survey of 6178 adult consumers regarding their financial attitudes and behavior. All four demographic variables were significantly related to refusal to respond; the primary reason given for refusal related to "invasion of privacy."

Introduction

Associated with every empirical survey is a certain amount of "error." This error can be conceptually decomposed into sampling error and nonsampling error. Sampling error is a function of data quantity--the greater the quantity of survey data, the smaller the magnitude of sampling error. Nonsampling error, on the other hand, relates to the quality of survey data, how "good" or "meaningful" the data are with regard to various criteria such as validity, reliability, and generalizability.

Traditionally most research on survey error has focused upon sampling error, perhaps due in part to reasons of analytical ease. This is in spite of the fact that non-sampling error (data quality) is likely to be more directly associated with the quality of inferences derived from the data. Recently, however, there has been an increasing interest in, and research relating to, the assessment and impact of data quality. This research has not been limited to marketing, but pervades all the behavioral sciences (see for instance, Bailar and Lanphier (1978) for several data quality issues currently being addressed).

A major nonsampling error is nonresponse or refusal to cooperate. Nonresponse may be complete, as when a sampled individual refuses to participate at all in a survey. Or, nonresponse may be partial, as when a study individual refuses to answer a specific question. This type of refusal is commonly termed an item nonresponse and is an "error of omission". It most frequently occurs for questions relating to sensitive issues such as personal hygiene or moral behavior. Item nonresponse is particularly acute in telephone interviews. This is because study individuals are neither offered the anonymity of a mail interview, nor do they experience the social pressure to answer of a personal interview.

The present research is concerned with investigating item nonresponse. While there has been voluminous research conducted on refusal to participate or cooperate in a survey per se, especially in the context of mail interviews, relatively little research has been conducted on individual item nonresponse. Still, what limited research does exist suggests that item nonresponse can significantly influence data quality in several ways (see Ferber 1966, Craig and McCann 1978, Wiseman and McDonald 1979).

Specifically, the present research consists of an investigation of respondent characteristics relating to refusal to answer a household income question in a telephone survey. The household income question is frequently a crucial question in consumer surveys for several reasons. Most obviously, income commonly serves as a dependent var-

iable (to be estimated) for certain classes or groups of consumers. Moreover, income is often used as an independent variable to predict other, often behavioral, variables. In addition, income data are typically used to evaluate the representativeness of a sample by comparing sample income data with criterion (usually population or census) income data. Finally, income is sometimes employed as a surrogate variable or in combination with other variables in index form.

Despite the importance of the income question, nonresponse to income questions tends to be relatively high. Indeed, among demographic or socio-economic variables commonly employed in marketing research surveys it tends to possess the highest refusal rate (Herriott 1977, Locander and Burton 1976). While anecdotal evidence would explain income nonresponse as a function of the "private nature" of income, little empirical evidence exists as to who refuses to answer the income question or why study individuals refuse to answer. The purpose of this paper is to provide insights into both the who and the why questions.

Methodology

Research Data

Data for this investigation were derived from telephone interviews with 6178 adult consumers in a southwestern metropolitan area. Interviews lasted 8-10 minutes and had as their subject matter financial attitudes and behavior. Data collection employed systematic random digit dialing, a dialing technique which gives each household in the specified exchanges an equal probability of being included in the sample (it is a non-directory approach). Two call-back attempts (three attempts in all) were made to contact study individuals, and all interviewing was conducted between the hours of 6-9 p.m. weekdays and 10 a.m. to 9 p.m. weekends. Across all study individuals interviewed the overall refusal rate for the income question was 12.6 percent (779 of the 6178 individuals interviewed).

The dependent variable was refusal to answer an income question. This question was asked in a straightforward manner as "What is your total annual household income?" with five response categories provided. Four demographic characteristics were employed as independent variables: sex, marital status, education, and age. To facilitate data analysis the latter three variables were dichotomized. Marital status response categories consisted of "married" and "not married" (the latter including individuals who were single, divorced, or widowed). The "low" education category included individuals who possessed no college education while "low age" study individuals were those 18-34 years of age. These variables were respectively coded as follows:

Variable	Category	Code
sex	male	1
	female	0
marital status	married	1
	not married	0
education	low	0
	high	1

Variable	Category	Code
Age	low	0
	high	1

Study data are presented in Table 1, which is essentially a four way (16 cell) contingency table. Observed proportions can be interpreted as refusal probabilities. Thus the probability was .087 that a low age, low education, married male refused to answer the household income question.

TABLE 1
Income Question Refusal Probabilities

		Male		Female	
		Married	Not Married	Married	Not Married
Low	Age	.087(160)*	.130(154)	.135(377)	.129(210)
	Educa- tion				
High	Age	.137(256)	.111(54)	.213(451)	.196(179)
	Educa- tion				
Low	Age	.076(618)	.055(814)	.108(741)	.074(717)
	Educa- tion				
High	Age	.153(528)	.120(83)	.181(574)	.183(262)
	Educa- tion				

*Figures in parentheses represent number of study individuals in cell.

Analytical Approach

Due to the nature of the research data (i.e., a discrete dependent variable and a relatively large sample size) a logit model approach was selected as appropriate for data analysis. The particular approach used ECTA, a likelihood ratio analytical model derived by Goodman (1972). Such an approach permits prediction of the probability of refusal to answer the income question (R) as a function of education (E), age (A), sex (S) and marital status (M). Technically the objective of logit analysis is to reduce a "saturated" model which incorporates all independent variable effects at all levels (and which obviously fits the observed data perfectly) into a more parsimonious model that still provides an acceptable fit to the data. The saturated model can be posed in standard ANOVA format as:

$$L_{ijkl}^{\bar{R}} = \mu^{\bar{R}} + \mu_{(i)}^{\bar{E}\bar{R}} + \mu_{(j)}^{\bar{A}\bar{R}} + \mu_{(k)}^{\bar{S}\bar{R}} + \mu_{(l)}^{\bar{M}\bar{R}} + \mu_{(ij)}^{\bar{E}\bar{A}\bar{R}} + \mu_{(ik)}^{\bar{E}\bar{S}\bar{R}} + \mu_{(il)}^{\bar{E}\bar{M}\bar{R}} + \mu_{(jk)}^{\bar{A}\bar{S}\bar{R}} + \mu_{(jl)}^{\bar{A}\bar{M}\bar{R}} + \mu_{(kl)}^{\bar{S}\bar{M}\bar{R}} + \mu_{(ijk)}^{\bar{E}\bar{A}\bar{S}\bar{R}} + \mu_{(ijl)}^{\bar{E}\bar{A}\bar{M}\bar{R}} + \mu_{(ikl)}^{\bar{E}\bar{S}\bar{M}\bar{R}} + \mu_{(jkl)}^{\bar{A}\bar{S}\bar{M}\bar{R}} + \mu_{(ijkl)}^{\bar{E}\bar{A}\bar{S}\bar{M}\bar{R}} \quad (1)$$

where $L_{ijkl}^{\bar{R}}$ is the natural logarithm of the odds ratio refusal probability/nonrefusal probability (the logit) and the μ 's represent the respective independent main and interaction effects.²

To simplify a saturated model M_1 the μ terms are systematically constrained to zero in a step-wise manner and the loss in explanatory power calculated at each step. Selection of terms to be included in the final model is done by simultaneously evaluating their standardized coefficients and explanatory power (variance accounted for). This approach proceeds in a hierarchical fashion with an

TABLE 2
Likelihood Ratio Results for Selected Logit Models

Model	Fitted Marginals	G ²	d.f.	p	Gain ₂ in G ²	d.f. of Gain	p	Relative Variance Explained
M ₁	EASMR	144.06	15	.000				
M ₂	+ AR	47.87	14	.000	96.19	1	<.001	66.8
M ₃	+ ER	36.21	13	.001	11.66	1	<.001	74.9
M ₄	+ SR	18.00	12	.116	18.21	1	<.001	87.5
M ₅	+ MR	13.59	11	.136	4.41	1	<.05	90.6
M ₆	+ EAR	8.46	10	>.500	5.13	1	<.05	94.1
M ₇	+ AMR	7.80	9	>.500	0.66	1	N.S.*	94.6
M ₈	+ EASMR	0.00	0	-	7.80	9	N.S.	100.0

*Not significant at .05.

²In general logit is the logarithm of the odds ratio $L = \ln \frac{p}{1-p}$ where $p = 1/(1 + e^{-x\beta})$.

independent variable's lower order terms being simultaneously evaluated with significant higher order terms.

Results

Table 2 contains a summary of the logit analysis. The saturated model M_1 has been reduced to the best fit model M_6 through the retention of significant independent variables. Thus model M_6 is given as:

$$\bar{L}R_{ijkl} = \mu \bar{R} + \mu \bar{E}R + \mu \bar{A}R + \mu \bar{S}R + \mu \bar{M}R + \mu \bar{E}A\bar{R}$$

This model explains the probability of income refusal as a function of variation in respondent education, age, sex, marital status and the interaction between education and age. As shown in the last column of Table 2, the model explains nearly 95 percent of the variance in income refusal probabilities. To further investigate model fit, a predicted probability for each cell in Table 1 was computed and compared with the cell's observed probability. Results from this comparison are presented in Table 3. Overall model M_6 fit the data relatively well; the average absolute difference was .013. The relative importance of the independent variables in predicting refusal can be assessed both by the column "Gain in G^2 " in Table 2 and by the estimated standardized effect coefficients in Table 4. The larger the absolute value of the coefficient the greater the magnitude of the effect.

Who Refused to Answer?

Table 5 presents the observed probabilities of refusal associated with the significant effects in model M_6 . In general:

- . respondents possessing lesser amounts of education were more likely to refuse than respondents possessing higher amounts of education;
- . older respondents were more likely to refuse than younger respondents;
- . females were more likely to refuse than males; and
- . married respondents were more likely to refuse than unmarried respondents.

Moreover, the interaction between age and education produced a significant relationship with refusal. Finally, the least likely respondent to refuse to answer the household income question was a young, highly educated, unmarried male. The most likely respondent to refuse to answer was an older, less educated, married female.

Hence, refusal to answer the income question in a telephone survey was found not to be randomly distributed across respondents. Rather it was significantly related to the demographic characteristics investigated.

TABLE 3
Observed Probabilities, Predicted Probabilities,
and Residuals for Model M_6

Cell	Observed Probability	Predicted Probability	Residual
Low education, low age, married male	.087	.106	-.019
" " high age, " "	.137	.155	-.108
High education, low age, " "	.076	.071	.005
" " , high age, " "	.153	.145	.008
Low education, low age, unmarried male	.130	.092	.038
" " , high age, " "	.111	.134	-.023
High education, low age, " "	.055	.060	-.005
" " , high age, " "	.120	.126	-.006
Low education, low age, married female	.135	.145	-.010
" " , high age, " "	.213	.206	.007
High education, low age, " "	.108	.097	.011
" " , high age, " "	.181	.194	-.013
Low education, low age, unmarried female	.129	.125	.004
" " , high age, " "	.196	.180	.016
High education, low age, " "	.074	.084	-.010
" " , high age, " "	.183	.169	.014
Mean absolute residual			.013

TABLE 4
Estimates of Model M₆ Effects

Predictor	Effect
$\mu \bar{R}$ (constant)	-1.952
$\mu \bar{ER}$ (education)	-.130
$\mu \bar{AR}$ (age)	.308
$\mu \bar{SR}$ (sex)	-.174
$\mu \bar{MR}$ (marital status)	.084
$\mu \bar{EAR}$ (age x education)	.094

TABLE 5
Observed Probability of Refusal
Associated With Each Effect

Effect	Refusal Probability
Education	
Low	.154
High	.108
Age	
Low	.089
High	.174
Sex	
Male	.097
Female	.141
Marital Status	
Married	.174
Not Married	.099
Age x Education	
Low age, low education	.124
Low age, high education	.078
High age, low education	.183
High age, high education	.168

Why Did They Refuse?

Ten percent of the study individuals refusing to answer the income question were reinterviewed by telephone and questioned as to the reason for their refusal. This resulted in 67 completed reinterviews. By far the most frequent refusal reason given related to "invasion of privacy": 76 percent of the refusals were attributable to perceived privacy invasion. While study individuals giving this answer were found in all demographic categories, nonmarried females were more likely than other study individuals to respond in this manner.

Twenty percent of the study individuals reinterviewed

stated they did not know their annual household incomes. These individuals tended to possess the characteristics of low education, high age, and be married females. Further exploration revealed that the married females were only provided a weekly allowance by their husbands; they were simply not informed as to their husbands' earnings.

Finally, 4 percent of the study individuals reinterviewed said they did not have time to calculate their household income since it would be necessary to take into account a spouse's income as well as "other" income sources. For them responding required too much effort.

In brief, reasons for refusing to answer the income question differed across respondents and also (albeit to a lesser degree) related to demographic characteristics. Thus the results lend support to those of Skelton (1963) with respect to why study individuals refuse to answer an income question.

Conclusion

The present research investigated the relationship between refusal to answer a household income question in a telephone survey and four demographic variables. Using the logit model approach of Goodman a significant relationship was found between each of the four independent variables and refusal to answer. Moreover, reasons given for not answering an income question also differed across the individuals studied.

The implications of these research findings are twofold. First, the mere existence of a relatively large refusal rate for the income question increases the standard error of that variable and impedes investigation of relationships between income and other variables. Second, because nonresponse was not randomly distributed, any estimate of population income from the sample will be biased unless corrections are made. Moreover, any calculated relationships involving income are likely to be distorted due to the differential refusal rates.

Still, while this paper documented income question nonresponse, more research needs to be conducted on item nonresponse in general. Ultimately strategies for effectively reducing as well as handling item response need to be developed and implemented.

References

- Bailar, Barbara and Lanphier, Michael C. (1978), Development of Survey Methods to Assess Survey Practices. Washington, D.C.: American Statistical Association.
- Craig, C. Samuel and McCann, John (1978), "Item Nonresponse in Mail Surveys: Extent and Correlates," Journal of Marketing Research, 15, 285-289.
- Ferber, Robert (1966), "Item Nonresponse in a Consumer Survey," Public Opinion Quarterly, 30, 399-415.
- Goodman, Leo (1972), "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables," American Sociological Review, 37, 28-46.
- Herriott, R. (1977), "Collecting Income Data on Sample Surveys: Evidence From Split-Panel Studies," Journal of Marketing Research, 14, 322-329.
- Locander, William and Burton, John (1976), "The Effect of Question Form on Gathering Income Data by Telephone," Journal of Marketing Research, 13, 189-192.
- Skelton, Vincent C. (1963) "Patterns Behind 'Income

Refusals'," Journal of Marketing, 27, 38-41.

Wiseman, Frederick and McDonald, Philip (1979), "Noncontact and Refusal Rates in Consumer Telephone Surveys," Journal of Marketing Research, 16, 478-484.

THE EFFECTS OF SALUTATION, MONETARY INCENTIVE, AND DEGREE OF URBANIZATION
ON MAIL QUESTIONNAIRE RESPONSE RATE, SPEED, AND QUALITY

Linda L. Golden, University of Texas at Austin
W. Thomas Anderson, University of Texas at Austin
Louis K. Sharpe, Sharpe's Department Store, Checotah, Oklahoma

Abstract

The effects of three cover letter personalization treatments, a cash drawing monetary incentive, and rural, suburban, and urban place of residence upon response rate, speed, and quality for a mail questionnaire were investigated. The results indicated that cover letter personalization treatments did not have significantly different effects on response rate, speed, or quality. The probabilistic monetary incentive lowered response rate, but degree of urbanization moderated the effects on response rate, speed, and quality. Rural, suburban, and urban respondents may be differentially misunderstood by mail survey researchers due to potentially differing degrees of response and nonresponse biases.

Introduction

One of the most commonly used and intensively researched data gathering techniques in the behavioral sciences is the mail questionnaire. However, as two of the most recent review articles point out (Kanuk and Berenson 1975, p. 451; Linsky 1975, p. 100), even with the large volume of research on mail questionnaires, very few systematic conclusions can be drawn about the effectiveness of the various mail survey techniques. This is partially due to inconsistent results across studies, a lack of replication and research on any one technique, a near absence of research designed specifically to compare mail survey techniques in a variety of purposively selected populations, and the lack of a systematic body of knowledge or conceptual framework developed from general theory (Kanuk and Berenson 1975, Linsky 1975).

While the present research does not constitute a replication, nor do the authors pretend to develop a theoretical framework for mail survey research, this investigation does provide additional information on the effectiveness of two mail survey techniques: the personalized cover letter and the cash drawing as a monetary incentive. Moreover, this research compares the effects for respondents from three purposively selected populations: urban, suburban, and rural.

Personalization of Cover Letters

While one review article (Linsky 1975, p. 92) reports 16 studies dealing with personalization of cover letters, a review of these articles and others reveals that the method of personalization varies widely across studies. Personalization treatments have ranged from individually typed letters versus form letters (Simon 1967) to handwritten postscripts forging reply versus no postscript (Frazier and Bird 1958). Similar to figures reported by Linsky (1975, p. 92), of the 20 studies reviewed, 11 reported higher response rates for personalized over impersonal cover letters (Carpenter 1974, Cox, Anderson, and Fulcher 1974, Frazier and Bird 1958, Linsky 1965, Longworth 1953, Moore 1941, Roehrer 1963, Simon 1967), five reported higher response rates for impersonal cover letters (Andreason 1970, Houston and Jefferson 1975, Simon 1967, Watson 1965, Weilbacher and Walsh 1952), and four studies reported no significant differences between personalized and impersonal cover letters (Clausen and Ford 1947, Kawash and Aleamoni 1971, Kimball 1961).

Thus, results for the various forms of cover letter personalization are mixed. Further, the results are also mixed for any particular form of personalization, although almost half as many different forms of personalization were tested as there are number of studies.

The present research focuses on personalization via the inside address and cover letter salutation. Five studies were found which investigated the effects of personalization via either the salutation or the inside address and salutation on response rate. One reported significant differences between personalized and impersonal cover letters favoring the personalized form (Cox, Anderson, and Fulcher 1974), one reported significant differences favoring the impersonal form (Andreason 1970), and three reported no significant differences between personalized and impersonal cover letters (Clausen and Ford 1947, Kimball 1961, and Weilbacher and Walsh 1952). While the results across studies generally indicate no significant effect of personalization, the inconsistent results imply that the research context and topic may be an important factor in the effects of personalization.

Monetary Incentives

Inducements offered to the respondent in an attempt to increase response rate have frequently been in the form of monetary incentives. While studies investigating the effectiveness of premiums (nonmonetary rewards such as a ballpoint pen or trading stamps) have unanimously reported that offering a premium produces a higher response rate than offering no premium at all (Linsky 1975, p. 99), one study found that inclusion of money significantly increased response rate over inclusion of a premium (Goodstadt, Chung, Kronitz, and Cook 1977). "In general, money seems to be the most effective and least biasing incentive, the easiest to obtain and mail, and the most useful to all recipients" (Kanuk and Berenson 1975, p. 447).

There are two broad categories of studies investigating response rates produced by monetary rewards: those investigating a prepaid monetary incentive and those investigating a promised monetary incentive. Of the 20 studies reviewed researching the effects of prepaid monetary incentives, all found that offering a prepaid incentive produced a higher response rate than offering no monetary incentive at all (Armstrong and Overton 1971, Bevis 1948, Blumberg, Buller, and Hare 1974, Erdos 1970, Goodstadt, Chung, Kronitz, and Cook 1977, Hackler and Bourgette 1973, Huck and Gleason 1974, Hancock 1940, Kephart and Bressler 1958, Kimball 1961, Maloney 1954, Newman 1962, Pressley and Tullar 1977, Shuttleworth 1931, Watson 1965, Wiseman 1973, Wotruba 1966). Not all of these studies employed significance tests or reported significant differences; however, it may be concluded unequivocally that a prepaid monetary incentive produces a higher percentage response than offering no monetary incentive. And, as one review article reports, the larger the prepaid monetary incentive, the greater the increase in response rate (Armstrong 1973, p. 111).

Of the five studies investigating the effects of a promised monetary incentive on response rate, three compared

promised, prepaid, and no incentive treatments (Blumberg, Fuller, and Hare 1974; Hancock 1940; Wotruba 1966). All three found small and insignificant differences between promised incentives and a no incentive control, but, as reported above, much larger differences between prepaid incentives and the control. In addition, while promised incentives increased the response rate only very slightly, the prepaid treatment produced a significantly higher response rate than did the promised incentive treatment. One study by Gelb (1975), which did not include a control group, found the prepaid monetary incentives significantly increased response rate over promised monetary rewards for middle-class subjects, while promised rewards produced a higher response rate than prepaid incentives for lower-class subjects. Schewe and Cournoyer (1976) found that against a control of no monetary incentive, the size of the promised monetary reward influenced its effectiveness. Thus, while promised incentives appear to be generally less effective than prepaid monetary rewards, the amount of the promised incentive (Schewe and Cournoyer 1976) and the population sampled (Gelb 1975) may alter or even reverse the general ineffectiveness of the promised monetary incentive.

Research Focus and Purpose

While response rate differences have been examined for a variety of populations, no studies were found in the literature comparing the possible differential responsiveness of rural, suburban, and urban residents to personalized cover letters and monetary incentives. Moreover, no study was identified which compared rural, suburban, and urban mail survey response rates, or any other measure of mail questionnaire response for these populations. These residential groups would be expected to have distinct socioeconomic and demographic profiles; such respondent differences have been shown to correlate with measures of response and nonresponse bias (see Kanuk and Berenson 1975).

The general purpose of this study is to investigate the effects of a cash drawing as a monetary incentive, cover letter personalization and degree of urbanization (area of residence) upon response rate, quality, and speed. Response speed and quality, or degree of questionnaire completion, not only provide additional insight into the response patterns related to the treatments and residential conditions but are also frequently-used measures of response bias. The cash drawing incentive and cover letter personalization treatments are expected to produce higher response rates, speed, and quality than no incentives or no personalization. And, it is expected that urban and suburban residents will have higher response rates, speed, and quality than rural residents. The research also investigates psychographic differences among respondents receiving the monetary incentive who entered and who did not enter the cash drawing, and psychographic differences associated with response speed and quality. It is expected that external locus of control and future-oriented respondents will be more likely to participate in the cash drawing and will have higher response, speed, and quality.

Method

The questionnaire was a three-page inquiry into shopping patterns for basic types of products: groceries, clothing, and furniture and appliances. Since the respondents were likely to be of varied educational backgrounds, the questionnaire was extensively pretested in the field for clarity.

Standard demographic information was obtained and four psychographic measures were included in the final sections of the questionnaire. The psychographic variables measured were: James' Internal-External Locus of Control

(Rotter 1966), Life Satisfaction (Robinson 1977), Attitude Toward the Past, Present and Future (Rokeach 1956), and Traditional Family Ideology (Levenson and Huffman 1955). Life satisfaction and traditional family ideology measures were included only for descriptive comparison among rural, suburban, and urban residents; however, it was expected that locus of control and attitude toward past, present, and future would influence responsiveness to the monetary incentive offered.

The questionnaire was mailed to a sample of 4,500 potential respondents in Eastern Oklahoma, randomly selected from current telephone directories of five trading areas: (Tulsa (n = 1,500), suburbs of Tulsa (n = 1,500), Muskogee (n = 500), Chocotah (n = 500), and Stigler (n = 500)). The last three communities represent the most rural communities sampled. The cover letter, each individually hand-signed, requested that the questionnaire be completed by the person who does most of the shopping in the household.

The two experimental treatments tested in this study, monetary incentive and salutation, were administered via the cover letter. Three thousand of the 4,500 potential respondents received a cover letter which described a drawing for cash prizes totaling \$75.00. One \$25.00 prize, five \$5.00 prizes, and 25 \$1.00 prizes were offered. The respondent was to return a cash drawing card in the postage-paid return envelope provided for the questionnaire if s/he wished to enter the drawing. The cash drawing card was used as the method for entry to allow the respondent to return the questionnaire anonymously without entering the contest. The dollar values of the prizes offered were chosen in order to maximize the respondents' probability of winning a potentially higher dollar-value return. Fifteen hundred potential respondents received a cover letter that made no mention of a monetary incentive but which, in all other aspects, paralleled the monetary incentive cover letter.

The cover letters also contained one of three different levels of salutation. Fifteen hundred potential respondents received a standardized "Dear Shopper" salutation with no inside address. Another 1,500 potential respondents were sent a cover letter with a formally typed complete inside address and "Dear Mr. and Ms. _____" salutation including the respondent's last name. The remaining 1,500 respondents received a handwritten "Dear Mr. or Ms. _____" salutation personalized with the respondent's last name, but with no inside address. The salutation treatments were designed to represent impersonal, formal, and personal salutations, respectively, although the envelopes were all typed and personalized with Mr. or Ms. _____. Respondents were randomly assigned to both the monetary incentive and salutation treatments; assignment was controlled so that each area of residence received the same proportion of each treatment.

Responses were evaluated on three criteria: response rate, speed, and quality. As the measure of success typically used by researchers, response rate was the percentage of the total mailing which was returned, regardless of degree of questionnaire completion. Response speed was measured continuously by the postmark on the respondents' return envelope; the earliest postmark was assigned a value of one. Questionnaires were mailed simultaneously from central Texas, and returns began within a week of the mailing; the range for the response speed variables was from one to 43 days. Response quality was evaluated according to the degree of completion of the questionnaire. Since the questionnaire contained four sections, response quality ranged from one to five: one being a fully-completed questionnaire, four representing only one part completed, and

five a questionnaire with only scattered responses and no sections complete.

Analysis and Results

A total of 833 questionnaires were returned, resulting in a response rate of 18.5 percent. Due to financial constraints, and in order to reach a broader segment of the population, the researchers opted to send questionnaires to a larger sample frame rather than reduce the sample frame size and use a follow-up. The response rate is reflective of what would be expected under no follow-up conditions.

Response Rate

Chi squares were computed for response and nonresponse frequencies for monetary incentive, salutation, and urbanization treatments separately and are presented in Table 1. Surprisingly, the nonmonetary level (control) produced a significantly higher response rate than did the monetary incentive. Slightly over 20 percent of the 1,500 potential respondents who were mailed the nonmonetary cover letter returned the survey, whereas 17.7 percent of the 3,000 potential respondents mailed the monetary incentive returned the survey. Thus, the cash drawing was significantly less effective than offering no monetary incentive.

There was no significant effect for the salutation treatment; all three levels yielded response rates approximating the expected 18.5 percent. However, response rate did vary significantly by urbanization. Urban residents had the highest response rate (21.3 percent) and were the only group whose frequency of response was higher than the expected rate (16.5 percent), while the rural residents had the lowest response rate (15.2 percent).

To investigate psychographic differences among participants and nonparticipants in the cash drawing, the data were submitted to one-way analysis of variance (Nie et al 1975, p. 422) for each psychographic variable separately. As indicated in Table 2 and as expected, locus of control and attitude toward past, present, and future were the only two psychographic variables on which participants and nonparticipants in the cash drawing differed. Respondents returning the cash drawing card were significantly more externally-oriented than respondents who did not return the cash drawing card. In addition, participants in the cash drawing were significantly more future-oriented than nonparticipants.

To investigate differential responsiveness to the personalization and monetary treatments among rural, suburban, and urban residents, chi squares were computed for response and nonresponse frequencies on each treatment separately. Of the six chi squares computed, only one was significant within an alpha level of five percent. The suburban response rate to the monetary incentive was

TABLE 1
Effects of Incentive, Salutation, and Urbanization on Response Rate

Variable and Level	Actual		Expected		X ²	n	d.f.
	n	%	n	%			
Incentive ^c					4.14 ^b	833	1
Monetary	530	17.7	555	18.5			
Control	303	20.2	278	18.5			
Salutation ^d					.18	833	2
Formal Typed	283	18.9	277	18.5			
Handwritten	277	18.5	278	18.5			
Dear Shopper	273	18.2	278	18.5			
Urbanization ^d					20.96 ^a	795	2
Rural	228	15.2	265	17.7			
Suburban	248	16.5	265	17.7			
Urban	319	21.3	265	17.7			

^a $\alpha < .01$

^b $\alpha < .05$

^c 3,000 monetary treatments and 1,500 nonmonetary treatments were mailed.

^d 1,500 of each treatment level were mailed.

12.6 percent, compared to a 16.6 percent response rate for the control group (chi square of 4.22 with one degree of freedom). Thus, suburbanites responded less frequently than expected (13.9 percent) to the monetary incentive and more frequently than expected (14.0 percent) when there was no monetary incentive. While the suburban response rate was significantly decreased by the monetary incentive, rural and urban response rates did not differ between the monetary incentive and the control. Urbanization groups did not respond significantly differently to the cover letter personalization treatment levels.

Response Speed and Quality

In order to investigate the effects of the monetary incentive and salutation treatments and the urbanization condition upon response speed, the data were submitted to three-way analysis of variance (Nie et al 1975, p. 410). Table 3 shows the results of these analyses.

Neither treatment produced a significant main effect, and there were no significant interaction effects. There was, however, a significant main effect of urbanization. Contrary to expectations, rural respondents returned the questionnaire more quickly than either the suburban or

TABLE 2
Psychographics for Respondents Returning and Not Returning the Cash Drawing Card

Variable and Level	Mean Squares	d.f.	F	Mean Scores	
				Return	No Return
Locus of Control ^c	278.35	1	3.94 ^b	26.86	25.17
Error	70.60	504			
Family Ideology ^d	89.00	1	.94	33.20	32.25
Error	94.99	505			
Attitude Toward Past, Present, and Future ^e	29.18	1	6.04 ^a	10.21	10.78
Error	4.84	488			
Life Satisfaction ^f	1.57	1	1.23	3.91	4.04
Error	1.27	505			

^a $\alpha \leq .01$

^b $\alpha \leq .05$

^c The higher the score, the more external; range: 11-52; n return = 373, n no return = 133.

^d The higher the score, the more traditional the orientation; range: 12-57; n return = 374, n no return = 133.

^e The higher the score, the more past oriented; range: 5-15; n return = 361, n no return = 129.

^f The higher the score, the more satisfied; range: 1-5; n return = 372, n no return = 135.

TABLE 3
Effects of Incentive, Salutation, and Urbanization on Response Speed^a

Variable and Level	Mean Squares	d.f.	F	n	Mean Response Speed
Incentive	.077	1	.00		
Monetary				530	8.02
Control				303	7.97
Salutation	48.09	2	.96		
Formal Typed				283	8.30
Handwritten				277	7.54
Dear Shopper				273	8.26
Urbanization	369.54	2	7.39 ^b		
Rural				228	6.98 ^c
Suburban				248	7.53 ^d
Urban				319	9.24 ^{c,d}
Explained	71.96	17	1.44		
Error	50.04	774			

^a Measured from the earliest postmark as one and continuously; range: 1-43.

^b $\alpha \leq .01$

^{c,d} Scheffé test: $\alpha \leq .05$

the urban respondents, respectively. Scheffé tests (Scheffé 1959, p. 477) indicated that both rural and suburban respondents responded significantly more quickly than did urban respondents, although there was no significant difference in mean response speed between rural and suburban respondents. Thus, while urban respondents had the highest response rate, they were the slowest to

respond. Rural respondents showed just the reverse pattern: the lowest response rate but the quickest response speed.

Table 4 presents the results of the three-way analysis of variance (Nie et al 1975, p. 410) for the effects of the monetary incentive and salutation treatments and the

TABLE 4
Effects of Incentive, Salutation, and Urbanization on Response Quality^a

Variable and Level	Mean Squares	d.f.	F	n	Mean Quality
Incentive	.76	1	1.63		
Monetary				527	1.21
Control				301	1.26
Salutation	.27	2	.57		
Formal Typed				282	1.23
Handwritten				274	1.27
Dear Shopper				272	1.18
Urbanization	4.05	2	8.66 ^b		
Rural				228	1.32 ^{d,e}
Suburban				247	1.11 ^d
Urban				317	1.09 ^e
Explained	.85	17	1.81 ^c		
Error	.47	774			

^a Measured from one to five; 1 = all sections completed, 5 = scattered responses with no sections completed.

^b $\alpha \leq .01$

^c $\alpha \leq .05$

^{d,e} Scheffé test: $\alpha \leq .05$

TABLE 5
Relationships Between Psychographics and Response Speed and Quality

Psychographic	Response Speed		Response Quality	
	r	n	r	n
Locus of Control	.07 ^b	833	-.20 ^a	828
Family Ideology	-.06 ^b	833	-.16 ^a	828
Attitude Toward Past, Present, and Future	.04	833	-.33 ^a	828
Life Satisfaction	-.04	833	.03	790

^a $\alpha \leq .01$

^b $\alpha \leq .05$

urbanization condition on response quality. Again, there were no significant main effects for monetary incentive or salutation treatments and no significant two- or three-way interactions. Urbanization was the only variable which produced a significant main effect. The urban respondents had the highest response quality, followed by the suburban and rural respondents, respectively. Scheffé tests indicated that the rural respondents had a significantly lower quality of response than both the suburban and the urban respondents. There was no significant difference between urban and suburban response quality. Thus, while urban respondents were relatively slower respondents, they were relatively thorough in their questionnaire completion. Rural respondents, on the other hand, were relatively more quick to respond, but questionnaire completion was lower. The suburban respondents continually fell in the middle, aligning themselves with the rural respondents in terms of speed and the urban respondents in terms of quality. Suburban response rate also tended toward the lower response rate of the rural residents.

The data for the psychographic variables and response speed and quality were submitted to correlational analyses (Nie et al 1975, p. 280). Table 5 presents the results of these analyses. Response speed was significantly correlated with locus of control and family ideology, although the Pearson correlations were not high. Response speed was significantly correlated with three of the four psychographic variables: locus of control, family ideology, and attitude toward the past, present, and future. Life satisfaction was not significantly related to either response speed or quality. The correlations of the significant psychographics with response quality were much higher than the correlations with response speed.

The more external the respondent's locus of control, the slower to respond and the lower the response quality. Respondents with a more internal locus of control responded more quickly and had a generally higher degree of questionnaire completion than did the respondents who were relatively more external. Further, the more

traditional the family ideology, the faster the response speed and the lower the quality of response. The respondents with the least traditional family ideologies responded more slowly but with a higher completion rate than did the respondents with a relatively more traditional family ideology. For response quality, the more past-oriented the respondent, the lower the quality. Thus, the more future-oriented respondents provided the most complete questionnaires; however, this psychographic was not significantly associated with response speed.

Conclusions and Implications

This study investigated the effects of cover letter personalization, a promised and uncertain monetary incentive, and degree of urbanization upon response rate, speed, and quality for a mail questionnaire. The results suggest that the additional costs of cover letter personalization are not warranted if the researcher's objective is to increase response rate, speed, or quality. While the conclusions drawn from this study must be restricted to the sample under investigation and the specific treatments used, these results are consistent with those obtained by three of the five studies using the salutation as the method of personalization (Clausen and Ford 1947, Kimball 1961, Weilbacher and Walsh 1952). In addition, this study indicates that rural, suburban, or urban residence does not influence the effects of personalization.

A cash drawing monetary incentive appears to decrease response rate in contrast with no monetary incentive. There may be a raffle effect which results from this type of monetary incentive, whereby potential respondents may regard the questionnaire as just another piece of "junk mail," decreasing the researcher's credibility. However, there were no differences between the monetary treatment and the control for response speed and quality. Thus, according to this study, the uncertain promised monetary incentive has no redeeming virtues.

Rural and urban respondents do not appear to be differentially responsive to uncertain monetary incentive treatments, but suburban respondents may be, as they returned the control treatment questionnaires at a significantly higher rate than the monetary incentive questionnaires. Thus, while rural and urban respondents did not tend to react negatively to the cash drawing, at least as evidenced by their response rate, suburban respondents evidenced the strongest "raffle effect reaction."

Respondents who received a "raffle" monetary incentive and returned the cash drawing card appear to be different from those who do not return the cash drawing card. Respondents returning the cash drawing card were significantly more external in their locus of control and significantly more future-oriented than nonparticipants in the cash drawing. Externally- and future-oriented persons may be more "raffle prone" and more likely to take a "gamble" than persons with other orientations, thus making them the most likely to be responsive to the type of incentive investigated here. If the researcher has prior knowledge of a sample's psychographic characteristics, an uncertain monetary incentive may be able to reduce the costs of a monetary "reward" and possibly increase response rate.

Urbanization appears to have an important and significant influence upon response rate, speed, and quality. Urban residents had the highest response rate, were slowest to respond but provided the highest quality responses. Rural residents were a mirror image of the urban residents, providing the lowest response rate, the quickest response, and the lowest response quality. Consistent with their mid-range urbanism, suburban respondents occupied the middle. Suburban respondents aligned themselves with rural respondents on response speed and with urban respondents on response quality.

Response rate and quality differences among rural, urban, and suburban residents may be partially explained by demographic differences, specifically education.¹ With the highest educational level, urban residents may be more experienced communicators resulting in a higher response rate and response quality than for either suburbanites or ruralites. In addition, the ruralites, with the lowest educational level, would be expected to be the least responsive to mail surveys both in terms of response rate and quality. Such was the case.

Response speed and quality provide some response bias information for the three residential groups. The late respondents to a questionnaire sampling the three urbanism groups were more likely to be urban; thus, the nonrespondents may tend to be more like the urban population sampled although the urban residents do respond at a higher rate. Because of their low quality of response, however, the researcher is likely to obtain the least information from rural samples. Coupled with their low response rate, the low quality of responses make these respondents the least understood, or most misunderstood, of the three groups. They are, in fact, the least represented in this study both in terms of response and nonresponse bias.

Response speed and quality tend to be associated with certain psychographic characteristics. Respondents who were the slowest to respond were more external in their locus of control and less traditional in their family ideology. In addition, respondents who provided the lowest response quality in terms of survey completion were more externally-oriented and more traditional in family ideology, as well as being more past-oriented than respondents with higher quality responses. Thus, without taking place of residence into consideration, the nonrespondents (to the extent that they are similar to late respondents) are likely to be more external in their locus of control and less traditional than respondents. However, the measure of response bias indicates that the least complete questionnaires are likely to come from more externally-oriented respondents and those with a more traditional family ideology.

The resounding implication of this study is the importance of the population sampled and the subsequent effect upon mail questionnaire response rate and potential bias in terms of response speed and quality. Researchers sampling across two or more of these populations are advised that they may have differing response patterns which should be taken into consideration in adjusting for any potential biases. Rural populations may be the most misunderstood, while urban populations may be the most represented and the least misunderstood. Wave mailing of questionnaires should also consider the unique characteristics of these populations when determining the timing of questionnaire waves. Rural, suburban, and urban samples cannot be considered to be drawn from the same populations. Although potentially the least costly, mail questionnaires may be the least effective means of sampling rural populations.

References

- Andreasen, Alan R. (1970), "Personalizing Mail Questionnaire Correspondence," *Public Opinion Quarterly*, 34, 273-7.
- Armstrong, J. Scott (1975), "Monetary Incentives in Mail Surveys," *Public Opinion Quarterly*, 39, 111-16.

¹ Although not formally reported in this paper, Chi Square analysis for 11 standard demographics resulted in significance at the .05 level for nine demographics. For educational level of chief wage earner (Chi Square = 75.07), urbanites were the most formally educated, ruralites the least, and suburbanites in between.

- Armstrong, J. Scott, and Overton, Terry S. (1971), "Brief vs. Comprehensive Descriptions in Measuring Intentions to Purchase," Journal of Marketing Research, 8, 114-17.
- Bevis, Joseph C. (1948), "Economic Incentives Used for Mail Questionnaires," Public Opinion Quarterly, 12, 492-3.
- Blumberg, Herbert H., Fuller, Carolyn and Hare, Paul (1974), "Response Rates in Postal Surveys," Public Opinion Quarterly, 38, 113-23.
- Carpenter, E. W. (1974), "Personalization in Mail Surveys of the General Public: A Reassessment in Light of Recent Innovations," Public Opinion Quarterly, 38, 116-20.
- Clausen, John A., and Ford, Robert N. (1947), "Controlling Bias in Mail Questionnaires," Journal of the American Statistical Association, 42, 497-511.
- Cox, Eli P. III, Anderson, Jr., W. Thomas and Fulcher, David G. (1974), "Reappraising Mail Survey Response Rates," Journal of Marketing Research, 11, 413-7.
- Erdos, Paul L. (1970). Professional Mail Surveys, New York: McGraw-Hill.
- Frazier, George, and Bird, Kermit (1958), "Increasing the Response of a Mail Questionnaire," Journal of Marketing, 22, 186-7.
- Gelb, Betsy D. (1975), "Incentives to Increase Survey Returns: Social Class Considerations," Journal of Marketing Research, 11, 107-9.
- Goodstadt, Michael S., Chung, Linda, Kronitz, Reena and Cook, Gaynoll (1977), "Mail Survey Response Rates: Their Manipulation and Impact," Journal of Marketing Research, 14, 391-5.
- Hackler, James C., and Bourgette, Patricia (1973), "Dollars, Dissonance, and Survey Returns," Public Opinion Quarterly, 37, 266-81.
- Hancock, John (1940), "An Experimental Study of Four Methods of Measuring Unit Costs on Obtaining Attitude Toward the Retail Store," Journal of Applied Psychology, 24, 213-30.
- Houston, Michael, and Jefferson, Robert (1975), "The Negative Effects of Personalization on Response Patterns in Mail Survey," Journal of Marketing Research, 7, 114-7.
- Huck, Schuyler W., and Gleason, Edwin M. (1974), "Using Monetary Inducements to Increase Response Rates from Mail Surveys," Journal of Applied Psychology, 59, 222-5.
- Kanuk, Leslie, and Berenson, Conrad (1975), "Mail Surveys and Response Rates: A Literature Review," Journal of Marketing Research, 12- 440-53.
- Kawash, Mary B., and Aleamoni, Lawrence M. (1971), "Effect of Personal Signature on the Initial Rate of Return of a Mailed Questionnaire," Journal of Applied Psychology, 55, 589-92.
- Kephart, William M., and Bressler, Marvin (1958), "Increasing the Response to Mail Questionnaires: A Research Study," Public Opinion Quarterly, 22, 123-32.
- Kimball, Andrew E. (1961), "Increasing the Rate of Return in Mail Surveys," Journal of Marketing, 25, 63-5.
- Levenson, Daniel, and Huffman, Phyllis (1955), "Traditional Family Ideology and Its Relation to Personality," Journal of Personality, 24, 251-73.
- Linsky, Arnold S. (1965), "A Factorial Experiment in Inducing Responses to a Mail Questionnaire," Sociology and Social Research, 49, 183-9.
- Linsky, Arnold S. (1975), "Stimulating Responses to Mail Questionnaires: A Review," Public Opinion Quarterly, 39, 82-101.
- Longworth, Donald S. (1953), "Use of a Mailed Questionnaire," American Sociology Review, 18, 310-13.
- Maloney, Paul W. (1954), "Comparability of Personal Attitude Scale Administration With Mail Administration With and Without Incentive," Journal of Applied Psychology, 38, 238-9.
- Newman, Sheldon W. (1962), "Differences Between Early and Late Respondents to a Mail Survey," Journal of Advertising Research, 2, 27-39.
- Nie, Norman H., Hull, C. Hadlai, Jenkins, Jean G., Steinbrenner, Karen and Bent, Dale H. (1975). Statistical Package for the Social Sciences, Second Edition, New York: McGraw-Hill.
- Pressley, Milton M., and Fullar, William L. (1977), "A Factor Interactive Investigation of Mail Survey Response Rates from a Commercial Population," Journal of Marketing Research, 14, 108-11.
- Robinson, John P. (1977). How Americans Use Time: A Social-Psychological Analysis of Everyday Behavior, New York: Praeger Press.
- Roehrer, Allan G. (1963), "Effective Techniques in Increasing Response to Mailed Questionnaires," Public Opinion Quarterly, 27, 299-302.
- Rokeach, Milton (1956), "Political and Religious Dogmatism: An Alternative to the Authoritarian Personality," Psychological Monographs, 70, No. 425.
- Rotter, J. B. (1966), "Generalized Expectancies for Internal Versus External Control of Reinforcement," Psychological Monographs, 80, No. 609.
- Scheffé, Henry (1959). Analysis of Variance, New York: John Wiley and Sons.
- Schewe, Charles D., and Cournoyer, Norman G. (1976), "Prepaid vs. Promised Monetary Incentives to Questionnaire Response: Further Evidence," Public Opinion Quarterly, 40, 105-7.
- Shuttleworth, Frank K. (1931), "A Study of Questionnaire Technique," Journal of Educational Psychology, 22, 652-8.
- Simon, Raymond (1967), "Response to Personal and Form Letters in Mail Surveys," Journal of Advertising Research, 7, 28-30.
- Watson, John J. (1965), "Improving the Response Rate in Mail Research," Journal of Advertising Research, 5, 48-50.
- Weilbacher, William, and Walsh, H. Robert (1952), "Mail Questionnaires and the Personalized Letter of Transmittal," Journal of Marketing, 16, 331-6.
- Wiseman, Frederick (1976), "A Reassessment of the Effects of Personalization on Response Patterns in Mail Surveys," Journal of Marketing Research, 13, 110-1.
- Wotruba, Thomas R. (1966), "Monetary Inducements and Mail Questionnaire Response," Journal of Marketing Research, 3, 398-400.

WHAT IF OPINION LEADERS DIDN'T KNOW MORE?
A QUESTION OF NOMOLOGICAL VALIDITY

Jacob Jacoby, Purdue University
Wayne D. Hoyer, Purdue University

Abstract

Given the numerous investigations involving opinion leadership, it is surprising to find that the construct validity of this important concept has yet to be firmly established. While a few studies present evidence for convergent and discriminant validity, no data appear to exist bearing on nomological validity. The present investigation addressed this issue by examining a predicted relationship between opinion leadership and expertise. Results show a strong positive correlation between these two concepts, thereby providing support for the concept of opinion leadership. While this study represents only a brief methodological note, had the hypothesis not been confirmed, the implications would have been substantial.

Introduction

The concept of opinion leadership has been the subject of numerous investigations in a variety of fields, including consumer behavior, social psychology, communications, marketing, and sociology. Most of these studies have attempted to isolate social activity, general attitudes, demographic, personality, and life style characteristics in order to gain a better understanding of just who opinion leaders are and how they influence others in society (Engel, Blackwell, & Kollat 1978).

A major problem with this body of research is that the concept of opinion leadership is accepted uncritically; not much empirical evidence exists to establish the construct validity of this important concept (Jacoby 1974). In order to establish construct validity, one needs to show the extent to which the measures of the construct "fit into a network of expected relations" (Nunnally 1978). This requires demonstrating that: (1) different measures of the same concept are highly related (convergent validity); (2) measures of the same concept are correlated more highly with each other than with measures of totally different concepts (discriminant validity); and (3) the measures of this concept bear some relationship to measures of other constructs which are hypothesized to be related to the concept in question (nomological validity; Campbell 1960).

Virtually all efforts at operationalizing opinion leadership can be reduced to three basic techniques. The self-designation method asks the respondent to judge whether or not he himself is an opinion leader. The key informant approach involves, first, the identification of individuals who are assumed to be knowledgeable of the social interactions within a particular group and, second, having these key informants indicate which members of the group are opinion leaders. Finally, the sociometric method involves interviewing all members of a particular group in order to determine the nature of the social interactions within the group and thus identify the opinion leaders.

The first step in construct validation is to establish the extent to which these measures are related (i.e., convergent validity). Some evidence for convergent validity already exists. Katz and Lazarsfeld (1955) found

a strong positive relationship between the self-designating and sociometric approaches. Rogers and Catarno (1962) described data from an unpublished doctoral dissertation which revealed a positive correlation across all three measures. The problem with these studies, however, is that evidence was only provided for convergent validity and no attempt was made to establish discriminant validity.

A subsequent study by Jacoby (1974) applied the Campbell and Fiske (1959) multi-method multi-trait approach to determine both the convergent and discriminant validity of opinion leadership as this was manifested by the women in four large campus sorority organizations. Results showed that "the four coefficients pertinent for assessing convergent validity were surprisingly high in view of the limits that reliability places on validity." Discriminant validity was adequately established for only two of the four groups tested in the study. In general, however, there would appear to be at least some evidence for both the convergent and discriminant validity of opinion leadership.

What still needs to be determined, however, is the extent to which measures of opinion leadership are related as predicted to measures of other conceptually specified constructs (i.e., nomological validity). For example, it seems to be generally assumed that opinion leaders actually possess more knowledge or expertise in the product category of interest than do members of the general population. That is, it seems logical that individuals who are sought out and considered useful sources of information regarding an issue or product category would, in fact, possess a somewhat higher level of knowledge regarding the issue or product category. Evidence for nomological validity would therefore be supplied if a measure of opinion leadership in a particular product category would correlate positively with a measure of actual knowledge regarding that product category. The purpose of the present study was to determine the extent of this relationship and to hopefully be able to provide some degree of evidence for nomological validity which had been previously lacking.

Consumer researchers (e.g., Abelson, et al. 1974, Valenzi and Andrews 1973) have usually measured "knowledge level" or expertise by asking respondents to provide a self estimate of the degree of their knowledge. Researchers are increasingly finding, however, that individuals are not accurate judges of their own knowledge level or ability. For example, De Nisi and Shaw (1977) found that self ratings of ability correlated very poorly with scores on an actual ability test and that self-ability ratings were unable to differentiate between those of high and low ability.

The present study avoided this problem by using a 30 item stereo knowledge test developed by Jacoby and Williams-Jones (1973). This instrument was pretested on 50 undergraduates. Results showed "that scores fell into 3 distinct groups, with approximately one tenth of the scores being clearly higher than the rest and another one fifth falling much lower than the majority of subjects whose scores formed the center of the distribution" (Williams-Jones 1974, p. 16). For purposes of validation, this

test was next administered to a sample of 12 stereo repairmen and technicians, all 12 of whom scored in the "expert" range. Given these results, it was felt that testing expertise via this instrument was preferable to asking respondents to provide a judgment regarding their level of expertise.

In summary, the present study attempted to examine the nomological validity of opinion leadership by testing the relationship between this construct and another construct believed to be logically related, namely expertise. The specific hypothesis being tested was: Opinion leadership in regard to stereo equipment is positively related to knowledge about the product category.

Method

Subjects - One hundred thirty one students (72 males and 59 females) enrolled in an introductory psychology course at a large midwestern university served as subjects in this study. Participation served as partial fulfillment of a course requirement.

Procedure and Measures - Each subject filled out a three part questionnaire. The first part consisted of Jacoby's (1972) seven item self-designating measure of opinion leadership. This index differs from the more traditional measures (e.g., King and Summers 1970) in having seven rather than six questions and more response options (for improved reliability) for each question. Apropos of our earlier comments regarding the frailty of verbal report instruments and our comments elsewhere regarding the necessity for multiple indicant research (Jacoby 1976, p. 7), it should be noted that the convergent validity between this index and separate sociometric and key informant indices was found to be quite high (cf. Jacoby 1974). This previous research provides the rationale for employing only a single indicant approach in the present investigation.

To provide additional perspective on the findings, the second section of the questionnaire consisted of five questions which asked subjects: whether they had ever owned any stereo equipment both in the past and presently, whether they had ever been in the process of looking for stereo equipment, the last time they purchased any equipment, and how much they felt they knew about stereos.

The final portion of the questionnaire contained the stereo knowledge test developed by Jacoby and Williams-Jones (1973). The test consisted of 16 multiple choice and 14 matching questions covering a variety of facts about stereo equipment (see Appendix A).

Results

The Pearson product-moment correlations between the measure of opinion leadership, stereo knowledge, and the other assessed variables are presented in Table 1.

Opinion Leadership and Expertise

The most significant finding is that opinion leadership scores correlate as one would expect with expertise ($r = .69, p < .001$). That is, the tendency to be an opinion leader for stereo equipment is highly related to one's knowledge about stereo equipment.

Opinion Leadership and Past Experience

Also as might be expected, a positive correlation ($r = .32, p < .01$) was found between past stereo ownership and opinion leadership. In other words, opinion leaders were

somewhat more likely to have owned stereo equipment in the past than were non-opinion leaders. Further, opinion leaders were more likely to currently be in the process of looking for stereo equipment than non-opinion leaders ($r = .50, p < .01$). There was either a low or no correlation between opinion leadership and either present ownership, number of years owned, years since most recent purchase, or whether the respondent believed he had greater knowledge at the time of the first purchase or now.

Expertise and Past Experience

A low positive correlation was found between expertise and past ownership ($r = .25, p < .05$). In addition, "experts" were somewhat more likely to be currently looking for stereo equipment than were "nonexperts" ($r = .33, p < .01$). Consistent with the findings for opinion leadership, the other dependent measures exhibited little or no correlation with expertise.

Sex Differences

As expected, expertise correlated negatively with sex of subject ($r = -.69, p < .001$). Females were far less likely to be "experts" for stereo equipment than were males.

Sex of subject was also correlated negatively with opinion leadership ($r = -.44, p < .01$). That is, females were also less likely to be opinion leaders with regard to stereo equipment.

In view of these findings, the main analysis was further subdivided according to sex. Table 2 presents the Pearson correlations for males and Table 3 for females. It can be seen from these Tables that, in the case of males, the correlation between opinion leadership and expertise was almost identical to the previous result ($r = .66, p < .001$). The female correlation was substantially lower ($r = .27, p < .05$). Thus, the relationship between opinion leadership and expertise for the stereo product class appears to be sex-specific.

Product Involvement

Finally, it is always possible that the link between opinion leadership and expertise is a reflection of a third variable. One logical possibility for such a third variable is involvement with the product category and several of the variables that were assessed (e.g., looking to purchase, ownership) appear to tap such a dimension. Table 4 presents partial correlations which control for these "involvement" variables. As shown in Table 4, the relationship between opinion leadership is not significantly lower when these variables are held constant. Thus, these data do not support the third variable hypothesis insofar as the two variables examined are concerned.

Discussion

As hypothesized, opinion leadership was found to be highly related to expertise. This finding is noteworthy for two reasons. First, it has long been assumed that opinion leaders were more knowledgeable in their area of leadership but this key assumption seems to have never been tested. The present findings provide the missing empirical support.

More importantly, this finding can be viewed as evidence for the nomological validity of opinion leadership. In other words, the measure of opinion leadership did in fact bear a predicted relationship to the conceptually related concept (expertise). If one also includes the findings that opinion leaders were more likely to have

TABLE 1

	OL	Sex	Ever Owned	Looking	Now Own	Yrs Owned	Yrs Buy	Know More
Expertise	.69***	-.69***	.25*	.33**	.00	.08	-.09	-.10
OL	-	-.44***	.32**	.50**	-.04	-.11	-.26*	-.22*
Sex		-	.06	-.25**	-.05	.05	.13	.02
Ever Owned			-	.21**	.65***			.12
Looking				-	-.37**	-.05	-.12	-.19
Now Own					-	.07	-.15	.11
Yrs Own						-	.49***	-.04
Yrs Buy							-	-.06

* = .05
 ** = .01
 *** = .001

Expertise = Score
 OL = Score
 Sex 1 = M; 2 = F
 Ever Owned
 Looking 1 = no; 2 = yes
 Now Own
 Yrs Owned Years
 Yrs Buy
 Know More 1 = Now; 2 = Then

TABLE 2

Males

	OL	Ever Owned	Look	Now Own	Yrs Owned	Yrs Buy	Know More
Expertise	.66***	.52***	.25*	.23	-.05	.08	-.21
OL	-	.45***	.52***	.05	.02	-.16	-.40**
Owned		-	.31**	.71***			.13
Looking			-	-.22	-.01	-.12	-.15
Now Own				-	-.09	-.31*	.05
Yrs Own					-	.32*	.05
Yrs Buy						-	.24

TABLE 3

Females

	OL	Ever Owned	Look	Now Own	Yrs Owned	Yrs Buy	Know More
Expertise	.27*	.21	.16	.20	-.08	-.05	-.04
OL	-	.32**	.26*	.18	-.33*	-.43**	.03
Ever Owned		-	.09	.55***			.11
Looking			-	-.62***	-.11	-.04	-.26
Now Own				-		.04	.18
Yrs Own					-	.66***	-.13
Yrs Buy						-	-.11

TABLE 4

Partial Correlations: OL Vs. Expertise

Controlling for:	r
Looking	.64***
Owned	.66***
Looking and Owned	.62***

owned stereo equipment in the past and that they were more likely to be presently looking for such equipment, even more confidence can be placed in the nomological validity of the concept. Since stereo equipment is a more male oriented area, the finding that opinion leadership and expertise are strongly related to sex can also be interpreted as support for nomological validity.

Several considerations, however, must be kept in mind when interpreting the results. First, this assessment was performed for only one product category. Research is needed to determine whether this relationship holds for other product categories and subject matters as well. More importantly, stereo equipment is an area which can be considered highly technical. As a result, the difference between expert and non-expert becomes much more distinct. There may be areas or product categories where this distinction is not clear and these areas may not produce as strong results. Relatedly, the product category is highly male oriented. Research is needed to determine if the same pattern exists in female dominated and more neutral or non-sex typed areas.

Another issue concerns the fact that technical knowledge may be only one of several factors which jointly determine expertise. Other expertise oriented factors which may be

equally or more important for the opinion leader to possess include such things as knowledge regarding where to purchase, how much to pay, suitability for various purposes, etc. These factors also warrant empirical attention.

In conclusion, the results of this study, when coupled with the results of the other studies mentioned earlier, seem to indicate a satisfactory degree of validity for the construct of opinion leadership. While it hardly comes as a surprise to find that those who rate themselves as a good source of advice regarding stereo knowledge also have a basis in knowledge for that advice, had this relationship not materialized, the implications would have been substantial. Thus, while only a brief methodological note, this study provides evidence for a critical, but heretofore empirically neglected assumption.

References

Abelson, H., Schroyer, D. and Gunzelman, S. (1974), Food and Nutrition -- Knowledge and Beliefs. Report prepared for the Division of Consumer Studies, Bureau of Foods, Food and Drug Administration. New Jersey, Response Analyses Corp.

Campbell, D. T. (1960), "Recommendations for Test Standards Regarding Construct, Trait, and Discriminant Validity." American Psychologist, 16, 546-553.

Campbell, D. T., & Fiske, D. W., (1959), "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix," Psychological Bulletin, 81-105.

De Nisi, A. S. & Shaw, J. B. (1977), "Investigation of the Uses of Self Reports of Abilities," Journal of Applied Psychology, 62(5), 641-644.

Engel, J. F., Blackwell, R. D., & Kollat, D. T. (1978), Consumer Behavior, Third Edition, Hinsdale, Illinois: The Dryden Press.

Jacoby, J. (1972) "Opinion Leadership and Innovativeness: Overlap and Validity." In M. Venkatesan (Ed.) Proceedings, Third Annual Conference, The Association for Consumer Research, 2, 632-649.

Jacoby, J. (1974), "The Construct Validity of Opinion Leadership," Public Opinion Quarterly, 26(1), 81-89.

Jacoby, J. (1976), "Consumer Research: Telling it Like it is." In B. B. Anderson (Ed.), Advances in Consumer Research, 3, 1-11.

Jacoby, J. & Williams-Jones, J. (1973), "Price Cue Utilization in Quality Judgments as a Function of Consumer Expertise." Unpublished manuscript.

Katz, E., & Lazarsfeld, P. F. (1955), Personal Influence: The Part Played by People in the Flow of Mass Communication, Illinois: The Free Press, 137-61.

King, C. W., & Summers, J. O. (1970), "Overlap of Opinion Leadership Across Consumer Product Categories." Journal of Marketing Research, 7, 43-50.

Nunnally, J. C. (1978), Psychometric Theory, Second Edition, New York: McGraw-Hill.

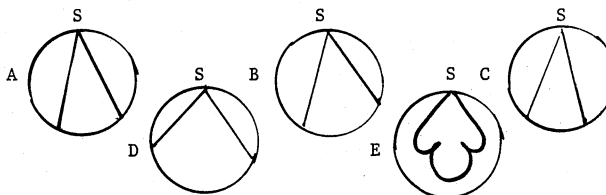
Rogers, E. M. & Catarno, D. G. (1962), "Methods of Measuring Opinion Leadership," Public Opinion Quarterly, 441.

Valenzi, E. R. & Eldridge, L. (1973), "Effect of Price Information, Composition Differences, Expected, and Rating Scales on Product Quality Ratings." Proceedings of the 81st Annual Convention of the American Psychological Association, 8, 820-830.

Williams-Jones, J. (1974), "Price Cue Utilization in Quality Judgements as a Function of Expertise: A Replication." Unpublished Master's thesis, Purdue University.

Appendix A: Stereo Knowledge Test*

- B 1. The best stylus (needle) material is:
a. sapphire b. diamond c. ceramic d. ruby
- D 2. The two most commonly used turntable speeds are:
a. 16 and 45 b. 33 and 45 c. 45 and 78
d. 33 and 78
- B 3. A receiver includes all but which of the following?
a. pre amp b. cartridge c. power amp d. tuner
- B 4. Stereo records should ideally be stored:
a. horizontally b. vertically c. near heat
d. without the abrasive dust cover
- C 5. Which of the tracking forces below would probably indicate the highest quality cartridge?
a. 3.0 grams b. 1.0 grams c. .5 grams
d. .75
- C 6. Distortion refers to:
a. unwanted random sound in electronic equipment often heard as a hiss
b. the extent to which a speaker distributes the acoustical power widely to the listening area
c. unwanted sound resulting from the inability of the stereo system to exactly reproduce the tonal qualities of the input
d. a wavering of pitch which is due to minute variations in turntable or tape speed
- D 7. A two-way speaker would not contain:
a. a cross-over network b. a woofer
c. a tweeter d. a mid-range
- C 8. Low efficiency speakers should ideally be paired with a:
a. lower powered amplifier
b. manual turntable
c. higher powered amplifier
d. automatic turntable with magnetic cartridge
- D 9. Which of the following would be the best speaker dispersion pattern? (S=location of the speaker)



*The correct answers are indicated in the response blanks.

- D 10. When turning the balance control to the right:
 a. it increases the right channel (speaker) volume
 b. it decreases the treble volume
 c. it increases the bass volume
 d. it decreases the left channel volume
- C 11. A preener is:
 a. an electrical network designed to eliminate selected frequencies
 b. a fine brush used to clean the cartridge
 c. a type of brush used to clean records
 d. a device which changes AC to DC
- C 12. Which of the following speakers would be most likely to contain a thin sheet of mylar as the diaphragm?
 a. bass reflex b. acoustic suspension
 c. electrostatic d. folded horn
 e. Dolby speaker
- C 13. If you are living in a major metropolitan area, which of the attributes of a tuner is least important to you?
 a. AM rejection b. selectivity c. sensitivity
 d. resistance to overload
- B 14. The purpose of the loudness (contour) control found on most receivers is to:
 a. increase or decrease volume to a greater extent
 b. mainly compensate for loss of bass at low volume settings
 c. mainly compensate for loss of treble at low volume settings
 d. increases the output of the mid-range speakers
- A 15. A system that reduces tape hiss is called a:
 a. Dolby b. tape monitor c. limiter
 d. rectifier e. Hirsch-Houk system
- D 16. There are three methods of rating power commonly used by manufacturers of high fidelity components. A typical amplifier rated under the three different methods would probably have which of the following?
 a. RMS rating 80 watts
 IHF rating 125 watts
 IHF + 1db rating 120 watts
 b. IHF rating 80 watts
 IHF + 1db rating 100 watts
 RMS rating 125 watts
 c. IHF + 1db rating 80 watts
 IHF rating 100 watts
 RMS rating 125 watts
 d. RMS rating 80 watts
 IHF rating 100 watts
 IHF + 1db rating 125 watts
- G 7. gain
J 8. AM tuning range
P 9. infinite baffle
B 10. adequate signal to noise ratio in a pre amp
C 11. bandwidth
Q 12. compliance
U 13. balance control
R 14. IPS
- E. 20 Hz to 20 kHz \pm 1 dB
 F. 88 mHz to 108 mHz
 G. amount of amplification
 H. distance between poles of a magnetic field
 I. speaker that reproduces bass frequencies
 J. 550 kHz to 1600 kHz
 K. 100 Hz to 10kHz \pm 3 dB
 L. reflection of sound (from walls, etc.)
 M. 3 dB
 N. a low frequency vibration mechanically transmitted to the tone arm and pick-up and thus superimposed in the reproduction
 O. wavering of pitch which is due to minute variations in turntable or tape speed
 P. air tight speaker enclosure containing a bass speaker with a very low open air resonance, plus a sealed mid-range and tweeter
 Q. how well a cartridge follows the record groove
 R. speed at which a tape moves
 S. a control which mechanically lifts the tonearm off the record
 T. a total confusion of frequency input
 U. governs relative volume from left-right speakers

Please match the following with the lettered phrases at the right.

- E 1. an adequate frequency response of a power amp
F 2. FM tuning range
O 3. WOW
I 4. woofer
D 5. image
S 6. cueing
- A. ability of a tuner to reject unwanted FM interference on same frequency
 B. 65 dB
 C. range of frequencies occupied by a signal
 D. any reception not occurring at the proper frequency

A DISCUSSION OF SURVEY RESEARCH

James H. Barnes, Jr., The University of Georgia

Introduction

The three papers presented in this session are loosely tied together since each deals with some aspect of survey type research. The differences among them, however, make each very difficult to compare and contrast. Consequently, after presenting some concerns about each paper, I shall attempt in a separate section of this discussion to offer some observations on the state of consumer survey research which I hope will be of some use to the audience of this section. No attempt is made to evaluate the net worth nor the relative contribution of the papers since each reader is probably a better judge of that given his or her particular frame of reference.

Peterson, Leone and Sabertehrani

The authors of this paper deal with the nagging problem of refusal of respondents to provide information relative to family income. This non-sampling error they point out is often encountered in telephone surveys. Three reasons for refusal emerged from the study. Foremost, respondents viewed the question as an invasion of privacy. Second was the fact that they lacked knowledge of their annual household income and finally, they felt that they did not have time to calculate their total income since it was made-up of several different components. The authors also point out that the refusal rate differed across respondents.

While the paper is interesting and generally well-written, the reader is given the impression that the authors are taking an analytical technique and looking for a set of data "crunch". This of course is not necessarily bad since researchers should have an interest in research methods per se. As a methods paper, it would be interesting to see how application of the logit analysis provides greater insight into the data than, for, example, simple cross-tabs of the data set. The use of the logistic function has been found to represent the data when the dependent variable is an indicator variable such as the probability used in the present case. The application of the logit approach presented in this paper is a standard application except perhaps for the conversion of the independent variables to be all 0-1 dummy variables. The authors fail to point out the problems of collapsing the independent variables into categories. While this approach is generally motivated around maintaining cell sizes, real analysis problems can and often do arise.

A second methodological concern is one of the choice of final model arrived at by the authors. The "saturated" model with which the authors begin is one which contains all independent variables as well as all interaction terms. From this model, the technique used reduces the model on the basis of loss in explanatory power. This however, presupposes that the variables education, age, sex, and marital status were the correct variables to begin with. The reader is not informed as to why these particular respondent characteristics were chosen.

Golden, Anderson and Sharpe

Response rate, speed of response and quality of response have long been of vital importance to users of mail survey instruments. In this article, the authors investigate the effects of monetary incentive, salutation and habitancy on

these response measures. The type of salutation is varied by three levels and a probabilistic incentive is offered to a portion of the respondents. Also, the authors evaluate the self-selection variable of habitancy and have the respondents answer four psychographic measures along with the standard demographic questions. The authors conclude that an apparent raffle effect of the monetary incentive tends to decrease response and that urbanization has an influence on response rate, speed, and quality of response.

An important task in any experiment is to control for external factors which can influence the dependent variable. In the present study, the authors use the postmark date on the respondent's return envelope as their measure of response speed (the surveys were sent from a single Texas location to four different eastern Oklahoma areas). The problem is, that this measure is probably influenced to a large extent by the time required for the Postal Service to deliver the mail to the respondent. While I am not familiar with the postal procedures of Texas-Oklahoma, I have had some experience with the variance of postal speed between and within Athens (rural) and Atlanta (metropolitan), Georgia. I suspect that a wide variance also exists in the subject study area of Oklahoma.

Response quality was measured on a one to five scale depending on how many of the four sections of the questionnaire were answered. The selection of this approach is not clear since most prior studies have selected percent of questions completed as a quality measure. This could affect the results reported in Table 4 since the mean quality varied only over the range 1.09 to 1.32. The assignment of a one to five value tends to mask small differences making them appear more significant than otherwise. For example, if a respondent completes only half of one part of the questionnaire does he get a value of four or five. The effect on the computation of the mean is obvious. Also, the authors fail to discuss what types of questions were asked in the four sections of the study. Based on the affiliation of the third author, one would suspect that the questions related to department store shopping. Could rural, urban and suburban residents have different views of department stores which would effect their interest in responding? This, of course, would have affected the measures of rate and quality. Stated differently, are these measured results questionnaire specific.

The theoretical connection between the rather global psychographic measures used and the dependent measures under study is not clear. For example, why does one expect a relationship between external locus of control and response speed? More specific measures of lifestyle would probably have been more useful.

Finally, does the procedure for entering the cash drawing force the respondent to give up anonymity? If both the questionnaire and the entry card are in the same envelope, then one would suspect that this could have some bearing on whether or not one chose to enter the drawing. Also, did many respondents attempt to enter the cash drawing without returning a questionnaire.

Jacoby and Hoyer

In this paper, the authors address the construct validity of opinion leaders. That is, they ask the question, what if so-called opinion leaders really don't know more?

Specifically, Jacoby and Hoyer test the relationship between the opinion leader construct and expertise, another construct believed logically related to opinion leadership. The authors found that as expected, opinion leadership was highly related to expertise.

The link between opinion leadership and expertise has been implied to be one of causality. That is, a person is an opinion leader because he is more of an expert. Although the authors do not deal with the causality issue directly, one could argue that both opinion leadership and expertise are products of another variable namely involvement with the product category. Thus, correlation between opinion leader and expertise would result. Several of the variables used in the present study could be proxies for product category involvement. In fact, by checking the correlation matrix in Table 1, one can note the high correlation between the involvement proxy looking and both the expertise and opinion leader variables. Also, one can see a high correlation between opinion leader, expertise and past ownership, but, low correlation between present owner. This would also suggest involvement since past owners may be in the looking process thus more involved than present owners who are not in the search process.

One would suspect that opinion leadership is a multi-dimensional concept. Such things as where to buy, price, and suitability for various uses and perhaps even more importantly, source credibility. The present study looks at only the link between technical expertise and opinion leaders. Future research should consider these other factors.

Some General Comments About the Area

More years ago than I sometimes care to remember, I was trained in the field of electronic engineering. A favorite training tool of my mentors was to give the student a black-box which contained some form of circuit. However, the student had only the opportunity to manipulate the input terminals of the box. He did not know what was inside. The approach that one must take in analyzing the box was to input various currents and voltage patterns then measure the resultant output. Based on certain principles or laws of electronic components, one could, through manipulating the input and measuring the outputs, determine, within limits, the internal make-up of the circuit components inside the box.

The above story holds, I think, a lesson for us in survey research. A very brief review of the literature reveals over 132 studies of things such as effects of cover letters, postage, incentives, offer of survey results, and other factors on response rates for mail surveys. My point is that we are really like the student above, but we have failed to accomplish the final important task. That is, we are really only looking at the output patterns of the black box as a result of various inputs (manipulations). Continued focus on only outputs puts this research topic into the realm of "learning more and more about less and less." What is needed, is for us to take that next step and infer something about inside the black box (survey respondent). In the studies reviewed above, I questioned several authors about their choice of independent variables. In fact, neither I nor the authors here really have very firm ground for discussing which variables are or are not important. In short, we really don't have a theory on which to base our studies. Thus, my final plea to the reader is that we proceed to combine our multitude of results and determine the circuit inside survey respondents. This, of course, is no easy task, but, a necessary and much needed one at this time in our development.

MULTIATTRIBUTE SCALING MODELS: SOME OBSERVATIONS

Naresh K. Malhotra, Georgia Institute of Technology

Abstract

Some observations are made on four issues which deserve more attention in conjoint analysis and multidimensional scaling research. These issues concern the use of metric vs. nonmetric techniques, robustness of multiattribute scaling models, cognitive processes underlying these models and the role of individual variables.

Introduction

Over the last decade multidimensional scaling and conjoint measurement have become extremely popular for assessing consumers' perceptions and preferences. Due to their popularity, the multiattribute scaling models have also filtered into the annual conferences of the Association for Consumer Research. The last several conferences have featured papers and/or sessions on this topic. Hence, it was only appropriate that the current conference also include a session on this important area.

However, several issues concerning the multiattribute scaling models remain unresolved or even virtually untouched. Some observations are offered on four such issues: the use of metric vs. nonmetric estimation procedures, the robustness of multiattribute scaling models, the cognitive processes underlying these models and individual differences impinging on these models. No attempt is made to offer comprehensive reviews on each of these issues. Rather, the observations made point to, in the opinion of the author, the crux of the issues involved. It is possible that some of the author's views may not be shared by others working in this area. Also, the choice of these specific issues was influenced more by the author's familiarity and experience with them and does not necessarily reflect their intrinsic importance. Some of these issues have also been dealt with in the papers presented in this session.

Metric vs. Nonmetric

The issue of concern is whether the nonmetric techniques are generally superior to the metric procedures for estimating parameters in multiattribute scaling models. The nonmetric techniques assume that the dependent variable or input data is ordinally scaled or weaker. The metric procedures, on the other hand, are based on the assumption that the dependent variable or input data is at least intervally scaled. A brief discussion of the popular metric and nonmetric estimation procedures for conjoint analysis may be found in Green and Srinivasan (1978). An excellent classification of alternative metric and nonmetric multidimensional scaling procedures for analyzing perceptions and preference data is provided by Green and Rao (1972, p. 13). In the context of conjoint analysis, simulation studies (Carmone, Green and Jain 1978; Cattin and Wittink 1976) as well as empirical studies (Jain et al. 1979; McCullough 1978; Montgomery, Wittink and Glaze 1977) have found that the metric OLS regression does as well as the other nonmetric estimation procedures. In the case of multidimensional scaling also, the nonmetric and metric techniques yield very similar results. This is particularly true if the scaling has been done in correct dimensionality (Green 1975). Moreover, certain flexibilities offered by the nonmetric procedures such as the use of alternative distance functions have turned out to be more

of illusory benefits. The recent findings by Hauser and Koppelman (1979) also should make us pause and take a hard look at the nonmetric multidimensional scaling techniques.

In terms of user-oriented criteria such as algorithm availability, adaptability in the user's system, extent of preprocessing of data, and computing costs, the simpler metric procedures may offer some relative advantages (Jain et al. 1979). Further empirical evidence on the relative performance of metric and nonmetric procedures is needed. However, at this stage, it is reasonable to conclude that the general superiority of the more complex nonmetric procedures over the simpler metric procedures has not been so far demonstrated. It is true that in certain situations, particularly where the input data should be treated as only ordinal, the use of nonmetric techniques would be more appropriate. However, consumer researchers must safeguard against the tendency to automatically employ nonmetric procedures in every situation just because they appear to be more sophisticated.

Robustness

The various issues related to reliability and validity of conjoint analysis have been dealt well by Green and Srinivasan (1978). This discussion would focus on only the robustness of conjoint analysis and multidimensional scaling solutions with respect to incomplete or missing data. As the percentage of data (based on all possible observations) obtained from the respondents increases, the reliability and stability of the estimated coefficients should increase and the error in prediction should decline. On the other hand, obtaining large amount of data from the respondent in a given unit of time could cause fatigue and information overload.

Carmone, Green and Jain (1978) undertook a Monte Carlo simulation to examine the effect of missing data on part worth recovery in conjoint analysis. They varied the number of observations at four levels 18, 27, 54, and 243 (full factorial). Thus, the percentage of missing data was varied from 92.6% to 0%. Their results indicated that the part worths recovery was almost as good with 18 observations as with the full set of 243 observations. In a recent empirical investigation, Leigh, Mackay and Summers (1981) investigated the relative performance of full, half and quarter factorial designs. Their study suggests that the less fractionated design tends to produce higher test-retest part worth reliabilities when the number of attributes is limited. Of course, with a larger number of attributes and levels, the less fractionated designs become impractical and the researcher may be forced to adopt the more fractionated designs accommodating a higher percentage of missing data.

In the context of multidimensional scaling, the empirical study by Jain, Malhotra and Mahajan (1978) revealed that configuration recovery became significantly poorer as the percentage of missing data increased from 20% to 40% to 60%. However, the effect of employing cyclical designs or random method of selecting judgmental pairs on which information was obtained was not found to be significant. Similar results were obtained in the Monte Carlo simulation performed by Spence and Domoney (1974).

In sum, it seems that the strategy of collecting incomplete data to minimize respondent task involves trade-offs.

While the collection of incomplete data from the subjects does reduce respondent fatigue, information overload and costs of data collection, it may not adequately capture the underlying perceptions and preferences. However, the question of how such trade-offs are to be made is far from settled. Further empirical research is needed in this area.

Cognitive Processes

The cognitive processes underlying the formation of consumer's perceptions and preferences have not received due attention from marketing researchers employing conjoint analysis and multidimensional scaling techniques. The major issue here is whether and to what extent the model assumed to estimate parameters represents the actual decision process employed by the individual. In conjoint analysis, for example, a particular decision rule, such as the additive or interactive, is assumed to estimate the parameters of the individual's preference function. This has generally been done without adequate regard to the actual choice rule employed by the respondent. Then, post estimation attempts are made to justify the model used as being appropriate. The additive model has been the most popular. The use of a goodness or badness of fit measure, for example Kruskal's stress (1965), as well as the prediction criteria, to determine the appropriateness of the linear model has been rather misleading. Studies have shown that the additive model can yield good fits even when the data corresponds to other decision models (Dawes and Corrigan 1974; Green 1968). Furthermore, the correlational methods used to assess predictive ability suffer from several problems which limit the usefulness of these methods for examining the appropriateness of alternative decision rules. A discussion of the problems involved may be found in Bettman (1979, pp. 191-193).

An algorithmic development which somewhat alleviates this problem is the multistage decision process model by Srinivasan (1977). This is a general model which captures many of the decision rules considered in information processing literature. By taking a multistage view of the consumer decision process, Srinivasan's model takes us a step closer to modelling the actual choice process.

While past attempts have been lacking in this respect, future use of multiattribute scaling models should attempt to formulate estimation models based on the actual decision process adopted by the individual. An approach to accomplish this has been advocated by Olshavsky and Acito (1980).

Individual Variables

The formation of perceptions and preferences is affected by individual differences (Bettman 1979). Simplistic as this statement may seem, conjoint analysis and multidimensional scaling analysis have, hithertofore, not directly focused on such differences. Individual differences related to information processing would seem to be natural candidates for investigation. Of particular import are consumer cognitive styles (Goldstein and Blackman 1978).

Consider, for illustration, the particular cognitive style of cognitive differentiation. Differentiation refers to the number of dimensions used by an individual in processing information (Bieri 1971). The more cognitively complex individual is assumed to have available a greater number of dimensions with which to construe his environment. Thus, in the context of multidimensional scaling, it is reasonable to expect that for the cognitively complex individuals, a higher dimensionality space would be required to adequately represent their perceptions. As individuals with high cognitive complexity form their percep-

tions in complex ways, a greater percentage of data (in the context of incomplete data) may be required to capture their perceptions and preferences. Likewise, the decision rules used by individuals are also likely to vary depending upon the cognitive complexity level. In similar vein, a number of other interesting hypotheses may be generated. Thus, the need to incorporate individual differences in multidimensional scaling and conjoint analysis of consumer perceptions and preference is apparent, and, to date, largely unmet.

Session on Multiattribute Scaling Models

Two of the three papers in this session deal with some of the issues which have been raised in the foregoing discussion. The paper by Green and DeSarbo describes and applies two models for representing unconstrained choice data in a multidimensional space. Acito and Olshavsky examine the robustness of conjoint analysis with respect to the levels per attribute used to construct the stimuli set. A protocol analysis of the decision rules employed by the respondents is also conducted. Leigh, Mackay and Summers examine the robustness of conjoint analysis with respect to the percentage of missing data and measurement scale for the dependent variable.

References

- Bettman, J. R. (1979), An Information Processing Theory of Consumer Choice, Reading, MA: Addison-Wesley.
- Bieri, J. (1971), "Cognitive Structures in Personality," in Personality Theory and Information Processing, H. M. Schroder and P. Suedfeld, eds., New York: The Ronald Press Co., 178-308.
- Carmone, F. J., Green, P. E., and Jain, A. K. (1978), "The Robustness of Conjoint Analysis: Some Monte Carlo Results," Journal of Marketing Research, 15, 300-3.
- Cattin, P. and Wittink, D. R. (1976), "A Monte Carlo Study of Metric and Nonmetric Estimation Methods for Multiattribute Models," Research Paper No. 341, Graduate School of Business, Stanford University.
- Dawes, R. M. and Corrigan B. (1974), "Linear Models in Decision Making," Psychological Bulletin 81 (February) 95-106.
- Green, B. F. (1968), "Descriptions and Explanations: A Comment on Papers by Hoffman and Edwards," in Formal Representation of Human Judgment, Benjamin Kleinmuntz, ed., New York: Wiley, 91-98.
- Green, Paul E. (1975), "Marketing Applications of MDS: Assessment and Outlook," Journal of Marketing, 39, 24-31.
- _____, and Rao, V. R. (1972), Applied Multidimensional Scaling, New York: Holt, Rinehart and Winston, Inc.
- _____, and Srinivasan, V. (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5 (September), 103-123.
- Goldstein, K. M. and Blackman, S. (1978), Cognitive Style: Five Approaches and Relevant Research, New York: John Wiley and Sons.
- Hauser, J. R. and Koppelman, F. S. (1979), "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness," Journal of Marketing Research, 16, 495-506.
- Jain, A. K., Acito, F., Malhotra, N. K., and Mahajan V. (1979), "A Comparison of the Internal Validity of

Alternative Parameter Estimation Methods in Decomposition
Multiattribute Preference Models," Journal of Marketing
Research, 16, 313-323.

_____, Malhotra, Naresh K. and Mahajan, Vijay (1978), "The
Effect of Missing Values in Similarity Judgments on
configuration Recovery in MDS Models: An Experimental
Examination," paper presented at the First Joint Meeting of
the Psychometric Society and The Society for Mathematical
Psychology, McMaster University, Hamilton, Canada,
August 25-27.

Kruskal, J. B. (1965), "Analysis of Factorial Experiments
by Estimating Monotone Transformation of the Data," Journal
of the Royal Statistical Society, Series B (March), 251-
63.

Leigh, T. W., MacKay, D. B., and Summers, J. O. (1981), "An
Alternative Experimental Methods for Conjoint Analysis,"
Advances in Consumer Research, 8 (forthcoming).

McCullough, J. M. (1978), "Identification of Preference
Through Conjoint Measurement: A Comparison of Data
Collection and Analytical Procedures," Working Paper,
College of Business Administration, The University of
Arizona.

Montgomery, D. E., Wittink, D. R. and Glaze, T. (1977), "A
Predictive Test of Individual Level Concept Evaluation and
Trade-Off Analysis," Research Paper No. 415, Graduate
School of Business, Stanford University.

Olshavsky, R. W. and Acito, F. (1980), "An Information
Processing Probe into Conjoint Analysis," Decision Sciences,
11 (July), 451-470.

Spence, I. and Domoney, D. W. (1974), "Single Subject
Incomplete Designs for Nonmetric Multidimensional Scaling,"
Psychometrika, 39, 469-490.

Srinivasan, V. (1977), "An Approach to the Modeling and
Estimation of Consumer Multi-Stage Decision Processes,"
Paper presented at the AMA/MSI/PIT Conference on Analytical
Approaches to Product and Marketing Planning, Pittsburgh,
Pennsylvania, November.

TWO MODELS FOR REPRESENTING UNRESTRICTED CHOICE DATA

Paul E. Green, University of Pennsylvania
Wayne S. DeSarbo, Bell Laboratories

Abstract

The collection of unconstrained choice data (in which respondents choose products, TV shows, etc. from a reference set that is left unspecified) is a common practice in marketing research. This paper describes and applies two models for representing such data as points in a multidimensional space. The models are illustrated in the context of preferences for sports cars.

In marketing research it is common practice to collect what may be called unrestricted choice data. By this is meant that respondents are neither instructed to choose a fixed number of alternatives nor is the set of possible alternatives explicitly listed. The interesting thing about this class of data is that ambiguity surrounds the alternatives not picked by a given respondent. One cannot tell whether nonchosen alternatives are viewed less favorably than chosen ones because the respondent is not required to list all of the items considered in choosing those that are preferred.

Recently, Levine (1979) has described an interesting model for scaling such data that leads to a joint space representation of stimuli (items chosen) and respondents (choosers). This is carried out by means of an internal MDS (multidimensional scaling) analysis (Carroll 1972) in which the stimulus space is not specified beforehand. Levine also mentions--but does not attempt to develop--a companion external analysis in which the stimulus space is prespecified.

The purpose of the present research note is twofold. We first show how Levine's technique can be applied to problems arising in marketing research. Second, we describe how his method can be adapted to external analysis problems, based on the inclusion of prespecified attribute ratings data and application of Carroll and Chang's PREFMAP-2 model (1971). The net result of this extension is the development of a joint space consisting of three sets of points--stimuli, attributes of stimuli, and respondents' ideal points. An empirical application of the two models is presented, followed by a brief discussion of potential industry applications.

The Models

We start out by describing the Levine model and PREFMAP-2. Since detailed descriptions of each model appear in the references cited above, our discussion is brief.

The Levine Model

Basic input data to the Levine model consist of I respondents' choices of J stimuli (e.g. brands, political candidates) in terms of preference, endorsement, appropriateness, or any other construct of interest to the researcher. The J stimuli are defined as the set union of all I individuals' choices. The basic idea of the Levine model is motivated by the following argument. If one had available an a priori stimulus configuration (as in external MDS analysis), a natural procedure would be to compute any given respondent's ideal point as the centroid of the stimuli he/she picks. By the same token, if an a priori respondent configuration were available one could position a stimulus point at the centroid of the points representing all respondents who happened to pick that stimulus.

What Levine's model does is to find two sets of coordinates, one set for respondents and one set for stimuli, that satisfy the centroid criterion. In Levine's notation, given respondents $1, 2, \dots, i, \dots, I$ and stimuli $1, 2, \dots, j, \dots, J$, let E denote a symmetric $(I+J) \times (I+J)$ matrix in which the entries are

$$e_{i, I+j} = e_{I+j, i} = 1 \quad (1)$$

if and only if respondent i picks stimulus j (with zero otherwise).

Next, let D denote an $(I+J) \times (I+J)$ diagonal matrix with general entry d_{kk} representing row sums of corresponding rows of E . Let \underline{x} denote a vector of $(I+J)$ coordinates corresponding to the I respondents and the J stimuli. The vector \underline{x} is a solution to the problem if for all x_k in \underline{x} , the entry x_k is proportional to the centroid of coordinates representing entities to which k is linked by virtue of either choosing or having been chosen. In scalar notation, the solution is:

$$\lambda x_k = \left[\begin{array}{c} d_{kk}^{-1} \\ \sum_{n=1}^{I+J} e_{kn} x_n \end{array} \right] \quad (2)$$

where $1 \leq k \leq (I+J)$ and $\lambda > 0$. Expressing (2) in matrix notation, we have:

$$\lambda \underline{x} = (D^{-1} E) \underline{x} \quad (3)$$

and, hence, the desired coordinates are the eigenvectors of the nonsymmetric matrix $D^{-1}E$. Various ways are available to expand $D^{-1}E$ to a symmetric form that permits standard eigenstructure computer routines to be used in solving for \underline{x} . As in any kind of eigenstructure decomposition, the researcher may wish to retain only the first few eigenvectors, accounting for the greatest proportion of the trace of $D^{-1}E$.

In sum, the Levine model takes an $(I+J) \times (I+J)$ matrix of 0-1 entries and solves for a set of coordinates \underline{x}_r in some $(I+J) \times r$ -dimensional space, according to the criterion of equation (2).

PREFMAP-2

Carroll and Chang's PREFMAP-2 algorithm was motivated by the desire to define a stimulus configuration from preference-like data alone. In PREFMAP-2 the preferences are typically interval-scaled (e.g. ratings), not 0-1 data. The key idea of the PREFMAP-2 model is to allow the preference data to be used twice. For example, if I respondents are asked to rate J stimuli with respect to preference, then it can be shown that a singular value decomposition of the doubly centered $I \times J$ matrix (i.e. the $I \times J$ matrix with both row and column means removed) yields both a stimulus space and a set of ideal points for respondents. Unfortunately, however, each is defined only up to a linear transformation. PREFMAP-2 finds the desired linear transformation of the stimulus space (which satisfies an explicit least squares criterion) so that the stimulus configuration is appropriately linked to the ideal point configuration.

Having found this linear transformation, the specific ideal point locations are found by quadratic regression in exactly the same manner used in the original PREFMAP algorithm. Somewhat more formally, let:

$$p_{ij} \approx a_i d_{ij}^2 + c_i \quad (4)$$

denote the preference scale value of respondent i for stimulus j ; d_{ij}^2 is the squared Euclidean distance between stimulus j and ideal point i ; a_i and c_i are parameters of a linear function for respondent i ; \approx denotes least squares approximation.

In PREFMAP-2, equation (4) is constrained to $a_1 = a_2 = \dots = a_I = a$, so that all respondents have the same slope coefficient. The first step in the computational procedure is to factor P , the preference data matrix with general entry p_{ij} , into the matrix product:

$$P \approx YX' \quad (5)$$

where X is the desired stimulus matrix, defined up to a linear transformation. The object is then to find that linear transformation T , so that:

$$X = X^*T. \quad (6)$$

PREFMAP-2 provides a least squares criterion that solves for T in the sense of best fitting (in a least squares sense) the preference values. Details can be found in Carroll and Chang (1971).

Comparing the Two Models

PREFMAP-2 and the Levine model were developed independently for different types of data--0-1 responses in the case of the Levine model and interval- (or ratio-) scaled data in the case of PREFMAP-2. What is proposed here is to use PREFMAP-2 in developing a joint space of stimuli and a third set of entities--attributes of the stimuli. In this case the list of attributes is prespecified. It is further assumed that the respondents represent a homogeneous sample with regard to their perceptions of the stimuli that they associate with specified attributes.

Having found a joint space of stimuli and attributes, each respondent's unconstrained stimulus choices (i.e. Levine-type pick k/n data) are used to compute that person's ideal point in the joint space of stimuli and attributes. Ideal points are computed as centroids of the stimuli that each respondent picks. This step leads to three sets of points in a common space--attributes, stimuli and ideal points. An attractive property of the approach is that, if all distances are comparable, one can consider ideal point to attribute distances. These latter distances can be substantively interpreted as the desirability that each respondent has for each stimulus attribute (as inferred from consideration of the respondent to stimulus associations and the stimulus to attribute associations).

In sum, we address the problem of modeling pick k/n data for the case of an externally supplied stimulus configuration (as mentioned by Levine). This is accomplished by the PREFMAP-2 procedure. We then apply the centroid feature of Levine's approach to modify PREFMAP-2, so as to include ideal points derived from 0-1 preference responses.

A Pilot Application

The two-stage approach, just described, was applied, on a pilot basis, to a set of data obtained from 35 respondents, drawn from a university population. Each respondent was asked to complete three tasks in the order shown:

1. List all foreign sports cars that come to mind that he/she would like to own, if money were not a major

- constraint. (No specified number or list of choices was provided.)
2. For each chosen car, pick as few or as many attributes as desired, from a specified list of 20 (see Table 1), that are most highly associated with the choice.
3. List three car attributes from the list of 20 that he/she would most like to have in a foreign sports car.

Table 1 shows the nine foreign sports cars whose names appeared most frequently across the set of 35 respondents.

TABLE 1
SPORTS CAR NAMES MENTIONED MOST FREQUENTLY
AND PRESPECIFIED CAR ATTRIBUTES

Sports Cars	BMW	Jaguar SJ6
	Datsun 280ZX	Mercedes 450SL
	Ferrari	MGB
	Fiat	Porsche
		Triumph
Attributes		
Economical to maintain		Plush interior
High acceleration		High reliability
High braking ability		Excellent cornering
Good for city driving		Sleek, racy lines
Easy-to-get parts		Conservative styling
Good gas mileage		Well engineered
High durability		Good service facilities
Beautiful lines		nearby
High resale value		Luxurious styling
Comfortable on long trips		Not temperamental
		Prestigious

Application of the Levine Model

Let us consider the Levine model first. In this case we focus only on the 35x9 matrix of 0-1 entries in which a 1 denotes stimulus j being picked by respondent i ; 0 denotes otherwise. (As described earlier, this matrix is then transformed to a 44x44 symmetric E matrix.) A computer program was prepared to solve for the respondent and stimulus matrix of coordinates, as described in equation (3).

The first 3 eigenvalues accounted for 55 percent of the trace. For illustrative purposes, Figure 1 shows a plot of dimensions 2 and 3.¹ Each eigenvector has been scaled by the square root of its associated eigenvalue. The horizontal axis might be described as a "value" axis that separates cars of lower price (e.g. MGB) from those of higher price (e.g. BMW). The vertical axis might be considered as an appearance axis that separates highly elegant cars such as Ferrari and Jaguar SJ6 from the more functional cars, like BMW and Porsche. (However, these interpretations are speculative, at best.)

Although omitted (for clarity) from Figure 1, the ideal points obtained from the Levine internal analysis showed a fairly high concentration around Mercedes, BMW, and Porsche; still, each car had one or more respondents for whom it represented the closest point.

The Two-Stage Analysis

The next step in the analysis was to apply the PREFMAP-2 model to the 9x20 matrix of associations data in which each cell entry denotes the frequency, across respondents, with which each attribute (column) is perceived to be associated with each sports car (row). Solutions, involving the fitting of the simple ideal point model to the

¹In eigenstructure solutions of the particular matrix product in equation (3), the first eigenvector is always a constant vector, and hence, is deleted in the spatial plots.

the optimally rotated and differentially stretched stimulus space, were obtained in both three and two dimensions. For illustrative purposes we comment only on the two-dimensional solution.²

All multiple correlations between the dependent variable (frequency with which each attribute was associated with a given sports car) and the stimulus coordinates of the attributes were significant at the 0.05 level or better. Figure 2 shows the joint space of sports cars and attributes.

Comparison of the sports car positions in Figure 1 and 2 shows the influence of the attribute-sports car associations on the placement of the sports car points.³ As noted, Mercedes, BMW, and Porsche are still relatively near each other (as are Jaguar and Ferrari) but Fiat and MGB are much more distant from each other in Figure 2 than in Figure 1.⁴

We note that such attributes are prestigious, sleek and racy, plush interior, beautiful lines, and comfortable describe the Ferrari and Jaguar, while the Porsche is described as having high resale value, good braking, high acceleration, good cornering and being well engineered.

Computing the Ideal Points

Respondent ideal points were next found, for the analysis summarized in Figure 2, by simply computing the centroid of all sports cars that each respondent picked. As a type of simple model validation, the next step in the analysis was to compute, for each respondent in turn, the average Euclidean distance of the three attributes that were picked as most highly desirable from his/her ideal point. This average was then compared to the average Euclidean distance of the 17 attributes that were not picked by that respondent.

Our hypothesis, of course, is that picked attributes will be closer, on the average, to a person's ideal point than those not picked. For 24 of the 35 respondents this was, in fact, the case. A sign test indicated that the result was significant beyond the 0.05 level. Hence, there is some evidence to support the reasonableness of the two-stage modeling approach proposed in this paper.

Potential Applications

As mentioned earlier, marketing research surveys abound with questions dealing with the choice of unconstrained options. For example, such data as top-of-mind brand recall (list those brands that you recall having tried at least once in the past six months), non-aided listings of product use occasions, free associations data, and reasons for liking (disliking) a new product concept lend themselves

²The root-mean-square fit across all nine sports cars was 0.76 for the two-dimensional solution and 0.83 for the three-dimensional case.

³In future studies the researcher may wish to rotate (or otherwise transform) the two solutions to maximal congruence via Cliff's (1966) algorithm or some other such procedure. Here, our interest was primarily to examine the character of the solutions just as they appeared in their respective computer printouts.

⁴Still, the canonical correlation between the sports car coordinates in Figure 1 and their counterparts in Figure 2 turned out to be 0.84, showing reasonably good correspondence between the two solutions.

well to analysis by the original Levine model.

Moreover, if respondents can be meaningfully classified on some type of a priori basis (e.g. favorite brand, heavy versus light product usage), additional analyses can be carried out to see if their ideal points occupy different regions of the joint space.

The two-stage approach, incorporating PREFMAP-2, followed by Levine-type external analysis (computing the centroid of picked items), is also applicable to a number of situations. For example, the stimuli could be print advertisements and the attributes could be characteristics of the ads--type of theme, believability of claims, copy points, etc. In other cases the stimuli could be political candidates and the attributes could be issues on which the various candidates are perceived to have a position. In still other cases the stimuli could be vendors of telecommunications services and the attributes could be various features of those services, such as excellent technical support, high equipment reliability, and so on.⁵

In short, it is not difficult to think of many classes of marketing problems for which the models described here may be applicable. It is hoped that additional empirical work will be carried out in the future regarding the pragmatic value of these models in consumer and industrial buyer research. Some initial research on the applicability of the Levine model has already been carried out by Holbrook, Moore, and Winer (1980).

References

- Carroll, J. Douglas (1972), "Individual Differences and Multidimensional Scaling," in R. N. Shepard, A. K. Romney, and S. B. Nerlove (eds.), Multidimensional Scaling. Vol. I. New York: Seminar Press, 105-155.
- Carroll, J. Douglas and Chang, Jih-Jie (1971), "An Alternate Solution to the Metric Unfolding Problem," paper presented at the annual meeting of the Psychometric Society.
- Cliff, Norman (1966), "Orthogonal Rotation to Congruence," Psychometrika, 31 (March), 33-42.
- Holbrook, Morris B., Moore, William L. and Winer, Russell (1980), "Using 'Pick Any' Data to Represent Competitive Positions," paper presented at the TIMS/ORSA Conference on Market Measurement in Austin, TX, March.
- Levine, Joel H. (1979), "Joint-Space Analysis of 'Pick-Any' Data: Analysis of Choices from an Unconstrained Set of Alternatives," Psychometrika, 44 (March), 85-92.

⁵It is also possible to apply PREFMAP-2 to cases where the attributes are obtained by (say) free association, conditioned on each object that the respondent evokes. In this case unrestricted choice can apply to both brands (or other stimulus objects) and the attributes of those brands.

FIGURE 1
JOINT SPACE CONFIGURATION OBTAINED FROM THE LEVINE MODEL

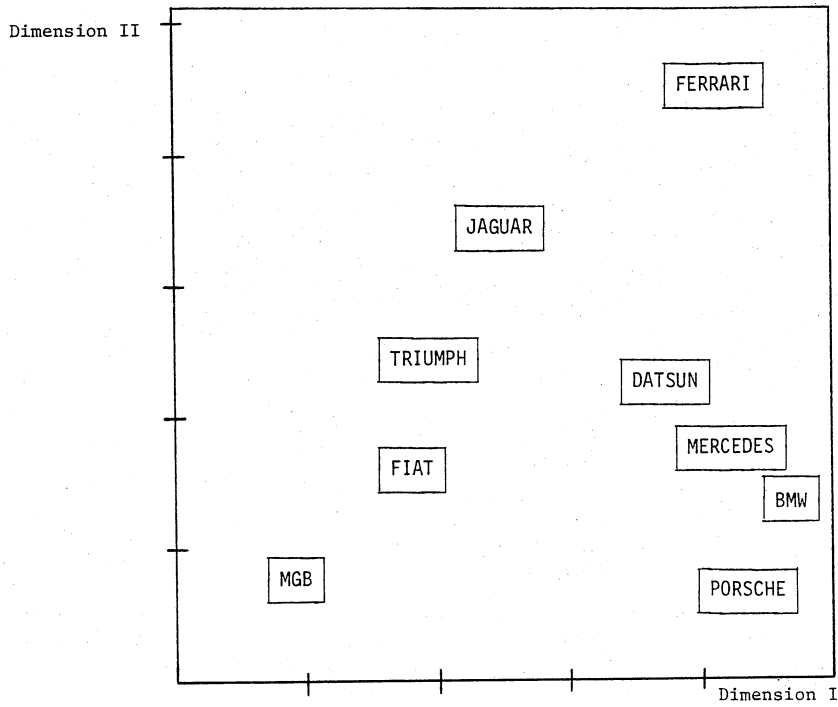
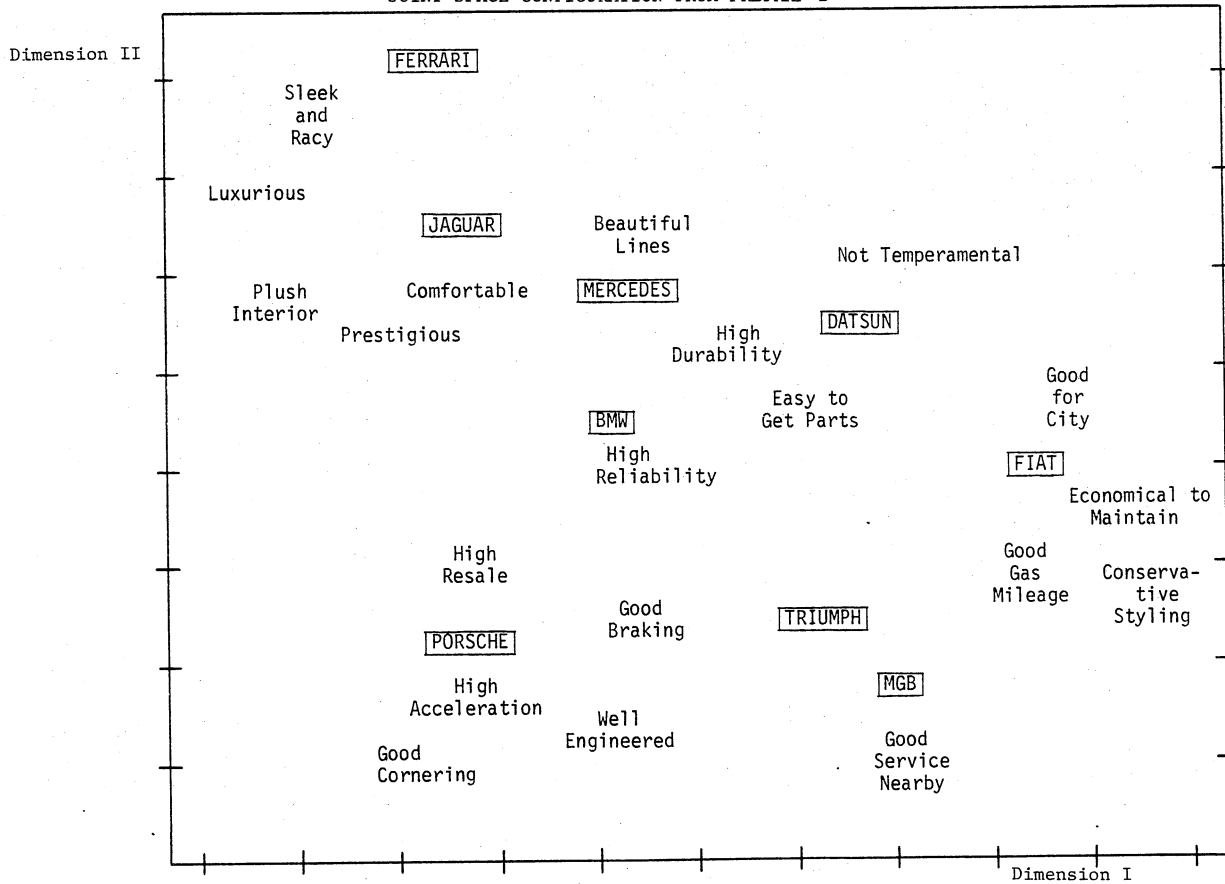


FIGURE 2
JOINT-SPACE CONFIGURATION FROM PREFMAP-2



LIMITS TO ACCURACY IN CONJOINT ANALYSIS

Franklin Acito, Indiana University
Richard W. Olshavsky, Indiana University

Abstract

The relative performance of two conjoint designs, one using two levels per attribute, and the other using three levels is compared. Although the three-level design can potentially provide more information to the researcher, the increased difficulty of the evaluation task results in poorer data than obtained in the two level design.

Introduction

Conjoint analysis, beginning with the expository article by Green and Rao (1971), has developed into a mainstream tool of the market researcher. Investigations of the reliability and validity of the technique (e.g., Green and Wind 1973, Acito 1977, Scott and Wright 1976, and McCullough and Best 1979) have produced generally favorable evaluations. Researchers, however, still face difficult design decisions in using conjoint analysis for which the existing literature provides little guidance. An example of such a design issue is the number of levels per attribute. Experimental designs (see Addelman 1962) exist for two, three, four or more levels per factor. In some cases the number of levels is dictated by the context of the problem. In other situations, the analyst has control over the number of levels.

Price is one attribute which allows flexibility in selecting number of levels. Two levels, P_A and P_B , can be used to derive utility values U_A and U_B . Utilities for intermediate price levels can be obtained by linear interpolation,

$$U_I = U_A + \frac{P_I - P_A}{P_B - P_A} \times (U_B - U_A). \quad (1)$$

The analyst may not be willing to make such an assumption of linearity, however. To investigate non-linear effects, a third price level can be included in the design. Inclusion of the third level provides, in principle, at least, more detailed estimation of the utility function. However, two issues arise in increasing the number of levels per attribute which may result in poorer information from the conjoint analysis.

First, increasing the number of levels results in a greater number of parameters to be estimated, increasing the chances for statistical error with a fixed number of profiles or assemblies. (This is true only if the attribute levels are treated as discrete, which would be the case if the analyst did not wish to restrict the utilities to an a priori functional form.) Second, the additional information processing burden on the respondent can cause confusion, carelessness, or the adoption of simplifying choice heuristics frustrating the researcher's attempt to obtain greater accuracy. This study is designed to determine if a reduction in accuracy does in fact occur as the number of levels is increased. This study is also designed to provide information, through the simultaneous use of a process tracing technique (protocol analysis), concerning the hypothesized deterioration in the subject's capacity to cope with the increased information processing burden.

Method

Subjects

Twenty MBA students served as subjects in this study. They were selectively recruited on the basis of their ownership

or usage of a product (typewriters) which was of interest in another, unrelated, part of the study. Since subjects were randomly assigned to the 2 level or 3 level condition, the lack of control for prior knowledge of stereo receivers should not be a problem. Subjects were paid for participation. The small sample size is typical of studies in which protocol analysis is performed.

Stimulus Materials

Stereo Hi-Fi receiver was selected as a product of interest to students for which the number of levels per attribute could be varied uniformly (i.e., most major attributes were inherently multi-level).

For this study, six attributes and three levels were selected: Power Output (45, 30, 50 percent); Sensitivity (2.0, 2.5, 3.0 microvolts); Signal to Noise Ratio (75, 70, 65 decibels); Price (200, 250, 300 dollars); Warranty (3, 2, 1 years). The number of levels per attribute was either 2 or 3, depending upon the condition (underlined values were used in the 2 value condition). Using an orthogonal array design, 23 profiles were developed for each of the two conditions (Addelman 1962, Plackett and Burman 1946).¹ The letters "A" to "W" were used to indicate alternatives of the 2 level condition, while the numbers "1" to "23" were used to designate the alternatives of the 3 level condition.

A second set of eight holdout profiles was developed (completely different from those in the previous sets) involving only 2 levels per attribute, also based upon an orthogonal array. (The fact that this set had only 2 levels is not a critical issue; all that was desired was a criterion set of preference ranks.) These were numbered "100" through "800."

A third set of ten actual stereo receivers was developed by selecting a representative set of ten "actual" models from a stereo catalog. For this set, actual brand names were used and warranty was dropped since no warranty information was provided in the catalog. (Subjects were told to assume that all models had the same two year warranty.) Additional descriptive information was also available, as was a photograph of each receiver.

Procedure

The experimental task required respondents to express preferences for the alternative receiver profiles. These profiles were printed on 3" x 5" cards with each card listing the value of each of the six attributes. In the actual brand condition, pages from the catalog of a local stereo equipment retailer were used. To ensure comprehension of all attributes, definitions were presented to each subject on a separate sheet; subjects were requested to read these definitions and to refer to them throughout the choice process as needed.

The 23 cards for the first task were shuffled (for each subject) and arranged in an array on a table. Ten subjects

¹The orthogonal design for the three level case given by Addelman (1962) contained 25 assemblies; however, only 23 were distinct. The design for 24 assemblies given by Plackett and Burman for two levels has 22 distinct combinations. Twenty-three were used, so that one duplicate was included in this design.

were randomly assigned to the 2 level condition and ten to the 3 level condition. The subjects were asked to examine the "receivers" and to select the one model from the set of alternatives offered that would be purchased. Subjects were further instructed that they could not move or in any other way manipulate the cards. This was done to simulate the situation encountered in a retail outlet in which customers could not physically rearrange the alternatives.² Finally, subjects were instructed to verbalize all of their thoughts as the receivers were considered. The experimenter, who was seated opposite the subject, constantly monitored the protocol and reminded (in simple, nondirective ways) the subject to articulate his/her thoughts whenever necessary to increase the amount of protocol data.³ Each session was tape recorded with the subject's knowledge.

After the subject made the first choice from the 23, that alternative was eliminated and he/she was asked to imagine it were no longer available for sale. The subject was instructed to again make a "purchase" decision from the remaining alternatives. This procedure was repeated until a preference ordering for all 23 alternatives was obtained.

To provide the data needed to test the predictive ability of conjoint analysis, subjects were asked to express their preferences for the set of eight holdout profiles with the same procedure used for the 23 alternative set. Finally, subjects were asked to indicate rank preferences for 10 models of receivers from the catalog pages. Data for each subject were submitted to MONANOVA (Kruskal 1965) to derive part-worth utility values.

Results

Prediction of "Holdout" Ranks

The ranks of the eight holdout profiles were predicted using the utility values derived from the conjoint analysis and the Spearman rank correlation coefficients were computed between the predicted and actual ranks for each respondent. Table 1 shows the mean correlations for the two and three level designs. The difference in the means was not significant. Moreover, the two level design produced more direct "hits" on the ranks than did the three level design (45 out of a possible 80 versus 31 out of 80). A χ^2 test for the 2 x 2 table resulting from this data indicated a significant difference at the .05 level.

Prediction of Actual Brands

The respondents' ranks for the ten actual brands were disaggregated according to the three brands used in the experiment. Three models each of Pioneer and Kenwood receivers were used while four models of the Technics brand were used. Appropriate utility values derived from the conjoint analysis were combined (using linear interpolations where necessary) to predict the preference ranks for each of the three sets of ranks. Table 1 shows the distribution of perfect prediction for each of the experimental conditions. The differences between the overall number of perfect predictions (24 out of 30 for the two level design versus 15 out of 30 for the three level design) was significant beyond the .05 level using a χ^2 test.

² Bettman (1979) has made a call for such realism in study design. Olshavsky and Acito (1979) have recently investigated the effects of changes in card sorting procedure on choice rule.

³ Evidence concerning the lack of interference of protocol analysis on cognitive tasks has been recently reviewed by Simon (1979).

Utility Value Violations

In many situations the directionality of respondent preferences for attribute levels can be stated a priori. In this experiment, each of the attributes had a clearly defined preference directionality (e.g., lower distortion should

TABLE 1
Summary of Stress and Predictive Accuracy
for 2 level and 3 level Conditions

Condition	Average Stress	Spearman Correlation between Predicted and Actual Ranks of Holdout Profile	Within Brand Predictive Accuracy*		
			Kenwood	Pioneer	Technics
2 levels per attribute	11.7%	.88	9	8	7
3 levels per attribute	14.0%	.82	6	5	4

* Number of cases where perfect prediction of within brand ranks was observed.

TABLE 2
Violations of A Priori Attribute Utilities
for Extreme Values

Condition	Attribute					
	Power 45w>15w	Price \$200> \$300	Distortion .1%>.5%	Sensitivity 2.0µv> 3.0µv	Noise 75db> 65db	Warranty 3 yrs.> 1 yr.
2 levels	0	3	1	1	1	0
3 levels	1	3	1	7	6	0

be preferred to higher distortion, lower price preferred to higher price, etc., with the usual ceteris paribus assumption). For both experimental conditions, the number of violations of a priori expectations was determined. For the three level condition, only the outer extreme levels of each attribute were used for this assessment. Table 2 shows the results. A total of six violations was observed for the two level condition while 11 violations were observed for the three level condition. (If the price utility sign is not considered, the two level had three violations and the three level design had 8 violations)

The utility values for the intermediate levels of each attribute in the three level condition were also examined. For each attribute used in this study, it seemed appropriate to assume that the utility values for the intermediate levels should lie between those for the extremes. In other words, utility functions for each attribute were assumed to be monotonically related to levels of the attribute. A number of violations of this monotonicity were observed: two violations for the power attribute; four for price; two for distortion; seven for FM sensitivity; five for noise; and six for warranty.

Error Estimates

The two-level design required the estimation of six parameters with MONANOVA while the three level design required

the estimation of twelve parameters. Since the number of profiles was 23 for both designs, the 3 level design is more susceptible to error. One perspective on this problem is gained by examining the stress values expected for both designs using random data. The average stress for the two level design with random data was .72, with a standard deviation of .09. Assuming the stress values to be normally distributed, 5% of the stress values for random data are expected to be below 57%. A similar procedure for the three level design resulted in an estimate that 5% of the stress values for random data would be below 29%.

Table 1 shows the average stress values observed for the two level and three designs. Since the random data results above indicate that it is "easier" to achieve a low stress value with the three level design, lower stress values would be expected for real data using that design. The averages indicate that, to the contrary, the stress for the two level design was somewhat lower, although the difference between the two designs was not significant at the .05 level.

Choice Rule Analysis

The protocol analysis was based on transcripts of the tape recordings for the set of 23 profiles only. The objective was to ascertain the decision rule used by each subject. The tape recorded protocols were transcribed and then broken down into a sequence of task relevant statements as is typically done with protocol data (Payne, Braunstein, and Carroll 1978). Subjects were then classified according to the type of choice rule they used (as inferred from the protocols). In the interest of consistency and to ensure reliable categorization, the same definitions of choice rule and coding criteria used by previous researchers were adopted (Olshavsky 1979, Payne 1976, Wright and Barbour 1977).⁴

Of the 10 subjects in the "two-level" condition, five used a lexicographic choice rule (by attribute). Three used a conjunctive rule (by brand) but modified the importance of the choice criteria according to a specific attribute priority order. And two used a modified conjunctive rule where priority of attribute was varied after all alternatives with the most desired attribute were depleted.

Of the 10 subjects in the three level condition, seven used a choice strategy that was so inconsistent and ambiguous that it was not possible to classify it as a single choice rule or even a combination of choice rules. In those seven cases a pair-wise comparison process (attribute dominance or additive difference) was used with little or no attempt to apply the pair-wise comparison strategy to all remaining alternatives. Of the remaining three subjects, one used a lexicographic rule but then switched to a pair-wise (attribute dominance) strategy.

Discussion

The major finding of this study is that increasing the number of levels per attribute does not necessarily increase accuracy in utility function estimation. The results suggest that, to the contrary, the utility values derived from the two level condition are somewhat superior in predictive ability when used in the manner described. This superiority persists even in cases where interpolations of utility values are needed (as in the catalog descriptions). If the three level design were superior, its superiority should be evident where interpolated utility values were used, since non-linearities in the utility functions would be captured. The utility values derived for the three level condition also were more likely to violate a priori assumptions about

⁴Details of the procedure for choice rule identification have been discussed in previous papers, e.g., Olshavsky and Acito 1980.

directionality of preferences. This result must of course be qualified given the small sample size and the unrepresentativeness of MBA students. Further research on this important issue is required using a larger sample of different types of subjects and less complex products.

These results could be due to the instability in parameter estimates resulting from fewer degrees of freedom remaining for error and/or they could be due to the confusion of the respondents. However, the error analysis performed here and the protocol analysis suggest that the more likely explanation is that most of the respondents in the three level condition were unable to cope with the increased information processing burden imposed upon them. This implies that great care must be given to the interpretation of conjoint analysis results based on designs involving a large number of levels per attribute.

References

- Acito, F. (1977), "An Investigation of Some Data Collection Issues in Conjoint Analysis," Educators' Proceedings, Chicago: American Marketing Association, 82-85.
- Addelman, S. (1962), "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," Technometrics, 4 (February), 21-46.
- Bettman, J. R. (1979), An Information Processing Theory of Consumer Choice, Reading, MA: Addison-Wesley Publishing Company.
- Green, P. E. and Rao, V. R. (1971), "Conjoint Measurement for Qualifying Judgmental Data," Journal of Marketing Research, 8 (August), 355-63.
- Green, P. E. and Wind, Y. (1973), Multiattribute Decisions in Marketing: A Measurement Approach, Hinsdale, Illinois: Dryden Press.
- Kruskal, J. B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformation of the Data," Journal of the Royal Statistical Society, Series B (March), 251-63.
- McCullough, J. and Best, R. (1979), "Conjoint Measurement: Temporal Stability and Structural Reliability," Journal of Marketing Research, 16 (February), 26-31.
- Olshavsky, R. W. (1979), "Task Complexity and Contingent Processing in Decision Making: A Replication and Extension," Organizational Behavior and Human Performance, 24, 300-316.
- Olshavsky, R. W. and Acito, F. (1980), "The Impact of Data Collection Procedure on Choice Rule," Advances in Consumer Research, 7, 729-732.
- Payne, John W., Braunstein, L. and Carroll, J. S. (1978), "Exploring Predecisional Behavior: An Alternative Approach to Decision Research," Organizational Behavior and Human Performance, 22, 17-44.
- Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analyses," Organizational Behavior and Human Performance, 16 (August), 366-387.
- Plackett, R. L. and Burman, J. P. (1946), "The Design of Optimum Multifactor Experiments," Biometrika, 33, 305-325.
- Scott, J. E. and Wright, P. (1976), "Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations," Journal of Marketing Research, 13 (August), 221-4.
- Simon, Herbert A. (1979), Information Processing Models of Cognition, Annual Review of Psychology, 30 (February), 363-96.

Wright, Peter and Barbour, Fredrick (1977), "Phased Decision Strategies: Sequels to an Initial Screening," in M. K. Starr and M. Zeleny (eds.), Multiple Criteria Decision Making: TIMS Studies in the Management Sciences, Vol. 6, Amsterdam: North-Holland.

Thomas W. Leigh, Pennsylvania State University
 David B. MacKay, Indiana University
 John O. Summers, Indiana University

Abstract

The rise in popularity of conjoint analysis as an approach to studying consumer preferences has been accompanied by the development of a variety of alternative methodological procedures. Interest in this paper is focused on the performance of "experimentally efficient" designs. Issues which face the researcher interested in deciding upon an appropriate design for conjoint analysis are discussed and results of a pilot study on the test-retest reliability and convergent validity of alternative procedures are presented. The results tend to favor the "less efficient" designs with full factorial and pair-comparisons data gathering procedures having the greater reliability and validity.

Introduction

Over the past decade, conjoint analysis has become an increasingly popular approach for the study of consumer preferences. In an excellent review of the literature, Green and Srinivasan (1978) have identified six major methodological steps involved in applying conjoint analysis: (1) selection of a model of preference; (2) data collection method; (3) stimulus set construction for the full profile method; (4) stimulus presentation; (5) measurement scale for the dependent variable; and (6) estimation method. The purpose of the paper is to investigate the effects of the methodological decisions in steps three and five on the reliability and validity of conjoint analysis.

Interest in this area is motivated by the recent emphasis on the use of "experimentally efficient" designs with conjoint measurement (Acito 1979, Green 1974, Green et al. 1978). Green and Srinivasan (1978) define an experimentally efficient design as one that returns a high amount of information per unit of experimental time. They suggest that efficient designs are likely to involve fractional factorials (stimulus set construction) and rank order (as opposed to pair-comparison) data (measurement scale). Of course, efficiency can also be used in a statistical sense to refer to estimates that have the property of minimum variance. By studying the reliability and validity of experimentally efficient designs, insight should be obtained as to whether experimentally efficient designs will lead to statistically efficient estimates of model parameters.

This issue is by no means trivial. Experimentally efficient designs reduce the amount of time required of a subject to provide preference data with respect to alternative product designs with a specific set of attributes. To the extent that reduction of subject time prevents fatigue, boredom, and stereotyped response patterns, random or error variance will be reduced and reliability and validity may be improved. However, the price of this savings in subject time is a reduction in the number of independent observations. Estimates of model parameters are thus more dependent on any given response. The self-compensating nature of random error has less opportunity to occur with a small number of observations and reliability and validity can be adversely affected. Experimental methods which are empirically shown to have higher reliability and validity have a better chance of providing statistically efficient estimates. However, experimentally efficient designs need not lead to increased reliability and validity.

The following sections expand upon the issues relevant to stimulus set construction, measurement scale, and reliability and validity. After this, a study in which the reliability and validity of 13 different conjoint measurement procedures based on combinations of stimulus sets and measurement scales is described. A presentation of the results and a discussion of the findings concludes the paper.

Stimulus Set Construction

The stimulus set construction step includes the issues of attribute range and variation, interattribute correlation, and full versus fractional factorial designs. Only the last issue is considered here. For full factorials, the number of responses required of a subject increases geometrically with the number of attributes. Even with a relatively small number of attributes, full factorial designs can become infeasible.

Even when the number of required responses for a full factorial design appears feasible to use with a given measurement scale (e.g., 5 attributes at 2 levels each; 32 stimuli) one might still find that a fractional factorial design produces more reliable and valid results than a full factorial design. The reduction of the subject's task with a fractional factorial may tend to increase the study's response rate and encourage respondents to put more time and effort into evaluating each individual stimulus. Conversely, the smaller number of responses provides less opportunity for random errors to cancel each other. Statistically, the basic issue is whether the lower degrees of freedom in a fractional factorial design are compensated for by higher quality (lower error) data. This is, of course, an empirical question and depends upon the particular stimuli being investigated, the subjects who participate, and the research setting. Caution is thus advised in generalizing the results from one study to another.

Another potential problem with fractional factorial designs is that they do not allow for the estimation of some higher order interaction effects. However, these will frequently be negligible and estimation of lower order effects may be sufficient.

Measurement Scale for the Dependent Variable

A variety of procedures for "defining a measurement scale for the dependent variable" have been used. Coombs (1966) refers to these as "methods of collecting data." The procedures fall into two basic categories, nonmetric (e.g., rank order and pair-comparisons) and metric (e.g., rating scales, graded pair-comparisons, and constant sum pair-comparisons). The nonmetric procedures only require the subject to make ordinal judgments concerning their preferences (or purchase intentions) while the metric approaches involve at least interval scale responses. The choice between these two basic approaches involves the experimental questions of whether subjects can provide metric responses and whether the longer time periods that usually accompany the greater number of responses frequently required with some nonmetric methods (e.g., pair-comparisons) result in boredom, fatigue and greater random error.

Of the two nonmetric approaches, rank order has been most

popular for collecting preference data for conjoint analysis (e.g., Green and Wind 1975; McCullough and Best 1979). This method generally requires less respondent time than pair-comparisons and, for a large number of stimuli, the pair-comparisons method becomes impractical. At a given level of fractionation, the number of responses required for pair-comparisons may greatly exceed that required with rank orders. Thus, a half factorial design for five attributes of two levels each would, for pair-comparisons, involve 120 comparative responses ($2^{5-1} = 16$; $C(16,2) = 120$), and for rank orders, it would involve 16 responses. Corresponding values for a full factorial design are 496 and 32 responses respectively. Unfortunately, when using the pair-comparisons method, incomplete block designs (Green 1974) often do not allow for the development of a complete rank order of the stimulus set for individual subjects (Cliff 1975).

Among the metric approaches, rating scales are the easiest to apply. Unlike the nonmetric methods of rank order and pair-comparisons, rating scales do not force a subject to use "comparative" judgments since the stimuli are evaluated individually, one at a time. This contributes to the anchoring effect difficulties that the method experiences. However, use of warmup trials and familiarization periods where the subject reviews all stimuli before making the rating responses can lessen the severity of this problem.

Graded pair-comparisons are a metric procedure that, because of its explicit requirement of comparative evaluation, would appear to have an advantage over rating scales with respect to potential anchoring effects. Its advantage over (binary) pair-comparisons is that it allows the respondent to specify a degree of preference for one stimulus over another. Another advantage over pair-comparisons is that graded pair-comparisons can use both fractional factorials and incomplete block designs to determine respectively the construction and selection of the stimuli that are presented to the subject.

Constant sum pair-comparisons, the third metric procedure, require ratio scaled responses and would seem to share some of the problems and advantages associated with graded pair-comparisons. The nature of the stimuli and familiarity of the subjects with the stimuli are important factors in deciding whether ratio scale responses are reasonable to require.

Reliability and Validity

Reliability can be conceptualized as consisting of temporal and structural components (McCullough and Best 1979). The focus of this paper is on temporal stability which is concerned with the effect of random (error) variance. Test-retest methods are a common way of measuring the temporal stability of consumers' responses. When fractional factorial designs are used, alternative forms should be used on the test and retest portions to reduce the possibility of memory producing a high test-retest correlation. This is of particular importance when the time interval between test and retest is of short duration. Product moment correlations are usually used to measure the similarity of test and retest (e.g., Acito 1977, Green and Srinivasan 1978) though other measures, such as sums of squared differences, are also appropriate. Correlations may be computed among the independent part-worths or among the estimated utilities that are derived from the part-worths. Most studies have employed the latter approach (e.g., Acito 1977, Green et al. 1972, Jain et al. 1975). Acito (1979) argues that the relatively small number of observations available when comparing the test and retest part-worths will lead to unstable results. However, it should be noted that the higher degrees of freedom associated with utility values is partially illusory since these values are functionally related through the part-worths. A design with five attributes at two levels each, for example, produces at most five "independent" part-

worths (four when the unit of measurement is arbitrary). Furthermore, if the variance of the part-worths is small (i.e., the attributes are all of relatively equal importance), the test and retest part-worths could be quite close (in an absolute sense) yet yield a low correlation. The severity of this problem can be tested by either comparing the test-retest correlations to the within test variance of the part-worths or by using a different measure of test-retest association, such as the sum of squared differences.

Validity considerations necessarily depend on the use for which the measure of interest is intended. Conjoint analysis of preference judgments has been frequently proposed as a procedure for "predicting" the success of potential new products (Green and Srinivasan 1978, Pekelman and Sen 1979). As such, predictive validity with actual choice behavior as the criterion would be of central interest (Wittink and Montgomery 1979). For many non-commercial research studies this is economically infeasible and a less demanding measure of future choice behavior, such as a raffle where subjects identify the product they would select if they were to win could be used to provide a measure of predictive validity. Alternatively, the researcher can obtain a second measure of preference or intentions to provide a measure of convergent rather than predictive validity (Scott and Wright 1976, Cattin and Weinberger 1979).

A number of articles have dealt with the reliability and validity of conjoint analysis. A review of these efforts has recently been provided by McCullough and Best (1979). Much of this work has dealt with the reliability and validity of alternative algorithms and/or individual conjoint measurement approaches. The comparative reliability and validity of experimentally efficient designs have generally been ignored.

Methodology

A convenience sample of 52 undergraduate students in an introductory marketing class was used in this pilot study. The stimuli were written descriptions of inexpensive pocket cameras. Pocket cameras were selected because they represent a moderately complex product category of reasonably high interest to the respondent population. An article on comparative ratings of pocket cameras in a consumer magazine was consulted in choosing the five camera attributes that were used in the study: (1) fixed versus zone focus, (2) maximum aperture--large versus small; (3) automatic exposure--"yes" or "no;" (4) electronic flash--"yes" or "no;" and (5) built-in telephoto--"yes" or "no." Many commercial applications consider a larger number of attributes and attribute levels. However, it was necessary to limit the number of attributes used in this study to insure that the less efficient procedures would be feasible for most of the stimulus sets considered. Five "measurement scales for the dependent variables," two nonmetric methods (rank order and pair-comparisons) and three metric methods (graded pair-comparisons, direct subjective estimates of the dollar values, and rating scales) were used.

The rank order instructions were patterned after those presented in Green and Wind (1973, pp. 261-62). Respondents in the pair-comparisons condition were merely asked to check the camera description (within each pair) they most preferred. For graded paired comparisons, respondents were also asked to specify (for each pair of stimuli) how much more they would be willing to pay for the camera of their choice. The "direct subjective estimates of the dollar value" of the cameras represented the "maximum the respondent was willing to pay (in dollars) for each of the cameras." Finally, the rating scale procedure involved evaluating each camera on an 11-point scale from "least prefer" to "most prefer."

Implementation of the above would suggest 15 approaches (3 levels of fractionation and 5 methods) to collecting preference data. However, two of these (combinations of the full factorial stimulus sets with pair-comparisons and graded pair-comparisons) were not deemed feasible because of the large number of responses required ($C(32,2) = 496$). These treatment combinations were thus excluded. Even with the 1/2 fractional factorial, 120 ($C(16,2)$) responses are required for both the graded and "ungraded" pair-comparisons. Because of the subject's added task of evaluating the degree of differential preference under the graded pair-comparisons method, 120 responses were considered infeasible for this method. Hence a partially balanced incomplete block design was utilized with the 1/2 fractional factorial for graded pair-comparisons. This was possible because of the interval scale nature of the data.

The 13 methodological treatments were randomly assigned to the 52 subjects (4 per cell). Each respondent was given a short statement explaining each of the five product attributes to eliminate potential problems with unfamiliarity concerning the terminology used. Following the completion of their respective preference judgment tasks each subject was asked to give a direct subjective estimate of the differential dollar value of the "higher" (versus the "lower") of the two levels for each attribute. This provided a second measure of each subject's part-worths. As such, it presented an opportunity to investigate the convergent validity of the part-worths. It also permitted the comparison of the test-retest reliability of the conjoint analysis procedures with that of a rather simple and less costly approach to analyzing consumer preferences.

After approximately one month the subjects repeated their respective tasks. The alternative form approach (i.e., a different set of stimulus descriptions) was not possible for the full factorial case and was not used for the fractional factorials to maintain consistency across treatments.

Analysis

Because of the small cell sizes and the relatively high variance within cells no statistical tests were performed. The results must be considered as only suggestive of the comparative viability of the alternative approaches. We will rely on examining trends in the data in interpreting the results.

Since most of the attributes were naturally dichotomous and all were represented at two levels, the "part-worth function" model was considered the most appropriate for representing the respondents' preferences. The more highly fractionated designs did not allow for the estimation of interaction effects and none were considered for any of the results presented in order to maintain comparability across data collection conditions. However, the existence of interaction effects was investigated for some of the metric models where possible (e.g., the rating scale used with a full factorial) and few (and in most cases none) were found to be significant for any of the subjects.

Monotone analysis of variance (Kruskal 1965) was used to estimate the part-worths for the nonmetric method (i.e., rank order and pair-comparisons). To convert the pair-comparisons into the predominant rank orders required by the MONANOVA algorithm, TRICON (Carmone et al. 1968) was used. Other procedures for the conversion of pair-comparisons to predominant rank orders are possible (Carroll 1972) but use of alternative conversion methods makes little difference.

Least squares estimators were used to derive the part-worths for the metric methods (i.e., graded pair-

comparisons, direct subjective estimates and rating scales). For the graded pair-comparisons data, each observation was classified by the differences in the first order parameters of the paired stimuli (see Dykstra 1958 for a discussion of the analysis of graded pair-comparisons data).

Pre-Analysis of the Data

The resultant part-worths were examined for a priori directionality and pareto optimality violations. In this study there appeared to be strong rationale for assuming that one level of each attribute should be preferred (e.g., a large aperture should be preferred to a small aperture). For other products and/or attributes such an assumption may not be warranted. However, whether or not this assumption can be made in a particular situation should not affect the comparative reliabilities of the various conjoint measurement approaches. Six subjects had multiple directionality violations on at least one of the "tests." The number of possible pareto optimality violations was also calculated for each task. Seven subjects were eliminated from the study because their data exhibited 15% or more of all possible pareto optimality violations. Five of these subjects were among the six having multiple directionality violations.

Test-Retest Reliability

Because of the small cell sizes and substantial within cell variance, the test-retest reliability results must be viewed as primarily suggestive rather than conclusive. Two measures of reliability were considered: (1) within subjects test-retest correlations of part-worths; and (2) within subjects test-retest correlations of the estimated utilities. Both the average adjusted (for sample size) and unadjusted correlations are presented in Table 1.

The average utility correlations were uniformly high (ranging from .799 to .988 "unadjusted" and from .791 to .988 "adjusted") while the average part-worth correlations varied widely across the data collection conditions (ranging from .074 to .924 "unadjusted" and from .000 to .879 "adjusted"). Similar results for utility correlations have been reported in the literature (Acito 1977, Green et al. 1972, McCullough and Best 1979). Over all conditions, the average stimulus value (utility) correlation exceeded the average for part-worths by more than .4 (for both "adjusted" and "unadjusted"). A relatively high test-retest stimulus value correlation does not necessarily suggest a substantial test-retest part-worth correlation. For example, the average adjusted test-retest stimulus value correlation for the subjective estimate (1/2 fractional factorial) condition was .885, while the corresponding average part-worth correlation was .000. The correlation between the two sets of reliability measures (stimulus value and part-worth) was only .479 (for the adjusted test-retest correlations).

Adjusting for sample size would seem particularly important for part-worth correlations since they typically involve small "sample" (number of part-worths) and the unadjusted correlations tend to be low to moderate. This results in a significant attenuation of the correlations. In the present study, the overall average part-worth correlation was reduced by .105 (from .490 to .385) compared with only .004 (.895 to .891) for the stimulus values.

The test-retest part-worth correlations would seem to be considerably more relevant than the stimulus value correlations in evaluating the reliability of conjoint analysis procedures. The primary objective of these procedures is to provide reasonably accurate estimates of these parameters. The utility value correlations may be somewhat inflated in the sense that merely observing pareto optimality in making the preference judgments can produce fairly substantial values on this measure. Also,

TABLE 1
ESTIMATED PART-WORTHS AND UTILITIES:
AVERAGE TEST-RETEST CORRELATIONS

Experimental Condition	Number of Responses	Unadjusted correlations		Adjusted correlations	
		Part-worths	Utilities	Part-worths	Utilities
Rank order					
Full (n=4)	32	.824	.912	.711	.909
Half (n=3)	16	.276	.881	.232	.877
Quarter (n=4)	8	.200	.799	.109	.791
Mean		.488	.862	.361	.857
Pair-comparisons					
Half (n=3)	120	.839	.988	.760	.988
Quarter (n=4)	28	.650	.821	.506	.814
Mean		.731	.892	.615	.889
Graded pair-comparisons					
Half-PBIB (n=4)	48	.924	.973	.897	.972
Quarter (n=3)	28	.532	.891	.270	.887
Mean		.756	.938	.628	.936
Subjective estimate					
Full (n=4)	32	.286	.919	.236	.916
Half (n=4)	16	.074	.890	.000	.885
Quarter (n=4)	8	.577	.918	.431	.915
Mean		.312	.909	.222	.905
Rating scale					
Full (n=3)	32	.657	.899	.450	.866
Half (n=2)	16	.092	.920	.000	.917
Quarter (n=3)	8	.268	.836	.204	.829
Mean		.370	.881	.245	.876
Average across methods (weighted by number of subjects)		.490	.895	.385	.891
Pearson correlation between part-worth and stimulus value test-retest reliabilities (n=45)		.468		.479	

the fact that the utility values are linear functions of the part-worths will tend to make these correlations high. Much the same argument could be made concerning rank order correlations between respondents' test and retest preference judgments which serve as the input for conjoint analysis.

Earlier it was argued that if the variance of an individual's part-worths was small, one might expect his test-retest reliability to be low. If the argument holds in this situation, one would expect that the correlation of the individual part-worth reliability measures with the "within subject" variances of the part-worths would be positive. This correlation was calculated across all subjects in two ways, using the average of the test and retest variances of the part-worths and using the minimum of the test and retest part-worth variances. The respective correlations were 0.10 and 0.11. While the above argument is intuitively plausible, it did not find substantial support in this study.

In comparing the results for the alternative approaches one notices that only small differences exist with respect to average test-retest stimulus value correlations, suggesting the lack of sensitivity of this measure. However,

the part-worth correlations display substantial variability. With the exception of the subjective estimate method, the less fractionated designs (full and half factorials) tend to produce more reliable part-worth estimates. Apparently, in this particular research context, the control over random error provided by the larger effective "sample sizes" of these designs more than offset any tendency for the increased effort demanded from the respondent to produce less stable judgments concerning individual stimuli.

Those methods involving "comparative" judgments (i.e., pair-comparisons, graded pair-comparisons, and rank order) tended to perform better than those requiring "individual" judgments (i.e., subjective estimate and rating scale). Graded pair-comparisons had the highest overall average test-retest part-worth correlation (.628) followed by paired comparisons (.615) and rank order (.315). Furthermore, the highest average correlation (.897) was obtained for the graded pair-comparisons (1/2 fractional factorial--incomplete block) design. Perhaps this method is more attractive, at least for situations involving a low to moderate number of attributes, than would be suggested from its limited use in previous marketing studies. However, in its present form, this approach may not be optimal for larger (sample size and number of attributes) commercial studies.

Convergent Validity

The direct dollar metric estimates of the part-worths were included in both the "test" and "retest" to provide a baseline for evaluating the test-retest reliabilities of the various conjoint analysis approaches and also for the purpose of examining convergent validity. These direct dollar metric estimates were obtained after the respondents provided their conjoint analysis data. Their reliability could be affected by fatigue or boredom created by the conjoint analysis tasks. However, this did not seem to be a significant problem. Except for the subjective estimate method, there was no systematic tendency for the less fractionated factorial designs (those requiring more judgments) to be associated with lower test-retest reliabilities for the direct dollar estimates of the part-worths (Table 2). Incidentally, the subjective estimate approach to conjoint analysis resulted in low test-retest reliabilities. It is not clear whether the respondents in this condition were unreliable per se or whether the difficulty of the subjective estimate task negatively affected these subjects' performances on the subsequent dollar metric task.

The direct dollar metric estimates of the part-worths demonstrated a higher average (over all conditions) test-retest reliability than the combined average obtained for the conjoint analysis approaches studied (.500 vs. .385) even though its reliability suffers from a lack of redundancy in the data collected. Furthermore, graded pair-comparisons which appears to share method variance with the direct dollar metric method was the only approach for which the test-retest reliability was higher for conjoint analysis than for the dollar metric given common subjects (.628 vs. .560). In this study at least, the less complex and time consuming task of providing direct dollar metric estimates of part-worths was superior (with respect to test-retest reliability) to most of the popular conjoint analysis approaches. While it could be argued that these direct estimates are more easily remembered, memory shouldn't be a significant factor in this study since the test and retest were separated by approximately one month and the subjects were unaware that a retest was to be conducted.

Convergent validity relates to the amount of agreement among maximally different methods of measuring the same construct. Our second method (the direct dollar metric estimate of the part-worths) like the conjoint analysis

TABLE 2
CONVERGENT VALIDITY: DERIVED PART-WORTHS
AND DOLLAR METRIC ESTIMATES^a

	Dollar metric: test-retest reliability	Convergent validity
Rank order		
Full (n=4)	.436	.516
Half (n=3)	.792	.425
Quarter (n=4)	.577	.413
Mean	.584	.453
Pair-comparisons		
Half (n=3)	.910	.777
Quarter (n=4)	.589	.266
Mean	.727	.436
Graded pair-comparisons		
Half-PRIB (n=4)	.632	.642
Quarter (n=3)	.463	.719
Mean	.560	.675
Subjective estimate		
Full (n=4)	.069	.436
Half (n=4)	.227	.109
Quarter (n=4)	.511	.278
Mean	.269	.274
Rating scale		
Full (n=3)	.784	.485
Half (n=2)	.094	.179
Quarter (n=3)	.441	.142
Mean	.482	.279
Average across methods (weighted by sample size)	.500	.406

^aAll correlations have been adjusted for sample size.

procedures, relies on stated preferences, and as such, shares some method variance with these procedures. Hence, only a weak test of convergent validity was possible.

Each subject's dollar metric estimates were correlated with their conjoint analysis results for both the test and retest and the correlations averaged (Table 2). The results show that this measure of convergent validity tends to be higher for the less fractionated designs (i.e., full and half factorials). The methods involving comparative judgments (i.e., rank order, pair-comparisons, and graded pair-comparisons) fared better than those requiring "absolute" judgments. This was expected since these designs previously demonstrated higher test-retest reliability. Convergent validity (with the direct dollar metric estimates) for graded pair-comparisons was substantially higher than that obtained for any of the other methods. However, this is probably due in part to the method variance they share (i.e., the similarity of the measurement tasks).

Summary

In spite of the increasing popularity of conjoint analysis among marketers, only limited published evidence exists regarding its reliability and validity within the various research contexts in which it has been applied. While the superiority of conjoint analysis over "direct subjective estimates" of the "part-worths" appears to be widely

accepted, substantial empirical evidence for this position is lacking. Furthermore, those wishing to apply conjoint analysis in a specific decision context have little objective information on which to base their selection of a methodological approach from the many alternatives available.

If conjoint analysis is to be more than a passing fad researchers must establish its test-retest reliability and predictive validity over a broad range of consumer choice situations. Moreover, it needs to be demonstrated that this set of techniques provides more accurate predictions (forecasts) than those obtainable from merely asking respondents to directly report their "part-worths." It is not sufficient to merely compare conjoint analysis with naive models which utilize no information concerning the relative importance of the various attributes. Finally, there is a need for empirical studies focusing on the comparative viability of the alternative methodological approaches under a variety of research conditions.

The present study has centered on a systematic investigation of the effects of two aspects of conjoint analysis (the use of fractional factorials in stimulus set construction and "measurement scales for the dependent variable") on the test-retest reliability and convergent validity of the estimated part-worths. Since the cell sizes were small, only tentative conclusions concerning the relative effectiveness of the alternative methodological approaches are appropriate.

The results suggest that the less fractionated design tends to produce higher test-retest part-worth reliabilities when the number of dichotomous attributes is limited, although these designs will become less feasible as the number of attributes is increased. Those methods involving "comparative" judgments (i.e., pair-comparisons, graded pair-comparisons, and ranking) were more reliable than those requiring "individual" judgments (i.e., direct subjective estimates of the dollar value of the stimuli and rating scales). In the case of rating scales, this may well have been partially due to problems with anchoring effects.

When compared to the test-retest reliability of the direct subjective estimates (dollar metric) of the part-worths, the conjoint analysis results were disappointing. Only graded pair-comparisons appeared to be superior on this criterion. This approach also showed the highest convergent validity with the direct part-worth estimates.

In other research settings, particularly those with extensive stimulus sets, experimentally efficient designs may be more desirable if not necessary. While some may argue that our evaluation of conjoint analysis is unfair in that it considers only individual reliability and validity and does not consider the reliability of the "average" (aggregate) subject, it must be remembered that direct part-worth estimates will also improve when across subject averages are used. The reliability of averages will always improve as sample sizes increase. Furthermore, if individual reliability and validity is weak, the ability of conjoint analysis to explore individual differences and define market segments must be questioned.

References

- Acito, Franklin (1977), "An Investigation of Some Data Collection Issues in Conjoint Measurement," *Proceedings*, American Marketing Association, 82-5.
- Acito, Franklin (1979), "An Investigation of the Reliability of Conjoint Measurement for Various Orthogonal Designs," *Proceedings*, Southern Marketing Association, 175-78.

Carmone, Frank J., Green, Paul E., and Robertson, D. J. (1968), "TRICON--An IBM 360-165 Fortran IV Program for the Triangularization of Conjoint Data," Journal of Marketing Research, 5, 219-20.

Carroll, J. Douglas (1972), "Individual Differences and Multidimensional Scaling," in Multidimensional Scaling, Vol. 1, Roger N. Shepard, A. Kimball Romney and Sara B. Nerlove, (eds.), New York: Seminar Press.

Cattin, Philippe and Weinberger, Marc (1979), "Some Validity and Reliability Issues in the Measurement of Attribute Utilities," Advances in Consumer Research, Vol. VII, 780-3.

Cliff, Norman (1975), "Complete Orders from Incomplete Data: Interactive Ordering and Tailored Testing," Psychological Bulletin, 82, 289-302.

Coombs, Clyde H. (1966), A Theory of Data, New York: John Wiley & Sons.

Dykstra, Otto Jr. (1958), "Factorial Experimentation in Scheffé's Analysis of Variance for Paired Comparisons," American Statistical Association Journal, 53, 529-42.

Green, Paul E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," Journal of Consumer Research, 1, 61-8.

Green, Paul E., Carmone, Frank J., and Wind, Yoram (1972), "Subjective Evaluation Models and Conjoint Measurement," Behavioral Science, 17, 288-99.

Green, Paul E., Carroll, J. Douglas, and Carmone, Frank J. (1978), "Some New Types of Fractional Factorial Designs for Marketing Experiments," in Research in Marketing, Vol. 1, J. N. Sheth, (ed.), Greenwich Ct.: JAI Press, 99-122.

Green, Paul E. and Srinivasan, V. (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5, 138-42.

Green, Paul E. and Wind, Yoram (1973), Multiattribute Decisions in Marketing: A Measurement Approach, Hinsdale, Ill.: The Dryden Press.

Green, Paul E. and Wind, Yoram (1975), "New Ways to Measure Consumers' Judgments," Harvard Business Review, 53, 107-17.

Jain, Arun K., Acito, Franklin C., Malhotra, Naresh K., and Mahajan, Vijay (1978), "A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decomposition Multiattribute Preference Models," Working Paper, School of Management, State University of New York at Buffalo.

Kruskal, Joseph E. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," Journal of the Royal Statistical Society, Series B, 27, 251-63.

McCullough, James and Best, Roger (1979), "Conjoint Measurement: Temporal Stability and Structural Reliability," Journal of Marketing Research, 16, 26-31.

Parker, Barnett R. and Srinivasan, V. (1976), "A Consumer Preference Approach to the Planning of Rural Primary Health Care Facilities," Operations Research, 24, 991-1025.

Fekelman, Dov and Sen, Subrat K. (1979), "Improving Prediction in Conjoint Measurement," Journal of Marketing Research, 16, 211-20.

Scott, Jerome E. and Wright, Peter (1976), "Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations," Journal of Marketing Research, 13, 211-24.

Wittink, Dick R. and Montgomery, David B. (1979), "Predictive Validity of Trade-off Analysis for Alternative Segmentation Schemes," Proceedings: American Marketing Association, 69-73.

MULTIATTRIBUTE CHOICE MODELS: A CRITICAL REVIEW

Jeffrey E. Danes, Virginia Polytechnic Institute and State University
Philippe Cattin, University of Connecticut

Abstract

This paper reviews and discusses the work by Green and Desarbo, "Two Models For Representing Unrestricted Choice Data," Leigh, Mackay, and Summers, "On Alternative Methods For Conjoint Analysis," and Acito and Olshavsky, "Limits To Accuracy in Conjoint Analysis."

Two Models For Representing Unrestricted Choice Data

Green and Desarbo's paper, "Two Models For Representing Choice Data," is one more advancement to be included in the rapid growth of multidimensional scaling models. Today market researchers may choose between a metric or non-metric algorithm, scale a 2-way matrix, a 3-way matrix, may employ longitudinal scaling, and may choose from a variety of metric axioms: Euclidean, Minkowski, as well as Riemannian (cf. Lindman & Caelli 1978, Piezko 1975). Further, market researchers have numerous data collection methods at their disposal. The Green and Desarbo paper makes an advance in this latter category. They first demonstrate the use of Levine's model (1979) for scaling unconstrained choice data--data obtained from respondents who are neither instructed to choose a fixed number of alternatives nor is the set of possible alternatives explicit. The primary contribution of the Green and Desarbo paper is the extension of Levine's (1979) model to unconstrained choice data in which the stimulus set is prespecified. The outcome of their effort is a multidimensional space that yields: ideal points, stimulus points, and attribute points. Although it is traditional to map attributes as vectors, the model provided by Green and Desarbo maps attributes as points in a multidimensional space. Let us focus our attention upon this feature.

Utility of Attributes As Points

One criticism of the Green and Desarbo paper is that they do not explain the role that dimensions play in their model. If attributes are represented as points, what do the dimensions represent?

It is now well known that the dimensions obtained in multidimensional scaling do not necessarily correspond to attributes. This has been found by numerous researchers who have regressed attribute vectors into multidimensional spaces. It is not uncommon to find six or more attribute vectors in a two or three dimensional space (cf., Schmidt 1976). The dimensions of a multidimensional space do not necessarily correspond to attributes; thus, it may be more meaningful to view the obtained dimensions as an arbitrary reference system in which points of all kinds are plotted. A common example of an arbitrary reference system in which dimensions have no inherent meaning is the letter-number grid on a city map. These dimensions may be translated and rotated with no harm done to the distance between any two homes. The coordinates simply tell us where "Oak Street" is relative to "Elm Street."

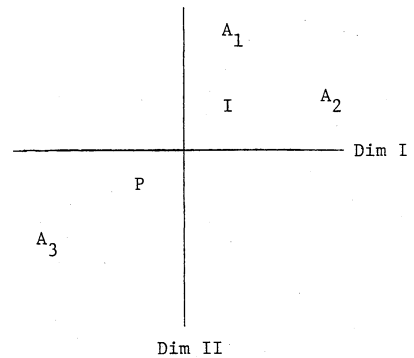
Green and Desarbo's treatment of the attribute as a point in space, however, gives new meaning and new potential applications of multidimensional scaling to market and consumer research. Let us now comment on one possible application of representing attributes as points.

Example Application: Advertising Strategy

The Green and Desarbo paper used association data for their example application. Hence, distance in the space is a function of association in the data--the highly associated points are close together. They also mapped ideal points. Without loss of generality, let us assume a homogeneous population with one ideal point, I; let it be mapped as follows (see Figure 1):

FIGURE 1

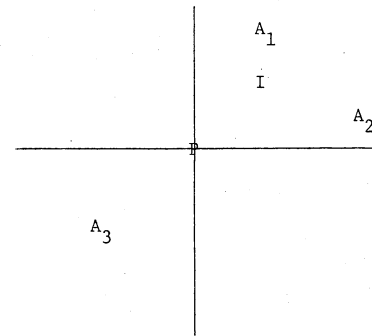
A Simple MDS Map With Attributes As Points



Let us denote the product as P, the ideal point as I, and three attributes as A_1 , A_2 and A_3 . Without changing the associations (or inter-distance relations), we may translate the dimensions of the space so that the product, P, has zero coordinates (see Figure 2).

FIGURE 2

Translated MDS Map With Attributes As Points



One marketing goal may be to "move" the product closer to the ideal point. How can this be accomplished?

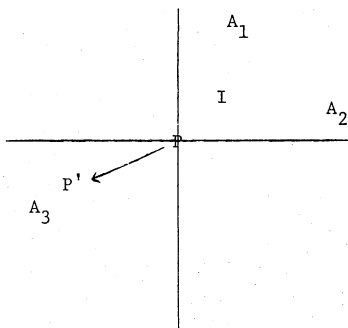
Communication researchers have long known that messages that repetitively associate two objects (e.g., Watergate and political crime) result in increased perceived similarity between the two objects. Likewise, the same appears to

be true for objects and attributes (Barnett, Serota, and Taylor, 1976). Marketers capitalize on this principle when they associate their product with intuitively desirable attributes.

Let us assume that a marketer, through communication, associates their product, P, with A_3 . Let us also assume that the communication campaign is relatively lengthy. What is the likely outcome? If the campaign successfully teaches the association the product, P, should "move" closer to A_3 . Let us denote the "new" location as P' in Figure 3. However, this

FIGURE 3

New Location of Product After Association of Product, P, With Attribute, A_3



is not consistent with the original goal--moving the product closer to the ideal point. A campaign that associates the product, P, with say, A_2 should be a much better strategy. But, the resulting direction of motion is slightly off target. However, if we assume that vector averaging is a viable cognitive principle, then the two-attribute message in which the product, P, is associated with, A_2 and A_3 , would be even better. This principle is diagrammed in Figure 4.

FIGURE 4

New Location of Product After Association of Product, P, With Attributes, A_1 and A_2

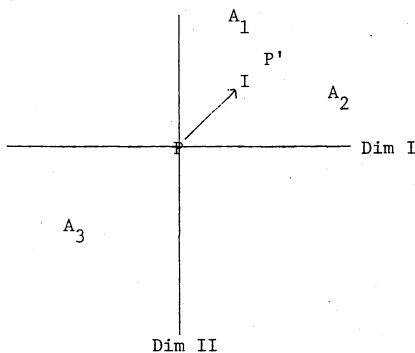


Figure 4 pictures the "final state" for the product given an advertising campaign that is successful in teaching the new associations, i.e., the product, P, is associated with the two attributes, A_1 and A_2 . The example we present here, of course, is over simplified. However, generalizations to spaces of larger dimensionality with more attributes is possible. Extended models and algorithms for using attributes as points (in multidimensional spaces) for the

measurement of communication and marketing processes may be found in Woelfel and Danes (1980).

Summary

Green and Desarbo's work is one more contribution to the rapidly growing arsenal of multidimensional scaling techniques. They provide a multidimensional scaling model derived from Levine's (1979) earlier model development. The primary contribution of their work is the extension of Levine's (1979) model to unconstrained choice data in which the stimulus set is prespecified. The outcome of their effort is a multidimensional space that yields: ideal points, stimulus points, and attribute points. The role of the dimension, which is usually treated as an attribute, is not made clear by Green and Desarbo. Our discussion, however, suggests that the dimensions in a multidimensional space may be usefully viewed as an arbitrary reference system, which by themselves, have no inherent meaning. Our discussion then moved to potential applications of the new model presented by Green and Desarbo; we discussed an advertising strategy application.

On Alternative Methods For Conjoint Analysis

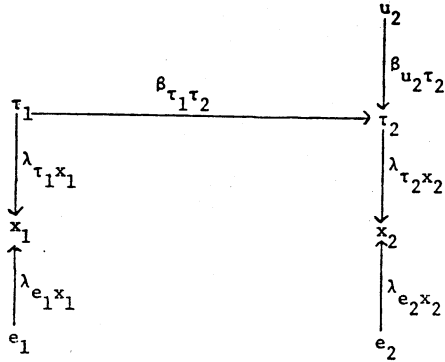
Leigh, Mackay, and Summers have presented an interesting study on alternative methods for conjoint analysis. The objective of their study was to evaluate full and fractional experimental and various methods of data collection: rank order, pair-comparisons, graded pair-comparisons, subjective estimates, and rating scales. The evaluative criteria were reliability and validity. The study concluded that the less fractionated designs were more reliable than the more fractionated designs; and that data collection methods involving comparative judgments (i.e., pair-comparisons, graded pair-comparisons, and ranking) were more reliable than profile stimulus sets. Based upon small sample sizes, however, the authors assert that the results must be tentively held. Hence, they agree that the findings are suggestive. The implication, of course, is that if the cell sample sizes had been larger, the results would be more meaningful estimates of reliability and validity. We show below, however, that this conclusion may be false.

Leigh, et al. use the test-retest correlation as a measure of temporal stability. They state that "Test-retest methods are a common way of measuring the temporal stability of Consumer's Responses (p. 318)." Their distinctions between temporal stability and structural reliability follow the work of McCullough and Best (1979). McCullough and Best (1979) have stated that "The stability of a measurement is related to the reproduction of measurement results at different points in time. This aspect of reliability generally has been inferred by the correlation between test-retest measurements over some specified time interval (p. 26-27)." Additionally, McCullough and Best (1976) have stated that structural reliability relates to systematic error. Let us first state that these definitions are not standard psychometric terms. Ordinarily when systematic changes are accounted for, the degree to which an instrument replicates its measurements is called the reliability of measurement. On the other hand, given measurement that contains no random error, the degree to which an instrument replicates its measurements is called the stability of measurement (Heise 1979, Wiley and Wiley 1971). Nonetheless, Leigh, et al. define temporal stability as reliability and structural reliability as stability; hence, one is not sure what they are really after. More importantly, Leigh, et al. used the test-retest correlation coefficient; it is shown below that one test-retest correlation does allow one to separate reliability from stability.

Separating Reliability From Stability

To demonstrate why a single test-retest correlation does not provide a good estimate of reliability or stability, we lean heavily upon the work of Heise (1971). His work stems from Coleman (1968), classical psychometrics (Lord and Novick 1968) and from path analysis. Let us begin with path diagram of standardized test-retest measurements given in Figure 5.

FIGURE 5



In this diagram, the coefficient $\beta_{\tau_1 \tau_2}$ is the standardized path coefficient representing the degree to which the true score at time 1, τ_1 , "causes" the true score, τ_2 , at time 2. In this context, $\beta_{\tau_1 \tau_2}$ is identical to the correlation between true scores, i.e., $r_{\tau_1 \tau_2}$. The coefficient $\beta_{\tau_1 \tau_2}$ (or $r_{\tau_1 \tau_2}$) is an estimate of stability, i.e., the degree to which test-retest measurements covary when measurement error is removed. The coefficients $\lambda_{\tau_1 x_1}$ and $\lambda_{\tau_2 x_2}$ are path coefficients reflecting the degree to which the true scores "cause" the observed measurements, x_1 and x_2 . These coefficients are identical to the correlation of the observed measurements with the true scores, i.e., $r_{x_1 \tau_1}$ and $r_{x_2 \tau_2}$. The square of $\lambda_{\tau_1 x_1}$ (or $r_{x_1 \tau_1}^2$) is defined as the reliability of measurement for x_1 (Lord and Novick 1968). The coefficients $\lambda_{e_1 x_1}$ and $\lambda_{e_2 x_2}$ are coefficients reflecting the degree to which random measurement errors "cause" the observed measurements. Since, we assume standardized variables, the variance of x_1 and x_2 is equal to one; hence,

$$\lambda_{e_1 x_1} = \sqrt{1 - \lambda_{\tau_1 x_1}^2}, \tag{1}$$

$$\lambda_{e_2 x_2} = \sqrt{1 - \lambda_{\tau_2 x_2}^2}.$$

The coefficient $\beta_{u_2 \tau_2}$ represents systematic changes produced in τ_2 ; it is given as:

$$\beta_{u_2 \tau_2} = \sqrt{1 - \beta_{\tau_1 \tau_2}^2}. \tag{2}$$

For the above path model we assume that measurement errors are mutually uncorrelated, and that measurement errors correlate only with the respective observed measurements. Following the rules of path analysis the test-retest correlation, $r_{x_1 x_2}$, is defined as:

$$r_{x_1 x_2} = \lambda_{\tau_1 x_1} \beta_{\tau_1 \tau_2} \lambda_{\tau_2 x_2}. \tag{3}$$

Hence, with three unknowns in one equation, we can easily see that test-retest correlation does not enable one to separate reliability ($\lambda_{\tau_1 x_1}^2$ or $\lambda_{\tau_2 x_2}^2$) from stability,

$\beta_{\tau_1 \tau_2}$. Even if we assume that the reliability of the instrument is constant over the two time periods,

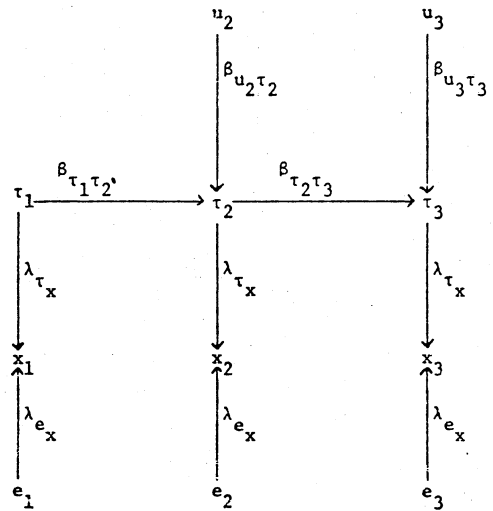
$$\lambda_{\tau_1 x_1}^2 = \lambda_{\tau_2 x_2}^2 = \lambda_{\tau x}^2, \tag{4}$$

we still have an unsatisfactory situation:

$$r_{x_1 x_2} = \lambda_{\tau x}^2 \beta_{\tau_1 \tau_2}. \tag{5}$$

Following Coleman's (1968) lead, Heise (1971) demonstrated that it takes at least three time periods to separate reliability from stability. Below we provide a path diagram for the test-retest paradigm--assuming that the reliability of the instrument is constant (see Figure 6).

FIGURE 6



If the assumptions stated above are preserved, and if it is assumed that the disturbances, u_2 and u_3 , are uncorrelated, we may use the rules of path analysis to write:

$$r_{x_1 x_2} = \lambda_{\tau x}^2 \beta_{\tau_1 \tau_2}$$

$$r_{x_2 x_3} = \lambda_{\tau x}^2 \beta_{\tau_2 \tau_3} \tag{6}$$

$$r_{x_1 x_3} = \lambda_{\tau x}^2 \beta_{\tau_1 \tau_2} \beta_{\tau_2 \tau_3}.$$

The first two equations may be written as:

$$\beta_{\tau_1\tau_2} = \frac{r_{x_1x_2}}{\lambda_{\tau_x}^2} \quad (7)$$

$$\beta_{\tau_2\tau_3} = \frac{r_{x_2x_3}}{\lambda_{\tau_x}^2}$$

Making the appropriate substitutions into the bottom of equation 6 and solving for $\lambda_{\tau_x}^2$ yields:

$$\lambda_{\tau_x}^2 = \frac{r_{x_1x_2} r_{x_2x_3}}{r_{x_1x_3}} \quad (8)$$

Once the reliability of measurement, $\lambda_{\tau_x}^2$, is found, the observed correlations may be corrected for attenuation. These corrected correlations, $\beta_{\tau_1\tau_2}$, $\beta_{\tau_2\tau_3}$, and $\beta_{\tau_1\tau_3}$, are the desired stability coefficients.

If the assumption that the reliability of three measurements is constant over time seems unrealistic, Wiley and Wiley (1971) present a path analytic model in which this assumption is relaxed.

Leigh, *et al.* used the test-retest correlation as a measure of "temporal stability." Our discussion, however, demonstrated that the test-retest correlation confounds stability with reliability. Furthermore, one test-retest correlation does not permit one to separate stability from reliability. We show that it takes at least three measurements (i.e., test-test-retest) for the estimation of stability.

Additional Comments

One interesting result obtained by Leigh, Mackay and Summers is that the average test-retest correlation of the direct dollar metric estimates of the part worths is greater than the average test-retest correlation of the part worths derived by conjoint analysis: .500 (Table 2) vs .385 (Table 1). The difference may not be significant because of the relatively small sample size.¹ In any event, one may wonder whether the direct part worth estimates have more predictive validity than those derived by conjoint analysis. Leigh *et al.* could have used the part worth estimates obtained (directly and by conjoint analysis) the first time to predict the rankings or ratings of the stimuli evaluated the second time. To be able to predict with the direct part worth estimates it is necessary to also know which level is preferred for each attribute and respondent. This can be obtained with one direct question for each attribute. Leigh *et al.* (apparently) did not have this information and thus could not estimate the predictive validity of the direct part worth estimates.

In a study that involved 41 respondents Cattin and Weinberger (1980, Tables 3 and 7, p. 782-783) found the predictive validity of direct part worth estimates to be very close to the predictive validity of part worths

¹In this case, the t-value would have to be taken with a grain of salt. As indicated by the authors themselves in their reply to a referee, a t-test is "strictly speaking not appropriate." This is because different data collection methods were used to obtain the part worth estimates derived by conjoint analysis. Hence, the variances of the test-retest correlations may vary across methods.

derived by conjoint analysis. Of course, direct part worth estimates can have biases. To reduce the biases it is necessary to tell the respondents which levels they should refer to and to think that the products are always exactly the same (on all attributes) except on the attribute they are evaluating. Even so, it has been found in several studies that respondents tend to underestimate the importance of important attributes and to overestimate the importance of lesser important attributes: see, for instance, Scott and Wright (1976, p. 214) and Cattin and Weinberger (1980, Table 6). However, Green (1979) has shown by simulation that the market share predictions obtained with such distorted part worth values are quite close to the market share predictions obtained with "true" non-distorted part worth values. Hence, the effect of this bias on market simulation results is relatively small. Actually, direct part worth estimates are used commercially: e.g., Aaker and Day (1980, p. 206).

If direct part worth estimates have relatively good predictive validity, combining them (somehow) with conjoint data might improve predictive validity. Bayesian regression procedures derived from Stein (1960) can be used to combine the two sets of data. Cattin and Danes (1981) have shown analytically and empirically that some improvement in predictive validity can thus be achieved.

Limits To Accuracy In Conjoint Analysis

Acito and Olshavsky's hypothesis is that the predictive validity obtained with conjoint data (using the full profile approach) in which each attribute is defined with two levels is superior to the predictive validity obtained if each attribute is defined with three levels. Acito and Olshavsky tested this hypothesis in the case where the number of stimuli evaluated by respondents is the same in both the two and three levels cases and where part worth attribute utilities (rather than utility functions) are estimated (i.e., one more parameter is estimated for each attribute in the three levels case compared to the two levels case).

The arguments in favor of the hypothesis, as stated by Acito and Olshavsky, are the following. First, the parameter estimates obtained in the three levels case are less reliable because there are less degrees of freedom (due to the fact that there is one more parameter estimated for each attribute). Second, the respondents are given more information in the three levels case (i.e., stimuli are defined on three levels instead of two) which produces more confusion, carelessness, and thus more noise in the data. It should be pointed out that the hypothesis applies primarily to the full profile approach (whereby the stimuli evaluated by the respondents are defined on all attributes). If one compares, for instance, the predictive validity obtained with ten (3 x 3) tradeoff matrices to the predictive validity obtained with ten (2 x 2) tradeoff matrices, the hypothesis is likely not to hold.

Acito and Olshavsky tested their hypothesis using twenty MBA students, ten of which were assigned to each of three levels and two levels designs. In both cases, the attribute utilities were estimated using MONANOVA on the rankings of 23 stimuli.² The predictions obtained on validation

²The 23 stimuli were taken from orthogonal design plans of 25 assemblies for the three level case and 24 for the two level case. However, the authors deleted two assemblies from the three level design and one assembly from the two-level design. Hence, the designs were not orthogonal.

data show that Acito and Olshavsky's hypothesis seems to be correct. Some of the results are significant even though the sample size was small.

One may wonder whether the superior predictive validity obtained with the two-levels design is due (a) to more degrees of freedom, or (b) to less noise in the data, or (c) to both. If (a) is the reason, one could improve the predictive validity in the three-levels case by increasing the number of stimuli (thus decreasing the number of degrees of freedom). If the amount of noise is the same, the predictive validities obtained with both designs would be the same if the "true" attribute utilities could be obtained. However, the estimated part worths have errors because they are based on a limited number of observations, and the error is expected to be greater with the three-levels design because there are less degrees of freedom. In other words, the resulting shrinkage in the correlation reflected in the predictive validity is expected to be greater in the three-levels case. Some insight in the difference in expected shrinkage (obtained with the two vs three levels designs) can be gained using equation 8 or 9 in (Cattin 1980, p. 410).

There is one way to find out whether there is more noise in the data in the three-levels case compared to the two-levels case.³ It involves assuming linear utility functions for each attribute in the three-levels case, thus estimating one parameter for each attribute (i.e., the same as in the two-levels case). In these conditions, the differences in predictive validity is due to the amount of noise and not to differences in the number of degrees of freedom. However, if linear utility functions are fitted in the three-levels case, the violations of the a priori attribute utilities would still be the same: i.e., higher in the three-levels case than in the two-levels case (Acito and Olshavsky 1980, Table 2). Hence, it appears that there might indeed be more noise in the three-levels case than in the two-levels case, and that the two-levels design would still produce higher predictive validities even if the number of stimuli (to be evaluated by respondents) is increased in the three-levels design.

Whether and by how much two-levels designs are found to have more predictive validity than three-levels designs depend upon several factors including the type of respondents and the type of attributes. If the respondents are more involved, there should be less noise and this could improve the results of a three-levels design compared to a two-levels design. However, whenever two-levels designs do produce data with less noise and increase (substantially enough) the predictive validity (when the full profile approach is used), it then becomes appropriate not to have more than two levels whenever possible. One way to proceed is to ask the respondents to position the levels of each attribute on (say) an 11-point scale where the least preferred and most preferred levels occupy the two ends (which can be asked of the respondents), and to then have the same respondents evaluate stimuli that are defined using only two levels for each attribute. See, for instance, Wind, Grashof and Goldhar (1978, p. 29-30). This procedure can be used with both discrete and continuous attributes. However, there

³In the three-levels case, Acito and Olshavsky used linear interpolation between part worths to estimate the utility of intermediate levels. Pekelman and Sen (1979) have shown that fitting a (quadratic) curve through the part worths improves predictive validity and is thus more appropriate than interpolation. However, this would have changed the results obtained by Acito and Olshavsky only in the case of the prediction of actual brands (and not in the case of the prediction of "holdout" ranks). It thus appears that the results obtained with the two-levels design would still have been superior to the results obtained with the three-levels designs.

might be some problems with continuous attributes especially if the attribute utility is expected to go through a minimum or a maximum (e.g., sugar content). In this case, it might still be best to have more than two levels in the conjoint design.

A final point regarding Acito and Olshavsky's study concerns the use of MONANOVA as the procedure for estimating part worths. MONANOVA is an iterative procedure that tries to minimize stress. However, it can end up in a local rather than global optimum. In a Monte Carlo simulation study, Cattin and Wittink (1976, Table 4, p. 20) found that an estimation procedure based on the maximum likelihood of the LOGIT model led to part worth estimates with lower stress values than MONANOVA more often than not. Hence, the simulation results obtained by Acito and Olshavsky to determine the stress values produced with random data are not very meaningful because there is no guarantee that MONANOVA reached the global optimum. With nonmetric data it is more appropriate to use LINMAP or a LOGIT approach, while regression can be used with metric data. For the tradeoffs between these methods, see, for instance, Green and Srinivasan (1978, p. 112-114).

References

- Aaker, D. A. and Day, G. S. (1980), Marketing Research, New York: John Wiley and Sons.
- Acito, F. and Olshavsky, W. (1981), "Limits to Accuracy in Conjoint Analysis" in Advances in Consumer Research, ed., Kent B. Monroe, Washington, D. C., Vol. 8, pp. 313-316.
- Barnett, G., Serota, L., and Taylor, J. (1976), "Campaign Communication and Attitude Change: A Multidimensional Analysis." Human Communication Research, 3, 277-44.
- Cattin, P. (1980), "Estimation of the Predictive Power of a Regression Model," Journal of Applied Psychology, 65 (August), 407-14.
- Cattin, P. and Wittink, D. R. (1976), "A Monte-Carlo Study of Metric and Nonmetric Estimation Methods for Multi-attribute Methods," Research Paper No. 341, Graduate School of Business, Stanford University.
- Cattin, P. and Weinberger, M. G. (1980), "Some Validity and Reliability Issues in the Measurement of Attribute Utilities," Advances in Consumer Research, ed., Jerry Olson, San Francisco, CA, Vol. 7, 780-3.
- Cattin, P. and Danes, J. E. (1981), "A Simple Bayesian Regression Procedure: Applications to the Estimation of Multiattribute Preference Models," Working paper, School of Business Administration, University of Connecticut.
- Coleman, J. S. (1968), "The Mathematical Study of Change," in Methodology in Social Research, eds., H. M. Blalock, Jr. and A. B. Blalock, New York: McGraw Hill, pp. 428-78.
- Green, P. E. (1979), "On the Insensitivity of Brand Choice Simulations to Attribute Importance Weights," Working paper, The Wharton School, University of Pennsylvania.
- Green, P. and Desarbo, W. (1981), "Two Models For Representing Unrestricted Choice Data," in Advances in Consumer Research, ed., Kent B. Monroe, Washington, D. C., Vol. 8, pp. 309-312.
- Green, P. E. and Srinivasan V. (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5 (September), 103-23.

Heise, D. R. (1971), "Separating Reliability and Stability in Test-Retest Correlation," in Causal Models in the Social Sciences, ed., H. M. Blalock, Jr., Chicago: Aldine-Atherton, pp. 348-63.

Leigh, T., Mackay, D., and Summers, J. (1981), "On Alternative Methods For Conjoint Analysis," in Advances in Consumer Research, ed., Kent B. Monroe, Washington, D. C., Vo. 8, pp. 317-322.

Levine, J. H. (1979), "Joint-Space Analysis of 'Pick-Any' Data: Analysis of Choices From An Unconstrained Set of Alternatives," Psychometrika, 44, 85-92.

Lindman, H. and Caelli (1978), "Constant Curvature Riemannian Scaling," Journal of Mathematical Psychology, 17, 89-109.

Lord, F. and Novick, M. (1968), Statistical Theories of Mental Test Scores, Reading, Mass.: Addison-Wesley.

McCullough, J. and Best, R. (1979), "Conjoint Measurement: Temporal Stability and Structural Reliability," Journal of Marketing Research, 16, 26-31.

Pekelman, D. and Sen, S. K. (1979), "Measurement and Estimation of Conjoint Utility Functions," Journal of Consumer Research, 5 (March), 263-71.

Piezko, H. (1975), "Multidimensional Scaling in Riemannian Space," Journal of Mathematical Psychology, 12, 449-77.

Schmidt, C. (1972), "Multidimensional Scaling Analysis of the Printed Media's Explanations of the Riots of the Summer of 1967," Journal of Personality and Social Psychology, 24, 59-67.

Scott, J. E. and Wright, P. (1976), "Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations," Journal of Marketing Research, 13 (August), 211-24.

Stein, C. (1960), "Multiple Regression," in Contribution to Probability and Statistics, I. Olkin et al., eds., Stanford, California: Stanford University Press.

Wiley, D. and Wiley, J. (1971), "The Estimation of Measurement Error in Panel Data," in Causal Models in the Social Sciences, ed., H. M. Blalock, Jr., Chicago: Aldine-Atherton, pp. 364-73.

Wind, Y., Grashof, J. F. and Goldhar, J. D. (1978), "Market-Based Guidelines for Design of Industrial Products," Journal of Marketing 42 (July), 27-37.

Woelfel, J. and Danes, J. (1980), "Multidimensional Scaling Models For Communication Research," in Multivariate Techniques For Communication Research, eds., Peter Monge and Joseph Cappella, Academic Press, in press.

INITIAL MARKET RESEARCH STEPS TOWARD
A MODEL OF THE MILITARY ENLISTMENT DECISION

Shel Feldman, Associates for Research in Behavior, Inc.
A. J. Martin, Office of the Secretary of Defense

Abstract

A large-scale tracking study of the propensity of young men and women to enlist in the National Guard or Reserves provided an opportunity to develop and test a regression-based model of a critical life decision. An Ordinary Least Squares (OLS) model comprising nine independent variables replicated in double cross-validations across randomly split subsamples in each of three independent samples of 1500 persons, in the baseline year. In each sample (non-veteran men, non-veteran women, veteran men), moreover, the model accounted for half the variance of the criterion, based on eight measures of military enlistment propensity. Follow-up interviews of 711 non-veteran respondents, a year later, demonstrated the predictive validity of the measure.

Problem

It often seems difficult to model important life decisions along the lines of problem-solving and consumer purchasing. Although choice of a career and choice of a mate are usually acknowledged to contain elements of rationality, for example, stress is often laid to chance factors and the unique occurrences that differentiate between outcomes. In part, this difficulty may reflect our need to assert uniqueness and free will in those areas most central to our self-concept. In part, it may also arise from a focus on the particular job or spouse chosen -- rather than on the class of jobs or of spouses -- and the consequent inability to obtain and analyze data on a large set of similar decisions. A large-scale study on the military enlistment propensity of young men and women offered the opportunity to develop and test a rational model of that important life decision.

Background

Successful manning of the All-Volunteer Force requires encouraging more than 650,000 young people to choose military service each year. The voluntary nature of the transaction with potential enlistees has focussed the Department of Defense on the need to consider the nature of its product -- military service --, its positioning, and its promotion efforts through advertising and recruiting, as well as on its pricing variable (enlistment standards). In Fiscal Year 1981, the budget for recruiting will reach \$970 million. Much market research will be aimed at understanding the perceptions of potential enlistees, and their decision process regarding military service, in order to carry out recruiting most effectively.

The data reported here are derived from a tracking study of awareness and attitudes toward enlistment into the Reserve Components (Associates for Research in Behavior, 1979). The National Guard and Reserves are part-time military units requiring a specified number of training periods each year, following initial basic training and specialty schools, and provide back-up capability for the Active (full-time) Forces. These components will require nearly 250,000 enlistments in Fiscal Year 1981, from among young men and women with no prior military service (NPS) and those who are veterans of the Active Forces.

The baseline data for this tracking study were collected in Fall, 1978, from a total of 4503 persons. One sample, of 1500, was composed of NPS males, between the ages of 17 1/2 and 26, who had completed no more than two years of college education. A second sample, of 1502, was composed of comparable NPS females. The third sample was composed of 1501 male veterans, separated from active service within the previous five years.

The data were collected by phone, in interviews of approximately 30 minutes duration, by trained professional interviewers. The NPS samples were obtained by random digit dialing within specified area codes and exchanges, chosen to form an interpenetrating block design weighted by accessibility to U.S. Army Reserve centers. The veterans were a stratified random sample of all men separated from the Active Forces in the designated time period and eligible for re-enlistment. Each sample was randomly divided into two equal subsamples to test the reliability of various results, including factor analyses and regressions.

Propensity to enlist in the Reserve Components will be reported in two ways. Each is a verbal measure of intention. The "standard" propensity measure indicates the respondent's highest expressed likelihood of serving in any of the six components (Army National Guard, Air Force Reserve, etc.), as indicated on a four-point scale. Thus, if a respondent says she will "definitely" serve in the Navy Reserve, but "probably not" in any of the others, she will be scored "definitely." And if a respondent says he will "probably" serve when asked about each of the components, he will be scored "probably." Responses to questions on each component are highly correlated, particularly for the NPS respondents, who tend not to differentiate between components. In the 1978 data, positive propensity ("definitely" or "probably" enlist responses to at least one component) was shown by 24.7 percent of NPS males, 12.9 percent of NPS females, and 21.0 percent of male veterans.

The "general" propensity measure was developed for use in the regression model, in order to remedy certain deficiencies of the "standard" measure. The "general" measure is a summative index of eight items reflecting military enlistment propensity under a number of conditions. These include the "standard" measure as well as responses to questions on propensity to enlist in the Active Forces, propensity to enlist in the Reserves under different lengths of enlistment contract, and propensity to enlist in the Active Forces or in the Reserves, were there a national service requirement. This summative index has an internal consistency index in the high .80's in each sample, much greater sensitivity than the standard measure, and a far less skewed response surface.

The independent variables on which "general" propensity was regressed were chosen to test a simple decision-making model of the enlistment process. The attempt was made to demonstrate that a single parsimonious model could account for a major share of the variance in each sample. Current work is directed at improving measures and the structural description of the model, given this initial demonstration.

The model posits that enlistment propensity is based on background attitudes, a benefit-cost valuation regarding the prospects of service, and factors peculiar to the

particular life situation of the respondent. Background attitudes include general acceptance or rejection of the military in American life, willingness or unwillingness to accept impositions on individual freedom in order to accomplish social goals (in the form of a national service requirement), and degree of need to be with others. Perceived benefits of service in a Reserve Component were measured with respect to opportunities to achieve life goals and opportunities to participate in a team. Perceived costs of service were measured with respect to loss of opportunities to be with family or friends, or to achieve further educational or occupational progress in the time allotted to service, and with respect to other negative outcomes, such as potential harassment by superiors. The respondent's current life situation was measured with respect to job commitment and perceived support for enlistment among current peers.

TABLE 1

"General" Propensity Regressed on Selected Variables

Variable	Male Veterans (N = 1248) ^{a/}	
	Est. Weight	St. Error
General predisposing attitudes		
Acceptance of the military ^{b/}	.976	.114
Acceptance of the imposition of a National Service Requirement ^{c/}	.333	.075
High need to be with other people ^{b/}	.997	.115
Anticipated gains from service		
Achievement of important individual life goals ^{b/}	.354	.031
Benefits of being a member of a team ^{b/}	1.380	.124
Anticipated costs of service		
Perceived opportunity costs ^{b/}	-.281	.122
The occurrence of negatively-valued situations ^{b/}	-.368	.114
Impacts of current life situations		
Commitment to present job situation ^{d/}	-.230	.066
Perceived social approval for enlistment ^{e/}	.870	.122
Constant term ^{f/}	20.429	
Multiple correlation	.678	
Coefficient of determination (R ²)	.460	

a/ Cases with missing data omitted.

b/ Factor scores; Mean = 0.

c/ 1 = strongly opposed; 5 = strongly in favor.

d/ Expects to look for job, or to change jobs: 1 = very likely; 5 = very unlikely.

e/ Others would be: 1 = very displeased; 5 = very pleased.

f/ 8 = least propensity to enlist; 34 = highest propensity.

Results

Perceived benefits and costs of service were measured with factorially-derived scales, as were background attitudes toward the military and regarding the need to be with others. The scales were derived from independent principal axis factoring in each subsample of each of the three samples. Comparability was tested using Schonemann's approach (1966; Schonemann & Carroll, 1970). Surviving factors were rotated to maximize interpretability. Resultant scales were constructed, and those with moderate to high internal consistency scores were retained for use in regression analyses of "general" propensity.

The results of the OLS regression analysis in the sample of male veterans are shown in Table 1. These results are clearly consonant with the model proposed. Each of the variables considered contributes significantly to the explanation of "general" propensity, and virtually half of the variance is explained, even without any correction for attenuation. (It should be noted, however, that none of the variables offers great leverage for altering "general" propensity, when evaluated at the point of means.)

The regression results for the male veterans sample were initially developed in each of the two random subsamples involved, and double cross-validated. Moreover, the same model applies to each of the NPS samples, with double cross-validation, accounting for 48.8 percent of the variance in "general" propensity among NPS males, and 47.5 percent among NPS females. With one exception, the variables involved are all significant in each of the other samples as well. (The relative importance of some are sample-dependent, however. For example, the estimated weight of "acceptance of the military" among NPS females is less than half as large as it is among NPS males or veteran males.)

The one exception to the general significance of estimated weights across samples occurs with respect to opportunity costs. The greater the perception of service in the Reserves as taking time from family, friends, job, or school, the lower the enlistment propensity of male veterans. However, this perception is not significantly related to "general" propensity among NPS males, and it is positively related to "general" propensity among NPS females. Presumably, these differences reflect differences in the life cycle position and the commitments of the members of the different samples, and should be explicitly incorporated in the interaction terms of more complete future structural models.

The regression equations were applied to the data from new independent samples in the Fall, 1979 wave of the tracking study. They accounted for between 20 and 30 percent of the variance in the several samples, despite changes in the wording of certain specific items in the predictors, changes in the composition of the "general" propensity measure, and changes in the social and political context (including the onset of the Iranian crisis).

Results of a Follow-up Study

The major implication of the regression analyses for this paper is that a single parsimonious OLS regression model can capture half the variance in a verbal measure of intention -- intention to undertake or not to undertake an important life change.

But is there any predictive validity to the verbal measure of intention? A follow-up study provided an opportunity to test this.

A year after the original interview, attempts were made to call every NPS respondent in the original sample who had indicated a definite or probable intention to join any of the Reserve Components in Fall, 1978. Attempts were also made to call a random sample of other NPS respondents, of equal size. In total, new interviews were obtained with 372 respondents who had expressed positive enlistment propensity (67 percent of the target group), and with 339 others (60 percent of that target group). These interviews, of about 15 minutes duration, obtained data for estimating the stability of various measures, but concentrated on obtaining reports of important behavior in the past year. The focus of behavioral reports included marriage, job changes, going to school, and military-related incidents, such as responding to military advertising, seeing a recruiter and actually applying to join the military.

Two analyses are of immediate interest. The proportion of respondents with positive enlistment propensity on the "standard" measure who engaged in each of several enlistment-related behaviors is shown in the left-hand column of Table 2. The proportion of others who engaged in each of those behaviors is shown in the center column. It is evident that the several behaviors form a continuum. (Guttman scalability is .85.) More important for present purposes, positive propensity respondents are clearly more likely to engage in each. In fact, the ratio of the proportion of positive propensity respondents to others increases at each step.

TABLE 2
Predictive Validity of Propensity Measures

Enlistment-Related Behavior	Proportion Reporting Designated Behavior, By "Standard" Propensity		r, 1978 "General" Propensity ^{a/}
	Positive (N = 372)	Negative (N = 339)	
Sent for literature or went to see recruiter	.323	.103	.29*
Attended an open house	.113	.032	.16*
Applied to a military service	.078	.012	.19*
Accepted for military service	.048	.006	.14*
Joined a military unit	.035	--	.10*

* $p < .05$

a/ Cases with missing data excluded.

The correlation of 1978 "general" propensity with 1979 reports of engaging in each of the relevant behaviors, over all respondents, is shown in the right-most column of Table 2. Each is statistically significant. The lowest coefficients are found in relation to behaviors that are (compared to applying) largely under the control of others (being accepted for service and then actually joining).

Discussion

Whatever its defects, the verbal measure of propensity to enlist in the military does predict whether or not a respondent is likely to take some steps toward enlistment. Not surprisingly, predictive validity is greater for more proximate behaviors, and for those more under the control of the respondent. This should be kept in mind in further measurement work and further modeling. It may also suggest some shifting of recruiting and advertising strategy away from stress on the final goal of enlistment. Given that a sequence of behaviors is clearly involved, it may be useful to shift more resources to encouraging more immediate behaviors, such as seeing a recruiter.

It cannot be said, at this time, that propensity is the outcome of the variables included in the OLS regression model. The last two decades of attitudinal research show that attitudinal changes, behavioral changes, and changes in the social context are related in a far more complex fashion than was formerly accepted. Ongoing work is focusing on more complex modeling of the enlistment process, including recognition of simultaneity effects. Nonetheless, it is clearly possible to develop a summary

index of enlistment propensity from measures of other attitudes, and to predict the likelihood of enlistment-related behavior from scores on that index. Further development of a rational model of this important life decision seems appropriate and useful.

References

Associates for Research in Behavior (1979), A Tracking Study Regarding Issues Related to Recruitment of Enlisted Personnel for the Reserve Components, Philadelphia, Pennsylvania.

Schonemann, P.H. (1966), "A Generalized Solution of the Orthogonal Procrustes Problem," Psychometrika, 31 (March), 1-10.

_____ and Carroll, R.M. (1970), "Fitting One Matrix to Another Under Choice of Central Dilation and a Rigid Motion," Psychometrika, 35 (June), 245-255.

MARKETING ACTION BASED ON CONSUMER DECISION PROCESSES:
THE CASE FOR MICRO-BEHAVIORAL SIMULATION

Richard A. Westwood, The Beaumont Organization, Ltd.
Harry S. Sunenshine, Heublein Inc.

Abstract

This case history describes how a number of sophisticated methodologies were brought to bear upon the strategic issues facing an established brand, and how results were eventually implemented by management. The research design was based on the SCIMITAR Research system. Techniques discussed include Conjoint Measurement, Correspondence Analysis, Brand Progress Analysis, Brand Impact Analysis, Hierarchical Cluster Analysis, and Micro-behavioral Simulation.

Introduction

Designing medium-term marketing strategy poses difficult challenges for researcher and manager alike. By its very nature, strategic design demands that several facets of the current and future market be synthesized within an action plan — including for example, immediate and more distant competitive frames; potential cannibalization between company brands; and the vulnerability of strategies to competitive response.

In attempting to deal with such broad ranges of issues, traditional research methods have generally adopted abstract views of the market. Psychographic segmentation provides a current, and popular, example of this type of abstraction. But, although abstract research results are often fascinating and insightful, their translation into action decisions tends to be vague and ambiguous. On the other hand, more specific and directly actionable methods run the risk of lacking generalizability, even of being trivial to the issues facing the strategic designer.

A number of technically advanced research methods, in particular micro-behavioral simulation, offer more powerful ways to approach strategic issues. The primary appeal of simulation lies in its ability to be sufficiently specific to relate directly to decision making yet, at the same time, sufficiently flexible to encompass a wide range of issues and contingencies.

This case history discusses how several sophisticated techniques, including simulation, were used to design strategy for a Heublein brand. The particular methodology described is Beaumont's SCIMITAR system, which comprises a suite of data gathering, analytic, and simulation approaches. Central to this system is the use of a model of the consumer brand choice process to integrate perceptual data with Conjoint Measurement, leading to several powerful analysis and simulation possibilities.

At the time of publication, early phases of the strategy have already been successfully implemented. Other aspects of the plan will be deployed in the future and, accordingly, the overall strategy remains confidential. In order to safeguard Heublein's security interests, while presenting data that have integrity for the research audience, this paper adopts the expedient of changing labels. The market under discussion, its attributes, and the key demographic groups, have been disguised. In the main, the actual results reported have not been changed, except in those few cases where they would permit the alert reader to infer which of Heublein's markets is the subject of this discussion.

Background

For a number of years, Heublein's brand had dominated the packaged goods category that was its primary source of business. Although two major national competitors existed, much of the remaining competition was limited to several small regional brands that were unable to dislodge the Heublein brand from its brand leader position in their markets.

More recently, the brand had begun to mature. Concurrently, its product category was becoming increasingly competitive. Where there had been one or two major brands, there were now four or five, some of whom posed serious competitive challenges.

Complicating matters, the competitive frame was by no means clearcut. Although Heublein's brand belonged to a product category that was clearly definable in manufacturer terms, it also appeared to compete with members of certain other categories. The fact that some of these adjacent product categories were growing more rapidly than Heublein's brand, or its immediate category as a whole, suggested that they might be limiting its opportunities.

Given the pressures upon its brand, management faced a two-fold challenge: to mount an effective defense against the recent, and vigorous, immediate competitors; and to create new growth at the expense of adjacent product categories. There was a need for research that would identify how Heublein might achieve these goals.

Behavioral Focus

It was clear that descriptive research methods would not be sufficient to address these objectives. Instead, it would be necessary to explain the process whereby consumers made their brand choice decisions, and to predict how these decisions might be changed by a variety of marketing actions.

The SCIMITAR research system was adopted for this research because it provided such a focus on the consumer decision process. The methodologies within the system were used to construct a three-component research program.

Research Program

The three components within the overall research program were as follows:

Preliminary Qualitative Research was conducted to provide the basis for questionnaire design.

An Immediate Category Study was the largest scale component within the research program, reflecting management's prioritization of the need to regain ground that had been lost to recent competitors, and to strengthen its brand against further attack.

A Total Market Perceptual Study was conducted in order to reveal the broader competitive perspective, and to be able to assess the more general implications of strategic actions.

Preliminary Qualitative Research

The preliminary phase of a strategic research program is the foundation upon which all subsequent results are built, and as such is crucial to project success.

As a first step, all pertinent past research was reviewed. This review was then augmented by qualitative Protocol Interviews with 16 individuals, including a variety of demographic and behavioral consumer types.

Protocol Interviews are a depth research methodology, designed to probe the brand choice process by having individual respondents perform a series of "choice tasks". For example, one task requires respondents to make a brand choice from seven blank packs, using the interviewer as a totally reliable information source. The interviewer uses each session to elicit the information used in brand choice decisions, the form in which information is cognitively structured, and the manner in which it is processed. The parameters appearing in individuals' choices are then generalized, and re-expressed in questionnaire form.

The semantic and conceptual structures within a strategic questionnaire act as a common language between the manufacturer and the market. To be effective, they must be capable of translating manufacturer concerns into consumer terms, and vice versa. Accordingly, consumer variables derived from the exploratory research were tested against management's own decision criteria. This resulted in some minor modifications to the questionnaires, including the addition of two attributes to the list of market descriptors.

Immediate Category Study

In this, the central component of the research program, 1,000 category users were interviewed in their own homes, using a 90 minute questionnaire. Based on the fact that differences between men and women were known to be important in the market, the sample was structured into two representative samples of 500 men and 500 women respectively.

The questionnaire was based on SCIMITAR methodology, which has been discussed in detail elsewhere (see Beazley and Westwood, 1976, Palmer and Westwood, 1976) and is therefore reviewed here only briefly.

In essence, the methodology reconstructs respondents' choice behavior, based on measurements of their requirements from the product field, and their perceptions of brands.

Conjoint Measurement is employed to measure requirements from the product field, using the approach discussed in Westwood (1973). The respondent states "trade-offs" between pairs of attributes, providing data that indicates the utility he derives from each attribute, and how he processes the attribute when making choices. The respondent also rates several brands across the same list of attributes. Conjoint and brand rating data are integrated at the individual respondent level, thereby explaining current choices, and simulating new choices under changing conditions.

22 attributes, derived from the Preliminary Qualitative Research, were investigated in the questionnaire. Additionally, extensive data were gathered regarding each respondent's purchase and usage behavior, demographics, and psychographics.

Role of the Heavy User

Analysis of usage patterns revealed heavy users to be extremely important to the product category, 75% of all consumption being accounted for by only 28% of consumers.

Furthermore, these heavy users tended to be demographically distinct and hence directly approachable through marketing action. For example, they were much more prevalent among men than among women. Based on this finding, much of the subsequent analysis focussed on the heavy user: in particular, how to attract him to Heublein's brand.

Brand Progress Analysis

The extent of the threat posed by recently launched national competitors was gauged using Brand Progress Analysis. This simple, but effective, analysis of brand strength pictures sales results in terms of a flow of consumers between various experiential states — from lack of awareness to awareness; from trial to acceptance; etc. In the current application, it was used to diagnose the limits to each major competitor's growth.

Among heavy category users, Heublein's brand compared to one of its newer national competitors as follows:

TABLE 1
BRAND PROGRESS ANALYSIS AMONG HEAVY USERS

	Unaware to Aware	Aware to Knowledge	Knowledge to Trial	Trial to Usage	Usage to Regular
Heublein brand	100%	100%	100%	84%	51%
Major competitor	88%	95%	95%	83%	61%

While Heublein's brand still benefitted from higher awareness levels, the competitor appeared to be able to match its trial and acceptance levels once awareness had been developed, and was superior in converting occasional users to regular users.

It appeared that this competitor had the potential eventually to gain brand leadership among heavy users, provided that it could increase its awareness level while simply maintaining its performance in other respects. Comparison of share patterns among consumers aware and unaware of the brand tended to confirm this finding:

TABLE 2
BRAND SHARES AMONG CONSUMERS AWARE AND UNAWARE OF COMPETITOR

	Unaware of competitor		Aware of competitor	
	Light Users	Heavy Users	Light Users	Heavy Users
Heublein brand	73%	81%	53%	36%
Major competitor	-	-	23%	47%
Other brands	27%	19%	24%	17%

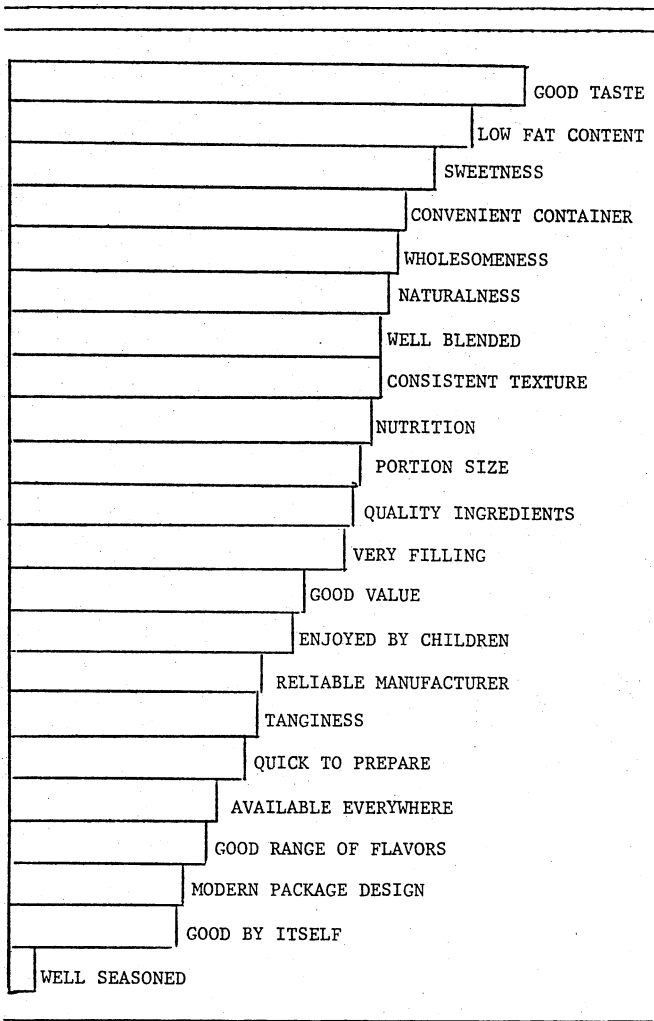
Brand Progress Analysis had revealed a major competitive threat, especially among the heavy users who had been shown to be of major importance to the category. Heavy users who became aware of the new competitor accorded it the majority of their usage. Unless countered, this brand could be expected to make significant inroads into the core of category volume, largely at Heublein's expense.

Management's prioritization of defensive strategy was thus strongly endorsed by research findings, and attention was now focussed on how such a strategy might be developed.

Consumer Wants

The logical starting point for strategic development is the examination of the factors that determine consumers' brand choices. In this study, the importances of choice criteria were measured in terms of utility scores derived from Conjoint Measurement. Figure 1 displays the overall utilities attached to various attributes by the average consumer:

FIGURE 1
IMPORTANCES OF ATTRIBUTES TO THE AVERAGE CONSUMER



Similar analyses were conducted within demographic and behavioral groups, and significant differences were found. For example, heavy users were shown to attach considerably more importance to children's enjoyment of the product — a factor that Heublein had not promoted actively, but that was a central aspect of a growing competitor's positioning. Attributes such as naturalness and wholesomeness, that had been the basis of recent Heublein advertising campaigns, were less important to heavy users.

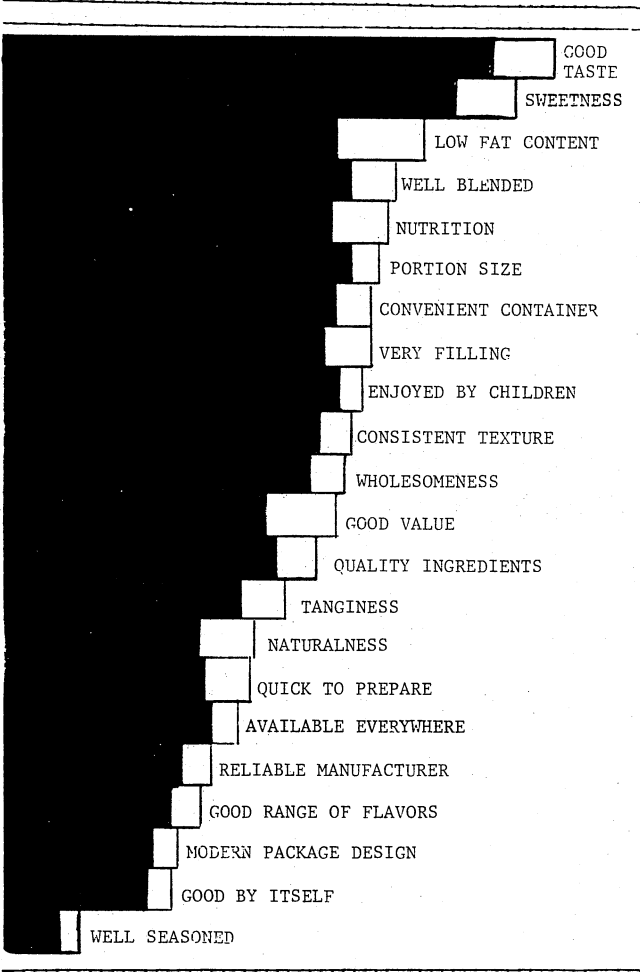
Satisfaction with Brands

While the analysis of consumer wants showed what was important to consumers, and provided first suggestions regarding how the market segmented, it did not reveal where leverage might exist against current competition. Accordingly,

Brand Impact Analysis was used to combine utility and branding data, and thereby to reveal the extent of "unfulfilled utility" remaining within each attribute.

Figure 2 provides an example of Brand Impact Analysis, referring specifically to Heavy Users' images of the brands they consume regularly. The overall length of each bar represents the total utility associated with a particular attribute; the shaded portion shows the extent to which brands currently satisfy consumers' wants; and the remaining unshaded area reveals the extent of unfulfilled utility within the attribute. Thus, an attribute displaying a sizeable unshaded area would comprise a vulnerability of current brands that might profitably be attacked.

FIGURE 2
BRAND IMPACT ANALYSIS: HEAVY CATEGORY USERS



Brand Impact Analysis revealed a high degree of product satisfaction among heavy users. It suggested that major share gains could not be expected to result from improvements on isolated individual attributes, since these would not provide a compelling point of superiority. Significant leverage might only be possible if images were shifted on several attributes in concert.

Other Brand Impact Analyses, among various brand user groups, showed that the users of major competitors perceived a number of disadvantages in Heublein's brand. These could, in the main, be attributed to known aspects of product formulation, and their rectification was made an immediate marketing priority.

Total Market Perceptual Study

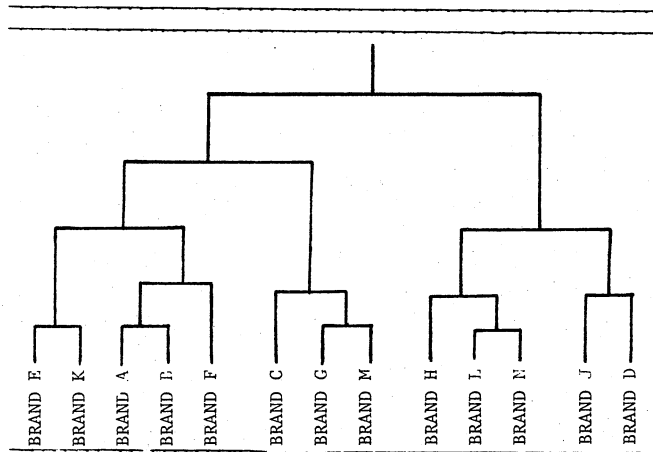
The Total Market Perceptual study was conducted to investigate the nature and extent of inter-category competition. In-home interviews were conducted among a representative sample of 350 respondents. The questionnaire required respondents to assemble 32 brands into groups of potentially substitutable items; to indicate which brand groups were suitable for use in each of a list of 21 situations; and to indicate which attributes were associated with each brand group. Volume usage data were also gathered for each situation.

The resultant data were subjected to a variety of multivariate and multidimensional scaling analyses, of which two are discussed below.

Market Structure

Johnson's Hierarchical Cluster Analysis was applied to the brand grouping data to determine the consumer's categorization of the market, providing a "family tree" of brand relationships as illustrated in Figure 3.

FIGURE 3
HIERARCHICAL CLUSTER ANALYSIS OF BRAND RELATIONSHIPS



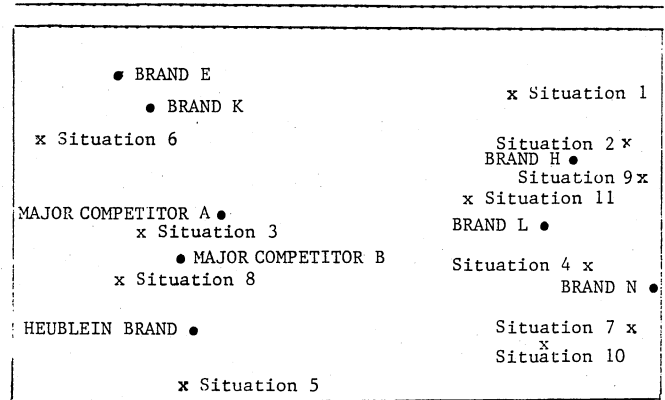
In general, the results suggested that consumers classified brands in much the same way as manufacturers. Thus, the primary competition to Heublein's brand was the immediate product category as the company had traditionally defined it. However, there were also some surprises elsewhere in the market structure. For example, two product categories that shared certain key ingredients, but that had previously been regarded as distinct, were shown to be highly substitutable, and virtually a single product category from the consumer standpoint.

Usage Patterns

The fact that brands and products belong to different perceived categories should not be taken to imply that they are non-competitive. Danish pastry and sausages belong to different categories, but compete for usage at breakfast. Taxis and bicycles are different products, but the consumer may choose between them during a transit strike. In general, categories that consumers regard as distinct may nonetheless compete within particular usage situations.

In the current study, patterns of cross-category competition were revealed by applying Correspondence Analysis to the 'brands by situations' data referred to above. This analytic method, analogous to both multidimensional scaling and principal components analysis, plotted brands and situations into a single space, as illustrated in Figure 4.

FIGURE 4
CORRESPONDENCE ANALYSIS OF USAGE COMPETITION



Like its major competitors A and B, Heublein's brand was found to be positioned towards a few high-volume situations (3, 8, and 5). However, within this positioning, its usage skewed towards the 'special purpose' Situation 5 rather than to the more mainstream Situations 3 and 8. Short term volume gains appeared to be possible if the brand could become more competitive in mainstream contexts.

Taking a longer term perspective, the Correspondence Analysis suggested that major gains might derive from a strategy that moved Heublein into more direct competition with Brands H, L, and N. These brands were regarded by both management and consumers as a separate product category. Their very heavy consumption derived from a wide range of usage situations which offered little volume usage individually, but a major opportunity cumulatively.

Approaches that might reposition the Heublein brand against new usage contexts were explored by superimposing attribute ratings data on the map shown in Figure 4. The resultant map suggested several approaches that could contribute to a more effective defensive positioning, including attention to those attributes where competitors' users perceived Heublein's brand to warrant improvement. A different set of considerations seemed relevant to expansion into new high volume opportunities.

Market Simulation

While the Immediate Category Study had revealed some defensive needs and some possible approaches to a heavy user strategy, it had also suggested that a successful strategy would probably require a number of mutually supportive elements. The Total Market Perceptual Study had reinforced these findings, and furthermore suggested that longer term expansion might imply different actions than shorter term defense.

In a traditional market study, these results would have concluded the project. Furthermore, the study would have been deemed a success in research terms, since it had hypothesized several promising directions for management consideration. However, in our view, data analysis — even sophisticated multivariate analysis — is not sufficient to address management concerns. As a case in point, the analyses in the current study had not provided a firm basis for evaluating alternative strategies, or for assembling them into coherent marketing programs. If it is to be truly actionable, strategic research must proceed one step beyond data analysis, and use other approaches to provide definitive statements regarding the corporate implications of each strategic possibility.

The methodology used for this study made such a next step possible. Through micro-behavioral simulation, it offered the ability to evaluate a virtual infinity of possible strategies. This simulation capability was an especially powerful feature of the research program, enabling broad directional hypotheses to be translated into specific action plans that could be acted upon directly.

The technical details of SCIMITAR's simulation approach are beyond the scope of the current paper, and are dealt with thoroughly elsewhere (see Beazley and Westwood, 1976, Palmer and Westwood, 1976). We shall therefore make only a few general explanatory remarks.

The SCIMITAR simulation model is a planning tool that tests strategic ideas through "what-if questions" addressed to a data base held within the computer. The user can create new brands, delete brands, and change brand perceptions inside the computer, which provides estimated results in terms of share potential, target group, and source of business.

Internally, the model operates by first "imitating" the ways in which individual respondents make brand choices, then aggregating individual responses into estimates of market potential. The simulation of individual respondent choice is based on the integration of brand image and utility data through a series of "decision rules" that replicate simple human thought processes:

1. Which brands is the respondent aware of?
2. Do any brands have completely unacceptable characteristics? (If so, reject them.)
3. Is any remaining brand unique in having an advantage on any important attribute? (If so, accept it as the predicted choice.)
4. Are any remaining brands inferior on the most important screening criterion; the second; and so on? (If so, screen them out of the set under consideration.)
5. What is the total utility associated with the attributes of each remaining brand? (Select brand with highest utility.)

Simulation runs were used to test the implications of a wide variety of image changes — first on single attributes, then on several attributes together. The results of some of the simulations based on single attributes are shown in Table 3.

TABLE 3
SIMULATION OF THE EFFECTS OF
UNIDIMENSIONAL BRAND IMAGE CHANGES

	Heublein brand	Major competitor
INITIAL SHARE	32%	17%
SINGLE ATTRIBUTE CHANGES		
A. Portion size	35%	16%
B. Consistent texture	35%	16%
C. Good value	35%	16%
D. Modern package design	36%	15%
E. Wholesomeness	36%	15%
F. Enjoyed by children	37%	15%
G. Good by itself	38%	16%
H. Low fat content	38%	16%

Several of the simulated changes replicated strategies that had previously been test marketed, and the results accorded well with the company's experience. As is customary in SCIMITAR projects, the simulation model had already been validated internally in terms of its ability to reconstruct the current brand choices of individuals interviewed, with satisfactory results. But the external validation against historical events, was, perhaps, a more

profound test of the model, and the management credibility it generated was of great assistance in implementing the model's recommendations concerning new strategic possibilities.

Simulation results provided clear confirmation of one of the major hypotheses previously emerging from data analysis, showing that Heublein's competitive position would not be improved dramatically by an isolated action on any single market dimension. In passing, we might note that this is by no means a typical result from the model, which often identifies individual attributes with significant competitive power (see, for example, Colvin, 1976, Gunter and Beazley, 1976).

Since no single attribute changes provided promising market potential, the simulation model was next used to test various attribute combinations that could be assembled within a brand strategy. Table 4 displays the results of some of these simulation runs.

TABLE 4
SIMULATION OF THE EFFECTS OF
MULTIDIMENSIONAL BRAND IMAGE CHANGES

	Heublein brand	Major competitor
INITIAL SHARE	32%	17%
MULTIPLE ATTRIBUTE CHANGES		
A + B + C + D	49%	8%
E + F + G + H	53%	9%
A + B + C + D + E + F + G + H	61%	6%

It was apparent that, although strategies based on single attributes were unlikely to be effective, certain combinations of attributes appeared to offer sizeable opportunities. This, of course, was a major finding in its own right, stressing the need for management to move away from individual tactical changes, and towards an integrated program of strategic activities.

More specifically, the model was able to point to the particular combinations of actions that would be expected to maximise market impact. For example, rectification of the defects that had been found in the brand's image among competitors' users was shown to offer good results, especially if incorporated with certain new benefits.

Simulation produced particular dramatic findings regarding overall business development, revealing that it was possible to merge shorter and longer term objectives within a single brand plan. Some of the attributes relevant to expansion beyond the current competitive frame were also found to be powerful within the more narrowly defined immediate category, provided that they were made part of a broader program.

Thus, through its use of market simulation, Heublein was able to develop a coherent multi-component strategy, that could be phased over time in a way that would build progressively more impact on the market. Early strategic phases would be concerned primarily with defensive considerations within the brand's immediate product category. These early phases, however, would pave the way for subsequent actions that would expand the brand's competitive horizons, and create the preconditions for new growth.

Conclusions

Although we have necessarily simplified and disguised this case history, we hope that we have communicated its considerable value to Heublein's strategy development.

For the future, we anticipate that Heublein's strategic research will continue to be firmly centered on consumer decision processes, and that attention will broaden beyond the issue of brand choice to embrace the full set of decisions that consumers make — including whether or not to purchase a category, how often to purchase it, and so on. This behavioral focus seems to us to be the natural and proper approach to strategy, since all marketing actions seek ultimately to influence consumer decisions in one respect or another. The recent advances in Information Processing Theory offer a framework, and to some extent a set of methodologies, that are suited to this purpose.

Furthermore, we expect that micro-behavioral simulation will continue to feature prominently in Heublein's attempts to understand how it can influence markets. Simulation promises to take research an important next step in its progress from description to explanation and eventually to prediction. It produces results in a form that is definitive, rather than vague and ambiguous, a form, furthermore, that is readily understood and implemented by management. Most importantly of all, it enables the strategy designer to move beyond the bounds of the problem at hand to a wider exploration of the multiple issues facing a brand in the real world.

References

Beazley, D., and Westwood, R.A. (1976), "Modelling Choice Behavior," Proceedings of the Market Research Society Annual Conference, 79-92.

Colvin, M. (1976), "Market Modelling at Ford," Paper presented at the Market Research Society Seminar on Market Modelling, Banbury, Oxon, UK. (Copies available from the authors.)

Gunter, Pieter, and Beazley, David (1976), "An Application of Micro-Simulation Modelling to the Marketing of a Consumer Durable," Proceedings of the ESOMAR Annual Conference, Special Groups, 629-649.

Palmer, J.B., and Westwood, R.A. (1976), "Designing Advertising Strategy - the New Technology," Proceedings of the Advertising Research Foundation Annual Conference, 51-81.

Westwood, R.A. (1973), "The Use of Models to Determine the Importances of Beliefs," Proceedings of the ESOMAR Seminar on Developments in Consumer Psychology, 165-189.

A CONSUMER BASED APPROACH FOR ESTABLISHING PRIORITIES IN
CONSUMER INFORMATION PROGRAMS: IMPLICATIONS FOR PUBLIC POLICY¹

Rohit Deshpande, University of Texas at Austin
S. Krishnan, Pennsylvania State University

Abstract

Although there exists substantial research on the style and format of consumer information programs, little attention has been devoted to the critical issue of whether consumers need new information at all. A conceptual approach to systematically assess this information need is described along with an empirical operationalization of the approach using data gathered from a sample of elderly consumers. Implications of this approach are discussed for public policy makers with suggestions for further research in the area.

Introduction

The past five years have seen an increasing amount of research being devoted to the public policy implications of consumer decision making. Paralleling the growth of this work has been a woeful disregard of the usefulness of the research in the actual policy making process. There have been some exceptions to this rule (Permut 1979), but consumer researchers have largely used a product orientation in their work, devoting themselves to tasks such as experimenting with different information presentation formats without undue emphasis on whether consumers desire the information at all. The result has been a subscription by both policy makers and marketing managers to a "more information is better" philosophy which has produced its obvious consequences discussed in the recent research controversy over information overload (Jacoby, Speller, and Kohn 1974, Russo 1974, Scammon 1977).

The dilemma has been recognized for some time now (Thorelli, Becker, and Engledow 1975); but, as McEwen (1978) notes, the suggested solutions stress what marketers or policy makers can do rather than what consumers need. The failure to use a marketing concept has produced information programs looking for people to use them. As Capon and Lutz (1979, p. 59) indicate: "Heretofore, policy makers have introduced consumer information programs without a clear idea of whether the particular information in the programs was desired by consumers."

The problem becomes even more serious when one considers the current trend toward reduced governmental involvement in regulatory and information providing capacities. Recent criticism of the FTC can be traced to feelings (within and outside government) that regulatory agencies are involving themselves in areas where no intervention was required. Clearly, there is a need to assess consumer desire for additional information or other support before programs are mounted to achieve this purpose.

Further, public agencies are similar to private organizations in that limited resources of money and manpower exist. The allocation of these scarce resources among information programs implies the establishment of a set of priorities. To use the Capon and Lutz (1979) analogy,

¹The authors wish to thank Dr. Gerald Zaltman and the University of Pittsburgh for providing the data for this study, and D. Sudharshan for his assistance. The research was supported by Administration on Aging, U. S. DHEW, Grant No. 5-38131.

if consumer information programs are looked at as products (or services), public policy makers can develop marketing programs detailing the alternative mixes of distribution, pricing (costing), and promotion. They also must decide which programs will be most effective.

Every marketing program begins with decisions on the market segment to which the product or program will be targeted. The development of the product itself lies in the predetermined need for that product by intended consumers. It is only after the determination of this need that the development of the product can be considered.

The discussion of this paper describes a consumer need based approach for deciding which products and services should be focused upon for consumer information programs. This approach is then operationalized for a specific segment of consumers, those over 64 years of age. Finally, the policy implications of using this approach are described by relating it to other current research in the area.

Conceptual Framework

The first task in the policy development process is to decide on a group of individuals to whom further information regarding products and services may be provided. Recent work in marketing literature suggests that the elderly (chronologically defined as over 64 years in age) are such a group. The lack of much empirical work describing the consumption habits of the elderly (Phillips and Sternthal 1977), their apparent disadvantages in the marketplace (Waddell 1975, Zaltman et al. 1978, Bearden et al. 1978, Deshpande and Krishnan 1979), and the growth of this age cohort as a proportion of the total U. S. population could be used as criteria for the selection of this consumer segment. These criteria are offered as illustrations, and several more could be developed to choose a target audience.

The next stage is an important one. Since information programs about a plethora of products and services cannot be practically developed (or even be needed), some method is required to determine the relative information demand for different products and services. One approach to determining this demand is obviously based on the perceived consumer need for information about a product or service. However, from an information program designer's perspective, this would be incomplete without a consideration of information availability. If consumers need certain kinds of information to make purchases, but this information is not difficult for them to acquire, then this is clearly not as critical a niche for program development as where both the need exists and the information is unavailable.² This may be called "information deficiency." When the degree of felt need for information for a product/service is high, as well as the difficulty in obtaining that information is high, a large information deficiency exists. Figure 1 shows an example of a scaling approach that can be used to describe an Information Deficiency Index for a product/service category.

²Both information need and availability were identified by elderly consumers during preliminary focus group interviews as being of major importance to them in making purchase decisions.

FIGURE 1
Information Deficiency Index* Scheme

		PERCEIVED NEED FOR INFORMATION ON PRODUCT	
		Yes	No
INFORMATION DIFFICULT TO OBTAIN	Yes	INDEX = 3**	INDEX = 2
	No	INDEX = 2	INDEX = 1

*Information Deficiency scores are unique only up to an ordinal level.

**Most critical niche for information program development.

On this index, it is assumed that the two dimensions (perceived need and difficulty in obtaining information) have equal weighting. This assumption has been made for lack of additional empirical information, although any alternative weighting scheme would as well be reflected in the indices given to each cell.

Information Deficiency scores for a set of "relevant"³ products and services can be developed using the above scheme. However, an approach is required for comparing the relative magnitudes of information deficiencies for different products for the consumer segment being considered. This implies that a method for interpreting the ordinal measurement in interval scale terms is needed. A procedure enabling this is the Thurstone Case V scaling model (Thurstone 1959), an excellent marketing application of which is described by Green and Tull (1978, pp. 180-87). The use of this model is provided in an empirical illustration below that follows the description of the sample used for this study.

Sample and Data

Data used in this study come from a larger research program focusing on the consumer problems of the elderly. The program investigates, among other issues, susceptibility to misleading marketing practices, information processing, and restitution mechanisms. Two structured mail questionnaires were developed from focus group discussions on these issues. The questionnaires were mailed six months apart to a national panel of 4,000 persons aged 25 to 80. The response-rates for each wave of the survey were 71.3% (2,853 responses) and 89.4% (2,551 responses from first-wave sample), respectively. In comparison with available statistics on the national level (U. S. DHEW 1978), the following were the characteristics of the overall sample:

- (1) a (naturally) disproportionate proportion of individuals over age 64 (61% to 11% nationally),
- (2) a lower representation of blacks and hispanics (3% and 0.1% versus 8% and 4% nationally) respectively,
- (3) a slight underrepresentation of high school graduates and a modest overrepresentation of college graduates, and
- (4) fewer individuals whose spouses were no longer living.

The analysis discussed in this paper deals with 1,747

³ A context specific definition of "relevance" is provided in the next section.

usable responses from the elderly (over 64) subsample or 68% of the final sample.

In the first wave of the questionnaire, respondents were presented with a list of over 40 products/services and asked to indicate those where they had felt cheated or taken advantage of during recent purchase experiences. This "bad buying experience" question also had an open-ended provision for products not included in the list (the original list had been arrived at through earlier focus group discussions). Table 1 presents the incidence of bad buying experiences as reported for the above question for the more frequently cited experiences.

TABLE 1
Bad Buying Experiences Reported
(Elderly Subsample, N = 1,163)

**Appliance Purchase/Repair	(28.8*)
**Home Repairs	(21.7)
**Insurance	(21.4)
**Auto Repair	(19.7)
**Prof. Services	(18.4)
**Utility Companies	(15.6)
**Auto Purchase	(14.4)
**Health Aids	(13.3)
Books/Mags./Encyclopedias	(10.2)
Home Decorator Services	(9.9)
Hospital Medical Care	(9.2)
OTC Drugs	(7.1)
Heavy Household Chores	(7.1)
Govt. Funded Programs	(6.5)
Pictures Taken at Store	(5.8)
Clothing	(5.4)
Landlord/Tenant Arrangements	(5.0)
Contractors/Builders	(4.9)
Investment Opportunities	(4.6)
Vacation/Travel Services	(4.5)
Contracts/Signed Papers	(4.0)

*figures in parentheses are percentages.

**major problem incidence areas.

After consultation with an advisory council of experts and policy makers in the elderly area, products and services with the most frequent citations in the above list were selected for more intensive scrutiny. These were (1) auto repair, (2) auto purchase, (3) home repair/improvement, (4) utility services, (5) insurance, (6) health care, (7) non-medical professional services, (8) appliance purchases, and (9) appliance repairs.

In the second wave of the study, the following question was asked:

I've listed below some different types of products or services. In Column A, "X" all those where you feel that you, personally, need information. In Column B, "X" the products or services where you feel the information you need is difficult to get.

	Column A Need Information	Column B Information Difficult To Get
Auto repair	<input type="checkbox"/> 1	<input type="checkbox"/> 1
An auto purchase	<input type="checkbox"/> 2	<input type="checkbox"/> 2
Home repair/improvement	<input type="checkbox"/> 3	<input type="checkbox"/> 3
Utility services	<input type="checkbox"/> 4	<input type="checkbox"/> 4
Insurance	<input type="checkbox"/> 5	<input type="checkbox"/> 5
Health care	<input type="checkbox"/> 6	<input type="checkbox"/> 6

	Column A Need Information	Column B Information Difficult To Get
Non-medical professional services	<input type="checkbox"/> 7	<input type="checkbox"/> 7
Appliances, purchase	<input type="checkbox"/> 8	<input type="checkbox"/> 8
Appliances, repair	<input type="checkbox"/> 9	<input type="checkbox"/> 9

The data obtained from 1,747 elderly consumer responses to this question form the basis for subsequent analysis reported below.

Analysis

For each respondent, an Information Deficiency (ID) score was determined as per the index described in Figure 1. Table 2 shows the median scores for each of the nine product categories being considered. Although these median values give some indication of the extent of information deficiencies present, as mentioned earlier, they still do not permit a meaningful comparison of the differences in relative magnitudes of the deficiencies for different product categories. This implies that the median values simply indicate that the information deficiency is greater for appliance repair than for home repair/improvement with no indication of how great the relative deficiency is. To do this, the ID scores can be converted using the Thurstone Case V model.

TABLE 2
Median Information Deficiency Scores

1. Auto Repair	1.584
2. Auto Purchase	1.246
3. Home Repair/Improvement	1.761
4. Utility Services	1.380
5. Insurance	1.359
6. Health Care	1.396
7. Non-medical Professional Services	1.477
8. Appliance Purchase	1.228
9. Appliance Repair	2.494

(N = 1,747)

This model is developed by first calculating the proportion of respondents for whom the ID score is higher for product category "j" than for category "k" for all possible category pairs. This is denoted by (j,k). The proportion of respondents having equal ID scores for a given category pair (j,k) were divided equally between proportions for (j,k) and (k,j). Table 3 shows the proportions matrix that results. For example, the entry in the second column and first row indicates that 37.8% of elderly consumers had a higher ID score for product category #2 (Auto Purchase) than for product category #1 (Auto Repair).

TABLE 3
Observed Proportions Matrix

		Product Category j								
		1	2	3	4	5	6	7	8	9
Product Category k	1	.500	.378*	.528	.460	.425	.450	.510	.367	.606
	2	.622	.500	.647	.570	.545	.568	.619	.487	.712
	3	.472	.353	.500	.432	.397	.423	.485	.340	.575
	4	.540	.430	.568	.500	.473	.492	.552	.417	.643
	5	.575	.455	.603	.527	.500	.524	.580	.442	.672
	6	.550	.432	.577	.508	.476	.500	.559	.418	.650
	7	.490	.381	.515	.448	.420	.441	.500	.372	.589
	8	.633	.513	.660	.583	.558	.582	.628	.500	.727
	9	.394	.288	.425	.357	.328	.350	.411	.273	.500

(N = 1,747)

Key to Product Categories

1. Auto Repair
2. Auto Purchase
3. Home Repairs/Improvement
4. Utility Services
5. Insurance
6. Health Care
7. Non-medical Professional Services
8. Appliance Purchase
9. Appliance Repair

*Table to be read as:

37.8% of respondents have a higher ID score for category 2 (Auto Purchase) than for category 1 (Auto Repair).

The Thurstone Case V model can be expressed simply as,

$$R_j - R_k = Z_{jk}$$

where $(R_j - R_k)$ is the relative distance expressing how much individuals discriminate between stimulus "j" and stimulus "k." In our case, this discriminial difference is the relative distance between the Information Deficiency for product category "j" and that for category "k." Also Z_{jk} is the standard variate (unit normal) associated with the observed proportion of cases in which the ID for "j" is perceived to be higher than that for "k." The model above assumes homogeneous perceptions of the stimuli across the respondent population. This assumption is supported to some extent in the split-half reliability test discussed below.

Table 4 reports the Z_{jk} scores matrix where each cell entry is the standard Z_{jk} variate associated with the corresponding entry in the Table 3 matrix (for computational procedure cf. Green and Tull 1978). The Thurstone Case V interval scale value (Information Deficiency) for each product category is then computed as the mean (\bar{Z}) of the column scores of the Z_{jk} matrix. Since this scale is unique up to an interval level, the ID values for the nine product categories can be rescaled with the lowest value (-0.254) being arbitrarily set to zero and the other values being correspondingly adjusted (R^*). Figure 2 expresses the rescaled ID values for the nine product categories on a unidimensional scale.

TABLE 4
Standard Variates (Z) Corresponding to Proportions Matrix

		Product Category j								
		1	2	3	4	5	6	7	8	9
Product Category k	1	0	-.31	.07	-.10	-.19	-.13	.03	-.34	.27
	2	.31	0	.38	.18	.11	.17	.31	-.03	.56
	3	-.07	-.38	0	-.17	-.26	-.20	-.04	-.41	.19
	4	.10	-.18	.17	0	-.07	-.02	.13	-.21	.37
	5	.19	-.11	.26	.07	0	.06	.20	-.15	.45
	6	.13	-.17	.30	.02	-.06	0	.15	-.21	.39
	7	-.03	-.31	.04	-.13	-.20	-.15	0	-.33	.23
	8	.34	.03	.41	.21	.15	.21	.33	0	.61
	9	.27	-.56	-.19	-.37	-.45	-.39	-.23	-.61	0
Total	.70	-1.99	1.34	-.29	-.97	-.45	.88	-2.29	3.07	
Mean (\bar{Z})	.078	-.221	.149	-.032	-.108	-.050	.098	-.254	.341	
R^*	.332	.033	.403	.222	.146	.204	.352	0	.595	

For key to Product Categories, see Table 3.

FIGURE 2
Unidimensional Interval Scale From
Thurstone Case V Model of ID Scores

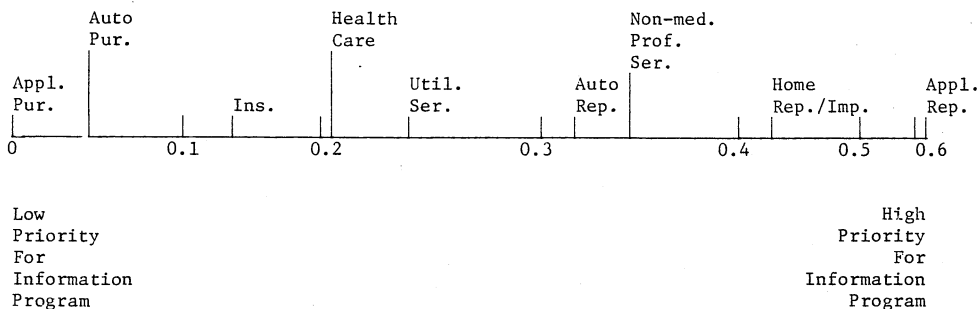


TABLE 5
Reliability Test for Thurstone Case V Scale
for Information Deficiency

	First Half	Second Half	Total Sample
1. Auto Repair	.382*	.306	.332
2. Auto Purchase	.022	.038	.033
3. Home Repair/Improvement	.394	.421	.403
4. Utility Services	.228	.196	.222
5. Insurance	.136	.182	.146
6. Health Care	.189	.218	.204
7. Non-med. Prof. Services	.360	.348	.352
8. Appliance Purchase	.000	.000	.000
9. Appliance Repair	.576	.615	.595
	(N = 873)	(N = 874)	(N = 1,747)

*Rescaled Thurstone ID score for first randomly split half of sample for Auto Repair category.

Discussion and Implications

As can be seen from Figure 2, the use of the Case V Approach provides a pictorial display of the Information Deficiency for different product categories. Since this is an interval scale, it can be used in a variety of manners for resource allocation in designing information programs. For example, we know that the difference in information deficiency between Appliance Repair and Non-medical Professional Services is approximately 4.8 times the difference between Non-medical Professional Services and Home Repair/Improvement since:

$$(.595 - .352) = 4.8(.403 - .352)$$

This provides an indication of the comparative resource allocation that can be made on information programs dealing with these specific product categories. Similar estimates of allocation can be developed for each of the nine categories described by the elderly consumer segment as requiring policy attention. Clearly, this can be a powerful tool when, as is usually the case, resources are limited and must be apportioned carefully. If the budget for an information program is extremely tight, it may even be necessary to expend resources on only three or four out of the nine categories described here.

Use of the approach outlined above then establishes the relative information demand priority for consumers. In an ordinal sense, it can be seen that more attention needs to be given to information programs on appliance repair, home repair, and non-medical professional services than, for instance, appliance purchase or automobile purchase at least as indicated by this sample of elderly consumers. More specific resource allocations can be determined using the pairwise differences rationale outlined above. Furthermore, other conceptual dimensions can also be used to determine information deficiencies or gaps. For instance, the cost of acquiring information and how understandable it is may replace the information need and availability dimensions used in our analysis.

As mentioned in the introductory comments, the use of this approach is based on consumer need for information rather than policy maker determinations of what information to provide to whom. It is possible that the use of such an approach can forestall both the criticisms of policy intervention where no intervention was desirable and that of consumers being provided with more information than they can efficiently process before making purchase decisions.

The conceptual approach suggested also has the advantage of allowing the comparison of different consumer segments

which need information-program attention. For example, it is conceivable that other demographic groups besides the elderly may need information regarding appliance and home repair. It may then be more cost-effective to devise programs targeted at all of these groups with similar information needs.

Conclusion

It remains to tie in the approach suggested here to some of the literature on public policy implications of consumer decision making referred to earlier. Responding to criticisms that research in consumer decision making takes an overly product orientation (McEwen 1978, Thorelli et al. 1975), this paper had advocated starting by focusing on consumer demand for information. In directly employing the marketing concept (Capon and Lutz 1979), policy makers would first select a target audience or segment and then determine what its information demands are (Day and Brandt 1974). This determination would be based on both information need and information availability dimensions. Additionally, it should allow policy makers to set priorities by product or service category.

Once these priorities have been set further research can be carried out to investigate the specifics of the information needed, such as unit pricing (Russo 1974) or restitution mechanisms for product complaints (Hunt 1977), and information presentation formats (Bettman 1975, Bettman and Kakkar 1977). But, to implement these latter information program specifics exclusively would be to deny the existence of consumer need for the program. This, in turn, would in all likelihood jeopardize the program in terms of its perceived relevance or usefulness to the consumers for whom it has been designed.

References

- Beardon, W.O., Gustafson, A.W. and Mason, J.B. (1978), "A Path Analytic Investigation of Life Satisfaction Among Elderly Consumers," Advances in Consumer Research, 6, 386-91.
- Bettman, J.R. (1975), "Issues in Designing Consumer Information Environments," Journal of Consumer Research, 2, 169-77.
- _____ and Kakkar, P. (1977), "Effects of Information Presentation Format on Consumer Information Acquisition Strategies," Journal of Consumer Research, 3, 233-40.

- Capon, N. and Lutz, R.J. (January 1979), "A Model and Methodology for the Development of Consumer Information Programs," Journal of Marketing, 58-67.
- Day, G.S. and Brandt, W.K. (1974), "Consumer Research and the Evaluation of Information Disclosure Requirements: The Case of Truth-in-Lending," Journal of Consumer Research, 1, 21-32.
- Deshpande, R. and Krishnan, S. (1979), "The Use of Buying Decision Rules by the Elderly: Public Policy Implications of Perceived Post-Purchase Dissatisfaction," AMA Proceedings, 44, 583-88.
- Green, P.E. and Tull, D.S. (1978), Research for Marketing Decisions, Fourth edition, Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Hunt, H.K., Ed. (1977), The Conceptualization of Consumer Satisfaction and Dissatisfaction, Cambridge, MA: Marketing Science Institute.
- Jacoby, J., Speller, D.E. and Kohn, C.A. (1974), "Brand Choice Behavior as a Function of Information Load: Replication and Extension," Journal of Consumer Research, 1, 33-42.
- McEwen, W.J. (1978), "Bridging the Information Gap," Journal of Consumer Research, 4, 247-57.
- Permut, S.E. (1979), "Research on Consumer Information: Public Sector Perspectives," Advances in Consumer Research, 6, 481-87.
- Phillips, L.W. and Sternthal, B. (1977), "Age Differences in Information Processing: A Perspective on the Aged Consumer," Journal of Marketing Research, 14, 444-54.
- Russo, J.E. (1974), "More Information is Better: A Reevaluation of Jacoby, Speller, and Kohn," Journal of Consumer Research, 1, 68-72.
- Scammon, D.L. (1977), "Information Load and Consumers," Journal of Consumer Research, 4, 148-55.
- Thorelli, H.B., Becker, H. and Engledow, J. (1975), The Information Seekers, Cambridge, MA: Ballinger Publishing Co.
- Thurstone, L.L. (1959), The Measurement of Values, Chicago: University of Chicago Press.
- U.S. Department of Health, Education and Welfare (1978). "Facts About Older Americans: 1977," Office of Human Development, Administration on Aging, DHEW Pub. No. (OHD) 78-20006.
- Waddell, F.E. (1975), "Consumer Research and Progress for the Elderly: The Forgotten Dimension," Journal of Consumer Affairs, 6, 146-75.
- Zaltman, G., Srivastava, R.K., and Deshpande, R. (1978), "Perceptions of Unfair Marketing Practices: Consumerism Implications," Advances in Consumer Research, 5, 247-53.

INFERENCE BELIEF FORMATION THROUGH THE USE OF
NON-INFORMATION: AN EXAMPLE

David W. Finn, Texas Christian University

Abstract

This paper reports on a surprising source of consumer information--the absence of information about salient attributes. This finding adds another dimension to research into the inference process. Not only do consumers combine information from the message, source, etc., to form beliefs, but they also seem to use information that is conspicuously absent.

Introduction

Perhaps the strongest single influence on consumers' behavior is the information they have about various choice alternatives--that is, the beliefs they possess about the alternatives on important dimensions. In fact, many attempts to model consumer behavior are structured around consumers' information and experience. For example, the Engel, Kollat, and Blackwell (1978) model discusses how information and experience are updated and changed (Information Processing), how information and beliefs are used to form attitudes and intentions toward alternatives (Product Brand Evaluations), and how existing levels of information influence the purchase process (Decision Process Stages). Similarly, the Howard-Sheth (1969) model places beliefs and brand experiences in crucial positions. Howard and Sheth offer explanations of how Perceptual Constructs mediate the effect of various inputs on beliefs (Brand Comprehension) and discuss how changes in beliefs related to brand knowledge and Choice Criteria influence attitude and intention. In summary, the central influence of beliefs on behavior is widely recognized.

If we accept these views that beliefs influence attitudes and intentions, and that attitudes and intentions, in turn, influence behavior, then the study of consumer belief formation and change becomes important in the study of consumer behavior. Three issues immediately present themselves regarding beliefs: 1) What are they? 2) What kinds are there? and 3) Where do they come from? The first two questions are handled rather easily. If all researchers can agree on a definitional basis for study, the explorational efforts into the third question will be more easily diffused.

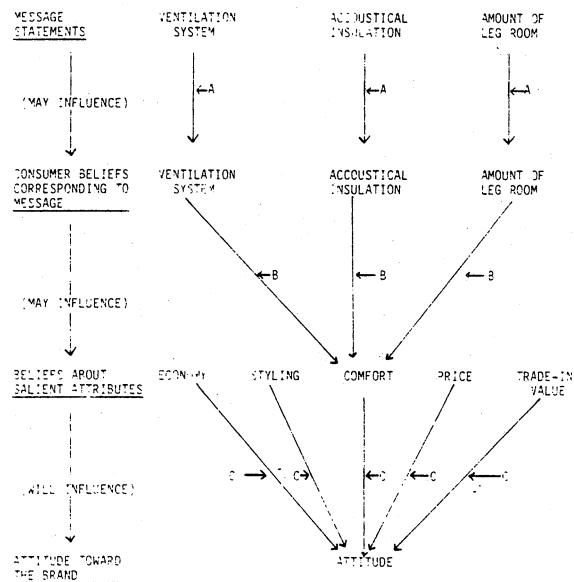
Fishbein and Ajzen (1975), in reviewing and condensing much of the literature related to beliefs and attitudes, have offered the definition of beliefs as representing the information a person has about an object. They further instruct that a belief can be expressed as a person's subjective probability that some object (or brand) has some specific attribute. This view of what beliefs are has built into it a method of measuring beliefs--namely via probabilistic scales.

They describe three basic kinds of beliefs, classified according to the informational determinants associated with belief formation. Beliefs formed as a result of direct observation and experience are called descriptive beliefs. A consumer forms descriptive beliefs about different brands mainly through use experience with them. For example, if a person test drives an automobile and feels the comfort and experiences the roominess, s/he will come to form beliefs about the presence or absence of these attributes. Predictably, beliefs formed in this manner are held with a great deal of confidence. Very little research has been

conducted on the nature or formation of descriptive beliefs although these beliefs are commonly tapped in various consumer behavior studies. Measures of recall, blind taste tests, and exploratory product concept tests all typically measure levels of consumer experience and preference that can be labeled "descriptive".

A second type of belief, one that is typically held with much less confidence, is an informational belief. Informational beliefs are formed by accepting information provided by an outside source. Most of the opinion formation literature, the literature on credibility and fear appeals, and other examples related to the work of the Yale group (Hovland, et al 1953), is concerned with the study of informational beliefs. In the consumer behavior literature, a study by Holbrook (1978) contains a clear example of informational belief formation. As is typical in message content research, Holbrook's study contained a message composed of belief statements related to the target beliefs that the message was meant to influence. For example, Holbrook listed characteristics of an automobile ventilation system, its acoustical insulation specifications, and its amount of leg room in order to influence receivers' beliefs about the comfort and roominess of the car in question. Clearly, before any change can take place in the target belief (comfort), the listener must first accept or reject the information about the ventilation system, etc. Belief links "A" in Figure A represent the location of this informational belief influence. The degree to which this information is accepted depends, in part, on the source, message, and channel performing the communication.

FIGURE A
ANTICIPATED INFLUENCE CHAIN FROM HOLBROOK (1978)



The third belief type is an inferential belief. Inferential beliefs are formed through the process of utilizing previously learned relationships. In the example above and in Figure A, consumers were expected to form beliefs about the comfort of the car based on their previous experience and knowledge that the ventilation system, insulation, and amount of leg room were all different aspects of the overall comfort of the vehicle. The message itself did not mention comfort, and formation of a belief that the vehicle is comfortable depends both on the level of acceptance of the message statements (informational belief formation) and on the inferences that receivers draw about the relation between comfort and these three attributes. (links "B" in Figure A). A successful change in this target belief will influence a consumer's attitude toward the car if comfort is a salient attribute. Figure A illustrates the influence links between message items and attitude.

Much of the knowledge we have of the inferential belief formation process is related to the literature on trait inferences started by Asch (1946). One of the more promising directions this literature has taken is toward the measurement of probabilistic relationships among beliefs--not only those beliefs related to message content, but also those beliefs residing in memory that can influence further belief formation. An early study of these probabilistic relationships is McGuire's (1960) analysis of logical syllogisms. Further work has been done by Wyer (1970; Wyer and Goldberg 1970), and an excellent positioning of this literature into the realm of consumer research was offered by Olson (1978). The Holbrook study is an important step in measuring these relationships in a consumer behavior setting.

Impact Effects

Not all inferential beliefs are formed due to planned influence by a communicator. Sometimes listeners take information from the communication and combine it with previously learned material in a completely unexpected manner. The result is inferential belief formation due to an unexpected impact effect. A good example of an impact effect is found in the corrective advertising study by Mazis and Adkinson (1976).

In their study, Mazis and Adkinson compared the effects of four different advertising messages about a brand of mouthwash. Three of the messages contained a germ killing appeal, while two added a corrective statement that "colds and sore throats could not be prevented, cured or benefited" by the use of the brand. Additionally, the corrective versions varied by apparent source of the corrective statement. The messages were:

- 1) noncorrective copy with germ killing appeal,
- 2) FTC required corrective copy with germ killing appeal,
- 3) corrective copy with germ killing appeal (no source mentioned), and
- 4) noncorrective copy with no germ killing appeal

All four versions made claims about the brand's usefulness in preventing bad breath. Version #1 differed from version #4 by also emphasizing the brand's germ killing ability. Version #3 differed from #1 only in the additional corrective copy that the brand could not prevent colds and sore throats. Finally, advertisement #2 added the phrase, "The previous statement was required by the Federal Trade Commission." This systematic inclusion of additional message statements allows us to measure belief change as influenced by incremental information. Table 1 repeats a portion of the Mazis and Adkinson data. Since Mazis and Adkinson found no difference in beliefs between messages 2 and 3, message #2 is deleted.

TABLE 1
BELIEF MEANS AND STANDARD DEVIATIONS
FOR EXPERIMENTAL ADVERTISEMENTS^a
[From Mazis and Adkinson (1976)]

DEPENDENT VARIABLES (Beliefs)	GERM KILLING NON-CORRECTIVE (Ad #1)		GERM KILLING CORRECTIVE (COMPANY) (Ad #3)		NON-GERM KILLING NON-CORRECTIVE (Ad #4)	
	M	SD	M	SD	M	SD
KILLS GERMS	1.42	.61	3.47	1.91	2.39	1.80
HELPS PREVENT COLD/SORE THROATS	2.16	1.74	4.67	1.83	3.04	2.01

^aBeliefs measured on seven-point probabilistic scales: "What is the probability that brand) has attribute)." "1" represented greatest amount of association (strong belief) and "7" represented smallest amount of association.

Comparing messages 1 and 4, we can isolate the effect of the germ-killing appeal. Because of the large difference in the sample variances, an approximation to the t-test must be used with an adjustment to the degrees of freedom (see Winkler and Hays, 1975 pg. 372). On a seven point scale (1 equals "probable", 7 equals "improbable"), the 23 respondents exposed to message #4 reported an average rating of 2.39 that the brand kills germs. The average rating for the 19 respondents exposed to message #1 was 1.42. The t-test approximation (28 d.f.) reveals that this difference is significant beyond the .025 level. In other words, the germ killing appeal, alone, had a positive influence. As expected, there is no significant difference on the "cold prevention" belief.

The influence of the corrective statement can be measured by comparing messages 1 and 3. Since they differ on only this piece of information, we should expect to see changes in only this belief. Table 1 reveals that this is not the case! Large and significant changes are present on the "cold prevention" belief (as expected), but large changes are also present in the "kills germs" belief. Introduction of the corrective statement had an unexpected impact on the belief that the brand kills germs.

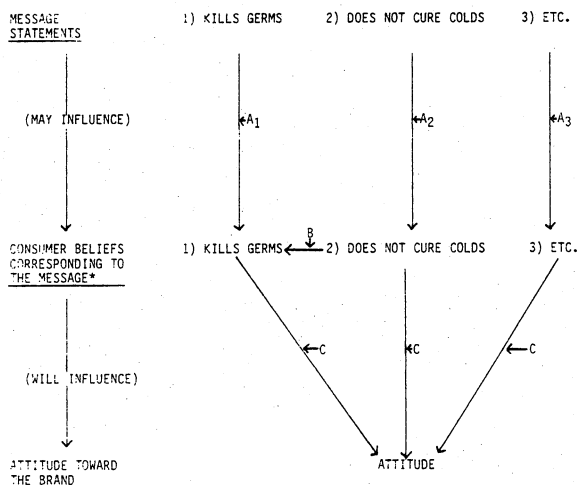
Even though message #3 emphasized the brand's germ killing properties, an audience exposed to it had significantly weaker beliefs that this was true than an audience exposed to the same message without the corrective statement (message #1). Figure B illustrates what may have happened:

- 1) as shown in Figure B, the germ killing emphasis of the message positively influenced receivers' beliefs that the brand kills germs (link "A1" of Figure B and the comparison of messages 1 and 4) while the corrective statement positively influenced their belief that it does not cure colds (link "A2"); and
- 2) the belief that the brand does not cure colds was inferentially related to the belief that the brand kills germs and impacted negatively on that belief (link "B" of Figure B).

Depending on the relative sizes of the positive and negative influences, the resulting "kills germs" belief could be stronger, weaker, or the same as the "kills germs" belief of an audience exposed to only message #4 (no mention). However, because of step #2 above, we would expect every audience exposed to message #3 to have weaker beliefs that the brand kills germs than an audience exposed to message #1. The Mazis and Adkinson data support this idea.

This example of an unexpected impact effect emphasizes the need for understanding the experience and belief systems of potential consumers before directing persuasive

FIGURE B
APPARENT INFLUENCE CHAIN FROM MAZIS AND ADKINSON (1976)



*Although Mazis & Adkinson did not isolate these particular attributes as salient, illustrating them as such will serve to demonstrate a two link chain from message to attitude.

communications at them. Where the Holbrook study showed that certain inferential processes can be expected, the Mazis and Adkinson paper can be interpreted as warning that inferential relationships can work against the intentions of the communicator. Both of these analyses demonstrate inferential belief formation due to the interaction between previously learned relationships and message-related beliefs--one planned and one unplanned. It is a logical step to expect that inferential beliefs might also result from an interaction between previously learned relationships and perceptions of the communication source or channel (e.g. one learns the characteristics associated with "trust" and forms message-related beliefs that vary in strength depending upon how the source matches those characteristics). We might conclude that the inferential process can distort any message-related belief or can use message-related beliefs in combination with any other learned material to create new beliefs. This logic, however, does not lead to any predictions concerning the absence of information. This paper presents evidence that impact effects resulting from the inference process can also occur in situations where certain information (particularly about salient product attributes) is totally absent.

Procedure

Two hundred and thirty-one junior and senior business students at a large Northeast university participated in the study. A posttest-only control group design was used (Campbell and Stanley, 1963), with one hundred and one of them serving as a control group. The others were exposed to an advertising-like message about a fictitious brand of toothpaste. The cautions voiced by Ferber (1977) were exercised in choosing the experimental product. The main consideration that guided the choice of the product was that subjects should have use experience with it and be the ones to make purchase decisions for it. Toothpaste was the ideal choice.

The use of an unknown brand was necessary because the study was concerned with belief change. Since beliefs about a brand are a function of information a person has about that brand, using an unknown brand allowed the assumption that the initial, or pre-exposure beliefs of both the control group and the experimental group were based on the same

information (little or none). Belief change could then be computed from the control group's position.

Accordingly, a cover story was invented that stated that a local professor (named) was engaged in consulting with a well known market research company (named) that specialized in test-marketing. The brand was said to be manufactured by a large, well known manufacturer (named) and to be in test market in the Mid-West. The members of the control group were told that the purpose of the questionnaire was to find out what their beliefs were concerning the new brand of fluoride toothpaste -- without any other information. They were to answer the questions using their knowledge and expectations about toothpaste in general.

After the cover story, the experimental group was told that the consultant had access to advertising information and that they were about to be shown the voice script of a television advertisement for the toothpaste. They were allowed to read the script (the experimental treatment) as many times as they wanted, but were not allowed to return to it once they started to answer later questions. The belief questions were the same as the control group's. The message is printed as Table 2.

TABLE 2
THE EXPERIMENTAL MESSAGE

VOICE SCRIPT OF A 30-SECOND TELEVISION ADVERTISEMENT FOR VISTA TOOTHPASTE

OFF-CAMERA VOICE: HEAR WHAT DOCTOR JOHN HUNT, A PRACTICING DENTIST IN CLEVELAND, OHIO, HAS TO SAY ABOUT NEW VISTA TOOTHPASTE.

DR. HUNT: HI! I'D LIKE TO TELL YOU ABOUT A NEW TOOTHPASTE CALLED VISTA. VISTA IS MADE WITH A NEW TYPE OF FLUORIDE. BUT IT'S NOT JUST ANOTHER FLUORIDE TOOTHPASTE.

THE FLUORIDE COMPOUND USED IN VISTA MADE IT POSSIBLE FOR (COMPANY) TO OFFER YOU A BETTER DECAY PREVENTIVE TOOTHPASTE THAT IS LOWER IN ABRASIVES, YET HIGHER IN WHITENING ABILITY, THAN OTHER BRANDS OF FLUORIDE TOOTHPASTE.

I REALLY LIKE THE TASTE OF VISTA. AFTER BRUSHING WITH VISTA MY MOUTH FEELS CLEAN. BEST OF ALL, WITH VISTA BOTH FRESH TASTE AND CLEANER TEETH ARE POSSIBLE, AND YOU DON'T HAVE TO SACRIFICE DECAY PREVENTION.

TRY VISTA

To measure subjects' beliefs about the brand, questions were asked about the probability of the brand being associated with various attributes. A seven-interval scale anchored by the phrases "very improbable" (1) and "very probable" (7) was used to record the response. Changes due to the message were measured from the control group mean.

TABLE 3
MEAN BELIEF CHANGES FROM CONTROL GROUP MEANS

DEPENDENT VARIABLE	CONTROL GROUP MEAN N = 101	CHANGE	SIGNIFICANCE ^a
MADE WITH NEW TYPE FLUORIDE	3.01	1.300	<.001
PREVENTS CAVITIES BETTER	3.63	-.080	NS
LOWER IN ABRASIVES	4.11	.453	.028
HIGHER IN WHITENING ABILITY	3.92	.402	.064
MAKES YOUR MOUTH FEEL CLEAN	5.49	-.139	NS
FRESH TASTE AND CLEANER TEETH	4.62	.253	NS
WORKS LIKE A MOUTHWASH	4.25	-.356	.080
SEAL OF APPROVAL	5.54	-2.391	<.001
ABOUT THE SAME PRICE	5.73	-1.153	<.001

^a Significance was computed using a two tailed t-test comparing control group to experimental group.

An earlier questionnaire had revealed that the salient attributes of toothpaste desired by members of the subject pool included; fluoride content, cavity prevention, low abrasiveness, whitening ability, good taste, the seal of approval of the American Dental Association, and a "reasonable" price. Many of these attributes were specifically claimed by the advertisement (Table 2) but no mention was made about the price or the seal of approval. It was expected that any message effects would be only on beliefs about those attributes specifically mentioned in the message (the "A" links) or on beliefs that could be inferentially linked to message content (for example, the "clean mouth" and "good taste" claims in the advertisement could result in a belief about "mouthwash effect"). Table 3 lists the beliefs measured. The beliefs above the dashed line are directly related to message elements; beliefs below that line were not contained in the message, but were measured to study possible impact effects.

Results

The results of the analysis are contained in Table 3. The first belief (made with a new type of fluoride) was the belief most dramatically affected by statements in the message. The control group (with no information) had an average belief of 3.01 on the seven interval scale. Members of the treatment group increased their belief to 4.31 on this ($P < .001$). Other changes in message related beliefs were less dramatic with only the third and fourth beliefs changing significantly.

Although changes in these early beliefs in the change chain from message to attitude would be necessary if targets lower in the chain (e.g., attitudes) were expected to change, the focus of this paper is on changes in beliefs not related to the message. The belief that the brand "works like a mouthwash" probably is inferentially related to message beliefs, and changes in it can be expected through the same processes as in the Holbrook paper. The last two beliefs however, appear to be unrelated to any message elements, yet change in them was among the largest of any belief changes. The message said nothing about the presence or absence of the seal of approval of the American Dental Association, yet the treatment group had a significantly weaker belief in its presence than the control group. A similar result is evident in the belief that the brand sells at about the same price as other brands.

Observation of the control group means revealed that, based simply on their past experiences with the toothpaste category, respondents had a fairly high expectation that any new fluoride toothpaste would possess the seal of approval and be priced in a competitive range. Unexpected impact effects of the message or the message environment changed these beliefs in a negative direction. Respondents exposed to the message did not report these strong beliefs.

Summary and Conclusions

The intent of this paper is to report on an interesting observation of impact effects of an advertisement. Other impact effects may be present, but to discover all of them, a complete inventory of beliefs would have to be measured - an impossible task.

Where did the large negative changes in beliefs about the seal of approval and price come from? Earlier in the paper it was demonstrated that impact effects occur on beliefs related to informational items in a message. It is very doubtful that "seal of approval" and "price" are related to informational items in the toothpaste message. However, it can be argued that they are related to information contained in the complete message environment. That is, the information processed by the treatment group included the observation

that nothing was said about these two salient attributes. If this is the case, then the schema proposed by Olson (1978) is applicable to non-information as well as to information. That is, the absence of any mention of "seal of approval" or "price" may have been perceived as a "cue" in the task environment, leading to the formation of a descriptive belief ("nothing was mentioned about ____"). This belief could then have combined with the expectation that a company would mention the attribute if the brand had it, resulting in an inference that the brand might or might not possess the attribute. Unfortunately, the results of this study were neither predicted nor expected, due to its exploratory nature, and the data were not collected in a manner that allows quantifying the possible relationships among beliefs. A tentative hypothesis is that "top of the mind" salient attributes are noticed in their absence as well as when mentioned.

The general findings of this study suggest the need for future research in at least two areas. First, what are the dynamics of inferential belief formation due to advertising? Holbrook (1978) attempted to explore this issue and has provided a meaningful framework for further study of the problem; the work of Mazis and Adkinson (1976) highlights the existence of unexpected impact effects related to inferential belief formation; and this study suggests that impact effects may result not only from the interaction of message elements, but also from the entire message environment.

The second area for future research evolves from the question, do consumers expect an advertiser to state product claims for those salient attributes for which he can make legitimate claims? This question has implications for identifying the more salient attributes of different product classes. That is, it may be that belief change as noted above is a good index of relative salience of various attributes. For example, if consumers make stronger inferences about salient than non-salient attributes, we could systematically leave out mention of various attributes and note the effect on beliefs. Those beliefs changing the most when not mentioned in an advertisement would then be identified as most salient.

Together, these two questions raise the issue, "what is the information content of non-information?"

References

- Asch, S.E. (1946), "Forming Impressions of Personality," Journal of Abnormal and Social Psychology, 41, 258-90.
- Campbell, Donald T., Stanley, Julian C. (1963), Experimental and Quasi-Experimental Designs for Research, Chicago: Rand McNally.
- Engel, James F., Blackwell, Roger D. and Kollat, David T. (1978), Consumer Behavior (3rd edition), Hinsdale, Illinois: The Dryden Press.
- Ferber, Robert (1977), "Research by Convenience," Journal of Consumer Research, 4, 57-58.
- Fishbein, Martin and Ajzen, Icek (1975), Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading, Massachusetts: Addison-Wesley.
- Holbrook, Morris B. (1978), "Beyond Attitude Structure: Toward the Informational Determinants of Attitude," Journal of Marketing Research, 15, 545-56.
- Hovland, C.I., Janis I.L. and Kelley, H.H. (1953), Communication and Persuasion, New Haven: Yale University Press.

Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: Wiley.

Mazis, Michael B. and Adkinson, Janice E. (1976), "An Experimental Evaluation of a Proposed Corrective Advertising Remedy," Journal of Marketing Research, 13, 178-83.

McGuire, William J. (1960), "A Syllogistic Analysis of Cognitive Relationships," in Hovland, C.I. and Rosenberg, M. (eds.), Attitude Organization and Change, New Haven: Yale University Press, 65-111.

Olson, Jerry C. (1978), "Inferential Belief Formation in the Cue Utilization Process," in Hunt, H.K. (ed.), Advances in Consumer Research, Vol 5, Association for Consumer Research, 706-13.

Winkler, Robert L. and Hays, William L. (1975), Statistics (2nd edition), New York: Holt, Rinehart and Winston.

Wyer, Robert S., Jr. (1970), "Quantitative Prediction of Belief and Opinion Change: A Further Test of a Subjective Probability Model," Journal of Personality and Social Psychology, 16, 559-70.

Wyer, Robert S., Jr. and Goldgerg, Lee (1970), "A Probabilistic Analysis of the Relationships Among Beliefs and Attitudes," Psychological Review, 77, (no. 2), 100-120.

AN EXAMINATION OF INFORMATION PROCESSING TRAITS: GENERAL SOCIAL
CONFIDENCE AND INFORMATION PROCESSING CONFIDENCE

James M. Munch, The Pennsylvania State University
John L. Swasy, The Pennsylvania State University¹

Abstract

This paper reviews Wright's (1975) findings concerning the role of the information processing traits General Social Confidence (GSC) and Information Processing Confidence (IPC). This review concludes that Wright's findings were an artifact of the experimental procedure. Also this paper presents suggestions for future research concerning the role of information processing traits as factors influencing cognitive resistance to persuasion.

Introduction

The systematic examination of persuasive communication is of great importance to both the public and private sectors of our economy. According to 1980 projections, U.S. total advertising expenditures will surpass \$55 billion. The ability to increase the effectiveness of such expenditures rests upon understanding audience responses to persuasive communications.

Early research in persuasion typically focused upon the direct measurement of outcome variables (e.g., attention, recall, beliefs, attitudes). More recent efforts have taken a process-oriented focus in an attempt to achieve a greater understanding of persuasion. One common approach is based upon evidence that receivers of a persuasive message actively process and subvocally react to the message content (Greenwald 1968). Reports of these reactions have been termed cognitive responses and are believed to reflect the results of the receiver's information processing. One typical categorization of such responses distinguishes counterarguments and support arguments to the message claims and also derogations or attacks against the message source (Wright 1973). Furthermore, these reported thoughts have also been categorized according to their origin (i.e., thoughts which were restatements or reflections of a message statement, or thoughts which were "original" or developed by the receiver, independent of message claims) (Greenwald 1968).

Marketing and psychology studies have shown that frequency counts of these response categories, taken from receivers' reports of their thoughts during a message, are good predictors of subsequent message acceptance and attitude change. This predictive ability, as well as physiological evidence supporting the mediating role of subvocal responses in attitude change (Cacioppo, Sandman and Walker 1978), provides strong evidence that cognitive responses such as counterarguing may underlie persuasion outcomes.

Much theorizing and research have been directed toward understanding the antecedents of message-evoked thoughts.

¹James Munch is a doctoral student in Marketing and John Swasy is an Assistant Professor of Marketing at The Pennsylvania State University, University Park, Pennsylvania 16802. This research was supported by a grant to John Swasy from the Center for Research, College of Business Administration, The Pennsylvania State University. The authors gratefully acknowledge the helpful comments from the ACR reviewers. All correspondence should be addressed to John Swasy.

Such differential response to persuasive message arguments has been attributed to: prior experiences (e.g., individual differences), differences in communications (e.g., source effects), and the social context (e.g., situational factors), (Eagley 1980). Research has been undertaken to examine such factors as the accessibility of various responses from long-term memory (Edell and Mitchell 1978), prior commitment on the message topic (Osterhouse and Brock 1970), response opportunity (Petty, Wells and Brock 1976) and message processing goals (Petty and Cacioppo 1979).

To date, however, there has been quite limited interest in research attempting to systematically examine these cognitive responses in relation to receivers' information processing traits. The only published article to address the relationship between receivers' information processing traits and cognitive responses is Wright (1975). The purposes of this paper are to position Wright's research within the framework of the existing personality research paradigms, to review Wright's findings and to suggest an alternative interpretation of his results. Additionally several suggestions for future research dealing with the study of information processing traits will be presented.

Fundamental Paradigms in Personality Research

The premise of the relative stability of personality variables represents one of the classic issues for debate in personality research. The purpose of this section of the paper is not to resolve the issue of stability in personality research. Rather, our attempt is to briefly examine three fundamental paradigms pertinent to this issue and to position Wright's conceptualization of General Social Confidence (GSC) and Information Processing Confidence (IPC) within this framework. The three paradigms to be reviewed are the situationalist, trait, and interactionist positions. For a more comprehensive discussion of these personality/stability issues see Epstein (1979).

Situationalist Position. The situationalists argue that behavior is determined almost exclusively by situational variables. This premise is based on three fundamental sources of evidence:

- (1) When behavior in one situation is correlated with behavior in another situation the results are extremely low--usually less than .30. Thus, Mischel (1968) notes, "A correlation of .30 leaves us understanding less than 10% of the relevant variance."
- (2) Analysis of variance results have demonstrated that the variance attributable to individual differences is usually much smaller than the variance attributable to the interaction of individuals and situations, Endler and Hunt 1968, 1969; Endler, Hunt and Rosenstein 1962).
- (3) Ratings of stability of personality by others have consistently overstated the stability of individuals across situations (Bem and Allen 1974; Jones and Nisbett 1971; Mischel 1968; Shweder 1975).

Trait Position. Trait theorists contend that the failure to find stable traits across situations is primarily due to inappropriate research procedures. A host of arguments are advanced to support such a contention, the more

compelling being:

- (1) The unit of analysis has not been adequately taken into account. What appears to be instability at a behavioral level of analysis may be stability at the trait level of analysis (Alker 1972; Bowers 1973). Stated differently, apparently unstable behaviors may be based upon relatively stable traits.
- (2) The use of moderate variables would consistently increase reliability coefficients (Alker 1972).
- (3) A failure to recognize that some people are more variable than others results in reporting generally low stability coefficients.
- (4) Stability in personality is mediated by an individual's cognitions; therefore, stability will only be found when ideographic (i.e., symbolic or graphic) procedures are used that take into account the subjective nature of perception (Alker 1972; Bem and Allen 1974; Bowers 1973; Mischel 1973).

Interactionist Position. Although the interactionist position does not resolve the issue of stability of personality, proponents of the interactionist position contend that the controversy between individual (i.e., trait) versus situation is meaningless, (i.e., behavior is always a function of the joint interaction of these variables.) Clearly, in any psychological situation the interaction of individuals and situations should account for more variance than either source of variance by itself. Thus it has been argued that an interactionist position should supplant both the trait and situationalist positions (Bowers 1973; Ekehammer 1974; Endler 1966; Endler and Hunt 1968).

As Epstein (1979) suggests, the situationalist position, trait position and the interactionist position are in fact all "correct" approaches to different problems. The situationalist focuses upon the general effects of situations over a sample of individuals. The trait theorist examines the consistency of behavioral tendencies in individuals over a sample of situations. The interactionist studies the behavior of people with certain attributes in situations with certain attributes.

The researcher's decision regarding which personality paradigm or conceptual framework to choose is thus dictated by the research purpose. Since Wright conceptualized General Social Confidence (GSC) and Information Processing Confidence (IPC) as "chronic" traits and examined these traits in four different reception environments, an interactionist position is presumed. Additional comments regarding the compatibility of Wright's conceptual framework with the measurement of that conceptual framework are presented in later discussion.

Wright (1975)

This section of the paper will briefly review Wright's hypotheses and methodology. The discussion will be limited to only those aspects most germane to the role of information processing traits. Following this overview, Wright's results will be presented and examined.

In his article, "Factors Affecting Cognitive Resistance to Advertising," Wright discusses:

- (1) the role of information processing traits, and
- (2) decision task involvement in audio and print message presentations.

The information processing traits under study were Information Processing Confidence (IPC) and General Social Confidence (GSC). Both GSC and IPC were conceptualized as "chronic" personal characteristics related to transsituational abilities and habits. More specifically, GSC was

regarded as one's degree of skill or self-confidence in assessing personal ability in coping with everyday problems (i.e., a characteristic which exhibits both social and cognitive dimensions). IPC was described as one's confidence regarding the ability to process information or think fast (i.e., a characteristic which is evidenced by quickness of mental reflexes or cognitive agility). It should be noted that both were self perceptions of one's ability and not necessarily equivalent to actual knowledge structures or processing performance levels.

The decision task involvement factor addressed a motivational determinant of a receiver's cognitive resistance to persuasion. Subjects in the "high processing involvement" treatment were informed they could expect to make a short-run decision about the product appearing in the impending advertisement. The relevance of this decision was emphasized in terms of their families, their own time and effort, and their personal finances. Subjects in the "low content involvement" treatment received no such information.

Wright's persuasive message treatment consisted of the presentation of either an audio or print advertisement. The audio message was defined to be a strainful information processing situation since subjects receiving the treatment had no control over the speed of message delivery or inflections of the speaker. Those subjects who received the print ad were by definition, engaged in a relatively leisurely information processing task.

Wright's Hypotheses

The focus of Wright's research was to address the form of the GSC-message acceptance relationship. Cognitive responses were also examined in order to test the hypothesis that people low in GSC might find it difficult to cope with both message encoding and critical message scrutinizing under strained processing conditions (i.e., audio mode) and therefore opt for source evaluation responses (e.g., source derogations) rather than message oriented responses (e.g., counterargumentation). Additionally, Wright introduced the information processing confidence (IPC) construct as an indicator of a person's mental reflexes.

Wright's hypothesized arguments regarding GSC and IPC are summarized as follows:

- (1) In a strained information processing environment (i.e., audio mode) both GSC and IPC will shape the nature of the responses. This can be contrasted with the unstrained (i.e., print) condition in which no strong relationship between GSC and IPC and message acceptance was expected.
- (2) In the audio presentation to a highly involved audience, a sense of social confrontation will be created. Therefore as in highly social situations, both GSC and IPC will be important determinants of a receiver's responses, however GSC will be relatively more important.
- (3) For audio presentations which reflect a less social confrontation (i.e., low involvement), both the GSC and IPC factors will again emerge, but those information processing capabilities (i.e., IPC) are posited to contribute more to individual differences in responses.

Figure 1 illustrates Wright's 2 x 2 crossed factorial design and highlights these hypothesized relationships.

Wright's Methodology

Subjects in Wright's study were 160 adult housewives. Each woman was randomly assigned to one of the four communication conditions in the 2 x 2 crossed-factorial design. A new type of food product (i.e., a soybean derivative) was chosen for the topic of the message to minimize product knowledge differences. The message arguments contained six favorable

FIGURE 1
Wright's 2 x 2 Cross Factorial Design and
Hypothesized Relationships

Task Involvement \ Media Mode	Audio Advertisement (strained information processing environment)	Print Advertisement (unstrained information processing environment)
High Involvement (high social confrontation)	Both GSC and IPC effect receiver's responses to the message with GSC > IPC	No relationships
Low Involvement (low social confrontation)	Both GSC and IPC effect receiver's responses to the message with IPC > GSC	No relationships

reasons for possible product adoption. In order to simulate as far as possible the natural conditions in which people encounter mass media advertising messages, Wright presented the ad following a short "editorial" statement. Immediately after exposure, cognitive responses were collected, followed by measures of message acceptance, content recall, and a check of the involvement manipulation. Subjects then rated each thought as "extremely important," "moderately important," or "slightly important." Also they rated each thought according to whether they felt it had originated directly in the message, had been a reaction on her part to an idea directly stated in the message text, or was an "original" thought in the sense that it was self initiated rather than a response to a direct statement in the message text (Wright 1975, p. 5). This thought judging task was then followed by a peer evaluation task² and the self report GSC and IPC measures. These are reported in Table 1.

Wright's Results

Wright's findings were generally consistent with the hypotheses presented above. The GSC and IPC scores were more highly correlated with counterarguing (CA) in the strained (i.e., audio) than in the less strained (i.e., print) conditions. As expected, for the audio presentation when involvement was low, the IPC-CA correlation was stronger than the GSC-CA correlation, while the GSC-CA correlation was stronger in the high-involvement audio condition. (See the selected results reproduced in Table 2). Wright also suggests that the same pattern (i.e., the flip-flop in importance for GSC and IPC) was observed for the trait message acceptance correlations.

Critique of Wright (1975)

A close examination of Wright's findings suggests that several conceptual and methodological issues need to be re-considered prior to the validation of GSC and IPC as information processing traits. This issues are:

- (1) Why did only the number of counterarguments classified by subjects as "original" evidence the hypothesized interactions with task involvement? and;
- (2) Is it appropriate to conceptualize GSC and IPC as "chronic" or enduring information processing traits and measure them in the manner used in this study?

² The peer-designated opinion leadership factor is not central to our analysis and will not be addressed in this review.

TABLE 1
General Social Confidence and Information Processing
Confidence Measures (Wright 1975)

- General Social Confidence
1. I do not spend much time worrying about what people think of me.
 2. It doesn't bother me to have to enter a room where people are gathered and already talking.
 3. When confronted by a group of strangers, my reaction is always one of shyness and inferiority.
 4. I am sure someday my companions will look up to me and respect me.
 5. I have frequently given up doing something because I thought too little of my ability.
 6. I feel capable of handling myself in most social situations.
 7. I don't make a very favorable first impression on people.
 8. I often fear my actions will cause other people to have a poor opinion of me.
 9. In group discussions I usually fear my opinions are inferior.
 10. I am not at all lacking in self confidence.

Information Processing Confidence

1. I have more trouble concentrating than most people.
2. I am able to solve riddles and puzzles rapidly.
3. My mind seems to work slowly compared to those around me.
4. I am totally confident about my ability to judge messages coming from the mass media.
5. I am certainly able to think quickly.
6. When I hear an argument being presented, I am quick to spot the weaknesses in it.
7. I usually have to stop and think for awhile before making up my mind even in unimportant matters.
8. My thoughts frequently race ahead faster than I can speak them.
9. I am never at a loss for words.
10. I don't seem to be very quick-witted.

Consistency of Results and Theory

The question of why only the number of counterarguments coded by subjects as being "original" counterarguments were related to personality traits can initially be discussed at a conceptual level. That is, one might have reasonably expected that the total number of counterarguments generated, irrespective of origin should have been related to these information processing traits. Indeed, as Wright suggests, prior theoretical discussions (Cohen 1958; McGuire 1969; Bither and Wright 1973) all imply that a negative linear relationship with counterargumentation and GSC and IPC should be expected. However these discussions do not address expected differences between "original" and other counterarguments' relationship to processing traits. Regarding this issue, the question also arises as to why the total number of thoughts irrespective of nature (i.e., counter or supportive) and origin should not be related to these information processing traits. Although Wright's premise suggests that individuals might hold a range of attitudes toward the content of a persuasive communication, existing theoretical and empirical evidence appears weak as to why the number of "original" counterarguments should be correlated more strongly than other cognitive response types with these personality traits. One plausible explanation for the resultant relationship between GSC, IPC and the number of "original" counterarguments may be the experimental procedure employed by Wright.

Methodological Considerations

This section of the paper attempts to address three important methodological issues in Wright's research:

- (1) The appropriateness of the data collection order.

TABLE 2
Selected Within Cell Correlations of GSC and IPC
With Response Measures (From Wright 1975)

	Audio		Print	
	High Involvement (n=40)	Low Involvement (n=40)	High Involvement (n=40)	Low Involvement (n=40)
Correlation of GSC with:				
Counterargument (total)	.51 ^a	.27 ^b	.17	.06
Counterargument (rebuttal)	.33 ^b	.35 ^b	.71	.08
Counterargument (original)	.35 ^b	.11	.09	.12
Support argument (total)	.13	.27	-.21	.12
Source derogation (total)	.02	-.17	.23	-.10
Acceptance (A ₁)	-.54 ^a	-.35 ^b	-.27	-.14
Acceptance (Intentions)	-.36 ^b	-.34 ^b	-.28	-.27
Acceptance (Index)	-.47 ^a	-.33 ^b	-.31	-.26
Correlation of IPC with:				
Counterargument (total)	.24 ^b	.45 ^a	.08	.16
Counterargument (rebuttal)	.34 ^b	.46 ^a	.09	.22
Counterargument (original)	-.07	.33 ^b	-.03	.08
Support argument (total)	.08	-.24	-.05	.19
Source derogation (total)	.20	-.05	.10	-.12
Acceptance (A ₁)	-.38 ^b	-.37 ^b	-.23	-.27
Acceptance (Intentions)	-.47 ^a	-.46 ^a	-.24	-.29
Acceptance (Index)	-.33 ^b	-.44 ^a	-.15	-.21

^a p < .01

^b p < .05

^c Acceptance (A₁): "The arguments about the product contained in the advertisement were very convincing"

Acceptance (BI): "Will you purchase when it becomes available in your local area?"

Acceptance (Index): Weighted index of subject's reported thoughts (i.e., $\sum_k w_k SA_k - \sum_j w_j CA_j - \sum_k w_k SD_k$)

where w_j s are subjective importance weights of each thought assigned by the subjects.

- (2) The problem of interpretational confounding.
- (3) The measurement issue concerning the validity of the GSC and IPC scales.

Order of Data Collection. In Wright's study the subjects reported their thoughts directly following their viewing of the ad. They were then given the task of indicating the "origin" of each thought (e.g., "Was this an idea or reaction which reflected a point presented in the ad message or was this one of your own?") and then given the IPC and GSC measures. Since one normally conceives of personality traits as being CAUSAL determinants of cognitive responses, the order of measurement for these variables should have been consistent with this conceptualization.

In all likelihood the experimental treatment which subjects had just experienced created a mental framework or set which influenced their responses on the IPC and GSC measures. It appears evident that subjects who have just experienced a task which called upon them to assess whether or not their "thoughts" were their "own" (rather than a reflection of the speaker's) should respond consistently when asked questions regarding their skills in a social confrontation situation (i.e., How well they can spot a weakness in an argument?). That is, subjects who have indicated that each counterargument was "original" (and therefore have a large number of original counterarguments), should express a high degree of "social confidence."

It should also be recognized that despite Wright's claim that the GSC/IPC-message acceptance measures exhibit the same pattern of results, a close examination of these correlations shows that only the "acceptance (index)" measure shows the hypothesized reversal in importance which parallels the GSC/IPC-"original" counterargument pattern. (See Table 2). It is apparent that this index should be closely related to the number of original counterargument cognitive response measure, since these responses contribute to the calculation of this index (i.e., $\sum w_j SA_j - \sum w_j CA_j - \sum w_k SD_k$). Note also that this index includes an additional perceptual influence stemming from the thought rating task (i.e., the subjective weighted importance ratings) which

might also affect the subsequent responses on IPC and GSC (e.g., "I rated my original thoughts as very important and therefore I guess I am "not at all lacking in self confidence."). The consistent relationships between both of these dependent variables (and not the others) with the GSC and IPC measures strongly suggests that the thoughts judging tasks affected the subsequent responses on the GSC and IPC measures. In summary, the experimental procedure employed by Wright offers an alternative explanation for the consistent relationships he observed.

So far this critique has argued that the likelihood of a carry-over effect (between the thought judging, i.e., classification of "originality" and "importance" and trait measures) was quite high. One question still remains as to why this effect only occurred in those particular conditions. The next portion of this paper will address this issue.

If one accepts Wright's argument that the high involvement-audio message is representative of a strainful and overt social influence situation (i.e., the uncontrolled message delivered by another person only once, after which the subject expects to make an important decision), then it seems clear that subjects would be experiencing a task which creates strong arousal potential. In this social comparison processing environment it appears likely that subjects' ratings of the number of "original" counterarguments and GSC would necessarily be correlated. On the other hand, consistent with Wright's hypotheses, IPC should not strongly relate to the number of "original" CA as the information-processing task orientation is subsumed by the pervasive social factors.

Compared with the high involvement treatment, the low involvement conditions exemplifies a less social and more task-oriented situation. Message receivers therefore place less emphasis on cognitive resistance activities inherent in the high involvement decision mode. The lack of personal decision making permits the information processing task environment to become more salient. Hence, the subjects' self-reports of IPC and number of "original" counterarguments are related in the low involvement condition. General Social Confidence is not a strongly related factor in this processing environment since the strong task orientation and lack of the personal decision mode diffuse the strength of a social orientation.

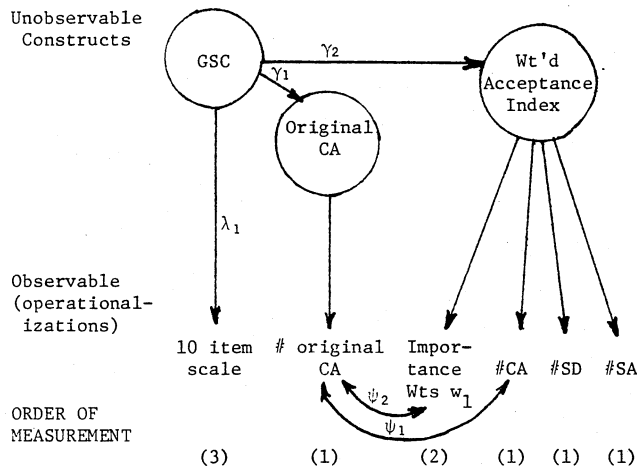
In summary, it appears that Wright's manipulations were effective in creating the desired processing environments. However, because the thought coding and importance weighting tasks immediately preceded the trait measurement, it appears that the conclusions regarding a "trait-affecting-cognitive response" causal relationship are unfounded.

The Issue of Interpretational Confounding. Figure 2 represents Wright's nomological network for the proposed General Social Confidence--"original" counterargument (i.e., GSC \rightarrow "original" CA) and General Social Confidence--Weighted message acceptance index (i.e., GSC \rightarrow Wt'd Acceptance Index) relationships. A similar nomological network exists and could be replicated for the IPC construct.

The model illustrates why one would necessarily expect significant correlations between the weighted message "acceptance (index)" and the GSC--"original" counterargument constructs. Because the number of "original" counterarguments has been utilized to operationalize both CA and the message acceptance index we have encountered a problem of interpretational confounding (Bagozzi 1980, p. 153). This general problem of ascertaining the meaning of theoretical variables i.e., interpretational confounding, "occurs as the assignment of empirical meaning to an unobservable variable which is other than the meaning assigned to it by an individual a priori to estimating unknown parameters" (Burt 1976, p. 4).

In Wright's model it appears that the γ_2 relationships is strongly influenced by the γ_1 relationship. Evidence for this premise can be obtained by examining the expected

FIGURE 2
Wright's Nomological Network



covariances ψ_1 and ψ_2 and by noting the temporal order of measurement.

The expected covariation between total number of counter-arguments (i.e., #CA) and the number of "original" counter-arguments (i.e., # original CA), (Noted as ψ_1), implies that the CA construct and weighted message acceptance construct are not conceptually distinct. The fact that other measures of message acceptance did not yield significant relationships with GSC/IPC provides additional credence for this premise.

The logic for the expected ψ_2 covariance is quite similar to that of ψ_1 . Additionally, because the # original CA and the importance weights observables represent self-reported measures subject to the earlier suggested biases of carry-over effects and arousal potential, these measures may share a considerable amount of covariance (i.e., ψ_2). A stronger test of theory would be to demonstrate a significant relationship between GSC and the weighted acceptance index while controlling for # original CA.

One additional rationale for the case of interpretational confounding involves the temporal order of measurement. As previously discussed, this ordering suggests that the causal relations γ_1 and γ_2 were actually reversed in the experimental setting. A final methodological issue concerns the validity of the GSC and IPC scales.

Factor Analysis of GSC/IPC. Measures of GSC and IPC were collected in two separate investigations by these researchers. In the first study (Swasy 1980) measures of GSC and IPC were collected for comparison with cognitive responses. Several aspects of this study closely resembled Wright's approach. These aspects included the use of an uncontrolled persuasive message (i.e., a videotaped sales presentation), similar sample of subjects (i.e., mothers, who were volunteer subjects and church group affiliates, $N = 249$), and who had a relatively high level of involvement with the topic (i.e., children's books). One important difference between this investigation and Wright's approach however, is that the GSC and IPC measures were collected one week prior to the experiment. Earlier arguments regarding possible carryover effects in Wright's design are eliminated with these pre-measures of GSC and IPC. No relationships between GSC and IPC with message acceptance or frequency of any cognitive responses were found. Subsequent orthogonal and oblique factor analyses of the GSC and IPC scales revealed that they were not unidimensional.

To reassess these findings a second investigation was conducted. A random sample ($N = 135$) of undergraduate college students completed a questionnaire which contained a number of attitude scales and the GSC and IPC measures. Factor analyses again revealed non-unidimensionality.

At a conceptual level, the oblique factor structure illustrates some interesting dimensions underlying the IPC and GSC scales. (See Tables 3 and 4). An assessment of these factors suggests that GSC contains both social and information processing subfactors. Although this is consistent with Wright's notion of the GSC construct, an examination of the IPC scales introduces some confusion. The IPC scale also appears to contain aspects of a "social" comparison (e.g., "I have more trouble concentrating than most people"). Although these factor analysis results are not a conclusive test of the scales' validity, they do suggest the need for a more systematic approach toward measuring individual differences.

One aspect of the scales' validity which should be addressed is the social desirability of the responses to these measures. In the survey of undergraduate college students, mentioned above, the Crowne/Marlowe Social Desirability Scale (1964) was also administered. For this sample, the GSC-SD correlation was .4449 ($p=.001$) and the IPC-SD correlation was .2094 ($p=.007$). These findings provide additional evidence of the weakness of the GSC and IPC measures and again suggest the possibility that a repertoire of subjective tasks such as judging the originality and importance of verbalized thoughts may also contain a high degree of social desirability bias which may be shared with the GSC/IPC scales.

TABLE 3
Oblique Factor Analysis
General Social Confidence

ITEM	FACTOR STRUCTURE		
	Factor I	Factor II	Factor III
ITEM 1	.03119	.86371	-.27201
ITEM 2	.49440	.49536	-.33789
ITEM 3	-.53069	-.47825	.49616
ITEM 4	.65029	.04886	-.29634
ITEM 5	-.24709	-.17745	.74853
ITEM 6	.74009	.41449	-.42885
ITEM 7	-.61770	-.03822	.26531
ITEM 8	-.50948	-.50197	.50295
ITEM 9	-.44565	-.29902	.57687
ITEM 10	.48906	.46091	-.79241

TABLE 4
Oblique Factor Analysis
Information Processing Confidence

ITEM	FACTOR STRUCTURE		
	Factor I	Factor II	Factor III
ITEM 1	-.00034	.75377	-.23300
ITEM 2	-.50493	-.19920	.10638
ITEM 3	.49453	.76980	-.48678
ITEM 4	-.03600	-.17519	-.41217
ITEM 5	-.41963	-.15794	.47789
ITEM 6	-.05482	-.24529	-.45639
ITEM 7	-.01098	.41214	-.23250
ITEM 8	-.25793	.13903	-.02990
ITEM 9	-.03344	-.15924	.47759
ITEM 10	.55593	.20741	-.65231

Discussion

This paper has reviewed Wright's (1975) findings concerning the role of information processing traits influencing cognitive resistance to advertising. At a conceptual level, an examination of the methodological procedures suggests

that the experimental design affected the GSC and IPC measures. A thorough examination of the pattern of results in the study also suggests that the thought coding and judgment tasks created a response set which carried over to the GSC and IPC measures. Finally, examination of the GSC and IPC scales in two separate investigations showed that the scales may not be unidimensional.

Although "chronic" traits as conceptualized by Wright represent an important determinant of cognitive responses and message acceptance, work by Epstein (1979) suggests that multiple measures and multiple observations yield more meaningful and consistent relationships between trait and behavioral variables. This approach is necessarily more demanding of the research, however an interactionist perspective appears to necessitate such rigor.

In summary, for the study of GSC/IPC future research should:

- (1) reassess the dimensionality of the GSC/IPC scales via factor analysis;
- (2) address the GSC/IPC-social desirability issue for evidence of possible format changes in GSC/IPC scales (e.g., forced choice responses);
- (3) investigate the social desirability aspects of thought reporting and coding tasks; and
- (4) following an interactionist's position:
 - (a) measure traits over several occasions prior to the experimental situation, and
 - (b) observe cognitive responses and attitude change in several stressful information processing situations. These multiple measurements and observations over several similar situations will provide a stronger test of the cognitive mediators of persuasion.

References

- Alker, H. A. (1972), "Is Personality Situationally Specific or Intrapsychically Consistent," Journal of Personality, 40, 1-16.
- Bagozzi, Richard P. (1980), Causal Models in Marketing, New York: John Wiley and Sons, Inc.
- Bem, Daryl J. and Allen, Andrea (1974), "On Predicting Some of the People Some of the Time: The Search for Cross-Situational Consistencies in Behavior," Psychological Review, 81, 506-20.
- Bither, Stewart W. and Wright, Peter (1973), "The Self Confidence - Advertising Response Relationship: A Function of Situational Distraction," Journal of Marketing Research, 10, 146-52.
- Bowers, K. S. (1973), "Situationism in Psychology: An Analysis and A Critique," Psychological Review, 80, 307-36.
- Burt, R. S. (1976), "Interpretational Confounding of Unobserved Variables in Structural Equation Models," Sociological Methods and Research, 5, 3-52.
- Cacioppa, John T., Sandman, C. A. and Walker, B. B. (1978), "The Effect of Operant Heart Rate Conditioning on Cognitive Elaboration and Attitude Change," Psychophysiology, 15, 330-8.
- Cohen, Arthur R. (1959), "Some Implications of Self-Esteem for Social Influence," in Personality and Persuasibility, eds. Carl I. Hovland and Irving L. Janis, New Haven: Yale University Press, 102-20.
- Eagley, Alice H. (1980), "Recipient Characteristics as Determinants of Responses to Persuasion," in Cognitive Responses to Persuasion, ed. T. C. Brock, New York: McGraw-Hill.
- Edell, Julie and Mitchell, Andrew (1978), "An Information Processing Approach to Cognitive Responses," in Research Frontiers in Marketing, ed. S. Jain, Chicago: American Marketing Association, 178-83.
- Endler, N. S. (1966), "Estimating Variance Components from Mean Squares for Random and Mixed Effects Analysis of Variance Models," Perceptual and Motor Skills, 22, 559-70.
- _____ and Hunt, J. McV. (1968), "S-R Inventories of Hostility and Comparisons of the Proportion of Variance from Persons, Responses, and Situations for Hostility and Anxiousness," Journal of Personality and Social Psychology, 9, 309-15.
- _____ and _____ (1969), "Generalizability of Contributions from Sources of Variance in the S-R Inventories of Anxiousness," Journal of Personality, 37, 1-24.
- _____, _____ and Rossenstein, A. J. (1962), "An S-R Inventory of Anxiousness," Psychological Monographs, 76, (17, Whole No. 526).
- Ekehammer, B. (1974), "Interactionism in Personality from a Historical Perspective," Psychological Bulletin, 81, 1026-48.
- Epstein, S. (1979), "The Stability of Behavior: I. On Predicting Most of the People Much of the Time," Journal of Personality and Social Psychology, 37, 1097-1126.
- Greenwald, Anthony G. (1968), "Cognitive Learning, Cognitive Response to Persuasion, and Attitude Change," in Psychological Foundations of Attitudes, eds. A. G. Greenwald, T. C. Brock, and T. M. Ostrom, New York: Academic Press.
- Jones, E. E. and Nisbett, R. E. (1971), The Actor and Observer: Divergent Perceptions of the Cases of Behavior, New York: General Learning Press.
- McGuire, William J. (1969), "Personality and Susceptibility to Social Influence," in Handbook of Personality Theory and Research, eds. E. F. Borgatta and W. W. Lambert, Chicago: Rand McNally.
- Mischel, W. (1968), "Toward a Cognitive Social Learning Reconceptualization of Personality," Psychological Review, 80, 252-83.
- Osterhouse, Robert A. and Brock, Timothy C. (1970), "Distraction Increases Yielding to Propaganda by Inhibiting Counterarguing," Journal of Personality and Social Psychology, 15, 344-58.
- Petty, Richard E., Wells, Gary and Brock, Timothy (1976), "Distraction Can Enhance or Reduce Yielding to Propaganda: Thought Disruption vs. Effort Justification," Journal of Personality and Social Psychology, 34, 874-84.
- _____ and Cacioppa, John (1979), "Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Responses," Journal of Personality and Social Psychology, 37, 1915-26.
- Shweder, R. A. (1975), "How Relevant is An Individual Difference Theory of Personality?" Journal of Personality, 43, 455-84.
- Swasy, J. L. (1980), "Conceptual and Modeling Issues in a Cognitive Response Perspective on Persuasion Processes," unpublished Ph.D. dissertation, University of California, Los Angeles.
- Wright, Peter (1973), "The Cognitive Processes Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-67.
- _____ (1975), "Factors Affecting Cognitive Resistance to Advertising," Journal of Consumer Research, 2, 60-7.

DISCUSSION

Jagdish N. Sheth, University of Illinois

Introduction

There is very little in common among the three research papers included in this session except that they all relate to very different aspects of consumer information as opposed to consumer choice behavior. Therefore, I will first discuss each research paper separately and then make some observations about consumer information as an area of research and theory. In order to conserve space, I will not summarize the findings and conclusions of each paper. Instead, I will simply discuss its strengths and weaknesses.

Consumer Information Programs

The first paper by Deshpande and Krishnan on establishing public policy priorities for consumer information programs based on market research findings is a very worthwhile and useful contribution to the discipline for at least three reasons. First, the research study adds to our very limited substantive knowledge about the information needs of the elderly consumers. Second, the authors provide a very neat conceptualization of information deficiency as a function of perceived need for information and perceived difficulty in obtaining it. In fact, the development of Information Deficiency Index may be the single most contribution of this paper. Finally, it represents a good application of Thurstone scaling procedures which have been neglected in the past in consumer research.

The paper, however, suffers from the "analysis overkill" syndrome. First of all, do policy makers really need or worse yet care about the relative magnitudes of information deficiency across the nine buying experiences? Would they not be just as happy and content to know the ranking of nine buying experiences? I seriously doubt that policy makers can even utilize the Thurstone interval scale points for budget allocation purposes since it is arbitrary, lacking in an absolute anchor point, and definitely subject to change by addition of other buying experiences to the list.

Ironically, the authors do have a more meaningful and useful quantitative scale in their Information Deficiency Index. It is an absolute score ranging in value from one to three. Average or median scores of nine buying experiences with respect to the index do represent aggregate market scale points which are at least interval scaled and, therefore, quantitative measures. In short, I fail to see any need for applying the Thurstone Case V Model to the data.

Finally, there are so many other more meaningful ways to generate relative magnitudes of information deficiency out of this data bank. For example, calculating the proportion of consumers who answered yes to both perceived need and access difficulty will provide a ratio scale which will remain invariant to addition of other buying experiences. Alternatively, one can calculate the mean deficiency score and perform a normal deviate analysis to generate standard scores of information deficiency among the nine buying experiences. Of course, this distribution will be subject to sampling problems but with a large number of buying experiences, it will reach limits and stabilize as a normal distribution.

A more interesting and useful analysis of the data is to identify those elderly consumers who really need information. For example, it would be very interesting to profile the "yes-yes" group (score three) in terms of demographic and socioeconomic variables: do they come from certain ethnic

minorities or are they concentrated in some regions of the country? If our past research is any guide, it is very likely that there is a hardcore segment within the elderly consumers which is most information deficient across all nine buying experiences. The public policy makers and the society as a whole will be better off if consumer information programs are targeted toward those who really need them. Otherwise, we will keep on repeating our past program mistakes such as unit pricing, nutrition labeling, and truth-in-packaging. As I have recently stated, unfortunately one law of consumer behavior is that those consumers who need information, do not use it (Sheth 1979) and, therefore, public policy efforts should be targeted toward them in terms of understanding and motivating them to use information.

Information Content of Non-Information

The second research paper by David Finn is an interesting study on how inferential beliefs enable consumers to add and subtract from what is communicated by the experimenter or the marketer. This study reminds of the neglected concept of stimulus-as-coded (s-a-c) proposed by Howard-Sheth theory of buyer behavior (Howard and Sheth 1969). According to their theory, prior attitudes and beliefs control overt search, attention and perceptual bias mechanisms with which stimulus-as-presented is modified into stimulus-as-coded. Recently, I have also discussed how advertising impacts on the consumer including many incidental, unintended and negative effects based on the stimulus-as-coded concept (Sheth 1974). I am, therefore, not as surprised as David Finn that consumers use phantom information.

There are also several methodological and conceptual problems with this study. First of all, post-test-only control group design simply cannot measure change as asserted by the author. It only measures differences between two groups, which is not the same thing as pre-post change. Second, although the control and test groups are randomly divided and belong to the demographically homogeneous population, it does not ensure that the two groups are psychologically homogeneous in terms of general toothpaste beliefs. Since the post-test control group is used as a benchmark, it is most critical to validate that the test group also has a comparable belief profile. One way to do this is to obtain population parameters about toothpaste beliefs and compare them with the sample parameters (control group). Otherwise, one can easily obtain perhaps a very different set of results by using another control group.

Finally, a closer examination of the mean differences among beliefs reveals that the greatest differences occurred for most probable (seal of approval, about the same price) and least probable beliefs (made with new type fluoride). Furthermore, lack of confirmation reduced the probability of most probable beliefs (-2.391 and -1.153), whereas highly assertive message increased the probability of least probable belief (1.300). I find this as evidence of what one would expect from all the cognitive consistency theories including the cognitive dissonance, congruence and assimilation-contrast models.

My own view is that projective techniques, rumor-transfer, story-telling and word association types of research may

be more relevant if we want to know the information content of non-information.

Information Processing Traits

It is amusing to critique a critique when you are not the author who is critiqued! I am sure Peter Wright does not need me to respond to the criticisms levied by Munch and Swazy. And, I am not eager to defend this research tradition either, simply because it is statistically uninteresting and frustrating by being loaded with correlations which gain significance from large samples rather than any evidence of true relationships between traits and information processing.

I do agree with Munch and Swazy about these three methodological and conceptual problems in Wright's study. First, there is a very definite social desirability bias in scales designed to measure General Social Confidence (GSC) and Information Processing Confidence (IPC). This is easy to detect from the skewed distributions of subject responses. In fact, I will go one step further and assert that most correlations based on these scales are likely to be biased and inflated and, therefore, encourage the experimenter to eagerly commit the type one error (rejecting the null hypothesis when it is true).

Second, there is conceptually strong support for the multidimensionality of constructs as complex as GSC and IPC. If they are multidimensional, as Munch and Swazy claim from their own research findings, it is very difficult to make causal inferences based on univariate analysis of variance techniques. You need to utilize path analysis or structural equation modeling procedures for testing the hypothesis.

Third, Munch and Swazy correctly point out the problem of intercorrelation between GSC and IPC. One can even argue that General Social Confidence (GSC) cannot be achieved by or attributed to someone who does not have Information Processing Confidence (IPC). In our society, individuals who think slowly, cannot concentrate, are at a loss for words or not quick-witted, do not emerge as either self-confident or socially confident individuals. I am, therefore, quite surprised that Wright chose to treat them as separate and orthogonal factors in his research study.

The Wright study and its critique by Munch and Swazy once again points out the age old admonition: Measure your independent constructs such as GSC and IPC more specific to the criterion situation. We must develop statements for both GSC and IPC which are directly related to advertising or at least consumer information oriented to ensure that both sides of the equation are at the same level of correspondence or specificity. Both personality researchers (Cohen 1967), and attitude researchers (Fishbein and Ajzen 1975) have belabored this point for so many years. However, we seem to ignore their admonition.

Observations on Consumer Information Research

Consumer information as an area of scientific research and theory is at its infancy in consumer behavior. Howard and Sheth (1969) pointed out, quite some time ago, that we know at least something about the learning constructs (attitudes, intention, beliefs and motives), but we virtually know very little about the perceptual constructs (attention, ambiguity, perceptual bias, overt search). It is, therefore, premature to conduct deductive research, either by developing theories or by borrowing constructs from other disciplines in this area. Instead, we must do considerable amount of empirical inductive research. In short, we must learn how to crawl before we start walking or worse yet, running.

Accordingly, we must lean toward more exploratory and qual-

itative research tools and tactics such as focused group interviews, projective techniques, clinical methodologies, and nonstructured surveys rather than experimental designs, construct development and measurement or laboratory studies.

In my opinion, consumer behavior as a discipline has already begun the shift from learning constructs such as attitudes and beliefs to the perceptual constructs such as consumer information's content, storage, retrieval and processing aspects. Contrast this year's ACR Proceedings with the proceedings of 1976 or 1977, for example.

References

- Cohen, Joel B. (1967), "An Interpersonal Orientation to the Study of Consumer Behavior," Journal of Marketing Research, 4, 270-80.
- Fishbein, M., and Ajzen, I. (1975), Belief, Attitude, Intention and Behavior, Reading, Mass: Addison-Wesley.
- Howard, J. A., and Sheth, J.N. (1969), The Theory of Buyer Behavior, New York: John Wiley & Sons.
- Sheth, J. N. (1979), "Surpluses and Shortages in Consumer Behavior Theory and Research," Journal of the Academy of Marketing Science, 7, 414-27.
- (1974), "Measuring Advertising Effectiveness: Some Theoretical Considerations," Journal of Advertising, 3, 6-11.

EVOKED SET FORMATION AND COMPOSITION:
AN EMPIRICAL INVESTIGATION UNDER A ROUTINIZED RESPONSE BEHAVIOR SITUATION

Jacques E. Brisoux, University of Quebec at Three-Rivers
Michel Laroche, Concordia University

Abstract

The purpose of this study is to test some hypotheses regarding evoked set formation under routinized response behavior for regular beer drinkers under a specific scenario. By using a decomposition approach of the perceptual product space, and individual configurations, it is found that the best decision model is the conjunctive one, and that all brands included in the evoked set are previously tried.

b. Evoked sets are formed by considering all brands in the awareness set at the same point in time, and in this case the most appropriate decision rules are the linear compensatory and lexicographic models. If a new brand enters the awareness set, the whole process is repeated. The results of Parkinson and Reilly are congruent with the second conceptualization, although the authors call for caution in interpreting their results due to a number of methodological factors: Low sample sizes, student respondents, halo effects, and measurement problems (1979 p. 229).

Introduction

Evoked Set

The proliferation of brands occurring in the consumer goods markets create information processing problems for the consumers. They must often devise means to simplify their purchase decisions. One of the results of this simplification process is the existence of what Howard has referred to as "evoked set," which include the brands the buyer considers when he contemplates purchasing a unit of the product class (Howard 1963, Howard and Sheth 1969). First empirically studied by Campbell (1969), the concept of evoked set has caused a growing interest from researchers in consumer behavior (Ostlund 1973, Jarvis and Wilcox 1973, Grønhaug 1973-1974, May and Homans 1976, Williams and Etzel 1976, Maddox et al. 1977, Belonax and Mittelstaedt 1977, Parkinson and Reilly 1979).

Inert and Inept Sets

Narayana and Markin (1975) have suggested a more complete conceptualization of the output of the consumer's simplification process: The individual may not be aware of all the brands available (total set) at time of purchase. This leads to the concepts of awareness set and unawareness set. Within the former any brand is categorized as either evoked, inept or inert. According to Narayana and Markin, a brand would belong to the evoked set if it were considered for purchase by the consumer and evaluated positively. It would be in the inept set if it were rejected from the consumer's purchase consideration and evaluated negatively. Finally, it would be in the inert set if it were neither accepted nor rejected by the consumer and evaluated neither positively nor negatively (p. 2). The measurement of the first two sets is done through direct questioning only on the buying consideration dimension, while the inert set is inferred by deduction of the first two from the awareness set (p. 3). As they noted later, the evaluations attached to the different sets are a precondition to the consumer choice behavior (Narayana and Markin 1976).

Evoked Set Formation and Composition

In his review of the papers by Belonax (1979), May (1979), and Parkinson and Reilly (1979), Myers (1979) raises a number of conceptual and methodological issues concerning the temporal aspects of evoked set formation and composition. He suggests two alternative conceptualizations of these processes:

a. "Evoked sets are formed over time by considering each new entry into the awareness set whenever it comes along. It would seem that a conjunctive or a disjunctive processing model would be much more appropriate for this type of evaluation." (p. 236).

Myers (1979) describes a phasing model by which consumers may establish their evoked set by using one cutoff model (i.e., conjunctive or disjunctive), followed by an evaluation for each brand within this set for purposes of choice by using a compensatory model. This conception is congruent with the findings by Best (1976), and Pras and Summers (1975), the propositions of Belonax and Mittelstaedt (1977) and Belonax (1979), as well as the paradigm developed by May (1979). This paradigm combines the information processing approach proposed by McGuire (1976) with the learning theoretical approach developed by Howard (1977). In particular, May suggests that Howard's three stages of concept learning, i.e., concept formation, attainment and utilization, correspond to the three levels of information processing complexity related to the exposure search and retrieval steps. The highest level of information processing complexity is required for the first learning stage (concept formation). The third learning stage (concept utilization) requires the least complexity.

Also, according to Howard (1977), at the concept formation stage, consumers identify their evoked set in terms of untried brands only. At the concept attainment stage they identify it in terms of mixed brands (i.e., untried and tried ones). At the concept utilization stage (Routinized Response Behavior), they identify it in terms of tried brands only.

The purpose of this research is to test what Myers called a phasing model, especially its first stage (i.e., a cutoff model is more appropriate), and to test some part of May's conceptualization of combining both learning and information processing hypotheses, including Howard's proposition of the evoked set composed of tried brands only in an RRB situation.

Methodology

In order to carry out this research and provide some control over possible sources of variation, a number of strategic decisions were made.

Purchase Scenario and Ideal Universe

The product class which was selected was beer to be purchased for regular and individual consumption by male beer drinkers in the evening before dinner within one major metropolitan area of Quebec (Canada), namely the Three-Rivers area. The scenario was selected because it is typical of beer consumption in that area. This decision is believed to be critical since the choice of a decision rule may be affected by situational factors (Hansen 1976), and

prior familiarity (Park 1976). The area selected is relatively homogeneous in terms of socio-demographic characteristics, and the ideal universe is defined as male drinkers residing within the region at the time of the study and consuming at least three 12 oz. bottles of beer a week. This operationalization of a regular drinker corresponds to the industry's definition. Finally, beer is a product at the maturity stage for most regular drinkers, and the purchase decision may be classified as Routinized Response Behavior (Howard 1977).

Sampling Method and Sample

The operational universe is restricted to the male regular beer drinkers, french speaking, 18 years old or more residing in the Three-Rivers, Shawinigan Metropolitan area. This area includes the cities of Three-Rivers, Cap-de-la-Madeleine, Shawinigan and Grand-Mère in the Province of Quebec and is located 90 miles east of Montreal. Its total population is about 150,000 people. In accordance with the rules of area sampling, residential streets were randomly selected from each city. On each street residences were then systematically selected. In each housing unit one regular male beer drinker was personally interviewed. In the case of several regular drinkers within a unit, only one was interviewed by randomly selecting by age. A total of 732 male subjects were contacted. Of these subjects 453 qualified as regular beer drinkers and 375 questionnaires were usable for the purpose of a much larger study than the one reported here and dealing with the content and size of evoked set (Brisoux 1980).

Data Collection

The data collection was done through personal structured interviews at the interviewee's residence. Each interviewer had an individual two hours training period and his work was checked regularly by a supervisor. Interviews were conducted mainly during the evening, and lasted about 80 minutes.

Questionnaire

The parts of the interview of concern here consisted of getting essentially three kinds of information:

- (1) Brand perceptions on selected scales: The selection of specific scales was made from the results of a factor analysis of the brand perceptions collected previously in a large scale (private) survey which used 37 scales. The decision to limit to 11 the number of scales was due to the fact that respondents had to rate 13 brands of beer available plus an ideal brand. Respondents were asked to rate all brands they were aware of one attribute at a time on a 9-point bipolar scale like the following one:

Light • Heavy

The other labels used were

Sweet/Bitter
 No/Much after taste
 Easy/Difficult to digest
 Does not create/Creates some discomfort (headache, stomach ache, etc.)
 Has less/more calories than the average beer
 Very/Little pronounced taste
 Very/Little tart taste
 Becoming less/more popular
 Would/Would not serve it to my friends
 Less/More salty taste than the average beer

Finally the wording of each scale was based on the most common semantic structure of regular beer drinkers

(Allaire 1972).

- (2) Brand classifications: During the interview, each respondent was given a set of thirteen cards, in random order, each with a different authentic beer label (aided recall procedure). The respondent was then asked to indicate the brands of which he was aware, and the others were discarded. Next, using the remaining cards, he was asked through a sequence of questions to classify each brand into one of three categories (Narayana and Markin 1975):
 - (a) The evoked set consisting of the brand which he would consider buying for his personal consumption in the evening before dinner, plus other acceptable brands to be bought for the same purpose;
 - (b) The inept set consisting of the brands the consumer would refuse to buy for the same consumption situation;
 - (c) The inert set consisting of the remaining brands, i.e., those which he would not consider nor reject.
- (3) Prior experience: The respondent was given a list of brands and asked to indicate which ones he had tried before.

Analysis

The purpose of the analysis was to develop a perceptual space for each segment within our sample and decompose this space according to some compensatory and non-compensatory decision models. The analysis proceeded in steps:

Step 1: By using the data on brand perceptions, determine if there are any segments of homogeneous perceptions.

Step 2: For each segment, derive a product space by using Johnson's procedure (1971). In this procedure a product space is built by means of multiple discriminant analysis. This technique permits to find linear orthogonal combinations of attributes which best separate the brands, maximizing the ratio of between-brand to within-brand variances. Average scores on each discriminant function permits to position each brand in the discriminant space.

Step 3: For each individual configuration, i.e., following a micro approach, decompose the product space a posteriori according to six decision models; three non-compensatory models: Conjunctive, disjunctive, lexicographic; and three compensatory models: One linear additive and two geometric ones (centered on the ideal point or the preferred brand). Figure 1 graphically represents how each model works: The shaded area would be the evoked set zone if each model were true. For each case, the evoked set zone is defined as the smallest area containing all evoked brands. A decision rule is considered successful if that zone does not contain any brand other than the evoked ones. If not, find out to which category these "offenders" belong.

Step 4: Determine the composition of the individual evoked sets in terms of prior trial of the brands.

Findings and Interpretations

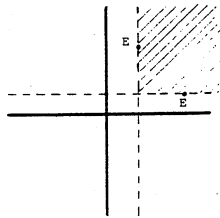
The findings will follow the four steps defined in the previous section.

Step 1: Identify segments with homogeneous perceptions. In order to identify clusters of respondents on brand perceptions, two different clustering techniques were used, that is the Howard-Harris cluster analysis, and the BMDP-2M

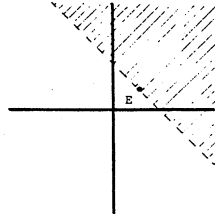
FIGURE 1

EVOKED SET ZONES AS DEFINED BY
THE LOGIC OF EACH DECISION RULE
(ASSUMING TWO DISCRIMINANT FUNCTIONS)

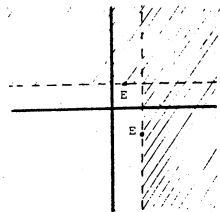
The shaded area is the smallest one which (1) is representative of the decision rule, and (2) contains all evoked brands



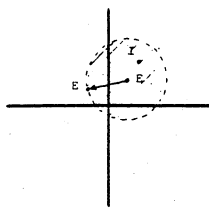
(a) Conjunctive Rule:
Predetermined cutoffs
on all attributes.



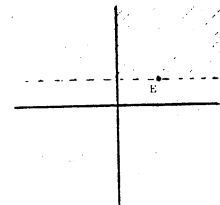
(d) Linear Compensatory Rule:



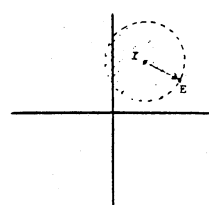
(b) Disjunctive Rule:
Predetermined cutoffs
on at least one
attribute.



(e) Geometric Compensatory Rule:
Centered on the most
preferred brand.



(c) Lexicographic Rule:
Rank order based on
most important attribute. If tied, use
second most important.



(f) Geometric Compensatory Rule:
Centered on the ideal
point.

hierarchical cluster analysis. The Howard-Harris program uses the criterion of minimum within-group variance at each level of clustering to form groups of objects using an objects-by-variable matrix (Howard & Harris 1966). The BMDP-2M procedure is also a hierarchical method. The two cases having the shortest distance between them are amalgamated and treated as one case and then, in turn, clustered with others. The procedure continues until all cases and clusters are amalgamated into one cluster.

Both techniques led to a single cluster: The whole group of 375 respondents, and, thus, the rest of the analysis is done on the total group.

Step 2: Generate a perceptual product space. The data on brand perception is then used for a multiple discriminant analysis of scale for 375 respondents, following the method developed by Johnson (1971). This method is particularly

suited to our needs in terms of stability, uniqueness of the solution and euclidian distances (Johnson 1971). It was found that three discriminant functions are significant and explain respectively 77.8%, 17.5% and 2.3% of the total variance, for a cumulative percentage of 97.6% of total variance. The first function may be interpreted as the social dimension of beer drinking (before), the second as the strength of beer (during), and the third as another taste dimension (after) concerning physiological effects (after taste, easy to digest).

Step 3: Determine the best decision rule for evoked set formation. In order to find out the best decision for each individual, we derived individual scores on the three significant discriminant functions. For each respondent, and each decision rule, we determined whether or not a particular decision rule leads to a zone containing only evoked brands. For example, in the geometric compensatory rule centered on the ideal point (Best 1976), the evoked set zone is defined as the sphere centered on the ideal point and the radius of which is the evoked brand farthest away. We then looked to see if one or more brand(s) other than an evoked one is contained within the zone (sphere). For the cutoff models, the cutoff point is inferred by taking the evoked brand with the lowest score on that dimension. The evoked set zone is then the area represented in dark in Figure 1, and an examination of that zone reveals whether or not it contains brands other than evoked ones.

The percentage of cases for which each rule is successful is presented in Table 1 as well as the classification of the brands belonging to each set violating that rule. It is clear that the best rule is the conjunctive model for which no other brand was found within the centroid defined by the three cutoff points for 72.0% of all respondents. Most violations of the rule occurred with inert brands, with only 5.3% of the cases for which an inept brand was found within the centroid. This result can also be viewed as follows: The percentage of success increases dramatically as we move from the largest area, i.e., the disjunctive rule with 9.6% success, to a smaller area, i.e., the lexicographic rule with 46.1% success, to an even smaller area, i.e., the conjunctive rule with 72.0% success. A similar reasoning with the three compensatory models reinforces the conclusion that the process of evoked set formulation is, in our case, one of inclusion based on cutoff points for all three major dimensions of the product space. This conclusion is consistent with Myers (1979) expectations, as well as those by Pras and Summers (1977). It is also consistent with the findings of Best (1976). On the other hand, our findings do not confirm those of Parkinson and Reilly (1979) although they followed a composition approach. They do report that the questions directly measuring the conjunctive or disjunctive cutoffs were fairly difficult for the subject to complete, and that these response errors were probably magnified in their simulation procedures.

Step 4: Determine prior trial of evoked brands. A comparison was made between the brands in the evoked set on one hand, and those which the respondent had indicated as tried brands. It was found that in 95.4% of all cases, all brands in the evoked set had been previously tried. The remainder of the cases, i.e., 4.6%, may be considered as an error term if we set a level of significance of $\alpha=5\%$, in which case the critical value is $p_c=.069$. Therefore we fail to reject the null hypothesis that all brands in the evoked set have been previously tried. This is consistent with Howard's theory of concept learning in a Routinized Response Behavior Situation (Howard 1977, May 1979).

Limitations

A first limitation of the present study deals with the problem of controlling for the homogeneity of respondents since many factors may affect the choice of the decision rule as reviewed by Belonax (1979, p. 232). Our decision to limit

TABLE 1
RESULTS OF THE DECOMPOSITION APPROACH

Decision rule	Percent of cases for which the zone of evoked set brands defined by the decision rule contains		
	No other brand	At least one inept brand	At least one inert brand
Non-compensatory			
Conjunctive	72.0%	5.3%	23.5%
Disjunctive	9.6%	60.0%	69.1%
Lexicographic	46.1%	17.9%	44.0%
Compensatory			
Linear	42.9%	6.7%	29.6%
Geometric - centered on ideal point	41.9%	21.9%	48.0%
Geometric - centered on most preferred brand	53.9%	14.7%	36.8%

our research to a RRB situation, a specific scenario, regular male beer drinkers, in a relatively homogeneous metropolitan area as well as two attempts at clustering respondents provided some measure of control. But other factors may be introduced such as breadth of categorization (Pettigrew 1956) which is one dimension of cognitive style, as well as perceived risk (Bettman 1973).

A second limitation may lie in our procedure of decomposing the product space into areas based on a specific decision rule on one hand and actual brand classification on the other hand. In effect, we tested the logic of applying a specific rule more than reconstructing the rule itself as in Parkinson and Reilly (1979).

Third, the boundaries of our evoked set zones were approximated by looking at the smallest area containing all evoked brands. There may be some value in doing some sensitivity analysis by varying the cutoff points to allow for a margin of error. Finally, we only looked at minimum cutoff points, but not at maximum ones. This may apply to only our second and third dimensions since the first dimension is the social image of beer which accounted for 77.8% of total discrimination.

Finally, the study is within a static context, and not in a dynamic one as suggested by Myers (1979).

Summary and Conclusion

This research has attempted to focus on specific questions regarding Myers' phasing model, as well as May's contention that both learning theory and information processing theory are necessary to explain evoked set formation and composition.

We selected a product in the maturity stage and consumers in an RRB situation, faced with a specific scenario, and homogeneous in their perception of the available brands. For each respondent, we generated his perceptual space, and we used a decomposition approach to test three non-compensatory rules and three compensatory rules. In a perfect criterion basis, the best rule was found to be the conjunctive model, which is consistent with previous expectations and findings. Finally, all brands in the evoked set were

found to have been previously tried, which is consistent with Howard's learning theory.

The decomposition approach followed by this study allowed us to examine and compare the output at the maturity stage (RRB situation) from both information processing activities and concept learning. From an information processing viewpoint, we gathered brand perceptions on several attributes and constructed individual product spaces. From a concept learning viewpoint, all brands in the evoked set were previously tried by most respondents. Thus a comparison of brand classifications and product space configurations leads to a perfect match for 72% of all respondents for the conjunctive rule. Thus, we may see here a convergence of both theories as hoped by May (1979).

References

- Allaire, Yvan (1972), "The Measurement of Heterogeneous Semantic, Perceptual and Preference Structures," unpublished Ph.D. dissertation, Massachusetts Institute of Technology.
- Belonax, Joseph J. (1978), "Decision Rule Uncertainty, Evoked Set Size, and Task Difficulty as a Function of Number of Choice Criteria and Information Variability," in *Advances in Consumer Research*, Vol. 6, ed. William L. Wilkie, Urbana, IL: Association for Consumer Research, 232-235.
- Belonax, Joseph J. Jr. and Middlestaedt, Robert A. (1978), "Evoked Set Size as a Function of Number of Choice Criteria and Information Variability," in *Advances in Consumer Research*, Vol. 5, ed. H. Keith Hunt, Urbana, IL: Association for Consumer Research, 48-51.
- Best, Roger J. (1976), "The Predictive Aspects of a Joint-Space Theory of Stochastic Choice," *Journal of Marketing Research*, 13, 198-204.
- Bettman, James (1973), "Perceived Risk and its Components: A Model and Empirical Test," *Journal of Marketing Research*, 10, 184-190.
- Brisoux, Jacques E. (1980), "Le Phénomène des Ensembles Evoqués: Une Etude des Dimensions Contenu et Taille," unpublished Ph.D. dissertation, Université Laval.
- Campbell, Brian M. (1969), "The Existence and Determinants of Evoked Set in Brand Choice Behavior," unpublished Ph.D. dissertation, Columbia University.
- Grønhaug, Kjell (1973), "Some Factors Influencing the Size of the Buyer's Evoked Set," *European Journal of Marketing*, 7, 3, 232-241.
- Hansen, Fleming (1976), "Psychological Theories of Consumer Choice," *Journal of Consumer Research*, 3, 117-142.
- Howard, John A. (1963), *Marketing Management, Analysis and Planning*, Irwin.
- Howard, John A. (1977), *Consumer Behavior: Application of Theory*, New York: McGraw-Hill.
- Howard, John A. and Sheth, Jagdish (1969), *The Theory of Buyer Behavior*, New York: Wiley.
- Jarvis, Lance P. and Wilcox, James B. (1973), "Evoked Set Size - Some Theoretical Foundations and Empirical Evidence," in *Combined Proceedings*, ed. Thomas V. Greer, Chicago: American Marketing Association, 236-240.
- Johnson, Richard M. (1971), "Market Segmentation: A Strategic Tool," *Journal of Marketing Research*, 5, 13-18.

- Maddox, R. Neil, Grønhaug, Kjell, Homans, Richard E., and May, Frederick E. (1978), "Correlates of Information Gathering and Evoked Sets Size for New Automobile Purchasers in Norway and the U.S.," in Advances in Consumer Research, Vol. 5, ed. H. Keith Hunt, Chicago: Association for Consumer Research, 167-170.
- May, Frederick E. (1979), "Evoked Set Formation and Composition: The Learning and Information Processing Hypotheses," in Advances in Consumer Research, Vol. 6, ed. Wilkie, Urbana, Il.: Association for Consumer Research, 222-226.
- May, Frederick E., and Homans, Richard E. (1976), "Evoked Set Size and the Level of Information Processing in Product Comprehension and Choice Criteria," in Advances in Consumer Research, Vol. 4, ed. William D. Perreault, Jr., Urbana Ill.: Association for Consumer Research, 172-175.
- McGuire, William J. (1976), "Some Internal Psychological Factors Influencing Consumer Choice," Journal of Consumer Research, 2, 302-319.
- Myers, James H. (1979), "Methodological Issues in Evoked Set Formation and Composition," in Advances in Consumer Research, 6, ed. William L. Wilkie, Urbana, Il.: Association for Consumer Research, 236-237.
- Narayana, Chem L., and Markin, Ron J., Jr. (1975), "Consumer Behavior and Product Performance: An Alternative Conceptualization," Journal of Marketing, 39, 1-6.
- Narayana, Chem L., and Markin, Ron J. Jr. (1976), "An Explanation," Journal of Marketing, 40, 60.
- Park, Whan C. (1976), "The Effect of Individual and Situation-Related Factors on Consumer Selection of Judgemental Models," Journal of Marketing Research, 13, 144-151.
- Parkinson, Thomas L., and Reilly, Michael (1979), "An Information Processing Approach to Evoked Set Formation," in Advances in Consumer Research, Vol. 6, ed. William L. Wilkie, Urbana, Il.: Association for Consumer Research, 227-231.
- Pettigrew, Thomas F. (1956), "The Measurement and Correlates of Category Width as a Cognitive Variable," Journal of Personality, 6.
- Pras, Bernard, and Summers, John (1975), "A Comparison of Linear and Nonlinear Evaluation Process Models," Journal of Marketing Research, 12, 276-281.
- Ostlund, Lyman E. (1973), "Evoked Set Size: Some Empirical Results," in Combined Proceedings, Series No. 35, ed. Thomas V. Greer, Chicago: American Marketing Association, 226-230.
- Sherif, Caroline W., Sherif, Muzafer and Nebergall, Robert E. (1965), "Attitude and Attitude Change: The Social Judgement - Involvement Approach," Philadelphia: Saunders.
- Williams, Robert G. and Etzel, Michael J. (1976) "An Investigation and Extension of the Evoked Set Concept Applied to Consumer Durables," Proceedings of the Southern Marketing Association.

THE COMPENSATORY DIMENSION IN
SUBJECTIVE EVALUATION PROCESSES:
A MULTIMETHOD VALIDATION

Ruby Roy Dholakia, Kansas State University

Abstract

One of the characteristics of evaluation processes of multi-attributed alternatives is that of compensatoriness which defines the nature and degree of trade-offs between attributes. An empirical investigation is reported in this study that utilizes multiple methods to analyze this dimension. The objective of the paper is to represent the compensatory dimension in terms of a single quantitative measure and find qualitative support for the reported findings. Judgments of three decision makers are analyzed using these methods and the findings are then used to select representation models of the evaluation processes.

Introduction

Early attempts to describe and analyze evaluation processes revolved around the model versus process controversy (Goldberg 1968, Wright 1975). While there is ample evidence to support differences in processes, both across individuals and situations, these differences are still viewed at the model level (Park 1976). Recent approaches have been attempting to isolate these differences at specific dimension level such as the influence of time horizon on process linearity (Wright and Weitz 1977).

When multiple attributes are used by individuals to evaluate alternatives, one component of the evaluation process is that of inter-attribute compensation. Two attributes are compensatory when a decrease in the value of one is balanced by an increase in the value of the second attribute. The amount required for balancing is likely to vary and will be determined by a variety of factors. Evaluation processes will differ, therefore, in terms of compensatoriness.

In traditional economic theory, the marginal rate of substitution between two goods was an index of compensation (Henderson and Quandt 1971). Recent formulations have shown the preferences for goods to be derived from the preferences for the goods' attributes (Lancaster 1966) and the marginal rate of substitution has been extended to relationships between attributes (Lancaster 1972). Raiffa (1968) showed through isopreference curves that the rate of substitution between an unit of attribute X and an unit of attribute Y can be constant or variable. A variable substitution rate depends on the specific attribute levels while a constant rate is independent of attribute values.

The compensatory dimension can, therefore, be represented as a continuum (degree) rather than as a dichotomy (compensatory). This implies that evaluation of multi-attributed objects rather than being categorically a compensatory or a non-compensatory process, can be characterized as a process with varying degrees of compensatoriness. Raiffa's (1968) variable rate of substitution supports this view. In the utility model proposed by Keeney (1974), the different values that the multiplicative constant W can assume is based on the variable rate of substitution.

There are two important implications of viewing the compensatory dimension as a continuum. When various models are used to describe a process or predict its outcome, the current practice is to distinguish these models as being compensatory or non-compensatory. However, if the compensatory dimension is viewed as a continuum, then greater

discriminations must be made among these models and their assumptions about the degree of compensatoriness.

Second, it allows conceptualization of the influence of various situational and individual variables on the evaluation process. For example, the interaction of a specific set of variables may be to make the evaluation process more or less compensatory. If a methodology can be adopted to quantify the degree of compensatoriness, then changes in magnitude can be hypothesized and measured.

Investigation of Compensatoriness

There is limited empirical evidence pertaining to the rate of compensation among attributes. Utility theorists have used goods rather than attributes and assumed the goods to be uni-attributed. MacCrimmon and Toda (1969) used money and ballpoint pens in one study and money and pastries in the second. Similarly, Rousseaus and Hart (1951) used goods rather than attributes when obtaining trade-offs between bacon and eggs.

There are some attempts by consumer behavior researchers to assess attribute trade-offs. Johnson (1974) reported a methodology which used paired comparisons to obtain trade-offs from which utilities of individual attributes as well as overall utility were derived. Fiedler (1972) used the methodology to obtain evaluation of condominiums and found it to be a good predictor of preferences. Hauser and Urban (1979) used an alternative methodology proposed by vonNeumann-Morgenstern and found considerable individual variation in trade-offs as well as inter-attribute differences. The focus of these papers has been to assess pair-wise trade-offs directly and to derive the overall utility of alternatives from the analysis of trade-offs.

An Empirical Investigation

A research study was designed to investigate the compensatory dimension of evaluation processes. Unlike earlier studies, the attempt was not to obtain overall utility but to focus on the compensatory dimension *per se*. The objective was to represent the dimension in terms of a single quantitative measure that can be used as an indicator of the degree of compensatoriness in a process. Since the methodological issues were of primary concern, the empirical investigation was confined to a study of career opportunities by MBA candidates.

Methodology

Students from a graduating MBA class were requested to participate in a decision making study. Since career placement loomed large in their lives at that point in time, job choices were used as the subject of investigation. This ensured cooperation and involvement from the participants.

Independent Variables. A set of attributes relevant to job choice among the graduating class was identified from group responses. In a classroom setting, the students were asked to list the five most important attributes in job selection. Five attributes were identified by taking into account frequency of mention and the average importance rating assigned to each. The choice of these five attributes were validated among a second group of students. From this emerged a reduced set of three attributes which was

considered most important by majority of the students.

These three attributes (initial salary, job location and responsibility) were used to develop alternative job opportunities. Twenty-seven jobs were generated from three levels of each attribute:

- Salary: \$14,000; \$16,000; \$20,000 annually
- Location: Northeast (NE), Midwest (MW), and Farwest (FW);
- Responsibility: Low, moderate and high.

A second set of 18 (2X3X3) jobs was composed for validation purposes using a subset of the attribute values: moderate and high levels of responsibility, the same locations and \$12,000; \$15,500 and \$18,000 salaries. Each job was described in terms of these attribute values on a 3x5 card.

Dependent Variables. A series of measures were obtained for the job evaluation task:

- a) Measure of liking for each job on a 11-point scale (0-10);
- b) Measure of preference order through successive sorting into most and least preferred categories;
- c) Verbalization of thoughts underlying the sorting procedure as the sorting took place. (This was tape recorded and later transcribed).

The first set of 27 jobs was evaluated along each of these measures. The validation set of 18 jobs was evaluated only in terms of (a) above. The entire evaluation task took an average of about 80 minutes to complete.

Participants. Twelve decision makers participated in the study. They were all graduating MBA students in a mid-western university, seeking permanent placement and had volunteered for the study. A token payment of five dollars was made to each participant.

Findings

Model Performance. Since one of the objectives for identifying the compensatory dimension was to improve the selection of representational models, the first analysis performed was to predict the job preferences by the use of different models. Three different models - regression (reg), additive conjoint (con) and lexicographic (lex) - were used. The three models assume different degrees of compensatoriness: in this research they could only be ordered in terms of this assumption, with the lexicographic model being the most compensatory (rank 1) and the regression model being the least compensatory (rank 3).

Table 1 provides the predictive validity of the three models using two criteria - recovery of the rank ordering of the entire job set and the accuracy in predicting the first choice.

TABLE I
PREDICTIVE VALIDITY OF THREE REPRESENTATION MODELS

Decisionmaker	Correlation Between Actual and Predicted Rank Order ^a			Prediction of First Choice ^b		
	Reg	Con	Lex	Reg	Con	Lex
1	0.54	0.81	0.78	—	✓	✓
2	0.98	0.99	0.99	✓	✓	✓
3	0.56	0.96	0.81	—	✓	✓
4	0.50	0.71	0.71	—	—	—
5	0.89	0.99	0.84	✓	✓	✓
6 DM A	0.75	0.94	0.78	—	✓	—
7	0.95	0.97	0.96	—	✓	✓
8	0.88	0.95	0.91	✓	✓	✓
9	0.83	0.89	0.68	✓	✓	✓
10 DM B	0.94	0.99	1.00	✓	✓	✓
11	0.82	0.93	0.69	✓	✓	✓
12 DM C	0.96	0.88	0.64	—	—	✓

^a Spearman rank order correlation coefficient

^b ✓ indicates first choice accurately predicted

As we can see there are some interesting differences in the validity of these models as representations of the process. Using both criteria of predicting the rank order and the most preferred job, four categories of model performances may be observed. In the first category, all the three models do equally well in predicting the rank orders and first choices for some decision makers (DM 2, 5, 8, 10 and 11). For DM 7, the competition is between two models - the conjoint and lexicographic. This category represents the maximum ambiguity since the models perform similarly although making different assumptions about the degree of compensatoriness in the underlying process.

Category two, represented by DM 12, generates some amount of ambiguity because the three models differ in their predictive validity - the regression model recovers the rank order the best but fails to predict the first choice while the lexicographic model predicts the first choice successfully but performs the worst in recovering the rank order.

The third category appears to be unambiguous because it is possible to select a model based on both criteria. For DMs 1, 3, 6 and 9, the model that predicts the rank order the best is also able to predict the first choice. Finally, the fourth category (which only includes DM 4) represents a situation where none of the three models perform satisfactorily on both criteria. To allow detailed exploration of the problem, further analysis and discussion is limited to three decision makers (labelled A, B, C) conveniently chosen as representatives of the three categories. Category four has been excluded because none of the models performed satisfactorily.

The Compensatory Dimension. Given the observed differences in model performances and the ambiguity in selecting a representational model, the objective of further analysis was to directly use the compensatory dimension in evaluating alternative models. To achieve this objective it is necessary to represent each decision maker on this dimension and a composite index of compensatoriness was devised for this purpose. (Technical details are given in the Appendix).

The values of the composite index are reported in Table 2 for the three selected decision makers. Higher the value, the more noncompensatory the process since it implies larger units of an attribute being required to compensate for another attribute. A perfectly compensatory process would be represented by a zero value in terms of the composite index of compensatoriness.

TABLE 2
A COMPOSITE INDEX OF COMPENSATORINESS

Decision maker	Composite Index	Interpretation
A	12.94	Moderately compensatory
B	49.25	Highly noncompensatory
C	3.53	Highly compensatory
Perfect Compensation	0	

According to Table 2, DM C exhibits the most compensatoriness in the job evaluation process while DM B is the least compensatory. Given this information, it would appear that the lexicographic model is a better representation of DM B's evaluation process but is not appropriate for DM C's process. On the other hand, the regression model which assumes a high degree of compensatoriness may be used to characterize DM C's process. By similar logic, the conjoint (additive) model, which allows for somewhat less compensatoriness in the process than the regression model, may be used to characterize DM A's process. Thus by using the compensatory dimension directly, it is possible to select representational models that are more appropriate for a decision maker based on both predictive criteria and the assumption about the level of compensatoriness in an

evaluation process.

Convergence. To validate the findings on the level of compensatoriness, one other analysis was performed. This was done to ensure that the values of the composite index was not an artifact of the computation process. Verbal protocols were analyzed to see if trade-offs were considered by the decision makers in evaluating the alternative jobs. The verbalizations took place while the first set of 27 jobs were being successively sorted into preference categories. They were taperecorded and later transcribed. Excerpts from the protocols are reported in Exhibit 1.

EXHIBIT I
EXCERPTS FROM VERBAL PROTOCOLS

Decisionmaker A

Job S (NE, LO, 20): Just don't think I can give in to the low responsibility. Salary is attractive and it is in northwest; but what I am doing is important; how much I am allowed to move within the job. I like it the least.

Job O (MW, MOD, 14): Moderate responsibility, in midwest, 14,000 - an attractive opportunity in the midwest even though salary is low. Salary at this point not as crucial as location and responsibility. I feel if I like something - you will be good at it and will want to work it; even though salary is low, I think I can work my way up.

Decisionmaker B

JOB M (MW, MOD, 20): Prefer Q; I would rather...FW sounds better than MW and the extra salary is not enough to compensate.

JOB Y (FW, LO, 20): Poor opportunity due to low responsibility...despite high salary and good location

Decisionmaker C

JOB J (NE, MOD, 20): Like somewhat - moderate responsibility but good salary, with a NE location.

JOB I (FW, HI, 14): Like somewhat - high responsibility but low starting salary with a good location.

Analysis of the verbal protocols support the quantitative descriptions of the evaluation process in terms of compensatoriness. As we can see from the excerpts, DM A allows some amount of compensation; for instance, between a midwest location and a low salary but not a lot, since low responsibility is not being balanced by a higher salary or an attractive location. On the other hand, DM B is quite strict about attributes desired in a job and is unwilling to like a job unless it meets the attribute preferences; even a \$4,000 extra in initial salary cannot compensate for a less desirable midwest location. Much greater balancing of attributes takes place in DM C's process, for instance, when a better salary is accepted even if the responsibility is not desirable.

Model Selection. To the extent that the three models - regression, conjoint and lexicographic - can be ordered in terms of their assumption regarding process compensatoriness, these models may then be used to represent the process of the three job candidates. The lexicographic model which assumes the highest level of noncompensatoriness is therefore selected as the best representational model for

DM B even though the other two models predict the job choice and rank order.

The predictive validity of the conjoint model is supported by the level of compensatoriness found in DM A's process. This level validates the conjoint (additive) model's assumption of a limited degree of compensatoriness. For DM C, the regression model is found to be the best representation because it assumes a high degree of compensatoriness which was found to characterize DM C's process.

Selection Validity. The descriptive and predictive validity of each model selected to be representative of the three processes are reported in Table 3. The correlations may be compared to those of the conveniently accessible regression model.

TABLE 3
DESCRIPTIVE AND PREDICTIVE VALIDITY OF SELECTED MODELS

Decisionmaker	Descriptive Validity*			Predictive Validity*		
	Con	Lex	Reg	Con	Lex	Reg
A	0.96	--	0.82	0.94	--	0.75
B	--	0.99	0.96	--	1.00	0.94
C	--	--	0.88	--	--	0.96

* Spearman rank order correlation

There is an improvement in both descriptive and predictive validity by the use of a specific model even though the improvement may not be statistically significant. However, the justifications for selecting a model to represent the process are built on different grounds. A valuable illustration of the limitation of statistical considerations is made by DM B; despite high predictive validity of the regression model, it makes a totally inaccurate assumption about the compensatory dimension of the underlying subjective process.

Discussion and Conclusion

It appears that focusing on the compensatory dimension of the evaluation process provides a specific criterion for discriminating and selecting among representational models. Despite the very high predictive validity of all the models, there is support for the noncompensatory assumption of the lexicographic model which makes it more valid for representing DM B. Similarly, the conflict created by the overall predictive validity of the regression model and the first choice accuracy of the lexicographic model for DM C's process could be resolved by an analysis of the compensatory dimension. The regression model is found to be a better representation of the underlying process because of the degree of compensatoriness. Superior predictive validity of the conjoint (additive) model for DM A is supported by the findings on process compensatoriness.

Even the use of multiple measures of predictive validity (overall order and the first choice) is seen to be insufficient for determining the nature of evaluation processes implied in the job preferences. Identifying and measuring the compensatory dimension of the evaluation process is found to contribute to its understanding and representation by models. The attempt to represent the magnitude of the compensatory dimension appears also to be fruitful.

There are substantial methodological differences in this paper as compared to other studies of process compensatoriness. Past studies that have focused on this dimension have used it to derive overall utility. While that objective has been successfully achieved, it often involves use of a methodology that is quite restrictive in its usefulness (Hauser and Urban 1979). It required a high degree of conceptual ability for the respondents to understand the procedure by which trade-offs were measured. In contrast,

this study has focused on the compensatory dimension without requiring the decisionmakers to undertake any task that is difficult or not normal. Instead, it has adopted a reverse route — moving from overall measures of evaluation to a specific measure of process compensatoriness.

The approach adopted in this paper for investigating subjective evaluation processes appears to have several other merits. By focusing on a specific dimension and by viewing the dimension on a continuum, it is possible to relate the effects of individual and situational variables on the evaluation process. Instead of just saying that these variables are likely to create differences, it is possible to say where the difference is likely to exist and the nature of the difference (i.e. more or less compensatory). One can hypothesize that increased involvement with a product or issue will reduce the compensatoriness of a process and can test the hypothesis. Such an approach is likely to be more productive in creating an understanding of subjective evaluation processes.

By restricting the study to three decisionmakers, the validity of the approach is open to question. However the rationale of the approach cannot be tested by numbers alone. The objective of the approach was to understand the subjective evaluation process in terms of a single dimension — compensatoriness. It is therefore necessary to directly portray evaluation processes on this dimension and to select models as representations when they satisfied the assumption on this dimension and predictive criteria. The approach offers an alternative to the model vs. process controversy by selecting models based on an understanding of underlying process dimensions.

To fully benefit from this approach, it is necessary to explore another implication, viz to generate other dimensions of the evaluation process and to determine their relationships to the compensatory dimension. The compensatory dimension is only one component of the evaluation process.

While there is some literature on other dimensions (e.g. linearity, configularity, etc.) of the evaluation process, there is a lack of concepts and theories that link these together and create an integrative framework for the conceptualization and understanding of subjective evaluation processes. There exists a need for developing such a framework that can put past research in perspective and identify future research issues. This need is strengthened by recent findings that individual evaluation processes are likely to be constructive and not rule or model applications (Bettman and Zins 1977).

Even though the investigation reported in this paper deals with job choice, it focuses on the compensatory dimension which is likely to be of even greater relevance to study of traditional consumer goods. In the present context of shortages and inflation when significant changes are likely in consumer processes, the impact on the compensatory dimension requires major attention.

Appendix

The compensatory dimension was represented by a set of values computed from the attribute utility scores obtained from conjoint analysis. In order to reach the overall index of compensatoriness, the following calculations were performed.

1. Compensation between pairs of attributes:
 - a. Responsibility and salary,
 - b. Location and salary,
 - c. Location and responsibility.

The rate of compensation is obtained for each change in the level of a pair of attribute, e.g. the change in utility associated with an increase in salary from \$14,000 to \$16,000 and from \$16,000 to \$20,000 as compared to change in utility with an increase in responsibility from low to

moderate and moderate to high. The ratio of this change $U(16)-(14)/U(\text{Mod})-U(\text{Lo})$, is taken as a measure of attribute compensation. Similarly, for the other two pairs.

2. Perfect rate of compensation is assumed to exist when the ratio equals one. An assumption here is that the unit of change in the attribute is same across different levels (e.g. moderate - low = high - moderate). For salary, an adjustment has been made to incorporate the different levels of change ($20-16=2(16-14)$) which leads to a perfect rate of 0.5)
3. The composite index of process compensatoriness is obtained by the following:

$$\sum |X_{ij} - X_j|$$

where X_{ij} = compensation between a pair of attribute level
 X_j = rate of perfect compensation for the same pair.

For example, the composite index for DM A is found by incorporating the following values:

$$12.94 = |(3.92-0.5)+(1.88-1.0)+(5.27-0.5) + (1.14-1.0)+(0.61-1.0)+(-1.34-1.0)|$$

The detailed values are given below:

COMPENSATORINESS IN JOB EVALUATION

Decision-maker	Compensation between Attribute Pairs						Composite Index of Compensatoriness
	Responsibility & Salary		Location & Salary		Location & Responsibility		
	1	2	1	2	1	2	
A	-3.92	1.88	5.27	1.14	0.61	-1.34	12.94
B	24.18	23.09	1.38	1.73	0.07	0.06	49.25
C	1.02	1.0	0.22	0.03	0.03	0.21	3.53
Rate of Perfect Compensation	0.50	1.0	0.5	1.0	1.0	1.0	0

References

Bettman, J. R. and Zins, M.A. (1977), "Constructive Processes in Consumer Choice," Journal of Consumer Research, 4, 2, 75-85.

Fielder, J. A. (1972), "Condominium Design and Pricing: A Case Study in Consumer Trade-off Analysis," Proceedings, Association for Consumer Research, 279-93.

Goldberg, L. R. (1968), "Simple Models or Simple Processes: Some Research on Clinical Judgment," American Psychologist, 23, 483-96.

Hauser, J. R. and Urban, G. L. (1979), "Assessment of Attribute Importances and Consumer Utility Functions: vonNeumann-Morgenstern Theory Applied to Consumer Behavior," Journal of Consumer Research, 5, 4, 251-62.

Henderson, J. M. and Quandt, R. E. (1971), Microeconomic Theory: A Mathematical Approach, Second Edition, New York: McGraw Hill.

Johnson, R. M. (1974), "Trade-off Analysis of Consumer Values," Journal of Marketing Research, 11, 121-27

Keeney, R. L. (1974), "Multiplicative Utility Functions," Operations Research, 22, 1, 22-3.

Lancaster, K. J. (1966), "A New Approach in Consumer Theory," The Journal of Political Economy, 74, 2, 132-57.

_____, (1972) Consumer Demand: A New Approach, New York: Columbia University Press.

MacCrimmon, K. R. and Toda, M. (1969), "The Experimental Determination of Indifference Curves," Review of Economic Studies, 36, 433-51.

Park, C. W. (1976), "The Effect of Individual and Situational-related Factors on Consumer Selection of Judgment Models," Journal of Marketing Research, 13, 144-51.

Raiffa, H. (1968), Decision Analysis, Reading, Mass: Addison-Wesley.

Rousseaus, S. N. and Hart, A. G. (1951), "Experimental Verification of a Composite Indifference Map," The Journal of Political Economy, 59, 288-318.

Wright, P. L. (1975), "Consumer Choice Strategies: Simplifying vs. Optimizing," Journal of Marketing Research, 11, 60-7.

DEVELOPING AN INSTRUMENT TO IDENTIFY INDIVIDUAL DIFFERENCES IN
THE PROCESSING OF PICTORIAL AND OTHER NON-VERBAL INFORMATION¹

Flemming Hansen, The Copenhagen School of Business Administration & Economics
Niels Erik Lundsgaard, Frederiksberg Hospital²

Abstract

The present paper discuss recent findings regarding brain lateralization and its implications for the study of consumer behaviour. This research suggests a new view on the way in which consumers react to different media and creative solutions. It should be possible to identify individual differences in the extent to which people rely upon left or right brain informational processing.

Background

Hemispherical lateralization implies that individuals use the left and the right brain differently. This has been studied for the last two decades. The work by psychiatrists (Wexler 1980), by psychologists (Gazzaniga 1977, Kimura 1973, and Deglin 1979) and by brain researchers (Witrock 1977) has improved our understanding of the way in which the right and the left brain differ in their handling of incoming information and in controlling behaviour.

Several aspects of this specialization of the brain halves are now generally agreed upon. Whereas the left hemisphere is primarily responsible for traditional cognitive activities relying upon verbal information, symbolic representation, sequential analysis, and with the ability to report consciously of what is going on; the right brain - without the individual being able to report verbally about it - is more concerned with pictorial, geometric, timeless, musical and other non-verbal information.³

Even though the differences often can be difficult to identify because of the intensive interaction going on between the right and the left brain half, it is well-established that some tasks lead to left brain activity, whereas others generate right brain processing. It is to be expected that particularly pictorial, musical and other non-verbal forms of communication give rise to right brain processing.

On this background, it has been proposed that even when attention - in the sense it is normally referred to in advertising research - is not present, it is still possible for the individual to receive some information, and to store this (Krugman 1977). Moreover, it is proposed that this process is particularly efficient with pictorial material and that the information is stored in a holistic unedited, non-verbal fashion very different from the way we normally store verbal and similar information, and is not recallable, but is easily triggered by recognition.

¹Sponsored by the Danish Social Science Research Council.

²Flemming Hansen is professor, ekon. dr., The Copenhagen School of Business Administration & Economics, 60 Howitzvej, DK-2000 Copenhagen, Denmark. Niels Erik Lundsgaard is cand. merc et mag. art. (psych), Frederiksberg Hospital. Request can be made to the first author.

³These conclusions do not apply to approximately 25% of all left-handed individuals who have the functioning of the left and right brain reversed. All subjects known to be left-handed are excluded from the study to be reported on the following pages.

Individual Differences

Various authors have pointed at the possibility of individual differences in the extent to which people rely upon the left or the right brain in coping with their environment. Bogen (1977) has suggested that such differences may depend upon the nature of the stimulation the child receives, while it grows up. Children growing up in a heavy verbally dominated culture, are more likely to rely more heavily upon left brain processing. Lindzay & Norman (1977) suggest that whereas intelligence as measured by existing IQ-testing procedures mostly reflect left brain activity, some aspects of creativity may imply important right brain processing. In line with this psychiatrists (Wexler 1980) are convinced that several psychiatric disorders are explainable in terms of unbalanced functioning of the brain halves. For example, schizophrenia is believed to imply reduced left brain activity.

Even consumer psychologists have touched upon the possibility of individual differences in hemispherical lateralization. Appel, Weinstein, and Weinstein (1979), in an attempt to find hemispherical specialization in handling various informational inputs from television commercials, accidentally find significant individual differences in the way in which the test material is being handled. (for a detailed discussion of their findings see Krugman 1980).

To the extent that such individual differences exist it would imply that the right brain dominated individuals would rely more on pictorial non-verbal information and that they handle information very differently from left brain dominated people.

Purpose

It is the purpose of the project being reported here to establish whether such differences exist among normal consumers in their use of left and right hemisphere. Moreover, the attempt is made to learn what implications these differences have for communication planning, segmentation, etc. The project attempts:

1. To establish the existence of individual differences in the extent to which different consumers rely upon left and/or right hemispherical processing.
2. To develop an operational procedure for measuring such differences.
3. To examine the implication of such differences in terms of differences in consumer's receiving, handling, storing, and reacting to different kinds of information.

The present paper concerns itself with tentative findings regarding the first two questions.

Measurements

Most studies of hemispherical lateralization have used complicated measurement procedures. Measurements have been carried out in a clinical setting and often advanced instruments have been used.

Some of the earliest findings came from studies of patients with various kinds of brain damages (Sperry 1973). Other studies have used the EEG (Electroencephalogram) technique. Here the pattern of α - and β -waves in the left and in the right brain half has been studied while the respondent is carrying out different tasks.

Also special tachistoscope procedures have been developed. Here the purpose has been to control to what side of the brain visual information is transmitted. Similarly, dichotic listening techniques have been used. Here different words, pieces of music, and other auditory stimulation are presented simultaneously to the left and to the right ear. By asking respondents about what they hear, it is possible to learn in what brain half information is being processed. Finally, psychological tests normally used for diagnostic purposes have been applied in an attempt to identify individual differences in lateralization.

All of these measurement techniques are, however, complicated, and almost impossible to apply in larger scale studies of consumer behaviour. Therefore, to be able to deal with individual differences in hemispherical lateralization in larger samples of consumers, it is necessary to develop a simpler measurement device.

Since the extent to which the individual relies on right vs. left brain informational processing can be expected to influence, his overall behaviour and possibly also his beliefs and values, it might be possible to have the individual himself reporting on the extent to which right or left brain functioning dominate. Of course, direct reporting cannot be expected to be valid, but the tendency of the individual may be revealed by his responses to constructively selected questions.

For example, one can speculate that a person relying more heavily upon information stored in the right brain tend to recognize faces better than a person relying more upon left brain information processing and storing. The latter on the other hand, might be expected to recall names better than faces. Replies to items reflecting the extent to which the individual find himself relying more or less upon the one or the other kind of information could be used for identifying individual differences. It is an important part of the present project to identify and validate such statements which could be used to identify individual differences.

Validation of such statements can be approached in two ways. It is possible to test the same items on two or more groups of individuals known to differ in the extent to which they rely upon left vs. right brain processing. Such a procedure is employed in a later stage of the present project, where measurements obtained with normal individuals are compared with similar measurements obtained from people who through clinical testing have been identified to be relatively right brain dominated. The present paper, however, reports findings alone based upon the second possible validation procedure.

With a sufficiently large number of test items developed, it is possible to validate these internally by comparing the response pattern among the items themselves. For this task factor analysis is a commonly used procedure. In addition to this, it is possible to validate the test items against various psychological tests supposed to reveal left and right brain specialization. In the present study, a selection of such tests are used against which the potential test items are validated.

Procedure

The validating experiment reported here has been carried out with 50 male and female subjects recruited at a student housing complex in Copenhagen. Each respondent has been paid approx. 20 \$ for his/hers participation. Since some of the tests employed are quite time-consuming each respondent went through a procedure lasting approx. 3 hours. Basically, the following three types of measures were used.

1. Reactions to 10 clinical psychological tests
2. Dichotic listening
3. Answer to 80 selected test items presented on a self-administered questionnaire.

Clinical Psychological Tests

A total of 10 clinical psychological tests were selected from different sources.

To make the administration easy and to facilitate the registration of the response to the tests in a quantitatively operational way, both the procedure and the scoring was transformed⁴, so that each respondent answered all 10 tests within reasonable amount of time, and so that the scores could be registered in a form prepared for computer processing.

In the order to which the tests were administered, the standardized forms are described below:

1. Word-Pair Learning (Theilgaard 1979)

10 different word-pairs, 5 logical such as street/car, see/fish, window/curtain, and 5 non-logical such as cap/hotdog, church/cotton, apple/pencil, were presented with each pair typed on a paper card. The respondents were instructed to try to learn all 10 word-pairs by running through the pile two times in a moderate to fast tempo. Hereafter the interviewer named the "first" word on each card, registered the number of right answers out of the 10 possible. This test was believed to measure left brain activity.

2. Copy Recall (Freeman 1965)

A meaningful story of 44 words was read aloud to the respondent, who was asked immediately to retell the story with the use of the same words and sentences. The copy was divided into 18 units, and the number of correct answers was registered as a measure of left brain activity.

3. Face Recognition (Theilgaard 1979)

The material for this test consists of 28 photos of 14 men and 14 young women. The young men are all dressed in a white shirt with a black tie and the women in white gowns. All were portraits. All pictures looked very much alike. All 28 portraits are shown as 4 rows each with 7 portraits. In the test 12 of the persons were chosen and first shown in individual presentation to the respondent. Each portrait was presented to the respondent in exactly 5 seconds, and immediately after the presentation of all 12 portraits, the 28 portraits were presented.

⁴Details of the transformation procedure can be obtained from the authors.

The respondents were now asked within approx. 60 seconds to identify the 12 persons they had previously seen among the 28 portraits. The interviewer registered the number of correctly identified persons, as a measure of right brain capacity.

4. Visual Gestalts (Andersen 1976)

The material for this test consists of 4 pieces of paper with 4 different geometrical figures: A circle, a rectangle, a triangle, and a half-circle. Four different cards each with four different gestalts printed in the geometrical figures are then presented to the respondent. Each presentation lasts 10 seconds. After each presentation the respondent is asked to draw as many of the gestalts as possible. The number of correctly drawn gestalts is taken as a measure of right brain capacity.

5. Subtraction (Theilgaard 1979)

The respondent is asked to subtract 7 from 100 and name his result, and then to continue by subtracting 7 from the result and to go on down to zero. This is done with a fixed time-limit of 30 seconds. Subsequently the interviewer registers the number of correct subtractions completed. The subtraction test reflects left brain capacity.

6. Cube Test (after Goldstein-Sheerer, Freeman 1965)

The material for this test consists of four cubes. They are all identical, and they have the following side colours:

1. blue
2. red
3. yellow
4. white
5. diagonally cut in white and red
6. diagonally cut in yellow and blue.

Figure VII, VIII, and IX from the original test are presented to the respondent in 10 seconds each. Thereafter the respondent is asked to reproduce the figures as fast as possible. The total number of seconds it takes to reproduce the three figures is registered and subtracted from 180. The resulting number is supposed to reflect right brain capacity.

7.-8. Word Mobilization Test I. & II. (Theilgaard 1979)

I: The respondent is asked within 60 seconds to name as many things as possible that one can see in the street. The interviewer registers how many things there are.

II: The respondent is asked to name as many animals he can think of within 60 seconds. The interviewer registers the number of animals named.

Both measures are taken as indicators of left brain activity.

9. Incomplete Pictures (Freeman 1965)

The 21 incomplete drawings from Wechsler's Adult Intelligence Scale are presented one by one. Because of the relative high intellectual standard of the respondents they were only permitted a few seconds to look at each drawing. As a measure of right brain capacity the interviewer registers how many errors the respondent correctly identified.

10. Face Identification (Benton 1978)

For this test map number 7 to map number 13 from the "Benton Test" are used. Each map consists of one portrait on the top and 6 at the bottom. Out of the latter 6, 3 are portraits of the same person, as the one on the top.

On each map the respondent is asked to identify those portraits at the bottom representing the same person as the one on the top. Each map is presented for a max. of 20 seconds and the interviewer registers how many correct identification the respondent makes. This is supposed to reflect right brain capacity.

Dichotic Listening

For the dichotic listening task 78 word-pairs were selected. Examples are: RUF/BOF, STA/GJO, etc.

These word-pairs were presented to respondents simultaneously with one word to the left ear and the other to the right ear. For this purpose a stereo recorder with normal head phones was used.

In line with Frumkin (1978) the dichotic listening score was computed as the number of correct items recalled among those heard with the right ear, less the number of items recalled from those heard with the left ear divided by the total number of items recalled.

The instrument was first tested with two groups. One comprising 10 business students, and the other 10 musical students. It was believed (Sperry 1973) that the musical students would show a higher right brain score than the business students. This also was found. Most of the 10 music students scored lower than any of the business students, and the average for the two groups differ significantly. In a later test/retest reliability check the correlation between the first and second measurement for a group of business students was .76.

In the subsequent application of the dichotic listening procedure some of the test items were left out to reduce the total time spent on the listening task. This was done by deleting those items from the original version which discriminated the least between the business, and the music students. An attempt was subsequently made to validate the dichotic listening tests with two other existing tests. The first of these is normally used for psychological testing at the Copenhagen State Hospital (Heshe, et.al 1978). This attempt was not successful. Neither of the three tests correlated very well with each other. Faced with this dilemma it was decided to use the original test, since this had performed well in the first test with music and business students. This version was also technically superior.

80 Self-Administered Test Items

80 items were formulated. All of these could be answered by expressing ones view on a 5 point scale, ranging from agrees completely to disagrees completely. Some of the statements were formulated departing in what is known about lateral specialization. Examples of statements are: "I enjoy listening to music", (supposed to reflect right brain activity), "I am very good at calculating", (supposed to reflect left brain activity), "I enjoy very much watching landscape painting" (right brain), "I plan carefully how to spend my time", (left brain), etc.

In addition, items were selected from existing psychological tests. These were included based upon the assumption that the mental disorders they were supposed to measure could be related to left vs. right brain dominance. For example, a number of items used in tests of neurotic fear tendencies were included. Examples are: "I like to be told what to do", "I feel uncertain when being together with people I have not met before", etc. A total of 41 items supposed to reflect right brain dominance and 39 items reflecting left brain dominance were selected. Scoring was carried out by adding the scores on the 41 right brain items separately, and on the 39 left brain items separately. Additionally, the items were treated individually in the subsequent analysis.

Finally, for each respondent age, sex, and a couple of other demographic variables were collected.

The Data Matrix

The initial data matrix can be described as a 50 respondents by 98 items large matrix. The 98 items being:

- . 10 test scores on the psychological clinical test
- . 1 dichotic listening score computed as described
- . 80 test items
- . 1 sum-score representing the 41 right brain items
- . 1 sum-score representing the 39 left brain items, and
- . 5 demographic variables (sex, age, etc.)

In the course of the data analysis additional variables were generated. Those variables are described subsequently.

Results

Clinical Psychological Tests

As it will be recalled the clinical tests were selected with five of them supposed to reflect right brain dominance, and five supposed to reflect left brain dominance. To test the interrelationship among the scores of the 10 variables, a principal component factor analysis was carried out with a subsequent vari-max rotation of the two first factors. These two factors explain 45% of the total variance. (A third factor explains additionally 10%). The results of the factor analysis is shown in table 1. It appears that 4 of the 5 test items supposed to reflect left brain dominance are associated with the first factor. Similarly, on the second factor, 4 of the right brain tests load significantly high, though 1 of them is also associated with the first factor. Neither of the remaining two tests correlate with either of the two factors.

TABLE 1.

Two factors, vari-max principal component solution for 10 clinical psychological tests of left vs. right brain dominance. (N = 50, explained variance = 45%)

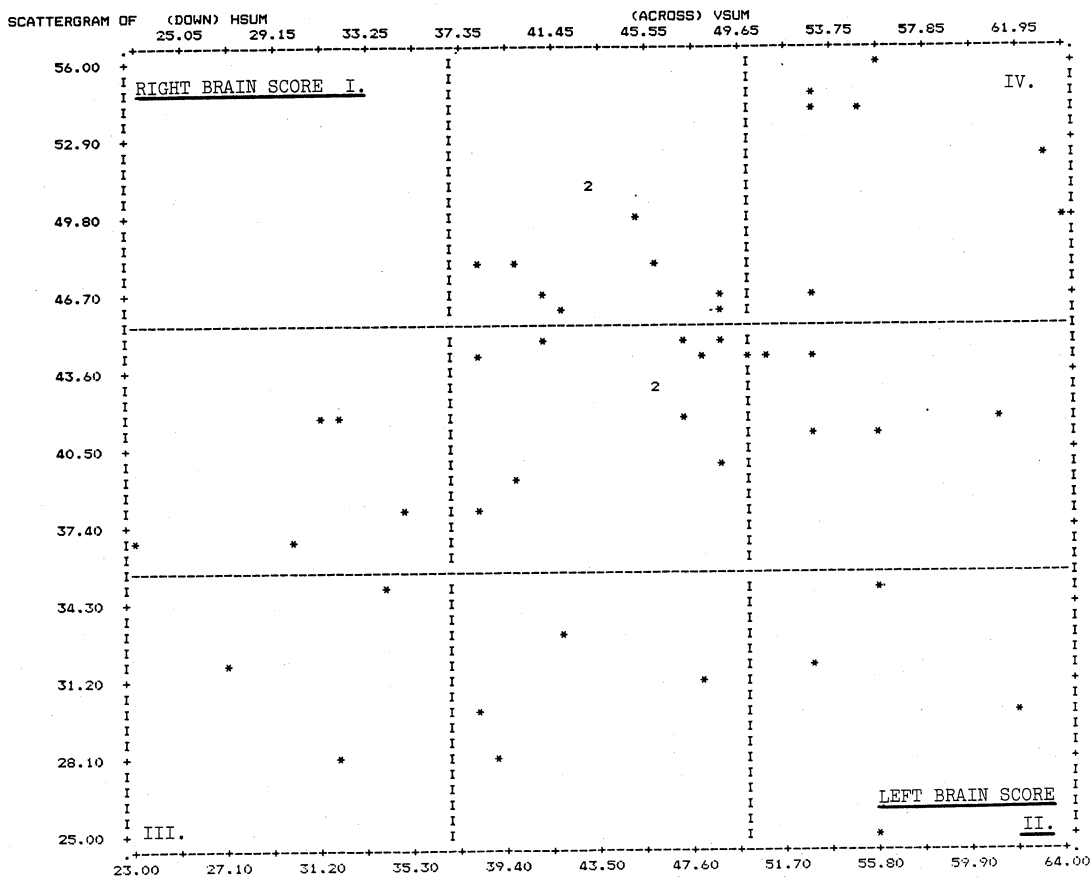
	Factor I.	Factor II.
Word mobilization II. (left)774	.076
Word mobilization I. (left)660	-.214
Copy recall (left)563	.221
Subtraction (left)487	.153
Word-pair learning (left)082	.045
Face identification (right)178	.243
Visual gestalts (right)	-.097	.512
Face recognition (right)	-.089	.419
Cube Test (right)167	.632
Incomplete pictures (right)537	.589

The failure of the face identification test to correlate with any of the factors is ascribable to the fact that this test turned out to be so easy that practically all of the participants scored maximum points. Thereby, the test becomes highly insensitive. With regard to the poor performance of the word-pair learning test, one possible explanation is that this test measures general intelligence, rather than anything else.

Based upon these results, it was decided in the subsequent analysis to combine the 4 tests scoring high on the first factor into a single score reflecting each individual's tendency to rely upon left brain processing. Similarly, four tests loading high on the second factor are used to compute a combined score for each respondent. This score reflect the individual's tendency to rely upon right brain processing. Since the intervals, within which the test results vary, differ between the tests, each test score was standardized and normalized before computing the combined score.

The interrelationship between the two scores is studied in figure A. It appears that the two scores - although slightly positively related (R = 0.301)- do measure two different dimensions.

FIGURE A.
The Relationship between Right and Left Brain Scores.



Since the purpose of the project is to differentiate between relative left and relative right brain dominance regardless of the absolute scores, a ratio was computed:

$$(I); \quad \frac{\text{Left brain score} \div \text{right brain score}}{\text{Left} + \text{right brain score}}$$

It will be seen that this is the same ratio as the one being used for the dichotic listening scores.

Dichotic Listening

The correlation between the dichotic listening score and the combined left brain activity score is .167 and only approaching significance.

The correlation between the dichotic listening score and the right brain score is -.328 and significant with $p \leq 0.01$. The resulting correlation with the ratio (I) is small and insignificant.

Even though there is some relationship between the dichotic score and the right brain score, the correlations are far below what was expected. Several reasons may be suggested for the almost complete failure of the dichotic score to validate the clinical test scores.

Basically, the dichotic listening score tests a left vs. right brain dominance. The data in figure A, however, suggested a four-way grouping rather than a two-way grouping.

A second reason for the problem may be the extreme difficulties encountered in presenting stimuli simultaneously. The same problem existed with the two other dichotic listening versions against which we originally tried to validate our own instrument. In conversations after the test, several respondents reported to have been able to hear with one ear before the other.

A third problem, of which we are aware, is more complicated. Even though information sent to the left ear is transmitted to the right brain and initially processed there, in normal individuals there is some information immediately being transmitted back to the left brain half. Moreover, it is possible that some information is picked up in the left brain, while the signals are being sent from the ear to the opposing right brain hemisphere. These complex interactions make it difficult to interpret the meaning of the dichotic listening scores. This is even more so since the disturbing processes may be more or less important depending upon whether verbal material, musical material, or nonsense words are used.

These experiences with the dichotic listening have not discouraged the research group completely, however. Further work with the dichotic listening procedure is being done in co-operation with the Frederiksberg Hospital in Copenhagen, where work is now going on in an attempt to develop more accurate procedures for the presentation of the stimuli.

The 80 Test Items

The 2 sum-scores did only correlate poorly with the clinical tests. Presumably too many irrelevant items were included. Therefore, a different analysis was carried out.

Like it was done with the clinical test the 80 test items were factor analyzed using a principal component procedure with the two first factors vari-max rotated. These two factors accounted for 18.3% of the total variance in the data. Out of the 80 items, 34 significantly correlated with the first factor, and of those 28 came from that subset of 41 which originally were believed to test left brain activity.

Similarly, among those 22 items significantly correlated with the second factor, 16 came from those 39 items originally believed to reflect right brain activity. Thus, there is some evidence that the first factor parallel the first factor of the clinical test analysis, and that the second factor reflect the same dimension as the one being measured with the second factor in the clinical test analysis. However, the low amount of variance accounted for, and the many questions not loading with any of the two factors again suggests, considerable noise in the data. Particularly, a number of items originally selected from tests of mental insecurity came out in contrast with the original hypothesis. For this reason it was decided to carry out a second analysis of the 80 scores. For this purpose the ratio (I) was used and a Pearson-moment correlation between each of the 80 tests and the score was computed.

In table 2 those 24 statements are reproduced which have significantly correlated with the ratio (I) with a probability of .85 or more. In the table the exact probabilities are reproduced.

TABLE 2.

24 out of 80 statements correlating with the combined left dominance score from the psychological clinical test. Positive correlations suggest left brain association, and negative suggest right brain dominance.

<u>Statements:</u>	<u>Correlation</u>	<u>Item No.</u>	<u>Exact probability</u>
I have a large vocabulary	0.26	3	.033
I read a lot	0.30	4	.017
I prefer that things are neat and orderly	-.17	7	.128
I always make sure that my home is neat and orderly	-.20	9	.084
In my work I do rather not progress after a fixed plan, rather I adjust to the particular circumstances19	14	.094
I prefer to relax with a magazine rather than with a book	-.27	20	.028
I find it more important to satisfy myself than others	-.16	21	.145
I do my best to behave so that others like me16	22	.130
I do not like when my work is planned too detailed in advance41	23	.002
When telling about a funny incident I often imitate the persons I am talking about by gesticulating25	24	.042
I like to say things that other people find witty and to the point31	25	.014
I like singing	-.22	29	.059
I work more carefully for others than for myself17	37	.121
I can easily adjust myself to a new situation18	39	.100
I rather avoid people who, one way or the other, are sad, helpless or in bad luck29	51	.021
I can definitely hear whether music is out of tune	-.18	52	.106
I talk a lot with other people21	53	.076
I watch a lot of television every day	-.21	55	.074
I get most of my information from television	-.44	56	.001
I do not care very much about cleanliness	-.22	60	.062
I can do things which are not quite decent or right without getting a bad conscious	-.21	61	.069
I have read in a cartoon magazine within the last three weeks15	69	.145
I am not the kind of person who plan ahead18	72	.102
I always feel lost and helpless when I am alone at a strange place	-.18	74	.103

Further Work

The findings reported so far are positive in the sense that they suggest that it may be possible to develop a test battery which will throw light on the extent to which an individual is relying more or less upon left and/or right brain information processing. There are also some suggestive findings that it may be possible to develop self-administered test items which will do the same job.

There are, however, also problems with the present results. The self-administered measurement instrument has not found its final form, and elements to go into it need further validation. Within the framework of the present project several attempts are being made to this effect.

A test is being carried out with patients from a major Copenhagen Hospital's psychological ward. For these patients other evidence suggest retarded left - relatively right brain dominated information processing. Along the same lines a test is being planned with a group matching the group of university students with respect to age, and sex, but not with regard to school training, verbal experience, etc. Both of these tests are hoped to show the ability of the revised self-administered test, as well as of the 8 clinical tests to discriminate among groups with supposed different hemispherical lateralization.

Also further work is being done with the dichotic listening instrument, and it is planned to develop a tachistoscope similar to that used by Kimura (1973). Also the possibilities for using EEG-measurements for validating the presently used tests are explored.

Finally, work is going on developing new statements based upon the experience gained with the present battery and based upon renewed screening of existing findings regarding the nature of specialized left and right brain processing.

If these activities result in an improved self-administered questionnaire, it becomes possible, with this in hand, to study a number of derived hypotheses regarding the nature of left and right brain information processing. For example, it will be possible to study the extent to which pictorial information communicates better with people with a tendency to use right brain dominated information processing. Similarly, various executions of verbal material can be tested in an attempt to identify sensitivity of the information processing procedures to variations in complexity of the material.

Also with a self-administered questionnaire available, it is possible to identify segments in society with more or less left vs. right brain dominated information processing.

Similarly, it is possible to carry out choice experiments with people known to differ in their hemispherical lateralization, whereby it becomes possible to study to what an extent different choice criteria influence choices for different individuals.

References

- Andersen, Ruth (1976), "Verbal and visuo-Spatial Memory", two clinical tests administered to a group of normal subjects. Scandinavian Journal of Psychology, 17, 196-204.
- Appel, Valentine, Weinstein, Sidney and Weinstein, Curt (1979), "Brain Activity and Recall of TV-Advertising", Journal of Advertising Research, 19, 7-15.
- Benton, A.L., Van Allen, M.W., Hansher, K. De S. and Levin, H.S. (1978), Test of Facial Recognition, manual, Benton Laboratory of Neuropsychology, Division of Behavioral Neurology, University, IOWA.
- Bogen, Joseph E. (1977) "Some Educational Implications of Hemispheric Specialization", in Wittrock, M.C. (ed): The Human Brain, 133-152.
- Deglin, Vadim L. (1979) "Journey through the Brain", Unesco Courier, April
- Freeman, Frank S. (1965) Psychological Testing, Holt, Rinehart and Winston, New York.
- Frumkin, Lynn R., Ripley, Herbert S. and Cox, Gary B. (1978), "Changes in Cerebral Hemispheric Lateralization with Hyponosis", Biological Psychiatry, 14, 741-750.
- Gazzaniga, Michael S. (1977) "Review of the Split Brain" in Wittrock, M.C. (ed): The Human Brain, 89-96.
- Heshe, Jørgen, Röder, Erik and Theilgaard, Alice (1978) "Unilateral and Bilateral ECC" Acta Psychiatrica Scandinavica, Supplementum 275.
- Kimura, Doreen (1973) "The Asymmetry of the Human Brain - Recent Progress in Perception", Readings from Scientific American, San Francisco.
- Krugman, Herbert E. (1977) "Memory without Recall, Exposure without Recognition", Journal of Advertising Research, 17, 7-12.
- Krugman, Herbert E. (1980) "Substained Viewing of Television", Paper presented at the Conference Board, Council on Marketing Research, New York, February.
- Lindzay, Peter H. and Norman, Donald A. (1977) Human Information Processing, Academic Press, New York.
- Sperry, Roger W. (1973) "Lateral Specialization of Cerebral Function in the Surgically Separated Hemispheres." In McGuigan, F.J. and R.A. Schoonorer (eds): The Psychophysiology of Thinking, New York, Academic Press, 209-229.
- Theilgaard, Alice (1979) "Demens - Psykologisk Set", Nordisk Psykiatrisk tidsskrift, 33
- Wexler, Bruce E. (1980), "Cerebral Laterality and Psychiatry: A Review of the Literature", The American Journal of Psychiatry, 137:3, 279-289.
- Wittrock, Merlin C. (1977) The Human Brain, Englewood Cliffs, New York: Prentice Hall.

CONSUMER INFORMATION AND WORKABILITY OF COMPETITION:
A THEORETICAL FRAMEWORK

Gunter Schoppe, University of Frankfurt/Main
Christine Czerwonka, University of Frankfurt/Main

Abstract

This paper deals with the question, whether the usual division of labor between competition and consumer information policy generates but positive effects, by using examples of price surveys and product tests that practice consumer information causes or intensify developments which counteract the common goal of supporting workable competition and improving the consumer's position. Consideration is also given to guidelines for the development of consumer information policy which are compatible with the aims of workable competition.

The "Ignorant" Consumer and his Godfathers

At the time the consumer was declared sovereign, the prevailing opinion was that he himself would best be able to rule his destiny without special assistance. Furthermore, it was felt that treating the consumer as primary would best serve social welfare. No further external interventions seemed to be necessary aside from establishing free markets and guaranteeing far-reaching personal liberty and equal rights for all members of society. Competition was not only expected to ensure consumer interest in better and reasonably priced goods, but it was also expected to achieve a compromise between conflicting consumer and seller interests.

The implementation of these ideas has resulted in consumer technological progress and a considerable increase in consumer welfare. However, in contrast to its promises, this development was accompanied by a concentration of economic power, by self-abolition of competition, by a partly disadvantageous technological evolution, and, last but not least, by a growing imbalance between producers and consumers. Search for the causes of these obvious insufficiencies has led to the recognition of a fundamental conceptual weakness of the liberal paradigm, namely that the market system is a vulnerable one and therefore needs organizational management. Thus, the aim to maintain the spontaneous self-controlling mechanism of the market became the guiding principle and led to the establishment and further development of competition policy. However, maintenance of the competitive system was not an end in itself; the ultimate aim was to increase the influence of buyers' interests and to abolish or prevent the predominance of sellers. In German literature this is well documented by such metaphorical slogans as: "Consumers and cartels are contradictions like fire and water; consumers and competition should be as thick as thieves" (Bock 1958, p. 51); "An effective antitrust law is the best consumer protection" (Bock 1958, p. 52); and "A law of this kind is first of all a law for consumer protection" (Erhard 1957, p. 172).

This intervention on behalf of the consumer represents the first prominent "godfather"; that is, one who takes care of the well-being of consumers by controlling producers' behavior via prohibitive general rules. This sponsor has tended to direct all his efforts -- in the scientific sphere as well as in practice -- towards furthering the development of the existing system of competition norms and the related catalogue of sanctions.

Most characteristic of this sponsor is the fact that, in accordance with the ruling opinion, he claimed an

exclusive role and only occasionally accepted the responsibility of other policies, (which in fact had a long tradition in social policy, the predecessor to today's consumer protection). Cases in example are when the "results of unrestricted free choice may be damaging to the individuals concerned or to others" (Hildebrand 1951, p. 22) and when there was the necessity for proper "distribution of scarcity". This claim to a sole agency which was based on a broad consensus originated from the conviction that the realization of consumers' interests, as well as the direction of the economy-- would be most satisfactorily guaranteed if the principle of competition was assured by an appropriate economic order. It was well known in this context that, among other things, the degree of competition depends on buyers' information, but, as Scitovsky (referring to Marshall and his contemporaries) has emphasized, one believed that "buyers' information in turn depends on the organization of the market" (1950, p. 48) and, accordingly, would be adjustable by means of public market organization. Both of the assumptions turned out to be false as time went on. However, in spite of this knowledge, no consequences were drawn as far as an extension of the catalogue of competition policy instruments is concerned. The only reaction was to assign occurring problems to the responsibility of others.

The impulse for abandoning an exclusive claim to sponsorship, which opened an "appreciated" field to modern consumer policy, was above all to be found in the debate about "workable competition".

During this discussion, the fact was stressed that, because of increasing product differentiation and other market imperfections (supposedly necessary for the dynamics of competition) one should not expect a given or a growing expertness of consumers (cf. Clark 1968, p. 13-14). Consumer ignorance (cf. Mitchell 1912, p. 269; Clark 1968, p. 20) would be an invitation to fraud unless adequate counter-measures were taken (Clark 1968, p. 466-467). In addition, their ignorance would prevent satisfactory resources allocation and, at the same time, would bring the danger of something like Gresham's law becoming operative, whereby these markets might drop to the lowest quality levels (cf. Clark 1968, p. 483). As already mentioned, these insufficiencies were not looked upon as severe challenge by the godfather competition policy. This was probably so, because, at least implicitly, the assumption of a monocausal relationship between market structure and buyers' information was kept up even though attention was more and more directed towards the interdependences between market structure, market conduct and market performance. Thus, neither the information deficit nor changes in consumers' state of information were expected to cause negative effects. This is a possible explanation for the fact that in modern literature on competition policy considerations of the type of consumer information that would be compatible with the aims of competition policy are missing. There have only been occasional hints that "... measures such as comprehensive grade-labelling..., plus, perhaps, wide public dissemination of product information would be salutary" (Bain 1967, p. 217; Clark 1968, p. 74-77; p. 86-87). This opinion has prevailed despite Scitovsky's (1950) convincingly expressed ideas that consumer ignorance reduces the necessity for price and quality competition, enforces the process of seller concentration and raises barriers to entry

which prevent potential competition. However, these are theses which suggest an intensive engagement with questions about adequate means of compensation on the part of competition policy.

Consumer policy is the younger sponsor of the consumer. It is similar to competition policy in that it is a one-sided action-program; however, it differs in that it is not as concerned with problem definition. Consumer policy has, with the help of its advocates in science and policy, not only aimed at improving the position of the individual consumer, who is overstrained in the market, but it has also aimed at achieving social advantages for the economy as a whole. This has been tied to the hint that only intelligent use of free choice can guarantee a selection of the best goods and the most efficient firms and thus lead to optimal allocation of resources (cf. Ferguson et al. 1971, p. 72).

Even after having been accepted as a policy area of high social importance (Kennedy 1962), consumer policy has tended to orient its market-directed actions (like information, advice, and education) to problems of the individual consumer's role rather than towards requirements for the functioning of the overall system. Additionally to this point, with exception of some rudimentary attempts (cf. among others Beem & Ewing 1954; Morris & Bronson 1969; Scherhorn 1977), there has been no result-control.

In exploring these neglects (cf. corresponding criticisms by Day 1976, p. 51; Chestnut 1977, p. 195), essentially two explanations can be found. First, consumer policy is concerned with itself as well as with those problems which arise as questions along the chosen way. Half of these are questions of proving that the consumer is overstrained and thus is not able to play the expected part. The other half is concentrated on description of activities, particularly on attempts to solve problems which are connected with insufficient use of consumer information. Second, this specific way of solving the task can be explained by the interpretation of competition's aims wherein the function of static allocation is dominating, in spite of occasional appreciation of the long-term advantages of the competitive innovation process. This view is closely connected with the conviction that removing information deficits is a problem to be solved separately, and that additionally provided information would automatically increase the efficiency of the competitive system.

This is the point where consumers' "godfathers" meet each other. It is quite understandable that each group looks upon its respective action programs as non-conflicting. This customary division of labor, however, does not necessarily best serve the common task of supporting a functioning competition and of adequately strengthening consumers' positions. When considering specifically Scitovsky's (1950) theses and the fact that consumer information changes the information environment of others too, it seems possible, according to the kind of information diffused, that consumer information might create problems not only for competition policy but for consumer policy as well. This idea suggests that more efforts be directed towards the evaluation and the further development of such instruments, within the frame of consumer research. It is the purpose of the following sections to point to questions which lead in this direction. Starting with a short outline of present consumer information policy in the Federal Republic of Germany (which is similar to American practice), we will discuss possible effects of price information and product tests on competition before questioning whether and to what extent there is a need for change.

Practiced Consumer Information and its Theoretical Foundation

The conditions under which consumers make buying decisions are the starting point for considering consumer information needs. These conditions include growing diversity and complexity of products, aggressive and manipulative selling techniques, and, last but not least, consumer deficiencies such as their comparatively small resources and their inability to make reflected decisions about their needs and to achieve a good price-quality relation. Thus, there is agreement that in order to correct consumer disadvantages and the resulting misallocations, the consumer should be given additional objective information, in order "to aid and foster a self-reliant, self-actualizing consumer who can make the most of decisions and play an equal role with sellers in the marketplace" (Thorelli & Engledow 1980, p. 9).

However, a good deal of the discussion in the area of consumer research and consumer policy is focused on the question of whether and to what extent information could contribute to the solution of consumer problems. Objections to the information model and the underlying image of the rationally acting consumer overlap with arguments against the corresponding claims of practiced information policy. These arguments include (1) that the provided information is not attained by those social groups who need most assistance; (2) that consumer information hunts a spectre by trying to improve market intelligence in that certain deficiencies are inherent characteristics of the market; (3) that these instruments start at a point where decisions concerning the fulfilment of needs have already been made and, thus, they contribute to the stabilization of existing social consumption patterns; and (4) that cognitive, motivational, temporal and spatial limitations of consumers is a hopeless undertaking since consumer behavior is more guided by hereditary S-R-patterns than by cognitive processes (Kroeber-Riel 1980, pp. 588-611). All these objections lead to the conclusion that the concept of consumer information needs to be legitimized. In fact, most of the research effort in the area is directed towards a search for confirmation, whether by discovering imminent efficiencies (e.g. with the help of improved lay-out and stronger incentives for users) or by discovering possibly positive effects on the sellers side which result from the mere existence of consumer information and not from the reactions of consumers. In particular, the discovery of indirect effects, called non-use benefits (Padberg 1977), was a great relief because this made it easier "...to bear our experience that the published test information usually does not reach those groups who need our special attention as far as consumer policy is concerned." (Neumann 1977).

The perception - or the rediscovery (cf. Beem & Ewing 1954) - of these effects upon suppliers shows an attribute which is very much neglected in the entire discussion: consumer information is a public good in the popular sense and, therefore, becomes known to producers and sellers as well, so it has not only effects on sellers' behavior indirectly through anticipation of possible reactions of consumers (Day 1976, p. 50), but also affects market performance and market structure because of the circular nature of the competition process (Sosnick 1958, p. 387). Consumer information is not only a matter of reactive adaptation but, because of its public character and its predictability, it can be used as an instrument to diminish future competitive pressure and, thus, indirectly change market structure. Both these aspects contribute to the fact that consumer information finally determines conditions of future competition as well as those structures with which consumers will be confronted when they are to make use of their free choice in the future. The same statement is valid for those effects which lead to direct changes in consumers' behavior because they also change

producers' and sellers' behavior as well as market performance and indirectly market structure.

In face of these interdependencies, consumer information is likely to create direct and indirect effects on competitive processes. Thus, it seems not to be sufficient to only look for affirming "use" or "non-use" benefits. It can be expected that for an adequate development of consumer information policy the opposite way would be more helpful, namely to look for "use-" and "non-use" costs'. The latter approach would not only uncover deficiencies and "traps" but would also result in consumer information which is adapted to the claims of workable competition. This approach is developed below.

The starting points of our considerations are two information instruments which are in widespread use in the US and in the FRG: price surveys and comparative product tests. Even though these instruments are essentially alike in both countries (product-related and seller-identifying price information on the one hand; quality ratings, connected with information on price differences on the other hand), the German procedure will be dealt with since it is characterized by peculiarities.

Price surveys as well as product tests are fully or partly financed by public means in Germany. Seventy-five percent of the costs for product tests are covered by the sales of the consumer magazine "test". Regional price surveys are distributed (with a circulation of about 500 - 30,000 samples) among consumers free of charge. They serve, like the "test" - magazine, as a base of information for local consumer advisory services. The dissemination of product test information is not restricted to the 'test' - magazine (circulation in 1980: more than 620,000 samples) but in a condensed form they are published by several other media.

Price surveys are carried out in a selected number of big and medium-sized cities (1979: 35 cities) as a non-disguised survey. The product categories concerned are especially electrical household appliances, audio - and video equipment as well as photographic articles, in some cases even food, detergents, and cleansers. They tend to be well selling brands and models. The results of the surveys are presented in brochures (title: "What to buy where cheapest?") in which, for every included model, the lowest price (optically accentuated), the highest price, the medium price, and the absolute as well as the relative price differences are shown. At the core of the tables is a compilation, specified by models, where the cheapest sellers are listed by name, address, and their demanded price. There is also a notice indicating whether the product was available at the moment of inquiry.

The quality rating published in the consumer magazine "test" is made by the "Stiftung Warentest" on the basis of tests that are carried out externally. "Stiftung Warentest" is a public foundation which is bound by charter provision to "...inform the public about objectifiable characteristics of utility and use-value of goods and services".

The specific organizational structure including controlling and technical bodies (members are sellers, representatives of consumer organizations, and "neutral" persons) implies much more transparency concerning testing intentions and operations than in the case of US-organizations. This transparency is enforced by the early public announcement of tests planned as well as several intended communications with the respective manufacturers. Similar to US-conditions, the tested product samples are limited to a certain number of nation-wide sold models whose market share is more than average. The testing programs are designed by the staff of the "Stiftung Warentest" and discussed by the bodies already mentioned. The same procedure takes place as far as the fixing of rating criteria and their weights are concerned.

The results are presented in detailed tables, where the separately rated characteristics and the subsumed test rates (five ratings from "very good" to "deficient") are shown. Usually there is a notice referring to minimum "medium" and maximum prices for the model as determined from a separate representative survey. Furthermore, these very detailed results are compressed to a short "compass" where the tested products (classified from the cheapest to the most expensive) are presented together with their most important ratings. In an additional text test results, characteristics, and trends of the concerned product market are explained and hints for adequate produce usage are given.

Comparatively high importance is given (as was found by our own investigations of a two years' edition of the 'test' - magazine) to information about prices, price differences and price trends. Frequently the "Stiftung Warentest" headline stresses the benefits of price active behavior and, in this context especially, the profitability of price comparison. They do not inform however about the dispersion of prices, because it is believed that even these simple diagrams will not be understood by the average consumer (Huttenrauch 1973).

The Effects of Consumer Information on the Competitive Process: Observations and Hypotheses

Advice, Direction, and Incentive to Buy by Means of Comparative Testing

As already shown, both comparative testing and price surveys have above all two characteristics in common: information is limited to a sample of renowned, well selling models and they are hardly suited to promote the proclaimed reflection on needs in so far as they are only able to give a certain assistance after the decision about needs has already been made. Comparative tests, however, have more advisory capacity, since they do not directly focus the attention on only one product. But the underlined final test results and price data give rise to a sympathy effect which, last but not least, results in a common desire to be relieved. Therefore, consumer information receives - in contrast to the original intentions - a partly uncritical confidence and works, in connection with the fact of product selection similar to commercial advertising.

Going even further than these directing effects, particularly comparative tests are expected to give buying incentives not only for individual products but also for product classes, because these tests create an effect of attention (especially when there are good rates resulting) and in this way influence decisions for one certain purchase out of many other product alternatives.

Selective Information about Price and Quality as an Investment to Strengthen Seller's Power

The fact that product test information and price surveys report primarily on well-known and widespread trademarked goods intensifies in principle those tendencies which would occur by themselves because of consumers' ignorance (Scitovsky 1950). The sales of large firms and well-known brands - which already serve as surrogate for information signals (cf. Nelson 1976; Spence 1976; Hauser 1979) - are thus promoted (when there are good test rates), whereas small and/or only locally selling firms, as well as newcomers suffer competitive disadvantages. Transmitted by retail's screening, this can in some cases go as far as threatening existence.

In spite of a possible expansion of the evoked set, consumer information makes only insufficient use of the opportunity to stimulate competition among producers. Beyond this, there is a further effect in the field of

retail: well-rated products and/or those which are shown as being moderately priced are, so to speak, presold, a fact, which is accommodating big low-price shops and furthermore contributes to retail's concentration because of the regularity of tests and price surveys.

Promotion of Lemon Markets

As already mentioned, consumers' ignorance tends to push away good quality and instead creates so-called lemon markets (cf. Akerlof 1970), if the market mechanism itself does not produce quality discriminating signals and appropriate information is lacking¹. However, the efficiency of these kinds of signals or respective information depends in how they are perceived. In case they are noticed by only a few segments of the entire consuming public, hybrid forms of lemon - and high quality markets may develop. Both of the phenomena are especially fostered by practiced consumer information which essentially reaches an elite only (Thorelli & Thorelli 1977), by pointing out "better" alternatives to just the quality conscious user only who is able to criticize and therefore initiate his exit (Hirschman 1970). But the majority of immobile consumers might remain loyal to slack firms (the poor pay more; cf. Caplovitz 1963) and secure their existence.

The Influence of Comparative Testing on Entrepreneurial Innovation Activities

It is known at least since the empirical research of Beem & Ewing (1954) that comparative testing (especially repeated testing of the same product classes) and the awarding of discriminating test rates have an influence on entrepreneurial R&D activities, because "...manufacturers begin to act as if this information may be an element in their competition with other manufacturers' (Padberg 1977, p. 9). The test criteria act as a minimum standard: As times goes on, they lead towards quality equalization and they influence methods of and technologies for quality control. They do so, not only after a test but - because of the common early announcements of tests in the FRG - also with respect to future tests. Critical points for the competitive process of innovation are that, on the one hand, they occupy the attention of entrepreneurial R&D activities and also prevent them from looking for alternatives. On the other hand, it is not sure that the quality standards requested by the experts are congruent with those of consumers (Kroeber-Riel 1977, p. 100). Finally, these influences create no pressure for innovation as soon as qualities are standardized. Thus, comparative testing has a tendency to direct innovation, but it is stimulating innovation only within the borders of a given technology.

Promotion of Model Change

The possibility or necessity for model changes are promoted by the now practiced comparative testing, as well as by price surveys - especially for goods which attract a high degree of attention. The latter is the case when tests of a certain product class are often repeated within short intervals of time. In this manner, the personally felt aging of products is increased and therefore increases the desire for replacements. Among other things, this takes place because consumers are often not offered a choice between unchanged and changed models at a price differentiated fully reflecting the cost-saving made possible by sticking to the former, and finally there is neglect of social costs (waste of resources and environment) which are caused by frequent model change.

¹This phenomenon has been discussed in economics recently; cf. especially Akerlof 1970; Nelson 1970; Nelson 1976; Spence 1973; 1976; 1977; Rothschild 1973; Hirshleifer 1973; Darby & Karnai 1973; Salop & Salop 1976; Rothschild and Stiglitz 1976; Jaffee & Russell 1976; Hauser 1979.

Price surveys in comparison do not only promote the possibility but even increase the necessity for frequent model change especially for mature and relatively high priced goods. Usually this is true for those products where there is an intensive price competition in agglomerated areas anyhow, which leads to short term price collapses that can only be caught up by price concessions on behalf of the manufacturers. Since this cannot go on infinitely, the way out of this dilemma is a regular change of models, which usually does not mean basic technological change, but which does offer the possibility to raise prices - at least temporarily - and so to warrant "healthy" mark-ups. The predictability of periodical price surveys of many considered products induces especially modern large scale retailers (at least at the time of the inquiries) to cut prices down to a level which is attractive for advertising. So the described process of price collapses and the resulting coercion for frequent model change are even intensified. In so far as currently practiced consumer information does not only contribute to the planned obsolescence which is so much criticized by consumer policy, it even contributes to tie up resources which then are missing for beneficial innovations and a socially desirable process of R&D.

Consumer Information as Substitute or Supplement of Market Information and Open-Price Systems

There was a discussion about supplier-sided market-information and open-price systems in the US as early as at the turn of the century (cf. Eddy 1913), and about 50 years later a lively debate on their evaluation started in the FRG, in the course of which in view of the requirements of dynamic competition and especially of incentives for competitive initiatives - essentially those conditions were developed as a profile, which would not allow to come to a conclusion about suppliers' and customers' identity or about individual transactions; which must be open to the supply as well as to the demand side; which report on completed transactions only; and which need to fulfill certain requirements as far as minimum-, maximum-, or average prices are concerned.² Likewise relevant for competition as market information - and open price systems is identifying price- and quality information prepared for consumers, because they produce mutual knowledge, the importance of which must be at least looked upon ambivalently. This is immediately understandable as far as identifying price surveys are concerned. Referring to the comparative product tests, this is similarly true. But being a source of information about the quality of competing products, this parameter hardly hinders active competition (cf. Beem & Ewing 1954, p. 117) in as far as the respective information is transmitted with a delay and so the imitation lag, which is necessary for workable competition, remains.

Bias Caused by Emphasizing Price as Competitive Parameter

There is a still increasing discussion going on about competitive effects as well as distortions caused by identifying price comparisons, and about the advantages of test-caused price active behavior. In general, one can expect the permanently repeated advice (in test reports) and the information about low-price sellers which is published in the price surveys to cooperate. One of the crucial points is that for several reasons it cannot be made sure that the consumer actually makes the "best buy" because of his price knowledge. On the one hand, price comparisons usually do not include technologically homogeneous goods (i.e., actual quality information which is necessary for "intelligent buying decisions" should be

²This corresponds closely to American jurisdiction concerning the Sherman Act Sect. 1; cf. Scherer 1971, p. 450-453.

published too as a premise for "true" price transparency). Since this does not happen, it may remain veiled that the cheaper product is possibly not the one with the more favorable price-quality-relation. On the other hand, identifications directing consumers to low-price-sellers do not necessarily guarantee the realization of good bargains, and they might not even warrant a buying environment which is profitable for all consumers.

On the long run, identifying price surveys are especially advantageous for aggressive distribution systems and "submarginal" retailers, and in directing distribution structures towards concentration. Transmitted by producers' increasing dependence on retail - which is a result of this concentration - there are feedbacks on the structure of production as well as on product supply itself. In the concrete case of a purchase, they might be favorable to consumers when those skim the cream, but perhaps not when they do all their shopping there. It should be added - at least in the case of complex products - that a buyer who is guided by price surveys could be subject to persuasion in the course of sales talks. He might buy a different product which is calculated more in favor of the seller and which might be cheaper in a different store. It is well known that price information is deliberately used in ads as a flat catcher in order to attract potential buyers. In a recorded case this method was abused to divert buying decisions to more profit-bearing products by means of bonus/malusincentives for salespersons; a case which most probably is no exception.

Reduction of Services and Lowering of "Marginal Morals" in Retail as a Result of Identifying Price Comparisons

"Last week I heard one Lady ask, 'Is this book O.K. for a nine year old?' and the clerk answered, 'Look, Lady, all I know is the list price is \$3.50 and our price is \$1.98'" (The New York Times, Dec. 1, 1961, p. 58, cited by Kahn 1966), p. 34. This example illustrates drastically the effects of structural changes in retail, which are supported particularly by identifying price surveys which increase the normal desire to make good bargains. These structural changes - among other things caused by price policy - are a result both of more favorable acquisition conditions of great distributors and of their neglect of traditional services like qualified advice, free delivery etc., a behavior which can usually be found by "armchair-retailers". Frequently this causes consumers to go to a qualified specialist retailer for advice, but to buy at a low-price shop in order to make noticeable savings. In economics this phenomenon is known as external effect. It is supposed to be one of the reasons for the failure of self-control in free-enterprise systems, in as much as there is no possibility to internalize such effects, which would mean in our case to separately charge these normally cost-free services. As a matter of fact, it would be possible to bring several of these services to an extra account, services which are characteristic for specialized trade and are claimed as being very desirable (cf. Kahn 1966, p. 37). This is not valid for consultancy by specialized personnel, because it is neither possible to tie these services to a purchase, nor can they be redemanded. So they are a quasi-public-good, the provision of which is dependent on the corresponding "part of subsidy" included in the price. Whenever this subsidy can no longer be held up because of the competitive pressure of submarginal traders, the competitors are forced to adjust this marginal behavior.³ In reality this result, in consumers being left

³ Such tendencies are imminent to the competition process which is based on enforcement of egoism. On the part of neo-liberal social scientists this has been recognized as a factor which can threaten an effective and (in favor of consumers) well functioning do-ut-des-principle. This is true whenever the ruling marginal moral is exposed to a submarginal pressure. (cf. especially Briefs 1957; Ropke 1961, p. 145-255)

alone if there are no public funded or commercial institutions which offer such services. In either case, the consumer would have to pay for such services.

Status Quo-Oriented Control Effects and Undesired Structural Changes as a Consequence of the Tyranny of Small Decisions

Considering the above impressions in the context, one can find a coherent pattern: On the one hand, its origin is to be found in the perspective of consumer policy which is oriented towards compensation of deficiencies of the present economic situation of consumers and towards an improvement of allocative efficiency. Consumer policy thus intends to contribute to an increase of economic efficiency in general. But - similar to what has been shown by Leibenstein (1966) at the example of allocative efficiency vs. X-efficiency - only marginal improvements could be made in comparison to what could have been achieved by an orientation according to the needs for a dynamic development of economy and society. On the other hand, consumer policy's orientation towards the problems of today, together with the provision of only selective information, may possibly lead to structures - by way of quantitative and temporal cumulation of effects of many (though individually advantageous) "small" measures and decisions - which are less favorable than today's ones, "in the sense that consumers might disapprove of the larger results thereby produced, if they were ever given the opportunity explicitly to vote for or against it" (Kahn 1966).

Guiding Principles for Consumer Information Which is Compatible with Workable Competition

Starting from the already outlined considerations and observations, the question arises - what kind of a profile a consumer information policy can have that is compatible with the aims of workable competition and which is the starting point for an obviously necessary re-orientation of consumer information. The above referenced deficiencies and "traps" give first hints, in as much as they indicate misdevelopments and possible one-way roads. In addition to a concept of a functionally appropriate consumer policy however, it is necessary to develop positive guidelines.

A first link is the idea, common to practically every concept of workable or dynamic competition, to understand competition as a process in which corporations develop initiative - they cut prices, improve product quality, create new products or processes - and in which imitators push from behind; not immediately (in order to preserve incentives for initiatives), but with not too much of a time lag either, so that the advantage can still be caught up and all the customers benefit from cheaper and better products. This assigns a key position to the time aspect (i.e. the central problem is to create or to preserve competitive situations (cf. Clark 1954, p. 324-328) so that the effective "imitation-lag" roughly corresponds with the time necessary for initiatives, and to prevent or abolish situations, in which the effective time is either much shorter or much longer than the necessary imitation lag (cf. particularly Prahl 1969, p. 26-48). Whereas in the latter case, which is characterized primarily by dominating positions and barriers to entry, potential competition is to be initiated, the first case is marked by a competitive stalemate-situation that can be found especially in oligopolies with only a few firms, in markets with comparatively homogeneous goods, and where participants are well informed. Such stalemate-situations are not broken up by more transparency however. They are even strengthened. An encouragement for the creation of advantageous big innovations, e.g. by pointing out the respective chances for innovations, would seem to be an adequate way. Consumer organizations could act as intermediating agencies between consumers' needs and possibilities of production. From the point of view of consumer policy, this does not mean that innovative activities

should be promoted at any price, but that only innovations with a positive trade-off between an increase in benefits and social costs should be initiated.

Another important aspect for the further development of consumer-information policy results from the necessity to get on distance from the image of the consumer as a selfish permanent-calculator with an irrational passion for rationality and to accept that consumers in a series of decisive situations do not want more intellectual efforts but relief (e.g. Kroeber-Riel 1977, p. 98). This makes sense, because this way the consumer keeps the capacity to continue to develop his needs in other, subjectively more attractive fields.

By using these considerations for the further development of consumer information, at first glance, the moderate transparency stressed by most concepts of workable competition seems to be an important guiding principle which suggests a certain reserve in the provision of consumer information. This may be opportune for some (especially mature) markets, but in many other markets (especially those with high rates of innovation) this will hardly create situations which would end in competition stalemate. Anyhow, the idea of "necessary" deficiencies results from the common perspective that suppliers can only be disciplined by buyers' exit. However, considering the fact that market reality is not only marked by exit, but also by loyalty (in connection with voice as soon as quality is diminished; Hirschman 1970), transparency may be evaluated differently. Voice does not only inform firms about CD and the related reasons, but it even diminishes - loyalty provided - the danger for existence, if such a firm is willing to take the evident chance for increased efforts. Improving the possibilities for articulation and transmitting voice (may be by initiating and evaluating complaints) would be a starting base for activities of consumer organizations which would not conflict with the transparency problem.

In spite of this diminishing of the transparency problem, identifying price surveys without simultaneous quality comparisons seem to be problematic and should be omitted in the frame of consumer information and consumer advice. As long as information about "best buys" as an instrument of consumer information is continued, adequate advice should be given as far as the implications of price and quality as well as differentiating characteristics of different types of shops are concerned. Easily available and consumer-conscious information about the buying-risks, which are related to individual product classes might be of a much higher value for buyers than regular price surveys, although the information does not give a "product-survey", but it enables one to avoid possible "traps". Adequate starting points for a reorientation of the requested kind are already being developed (cf. Schoppe & Czerwonka 1980, p. 69-70). After their introduction and further development they would be able to contribute to an increase in consciousness of consumers. However, facing scarce resources, it is necessary to set priorities with which the categorization which was introduced by Nelson (1976) in "search-" and "experience qualities" might be of some help. Product classes with a high portion of "experience qualities" which are unveiled to the consumer only by the way of usage and which are reported about only insufficiently by other sources should thereby be preferred.

Considering the further development of comparative product tests, it seems to be urgent on the one hand not only to raise but to lower barriers to market entrance for small sellers and newcomers, and, on the other hand, to improve the chances for product innovation by adequate reports.

It also seems necessary to have a stronger feedback between the choice of product classes, criteria for tests and ratings, and consumers. Only in this manner could it be

assured that dynamics of competition would be promoted and that the innovative process receives adequate signals as well as new impulses.

For an improved coordination of test-criteria and their importance to consumers' needs there are two different procedures, which could be tried either alternatively or - possibly - complementarily. The first procedure follows the conception of "perceived buying-risk", whereby a non-fixed number of buyers are interviewed about buying-risks with the help of semi-standardized questionnaires. In a second run the so-won risk-catalogue is presented to a random sample of consumers once more, in order to measure the relative importance of risks. In our opinion, this starting base offers a feasible approach for the development of consumer-oriented product test criteria. Furthermore, it enables one to find out changes in the desired quality standards by way of inquiries, and to feed this information back to producers by means of subsequent comparative tests.

The second method is risk-oriented too, but in this case it is tried to make use of group-dynamic effects by not only inquiring "isolated" consumers about deficiencies and suggestions, but by gathering and inquiring small groups of consumer "experts". In this case too, perceived risks and their relative importance should be evaluated separately in a two stage process. Possibly - at least some experience lead this way - group discussions may contribute not only to summarizing perceived deficiencies, but even to detect uncovered needs (which means desirable innovation in this case) and by doing so give information which could be passed on in the frame or test reports - information which would then become public but nevertheless stimulate competing firms to even more competition about better solutions.

References

- Akerlof, George (1970), "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," Quarterly Journal of Economics, 84, 488-500.
- _____ (1976), "The Economics of Caste and the Rat Race and Other Woeful Tales," Quarterly Journal of Economics, 90, 599-617.
- Bain, Joe S. (1967), Barriers to New Competition. Their Character and Consequences in Manufacturing Industries, 4th ed., Cambridge, Mass.: Harvard University Press.
- Beem, Eugene R. and Ewing, John S. (1954), "Business Appraises Consumer Testing Agencies," Harvard Business Review, 33 (2), 113-126.
- Bock, Josef (1958), "Verbraucherpolitik und Wirtschaftsordnung," in Verbraucherpolitik, ed. Josef Bock and Karl G. Specht, Koln and Opladen: Westdeutscher Verlag, 50-69.
- Briefs, Goetz (1957), "Grenzmoral in der pluralistischen Gesellschaft," in Wirtschaftsfragen der freien Welt, eds. Erwin v. Beckerath et al., Frankfurt/M.: Knapp, 97-108.
- Caplovitz, D. (1963), The Poor Pay More, New York: Free Press.
- Chestnut, Robert W. (1977), "The Delineation of Consumer Research. A Comment," Journal of Consumer Policy, 1, 195-196.
- Clark, John M. (1954), "Competition and the Objectives of Government Policy," in Monopoly and Competition and Their Regulation, ed. E. H. Chamberline, London and New York: MacMillan, 317-337.

Clark, John M. (1968), Competition as a Dynamic Process, 3rd ed., Washington, D.C.: The Brookings Institution.

Darby, Richard R. and Karni, E. (1973), "Free Competition and the Optimal Amount of Fraud," Journal of Law and Economics, 41, 67-88.

Day, George S. (1976), "Assessing the Effects of Information Disclosure Requirements," Journal of Marketing, 40, 45-52.

Eddy, Arthur J. (1913), The New Competition, 2nd ed., Chicago: McClurg.

Erhard, Ludwig. (1957), Wohlstand für Alle, Düsseldorf: Econ.

Ferguson, John R. et al. (1971/72), "Consumer Ignorance as a Source of Monopoly Power: FTC Staff Report on Self Regulation, Standardization, and Product Differentiation," Antitrust Law & Economic Review, 5 (2), 77-102; 5 (3), 35-74.

Hauser, Heinz. (1979), "Qualitätsinformationen und Marktstrukturen," Kyklos, 32, 739-763.

Hildebrand, George H. (1951), "Consumer Sovereignty in Modern Times," American Economic Review, Papers and Proceedings, 41, 19-23.

Hirschman, Albert O. (1970), Exit, Voice, and Loyalty. Responses to Decline in Firms, Organizations, and States, Cambridge, Mass.: Harvard University Press.

Hirshleifer, J. (1973), "Where are we in the Theory of Information," American Economic Review, Papers and Proceedings, 63, 31-39.

Hüttenrauch, Roland. (1973), "Probleme um Qualität und Preis beim Warentest," Markenartikel, 35, 434-444.

_____. (1977), "Zur Methodik des vergleichenden Warentests," Journal of Consumer Policy, 1, 143-150.

Jaffee, Dwight M. and Russell, Thomas. (1976), "Imperfect Information, Uncertainty, and Credit Rationing," Quarterly Journal of Economics, 90, 651-666.

Kahn, Alfred E. (1966), "The Tyranny of Small Decisions: Market Failures, Imperfections, and the Limits of Economics," Kyklos, 19, 23-47.

Kennedy, John F. (1963), "Special Message to the Congress on Protecting the Consumer Interest," Public Papers of the Presidents of the United States, 235-243.

Kroeber-Riel, Werner (1977), "Kritik und Neuformulierung der Verbraucherpolitik auf verhaltenswissenschaftlicher Grundlage," Die Betriebswirtschaft, 37, 89-103.

_____. (1980), Konsumentenverhalten, 2nd ed., München: Vahlen.

Leibenstein, Harvey (1966), "Allocative Efficiency vs. "X-Efficiency," American Economic Review, 56, 392-413.

Mitchell, Wesley C. (1912), "The Backward Art of Spending Money," American Economic Review, 2, 269-281.

Morris, Ruby T. and Bronson, Claire S. (1969), "The Chaos of Competition Indicated by Consumer Reports," Journal of Marketing, 33 (3), 26-34.

Nelson, Phillip (1970), "Information and Consumer Behavior," Journal of Political Economy, 78, 311-329.

Nelson, Robert H. (1976), "The Economics of Honest Trade Practices," Journal of Industrial Economics, 24, 281-293.

Neumann, Lotte F. (1977), "Kommentar zu Padbergs Beitrag," Journal of Consumer Policy, 1, 15-16.

Padberg, Daniel I. (1977), "Non-Use Benefits of Mandatory Consumer Information Programs," Journal of Consumer Policy, 1, 5-14.

Prahl, Klaus (1969), Patentschutz und Wettbewerb, Göttingen: Vandenhoeck & Ruprecht.

Röpke, Wilhelm (1961), Jenseits von Angebot und Nachfrage, Bern: Haupt.

Rothschild, Michael (1973), "Models of Market Organization with Imperfect Information: A Survey," Journal of Political Information, 81, 1283-1308.

_____. and Stiglitz, Joseph (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," Quarterly Journal of Economics, 90, 629-643.

Salop, Joanne and Salop, Steven (1976), "Self-Selection and Turnover in the Labor Market," Quarterly Journal of Economics, 90, 619-627.

Scherhorn, Gerhard (1977), "On Consumer Conduct and its Importance for the Workability of Markets," Journal of Consumer Policy, 1, 20-31.

_____. (1979), Verbraucherpolitische Einordnung und Begründung des Projekts 'Strategien der Verbraucherinformation', Hohenheim: Working Paper.

_____. (1980), "Keine Zukunftsprobleme im Konsum?," Wirtschaftsdienst, 60, 437-441.

Schöppe, Günter and Czerwonka, Christine (1980), "Empirical Consumer Research in the Federal Republic of Germany: The Research Projects of the Federal Department of Research and Technology," in Advances in Consumer Research, Vol. VII, ed. Jerry C. Olson, Ann Arbor: Association for Consumer Research, 66-76.

Scitovsky, Tibor (1950), "Ignorance as a Source of Oligopoly Power," American Economic Review, Papers and Proceedings, 40 (2), 48-53.

Sosnick, Stephan H. (1958), "A Critique of Concepts of Workable Competition," Quarterly Journal of Economics, 72, 380-423.

Spence, Michael (1973), "Job Market Signaling," Quarterly Journal of Economics, 87, 355-374.

_____. (1976), "Informational Aspects of Market Structure: An Introduction," Quarterly Journal of Economics, 90, 591-597.

_____. (1977), "Consumer Misperceptions, Product Failure, and Producer Liability," Review of Economic Studies, 44, 561-572.

Thorelli, Hans B. and Thorelli, Sarah V. (1977), Consumer Information Systems and Consumer Policy, Cambridge, Mass: Ballinger.

_____. and Engledow, Jack L. (1980), "Information Seekers and Information Systems: A Policy Perspective," Journal of Marketing, 44, 9-27.

Wärneryd, Karl-Erik (1980), "The Limits of Public Consumer Information," Journal of Consumer Policy, 4, 127-141.

THE IMPACT OF NON-COMMERCIAL PRODUCT TEST INFORMATION
ON COMMERCE AND INDUSTRY
- OUTLINE OF THE STUDY AND PRELIMINARY FINDINGS -

Wolfgang Fritz, University of Mannheim
Harald Hilger, University of Mannheim
Günter Silberer, University of Mannheim
Hans Raffée, University of Mannheim¹

Abstract

This paper presents an outline of the research program and preliminary findings of an empirical investigation, concerning the impact on the consumer-goods industry of product-test information in West Germany, published by the most important non-business test institution, the "Stiftung Warentest".

Introduction

Of the more than one dozen commercial and non-commercial product-testing bodies and institutions in the Federal Republic of Germany, the most important, in terms of consumer policy, is the non-commercial "Stiftung Warentest" in Berlin (see Meiners 1968, pp. 64-69; Thorelli and Thorelli 1974, pp. 246 - 288; Brinkmann 1976, pp. 30-43; Silberer 1979, pp. 65-87; Kandler 1980). The importance of this independent testing institution, which is sponsored by public funds, is evident due to the high degree of diffusion of its test results and its well-known status in West Germany.

Following a representative survey, the "Stiftung Warentest" was known in 1978 to 79% of the population in West Germany. Its most important publication is the monthly journal "test" which, in 1978, achieved a total edition of 786,000 per issue, reaching 15% of the population for a readership of 7 million consumers per issue. The readership, in the widest sense, taking those into account also who seldom read the journal, reached 34% of the population in 1978; this was approximately one-third of the total population of West Germany (see Stiftung Warentest 1978, pp. 5-6 and p. 49).

The fact that the "Stiftung Warentest" is so very well known is not due merely to the growth in the size of editions of the journal "test", but more especially to the increasing diffusion of the test results; these are published as short reports in other newspapers and publications and are broadcast on radio and television (see Silberer 1979, p. 67-75; Stiftung Warentest 1978, p. 5, p. 56 and pp. 69-81).

The fact that the "Stiftung Warentest" is so well known and has gained such a wide diffusion is not sufficient to evaluate the consumer-political effects of the "Stiftung Warentest". However the question must be examined, whether and how far the activities of the "Stiftung Warentest" help realize consumer-political goals (see Silberer 1979a, pp. 110 f.).

¹The authors are members of the "Forschungsgruppe Konsumenteninformation" at the University of Mannheim - a research project that belongs to the research program "Empirische Verbraucherforschung", sponsored by the Federal Department of Research and Technology, West Germany.

The "Stiftung Warentest" is, according to its constitution, bound to inform the public on the quality of widely used consumer goods which can be objectively measured. According to the aims of the federal government (the founder of the institution), the "Stiftung Warentest" is designed to augment the freedom of consumption for the consumer and to strengthen his market position vis à vis the suppliers of consumer goods, thus contributing to the maintenance and expansion of a workable competition (Friedrichs 1974, pp. 6-9; Schachtschabel 1976, p. 101; see Silberer 1979a, p. 111; Czerwonka, Schöppe and Weckbach 1977, pp. 279 f.; Biervert et al. 1977, p. 135; Scherhorn 1975, pp. 132 f. and pp. 211 ff.; Meiners 1968). Consequently the efficiency of the "Stiftung Warentest" must be judged not only by its impact on the consumer but also on the supplier, since this is the only way that one may evaluate the "non-use benefits" of the product test.

The Present State of Research

The systematic investigation of consumer test effects, as far as suppliers are concerned, has been considerably neglected both in Europe and in the United States. Only few empirical studies have been carried out in this field; these can be roughly classified into three groups:

- 1 The first group deals solely with the content analysis of the test reports and more particularly with the changes of the average quality level and/or price-quality relationship (USA: Morris and Block 1968; Morris and Bronson 1969; 1970; West Germany: Beier 1977; 1978; Diller 1977; Mensch 1975, pp. 68 f.).
- 2 Another group reports on investigations dealing specifically with single aspects of the effects of the consumer reports in the field of commerce, taking personal selling and assortment policy especially into account. Questionnaires and observation methods are used (see Beyss 1977; Eiermann 1977; IfaV 1977; Landes 1978; Landes & Specht 1979; Silberer 1980).
- 3 The last group deals with the investigation of the effects of the consumer product test on industry and commerce. Only a pilot study has been carried out in West Germany (Arnold 1978). The most comprehensive study in the US was conducted by Beem and Ewing in 1954.

None of these investigations are either systematic, comprehensive or a reliable analysis of the test effects on industry and commerce (see Silberer 1979a, p. 119). These investigations were based on insufficient samples and, therefore, often represent merely case studies. Furthermore, the studies build upon no theoretical background, and no theoretically oriented explanations were given (e. g. Arnold 1978). The methods used were often ill-suited for revealing causalities between the product-test results and certain assumed effects, eg. the changes of price-quality relationship. This applies especially to the content analysis. The investigations are therefore not generally representative, and have merely a descriptive character, possibly offering weak connections for, in our opinion, an overdue systematical analysis of product-test effects in the field of industry and commerce. The figures

indicated by the "Stiftung Warentest" concerning the changes in sales volume caused by the product tests refer to a few examples which have not been systematically investigated (see Hüttenrauch 1978; Silberer 1979a, p. 119). They thus have merely an anecdotal form (Thorelli and Thorelli 1974, p. 256).

Objectives, Theoretical Background, and Methods Implemented in the Investigation

The objective of the research project is the systematic analysis of the consumer-political effects and the deficiencies of product test results. The research project seeks to analyze the effects of product tests on industry and commerce as well as the effects on consumers. This combined observation makes it possible also to investigate the so-called non-use benefits of product tests. Such non-use benefits can be caused as a result of producers improving the quality of poorly evaluated products, these quality improvements, therefore, also benefiting the non-users of the product test. Furthermore, we have the opportunity more closely to analyze the interaction between suppliers and consumers.

This paper deals with the preliminary analysis of the effects and deficiencies of product-test information in the field of industry and commerce.

The first elementary objective of the study is to describe the possible effects and deficiencies of product test information, in particular in the following fields

- marketing activities (product policy, assortment policy, pricing, distribution, communication policy) of the suppliers;
- changes in sales volume;
- market structures.

It is planned to investigate whether and to what extent the product-test results lead to improvements in the product qualities, sales programs and price-quality relationships, as well as consumer-oriented product innovations. Furthermore planned is to investigate whether the diffusion of positively evaluated products is increased by the use of product tests in advertising, sales promotion and personal selling.

Another question to be examined is whether and to what extent product-test results are able to change the mobility of demand, which can be noted in changes in the sales volume.

Finally, note should also be taken of the changes in the market structure, due to product-test information. The question as to whether product-test results limit the degree of competition, or whether they influence the distribution of power between industry and commerce needs to be studied.

The second elementary objective of the investigation is to explain the described effects and deficiencies of product tests using theoretical approaches. The theoretical framework consists of two substantial principles, the gratification and capacity principle according to which individual and organizational behavior can be explained by expected or real gratifications and the restrictions of the behavior (see Silberer 1979, pp. 50-60; Schanz 1977, pp. 97-178). These principles help discover, classify, and integrate the theoretical approaches from which the workable hypotheses can be derived. In this context, approaches of communication, power and organization theory are especially important.

Product-test information influence the system of suppliers within a multi-step flow of communication. On the one hand, they influence the supplier system directly, i.e. without using information processing systems, and produce

effects, e.g. merely due to the anticipated test reaction of the buyer. On the other hand, test information actually cause changes in buyer behavior and in the behavior of other relevant persons or bodies (e.g. standardization institutes) and thus have an indirect influence on the suppliers (see Silberer 1979a; p. 112). In this context the hypothesis of communication theory, in particular dealing with multi-step flow and the efficiency of communication are of fundamental importance when explaining the test effects on the supplier system (see e.g. Specht 1979, pp. 139-155, 175-178; Silberer 1979, pp. 40-47; Kroeber-Riel 1980).

The type and extent of the test effects depend a great deal on the real or supposed distribution of power between buyers and suppliers, and this again can be altered by the test results (see e.g. Box 1977, pp. 190-196; Specht 1979, p. 159). Product-test information increases the independence of the consumers and makes it easier for the consumers to implement sanctions, such as "voice" and "exit" vis à vis the suppliers, and increase the bargaining power of the consumers (see Hirschman 1974, pp. 17-45; Scherhorn 1975, pp. 36-40, 76 f., 211-213). Product-test information improves mobility and articulation possibilities, therefore increasing the potential power that the consumer has against the suppliers (vgl. Box 1977, pp. 190-196; Specht 1979, pp. 67-74). Furthermore, the product-test information limits the power of the suppliers by increasing the degree of competition (see Müller 1965; Meiners 1968). Product-test information can alter the distribution of power between the suppliers by altering the articulation and mobility possibilities of commerce compared to those of industry. Thus not only approaches of communication theory but also those of power theory are extremely important for the explanation of the test effects.

Product test information often has merely an effect upon the supplier system if it brings about decision processes within organizations, resulting in changes in the behavior of these organizations. Behavioral theories of organization have to be taken into account when explaining the effects of tests on the supplier system; communication and power theoretical aspects also have to be considered (see Müller 1965; Schanz 1977; Kieser & Kubicek 1978; Wilpert 1980). Organizational reactions to test results depend, on the one hand, on the restrictions of behavior possibilities, e.g. the size or structure of the organization, and on the other, on the gratification expectations, e.g. perceived opportunities and risks (see Silberer 1979, pp. 148-155). Thus special note has to be taken of the so-called situational approach (see e.g. Kieser & Kubicek 1976), the behavioral theory of the firm (see e.g. Cyert & March 1963; March & Simon 1958), and the information processing approach (see e.g. Kirsch 1970-1971; Newell & Simon 1972) when explaining the effects of tests on the supplier system.

The third elementary objective is to propose practical measures for the marketing of the suppliers and especially for the product testing and consumer policy.

The investigation methods used include representative interviews using standardized questionnaires. The target is to interview at least one hundred firms, in industry and commerce. In industry, as the number of firms in specific branches to be interviewed is relatively small, complete investigations are to be carried out. Samples should also be conducted, above all among retailers.

The interviews in commerce and industry are designed to complement one another, and also to control the respective results.

The interviews are, as far as possible, to be carried out by the directors and members of the research group.

Preliminary Empirical Results

As the field work in the supplier system began only in August 1980 and is scheduled to last until the middle of 1981, we have as yet only few preliminary findings. These are based on extensive interviews lasting between 1 1/2 to 2 1/2 hours and were carried out with members of the top management of electrical appliance producers, especially those producing washing machines. Until now we have interviewed half of the producers selling washing machines in West Germany (both foreign and German makes), whose products were tested by the "Stiftung Warentest" (17 in all).

The survey method has proven to be very useful. The use of standardized questionnaires has presented no problems, nor have the interviews, which last between 1 1/2 and 2 1/2 hours. The willingness of industry to provide information has, until now, been considerable.

According to the results at present available, the marketing activities of the producers have been influenced in a number of ways by the product tests. With respect to the product policy most producers admitted taking note of the examining criteria of "Stiftung Warentest" when developing new products. Several top managers have even reported that their companies are prepared only to produce those products meeting the requirements of the "Stiftung Warentest". This means that in many cases it is not the published findings, but the anticipated inclusion of products in a test and the test requirements that influence the product policy.

The information from the sales companies of several foreign producers has been especially interesting; they stated that the test requirements of the "Stiftung Warentest" had been taken over by the foreign parent company for the development of new products for the foreign market. One of these firms is, in fact, one of the largest European companies in its field. This result can be considered to be a first indication of the international technological effects of product-test results in West Germany. Such effects can be expected especially for those foreign markets which have a large technological gap, which has to be compensated for by raising the level to that of the domestic market.

These international effects are limited not only by the different technical terms of delivery and stipulations on building that exist abroad, but also due to an obvious lack of international cooperation of the testing institutions. As several producers reported, the quality requirements of various foreign testing institutions, which differ to those of the "Stiftung Warentest", often stipulate that the production must be specifically adjusted to the quality requirements of testing institutions in foreign markets. At the same time however, the different requirements present sales possibilities abroad for those products which can no longer be sold at home due to the negative test results obtained.

Compared to the consideration of the test information within the field of product development, other product-political consequences of product-test information have hardly been mentioned. Some firms have mentioned the elimination of products and the introduction of new products, due to the test results. Very few firms have reported that product changes occurred because of the test results obtained.

In this context it is interesting to note that in some cases, the companies were of the opinion that the present product test had an inhibiting effect on innovations, above all new product ideas were either not sufficiently taken into account by the "Stiftung Warentest" or if notice was taken, this was usually too late.

An important area of test effects is, according to present results, the communication policy of the respective producers. About half the firms interviewed mentioned regular advertising and sales promoting activities, with positive test results, e.g. stickers or circulars sent to commerce. This indicates a difference in the effects of American tests, since these are not allowed to be used for advertising purposes (Thorelli & Thorelli 1974, p. 432).

Upon negative test results most firms try to ignore the results or to complain to the "Stiftung Warentest". Corrections, counter-advertising, or legal actions against the "Stiftung Warentest" have hardly ever occurred.

As far as the distribution policy is concerned, the test results above all influence the information used by the salesman. Until now all firms interviewed have admitted that they always inform their salesmen about the relevant test results. The salesmen often use these test results in sales discussions; moreover it has also been mentioned that commerce often initiates discussions based on the test results, and following negative test results special conditions or even the returning of products in question is demanded.

Nevertheless, according to the present findings, few test reactions of the producers deal with pricing. Price increases, due to positive test results, or price reductions, due to negative test results, have been carried out in only a few cases.

According to available information, most producers do not believe price reductions to be suitable to compensate for feared sales reductions due to negative test results, as these price reductions seem to "confirm" the poor test results.

The findings available to the present indicate that product-test information exerts a large influence on the sales of the firms interviewed. Most firms registered an increase in sales, following positive test results, and/or a drop in sales following negative test results.

These sales changes are, according to the producers concerned, however only of short duration, and after eight to ten months sales are back to normal. During this period a doubling of sales for products which received good results and/or a drop in sales by about 50% for products which received poor results was often the case. Spill-over effects of the positive test results on the sales of products which were not tested, however, have hardly ever been noted.

Not merely the actual, but also the anticipated reaction of sales to the test was considered particularly important by the firms interviewed. Virtually all the producers questioned took high sales risks into account following possible negative test results, and following possible positive test results, high sales chances were perceived. This is probably one of the decisive determinants of the indirect effects of the test results on industry.

To have an idea as to the relevant effects of the product-test on the market structures, we evaluated the appropriate statements made by the producers. Most of the producers interviewed were of the opinion that the product test increases quality competition, and at the same time favors large firms to such an extent that small and medium-sized firms are threatened with closure. Thus the power potential of the "Stiftung Warentest" vis à vis industry was considered to be especially large and insufficiently controlled. It was also considered to be superfluous as far as competition is concerned.

In summing up, it can be stated, that the test information of the "Stiftung Warentest" produces extensive effects within industry. These affect in particular product development, advertising, sales promotion, information of sales personnel, and above all the sales volume of the producers in question. It is important to note that many of these effects are brought about by anticipated demand behavior due to the test results and the resulting sales expectations. Thus, without a doubt the power and influence of the "Stiftung Warentest" on the market is considered very great. The proven results can confirm the assumption that the product test in the Federal Republic of Germany results in "non-use benefits" for the consumers.

These findings are, however, to be considered merely as preliminary results. The small number of interviews carried out up to now does not yet permit a determinant analysis nor hypothesis test.

All the same, due to the often mentioned statements, we can perhaps assume that the test effects are limited by - presumably - incorrect test results and above all by the low degree of topicality of the test results. In many cases, by the time the test report is published, new models or altered models of the goods tested are on the market. This can possibly have a neutralizing effect on the test results. The low degree of topicality could also possibly lead to the fact that potential opportunities to promote product innovations can often not be made use of.

If this is confirmed by the investigations still to be carried out, the current testing practice may be required to expand its testing capacities and possibly use new, time-saving media so as to alter this situation. The additional question should then be raised as to whether a test of proto-types, taking into account all the problems this involves, could contribute to increasing the innovative impulses of product-tests.

Furthermore, based on existing information, we can assume that due to a reduction of the international differences in quality requirements of the testing institutions, the international technological effects of product tests could be increased. This could be achieved by increasing the cooperation of the "Stiftung Warentest" with other foreign testing institutions.

To summarize the proven effects of the product tests on industry, it can be concluded that consumer policy should pay more attention to the supplier reactions on product test results than has to now been the case, and should not concentrate its activities solely on the consumers.

Summary

The "Stiftung Warentest" is the most important non-commercial product-testing institution in the Federal Republic of Germany. As its economic functions are directed towards maintaining and expanding a workable competition, its efficiency must be evaluated not only due to its effects on the consumers but also by its effects on the suppliers. The effects on suppliers have up until now not been empirically investigated to a satisfactory extent.

The objective of the reported research project lies in the description and explanation of the effects and deficiencies of product tests in the field of marketing activities, sales volume, and the markets of the suppliers of consumer goods, as well as in the presentation of practical suggestions.

Following the first extensive interviews with members of top management of electrical appliance producers, the preliminary results show that product tests have an especially marked influence on product development, advertising, sales promotions, information of sales personnel and sales volume. The power and influence of the "Stiftung Warentest" on the market can therefore be considered to be substantial.

References

Arnold, Ulli (1978), Warentestergebnisse und Absatzpolitik von Unternehmen, Bericht aus dem Seminar für Betriebswirtschaftliche Marktlehre der Universität Göttingen.

Beem, Eugene R., and Ewing, John S. (1954), "Business Appraises Consumer Testing Agencies", Harvard Business Review, 2, 113-126.

Beier, Udo (1977), "Vergleichende Warentests. Aufgaben, Ergebnisse und Möglichkeiten", Forum Ware - Wissenschaft und Praxis, 5, 137-161.

_____ (1978), "Entscheidungsbedingte Kaufkraftverluste: Formen, Umfang und verbraucherpolitische Relevanz", Zeitschrift für Verbraucherpolitik, 2, 159-171.

Beys, Bernd (1977), Einflüsse von Warentestergebnissen auf das Beratungsgespräch im Einzelhandel. Eine empirische Untersuchung. Unveröffentlichte Diplomarbeit, Universität Mannheim.

Biertvert, Bernd, Fischer-Winkelmann, Wolf F., and Rock, Reinhard (1977), Grundlagen der Verbraucherpolitik. Reinbek: Rowohlt.

Box, J. M. F. (1977), "Die Macht der Verbraucher auf dem Markt in Verbindung mit vergleichenden Warentest", ed. Kommission der europäischen Gemeinschaft, Kolloquium über Verbraucherinformation, Kommission 3, Brüssel, 165-200.

Brinkmann, Werner (1976), Die Verbraucherorganisationen in der Bundesrepublik Deutschland und ihre Tätigkeit bei der überbetrieblichen technischen Normung. Köln u. a.: Carl Heymanns.

Cyert, R. M., and March, J. G. (1963), A Behavioral Theory of the Firm. Englewood Cliffs, N.J.: Prentice-Hall.

Czerwonka, Christine, Schöppe, Günter, and Weckbach, Stefan (1976), Der aktive Konsument: Kommunikation und Kooperation. Göttingen: Schwartz & Co.

Diller, Hermann (1977), "Der Preis als Qualitätsindikator", Die Betriebswirtschaft, 37, 219-234.

Eiermann, Gudrun (1977), Der Einfluß von Warentestergebnissen auf die Sortiments- und Beschaffungspolitik von Handelsbetrieben. Unveröffentlichte Diplomarbeit, Universität Mannheim.

Friedrichs, Hans (1974), "Verbraucherpolitik oder Politik für den Verbraucher?", Stiftung Warentest, Jahresbericht 1974, Berlin, 3-9.

Hirschman, Albert O. (1974), Abwanderung und Widerspruch, Tübingen: Mohr

Hüttenrauch, Roland (1978), "Die Stiftung Warentest nutzt auch der Industrie", Markenartikel, 237-240.

Institut für angewandte Verbraucherforschung e. V. (JfaV) (1977), Die Angebotshäufigkeit von Produkten mit dem Testurteil "gut und sehr gut" im Einzelhandel Nordrhein-Westfalens. Eine empirische Untersuchung im Auftrage der Verbraucherzentrale Nordrhein-Westfalen e. V., Köln.

- Kandler, Cornelia (1980), Marketing-Management von Warentestinstitutionen, Berlin: Nicolai.
- Kieser, Alfred, and Kubicek, Herbert (1976), Organisation, Berlin: DeGruyter.
- _____ (1978), Organisationstheorien, II, Stuttgart u.a.: Kohlhammer.
- Kirsch, Werner (1970-1971), Entscheidungsprozesse, Vol. I/II/III, Wiesbaden: Gabler.
- Kroeber-Riel, Werner (1980), Konsumentenverhalten, München: Vahlen.
- Landes, Friedrich (1978), Warentestinformationen in ihrer Wirkung auf Einkaufsentscheidungen, Absatz und Absatzpolitik mittelständischer Facheinzelhandlungen. Eine empirische Untersuchung, Unveröffentlichte Studienarbeit, TH Darmstadt.
- _____, and Specht, Günter (1979), "Beachtung von Warentest-Ergebnissen vorteilhaft", Nachrichten der Industrie- und Handelskammer Darmstadt, 34, 93-95.
- March, James G., and Simon, Herbert A. (1958), Organizations, New York: John Wiley & Sons.
- Meiners, Dieter (1968), Ordnungspolitische Probleme des Warentests, Berlin: Duncker & Humblot.
- Mensch, Gerhard (1975), Das technologische Patt. Innovationen überwinden die Depression, Frankfurt/M.: Umschau.
- Morris, Ruby T., and Block, Betty (1968), "The Instability of Quality: As Revealed in 10 Consumer Union Studies of Sunburn Preventatives 1936 - 1966", Journal of Consumer Affairs, 2, 39-60.
- _____, and Bronson, Claire S. (1969), "The Chaos of Competition indicated by Consumer Reports", Journal of Marketing, 33, 26-34.
- _____ (1970), "The Potential Loss in Money Income to the American People by Haphazard Purchasing", Journal of Consumer Affairs, 4, 103-112.
- Müller, Udo (1965), Ökonomische Probleme und Wirkungen vergleichender Konsumgütertests, Diss., Würzburg.
- Newell, A. & Simon, Herbert A. (1972), Human Problem Solving, Englewood Cliffs, N.J.: Prentice-Hall.
- Schachtschabel, Hans G. (1976), Wirtschaftspolitische Konzeptionen, 3rd. ed., Stuttgart u.a.: Kohlhammer.
- Scherhorn, Gerhard (1975), Verbraucherinteresse und Verbraucherpolitik, Göttingen: Schwartz & Co.
- Schanz, Günter (1977), Grundlagen der verhaltenstheoretischen Betriebswirtschaftslehre, Tübingen: Mohr.
- Silberer, Günter (1979), Warentest - Informationsmarketing - Verbraucherverhalten. Die Verbreitung von Gütertestinformationen und deren Verwendung im Konsumentenbereich, Berlin: Nicolai.
- _____ (1979a), "Ansatzpunkte und Probleme einer Erfolgsbeurteilung beim neutralen Gütertest", Zeitschrift für Verbraucherpolitik, 3, 110-126.
- _____ (1980), "Die Verwendung von Testinformationen im Verkaufsgespräch des Konsumgütereinzelhandels. Ergebnisse verschiedener Pilotstudien", Die Betriebswirtschaft, 40, 81-93.
- Stiftung Warentest (1978), Jahresbericht 1978, Berlin.
- Thorelli, Hans B., and Thorelli, Sarah V. (1974), Consumer Information Handbook: Europe and North America, New York et al.: Praeger Publishers.
- Wilpert, Bernhard (1980), "Verhaltensorientierte Organisationstheorien", Zeitschrift für Betriebswirtschaft, 5,

SOURCES FOR PRODUCT IDEAS: A PROACTIVE
VIEW ON THE CONSUMER

Gerd Fleischmann, University of Frankfurt¹

Abstract

Normally consumers are not sources for product ideas both because for the individual consumer exit is cheaper than voice, and because producers for mass markets prefer process innovations to product innovations. But in affluent societies households are no longer constrained to be general producer units: they get the chance to specialize and to become expert partners to smaller firms which in order to survive show more interest in product innovation.

Why Consumers Are Normally Not Sources For
Product Ideas

Consumers usually feel powerless vis-à-vis firms. They believe all they can do with consumer goods, offered by a given firm, is take it or leave it. The firm too, be it a producer or a retailer, also feels pressured by the market. It believes that consumers easily switch to another supplier. Households and firms are anonymous to each other. Consumers do not suggest product changes or new products to firms, and firms do not ask households for new product ideas. This is not out of any ill feelings on either side, but, under certain conditions, the rational way to behave.

Consumers Prefer Exit to Voice: Exit Is Cheaper

If a consumer is dissatisfied with a consumer good, produced by one of several competing firms, he can choose either to induce the producer to change the good according to his preferences--this option is called "voice" by Hirschman (1970)--or to look for another supplier, i.e. to choose what Hirschman calls the exit-option.

In general the cost of exit is low. The consumer just has to find out, which are the other suppliers, to get from neighbors, retailers, consumer unions, or state agencies, some information about the qualities of the competitors' goods, and to buy next time one of these goods. Of course, there is the risk that experience with some other product is also disappointing. But even then, the difference between the expected and the experienced quality will hardly be larger, because the quality distribution of mass produced consumer goods is generally very narrow.

By contrast, the cost of effective voice is very high for the individual consumer. Admittedly, it is not expensive to write a letter to the producer. But chances are very poor that the complaints of a single consumer will induce the firm to change the good or to produce a new good. For, why should a firm with mass-produced goods follow the advice of an individual consumer? It is rational to do so only if there are indicators favoring the conjecture that the advice comes from a representative consumer. But it has to be suspected in the first place that consumers writing informed letters on how to change the quality of a product are not representative for the average consumer

of mass products. One effective way to convince the firm that it will pay to change the product is to induce a large number of other consumers to advise the firm concerning the same kind of change. But this is a very expensive endeavor, and even if a consumer tries it and succeeds, it is still uncertain, whether a firm will make the recommended change.

Seen from the point of view of the economic theory of collective action (Olson 1965), the choice of the voice option by an individual consumer is a contribution to a public good, in this case a contribution to the common goal of consumers with similar preferences regarding the kind of goods the individual consumes is complaining about. Generally, in large groups an individual has no incentives to make such contributions. A "political entrepreneur" is needed who organizes the group by preparing selective incentives for group membership, for instance in form of private goods, which are available for members only, so that free-riders can be excluded.

To sum up, if the consumer of a certain mass-product is dissatisfied with it, he will normally prefer the exit option, because it is cheaper than the voice option, if voice is to be effective. But in the case that consumers with similar tastes get organized by a political entrepreneur, the voice option may become cheaper.

Producers of Consumer Goods Prefer Process Innovations to
Product Innovations: Process Innovations Are Less Risky

My starting point is the general economic hypothesis that a firm facing alternatives with the same profit per dollar capital invested will choose the least risky. Now it is easy to see that for mass producers of consumer goods it is generally less risky to invest into a new production process than into a new product. The main reason for this is that the suppliers of investment goods, which are needed for the process innovation, have to supply their own investors with calculations and details of the know-how concerning the new process: in order to sell an investment good, one has to show that the good to be used will reduce production costs. There is no comparable activity on the side of the new consumer product. Even in cases in which the firm incurs the cost of getting informed about consumers' preferences of goods and of changes of goods, the probability that a new consumer good will be a success is relatively low.

The risk of innovation is controllable to some extent. One important way to reduce the risk is to have more communication with the main partners in the innovation process, namely the suppliers of new investment goods and the customers of new consumer goods. But here again it is more efficient for a firm with mass production of consumers goods to intensify the communication with firms supplying investment goods than with households buying the consumer goods; this is so mainly because in the case of communication between firms there are experts on both sides, but also because of the existing network of communication. Experts can easily agree, then, upon what is needed and what can be done about it, whereas households have in general no expertise on mass produced consumer goods and no easy access to the industry.

¹Professor of economics, co-ordinator of consumer research projects for the German Federal Department of Research and Technology. Address: Schumannstrasse 60, D 6000 Frankfurt/M. I thank Professor Joseph Agassi for his kind advice.

An additional reason for mass producers of consumer goods to concentrate on process innovation rather than on product innovation is that the loss of customers due to exit towards a producer of a new consumer good is usually a relatively slow process. Consumer habits usually do not change very quickly. But if a competitor succeeds in cutting down markedly production costs and thus prices, then there is the danger of a sudden loss of buyers. To put it differently, households are more accustomed to a kind of consumer good, for instance, to an automobile, a refrigerator or a washer, than to a special brand of that good.

In sum, mass producers of consumer goods generally prefer process innovations to product innovations because process innovations are easier to compute, because there is a better communication between firms than between firms and households, and because a successful process innovation by a competitor threatens the survival of the firm.

Consequences of the Dominance of the Combination of Exit and Process Innovation: More Goods and More Free Time for Everybody

The dominance of exit and of process innovations has resulted, on the positive side, in the rapid growth of productivity of labor and in rapid economic growth in general. On the negative side, the fine tuning of the qualities of consumer goods with the tastes of the consumers has been somewhat neglected. From an elitist standpoint this development may be deplored. Compared to the high standards of craftsmanship of former generations, many mass products of the present are of comparatively lower quality. But there are lots of consumer goods which even the most ardent admirers of pre-industrial craftsmanship would not miss: if not railroads, automobiles, airplanes, refrigerators, then cameras for reproductions of art work, hi-fi records, etc.

Seen from an egalitarian standpoint, the consequences of the combination of exit and process innovation seem more positive. The emphasis on process innovation resulted in transforming the luxuries of the rich of yesterday into mass consumption of the man in the street today. The most impressive example in the history of this development may be Ford's famous model T. It has not been a marked progress in automobile quality or style, but a very marked progress in automobile production. It was the breakthrough to the automobile using society. Even though it is easy to see today the negative consequences of automobilization--smog, noise, cancerous growth of suburbia, etc.--the automobile has doubtless also enhanced the freedom of action of the average household very effectively.

The emphasis on process innovations resulted not only in rapid economic growth; a part of the gain in productivity has been transformed into the acquisition of more free time for all members of the household: their role as income earners has been better accomplished by a shortening of their work day, and their roles as producers in the household was made much lighter by household investment goods which raise the productivity of household work; these are electric washers and irons, wash-and-wear clothing, vacuum cleaners, deep freezers, etc. As will be shown later on in more detail, it is just this gain in free time which hopefully will alter the relationship between firms and households, at least between some kinds of firms and households, from exit toward more voice.

Incentives and Capacities for the Development of Product Ideas in Households of Affluent Societies

Abundant free time in the household allows members to specialize within the household. This leads to a growing

demand for specialized consumer goods coupled with the required higher ability of the consumer to communicate with firms.

Households as Producers

Households are not only institutions offering the services of their work-force in the labour market and demanding products and services in consumer goods markets; they are also institutions in which market goods are transformed into final goods and services to satisfy its own members. Except for goods purchased for their function as status symbols, goods bought in the market serve as inputs into the household's own production process, where the other inputs are the (household work-) time and the human capital of its members. The output is the satisfaction of a variety of demands. Because there are many of those which have to be satisfied almost simultaneously, the household performs diverse production processes in parallel. The household can thus be characterized as a generalized producer. As long as the household has only enough income and free time to perform a multitude of the more urgent tasks, the members have little chance to develop expertise in one or a few of the household's production processes.

The scene alters as soon as the household has a growing income and free time, which enables individual members to specialize on one or few production processes.

The most important features of this specialization are the cumulative effects in knowledge and motivation. The specialized household has the chance to accumulate knowledge or--as economists say--human capital in a way similar to that of the specialized firm. In contrast to physical capital human capital has the nice property that use does not diminish it; on the contrary, use, above all intelligent and intensive use, gives the chance of growth of knowledge, through criticism, innovation, and adaptation. In this way members of the household may become experts in cooking, hiking, fly fishing, photography, even painting, etc.--with the result of ever-increasing standards of quality of the household production.

The cumulative effect of specialization is reinforced by a cumulative motivation effect. In humanistic psychology three kinds of needs are distinguished: existence needs (i.e. need for food, shelter, etc.), social needs (i.e. needs for communication and cooperation with other people) and growth needs (i.e. needs for personal development or ego growth). Whereas the satisfaction of existence needs intensifies social needs and the satisfaction of social needs further strengthens growth needs (c.f. Alderfer 1972). Thus it is to be expected that households with little income will concentrate mainly on existence needs--whereas those with income and free time at their disposal, will in the first place try to satisfy more growth needs. Specialization in households in the process of production intended to satisfy growth needs will therefore stabilize these processes. The cumulative process of the development of human capital or the growth of knowledge and the motivational cumulative process of growth needs, are thus mutually supporting.

Growing Demand for Expert Consumer Goods as Consequence of Growth of Human Capital in Household Production

An important consequence of the accumulation of human capital is that the productivity of the specialized production process in the household grows. Therefore the shadow prices of the input-goods for this production process will fall relative to the process of the other input-goods of the household production processes. *Ceteris paribus*, as prices fall the household will buy more of these input-goods (c.f.

Becker and Michael 1973). For instance, when a consumer becomes expert on baroque music, records with this music are relatively cheap considering the time invested to become an expert and the time necessary to hear the records repeatedly. The growth of knowledge in the realm of specialized household production reduces the significance of the market price of the relevant input-goods, though their quality will become more important; this is so because these goods have to fit the higher quality of human capital.

Higher Communication Ability and Interest as Consequence of Growth of Human Capital in Household Production

By specialization in a line of household production the consumer also will be an expert on those input-goods which he buys from firms. And because the quality of these goods is increasingly important to such a consumer, he is not only able, but also has increasingly greater interest in communicating with the supplying firm.

As distinguished from his interests in his role as generalized household producer, the preference of the specialized consumer is of voice over exit. The specialized consumer has as consequence of his expertise and of the minor importance of market prices generally input-goods of the highest attainable quality or input-goods which he finds best suited for his specialized production. Therefore exit is of little use for him: he has already left most of the market. All he can do in order to improve the input-goods is to convince the supplying firm to upgrade its product, or to improve his own specialization. Yet there is an important question here: is not the specialized consumer better off when he behaves as a free-rider, hoping that other consumers with the same specialization communicate with the firms in order to change the quality of the input-goods? Does the economic theory of collective action also apply here, because consumers specialized in certain kinds of household production form large (latent) interest groups? Obviously, many specialized household producers will choose the free-rider position. But there are also many consumers who want to communicate because this process is self-rewarding in the sense that exchange of information between experts has always the chance of promoting the knowledge of both sides.

Incentives for Firms to Adopt Product Ideas from Households

Communication requires at least two partners. It is not enough that specialized consumers are interested in it; also the supplying firm has to be ready for communication. It is therefore advisable to look for the consequences of household specialization for the firm, in order to find incentives for it to communicate.

Specialized Small Firms as Complements to Specialized Households

One result of household specialization is the mushrooming of new small markets for expert consumer goods in addition to the large markets of mass-produced consumer goods for the generalized part of household production. Small markets imply small series of production and therefore higher costs and higher prices. But as has been demonstrated, for expert consumers quality is more important than price. Therefore small enterprises producing expert consumer goods may be viable, whether as independent economic firms or as small activity units within larger corporations.

Small Production Series Favorable for Product Innovation

Just as large scale production is accompanied by a tendency to stress production innovations, so small scale

production induces entrepreneurs to use product innovation as main instrument of competition. This does not mean product innovation just for the sake of novelty. For an expert consumer the new is not always the better. More important for him are high standards of product quality and fit for the specialized household production process. The better the fit between qualities of the product and the requirements of the production process in the household, the better and the clearer will be the market position of the firm.

The hypothesis that mass production favors process innovation and small scale production favors product innovation could be challenged by the wave of revolutionary product innovations. The latest was caused by microelectronics in mass-produced consumer goods. Obviously, quartz watches, pocket computers, electronic programming of household appliances etc., are all instances of a large set of important product innovations of mass produced consumer goods. But all product innovations of the last century will do as examples. This is why the hypothesis says mass-production favors, not entirely insists on process innovations. The rapid diffusion of product innovations can only be explained by the fact that it resulted in even more revolutionary process innovations. Therefore if process innovation is understood in a generalized sense, the hypothesis can be stated strongly: Mass production insists on process innovation. Thus, about ten parts of an electronic watch, for instance, are cheaper to produce and to assemble than the about hundred parts of a mechanical one. Similarly, the contribution of microelectronics to large productivity gain can best be recognized by the change of the production location of pocket computers. Invented in the USA it has been first produced there. Lower wages in Japan and in other Asian countries attracted production to these countries. But with the invention of microelectronics, that is to say of the chips, the United States regained international competitiveness.

The importance of product innovation and quality adaptation to the demands of expert consumers are incentives for entrepreneurs to be responsive to suggestions of consumers. Whereas a complaining consumer in the case of mass products may be viewed as unrepresentative, in the case of specialized consumer goods the hints of a consumer are taken seriously, since his suggestions are from the start backed by the expertise that prompts them; and they depend to a much lesser degree on emotions than those of the mass-product consumer's complaint. This way newer and even higher standards of both production and consumption emerge.

Smaller Hierarchy Distances in Small Firms Facilitate Communication with Households

Not only the incentive to communicate with consumers is higher in small specialized enterprises than in large, mass producing firms, there are also higher chances that suggestions made by consumers will not be buried in the files of a large bureaucracy. In small enterprises it is easier to have a shortcut between gatekeepers in the organization, who communicate with the outside, and those persons, who initiate and support the processes of innovation. This is not to say that large firms necessarily are less innovative. They may decentralize and build small subsidiary enterprises which have in principle the same possibilities as independent small firms.

The New Combination of Voice and Product Innovation

I shall now present a brief historical case study to show how communication between households and firms in the process of product innovation may work and close with a word about consequences for the future. But first let me sum up.

Voice and Product Innovation as Consequences of Specialized Household Production

Household's disposable income and free time grows; money and time are no longer used for necessities of life only; an increasing part of the household's budget is spent on activities that satisfy personality growth needs; an important change occurs in household production. The generalized part of household production is complemented by a specialized one, characterizable by a cumulative process of human capital growth and of motivation for further growth. Consumers become experts interested and capable of influencing the process of product innovation in firms. In their role of specialized household producer, they prefer voice to exit. Consequent to specialized households small specialized enterprises evolve, at times viable in spite of small scale-high cost production processes, because the specialized households, for whom they cater prefer high quality to low price. Instruments of competition are therefore in the first place product innovations. This corresponds to the interest of the expert consumer to influence the quality of the goods, made by the firms. This way the firms become susceptible for new product ideas from consumers. A paradigmatic case is the following.

A Paradigmatic Historical Case Study: The Amateur Photographer Heinrich Kuehn, the Optical Engineer Franz Staebler, and their softlens "Imagon"

Heinrich Kuehn (1866-1944), who lived in Austria most of his life, was a famous photographer (c.f. Scholz 1980). He was an amateur photographer both because he was a person of independent means, and he objected to the professionalization of this new art. His artistic ideal was photography imitating painting. He wanted to be an artist who produces paintings not with brush and paint, but with camera and film. One of the techniques he invented is to superimpose images on one frame of film, one hard and one soft. This was effected by the use of the center of a lens, which gets a sharp image; the soft image was effected by the use of the periphery of a lens. The superimposition of a soft image on the sharp one had the result of a clear image with slight diffusion and sparkling light spots, so as to resemble a painting. This technique was very difficult and time-consuming and only fitted for immobile objects. Kuehn wanted to have a lens which at once produced the wanted result of the two images. He had the ingenious simple idea to put in front of the lens a diaphragm which has a perforated periphery. The normal small iris opening now yields the sharp core image and the marginal rays, which are reduced by the variable perforation of the periphery, form the the soft, but at once superimposed image.

In Franz Staebler Kuehn found an owner of a small optical firm who was ready to design and produce the lens he wanted. For years they communicated and cooperated. Staebler produced many models and Kuehn tested them till he had the model which met his demands. The lens first appeared on the market in the year 1928. Ambitious photographers still seek the "imagon", produced today nearly unchanged and never outperformed by any other softlens producer.

This case exhibits all the elements required for the household to become a source for new product ideas.

- a consumer with sufficient disposable income and free time to enable him to work in a specialized household production;
- the consumer accumulates human capital and satisfies his own growth needs;
- a small enterprise, primarily interested in product innovation, and with the entrepreneur at the same time

gatekeeper and promoter of innovations;

- a small, but not too small market with expert consumers, who appreciate the quality of a consumer good for specialized household production.

Conclusion: Consequences for the Future-the Household as Growth Sector

The problem of consumers as sources for product ideas may be of limited interest. But the analysis of this problem leads directly to a proposition which is of general interest, namely the thesis that in comparison with the household, markets and firms will lose their importance as sources of welfare. It is thus not necessary to combat the mass-market but to use it in order to supersede it.

Economic and Social Limits to Growth

Many world crises-nutrition, energy, raw materials-indicates the end of unmitigated economic growth. Perhaps technological progress will shift these limits outward. But there are also social limits to growth (c.f. Hirsch 1976). Perhaps one day every household may have its refrigerator, washer, car, etc. But it is impossible for everybody to have the proverbial lonely vacation house or to be a highly paid executive. The very success of economic growth in the past obscured for a while the social limits to growth. But when economic growth ends or slows down, the social limits will be increasingly felt. This will inevitably sharpen social conflicts. To contain these social conflicts, Hirsch proposes more equal income distribution. This may be one possible response to this challenge. But there are others.

No Limits to Growth in Specialized Household Production

Production in the general sector of the household seems to be one of the economic sectors with the largest growth of welfare today (c.f. Burns 1975 and Gershuny 1978). This growth depends largely on material investment goods, bought in the market, and therefore indirectly on economic growth. But growth in the specialized sector or household production depends more on the availability of free time which the consumer may devote to these activities, and on accumulated human capital; material goods bought in the market are of decreasing importance. Therefore the growth of welfare in this sector is not subject to the same material limits of growth. And growth in the specialized sector of the household is also independent of social limits to growth. For, to be a photographer or a fisherman who amply satisfies his growth needs, one need not have an expensive camera outfit or a motor yacht. Expensive material investment goods only compensate for lack of time and for lack of human capital. This therefore may satisfy the social needs of a busy executive-it may gain him social status by conspicuous consumption-but it will not satisfy his growth needs (in the sense of ego growth or personality growth). It may be one of the most important tasks of socio-economic theory and of welfare policy to explore and to strengthen the unlimited growth possibilities of specialized household production that go with the growth of human capital.

References

- Alderfer, Clayton P. (1972), Existence, Relatedness, and Growth, New York and London.
- Becker, Gary S., and Michael, R.T. (1973), "On the New Theory of Consumer Behavior", in Swedish Journal of Economics, 75, 378 ff.

Burns, Scott (1975), Home, Inc. The Hidden Wealth and Power of the American Household, Garden City, New York: Doubleday & Co.

Gershuny, Janathan (1978), After Industrial Society-The Emerging Self-Service Economy, London and Basingstoke: The MacMillan Press Ltd.

Hirsch, Fred (1976), Social Limits to Growth, Cambridge, Mass.: Harvard University Press.

Hirschman, Albert O. (1970), Exit, Voice and Loyalty, Cambridge, Mass.: Harvard University Press.

Olson, Mancur, Jr. (1965), The Logic of Collective Action Public Goods and the Theory of Groups, Cambridge, Mass.: Harvard University Press.

Scholz, Alfons (1980), Lichtbilder mit dem Imagon, Herrsching/Ammersee: vwi Verlag Gerhard Knuelle.

PERCEIVED RISK AS A HINT FOR BETTER CONSUMER INFORMATION AND BETTER PRODUCTS:
SOME NEW APPLICATIONS OF AN OLD CONCEPT

Konrad Dedler, University of Hohenheim
Ingrid Gottschalk, University of Hohenheim
Klaus G. Grunert, University of Hohenheim

Abstract

This paper addresses several aspects of deriving a standard for a workable information system. Basic elements of the approach presented are the analysis of consumer information demand derived from perceived risks, as well as the examination of the information supply by business and neutral sources by content analysis. Some illustrative empirical results from a study on consumer information strategies in the German automobile market are presented.

Introduction

Basically, there is no controversy about the general importance of consumer information in achieving an efficient competitive market system. All kinds of public consumer information programs, whether in Europe or in the United States, start right out from this assumption (e.g. FTC 1979). However, not all measures of consumer information policy have proved to be effective. Consumer researchers have investigated such inefficiencies mainly on the grounds that information not tailored to the information processing capabilities of consumers might not be very helpful for them after all. But information does not only have individual effects to be investigated at the level of the individual consumer, but a host of more indirect effects which result from the interaction of consumers and suppliers in the market. The nature of these effects resulting from a raise in the level of information in a market has not received much attention in the literature.

In this paper we want to present an integrated framework which makes it possible to trace both kinds of effects. From this we shall conclude that the complex interrelationship of positive and negative effects of information policy calls for the concept of a "workable information system", and for means of measuring this workability. We then develop such a measurement concept, using perceived risk as a central unit of analysis. Problems in applying this concept are discussed and illustrative results from a study of the German automobile market are presented.

The Effects of Information Policy and the Need for a "Workable" Information System

Information Policy and Welfare: A Consumer-Oriented Approach

Most people would probably agree that the basic aim of consumer information policy is to enhance consumers' welfare. It seems therefore appropriate to evaluate effects of information policy from a consumer-oriented viewpoint, which conceptualizes welfare as the aggregated utility consumers derive from the consumption of goods and services. We can broadly distinguish two ways in which information policy can enhance welfare in this sense:

- on a first level welfare is enhanced if the new information enables consumers to improve their purchase decisions, i.e. helps more consumers to choose among the available alternatives the one that suits best their individual preferences;
- on a second level more information can raise welfare by inducing changes in the supply of goods and services so that these conform better to the preferences of the consumers.

Later we will have to say more about how these levels interact and how changes at the second level are induced. Here it is important to note that the kinds of effects in question involve not only consumers but also suppliers: changes in consumer demand at the first level will most certainly lead to reactions on the supply side, and changes at the second level presuppose effects on the supplier side. Hence, although we take a consumer-oriented stand in analysing the effects of information policy, such effects always involve the supplier side as well, and all effects of information policy are therefore also relevant from the viewpoint of competition policy. Thus, the interaction between consumers and suppliers is crucial for the analysis.

How informationally induced changes in the relations between consumers and suppliers can affect welfare can be demonstrated by means of a simple model. Using a multi-attribute framework, we assume that consumer i considers some attributes as important in evaluating some particular product. Then his informedness concerning this particular market can be characterized by

$$IC_i = \sum_{k=1}^{p_i} z_{ik} \quad (1)$$

where: IC_i = Informedness of consumer i

z_{ik} = Number of brands where consumer i knows attribute k

p_i = Number of attributes that consumer i considers as important.

If there is a total of m brands in this market, we have

$$0 \leq IC_i \leq m \cdot p_i \quad (2)$$

Usually, the consumer will not attain maximal information due to information cost (C) and his own propensity to acquire information (AC_i). Thus, we have

$$IC_i = f(C, AC_i, m, p_i) \quad (3)$$

The informedness of suppliers can be characterized in a similar way. We assume that consumers can be divided according to their preferences into some segments, and that the total number of attributes used by consumers in evaluating the particular product in question is q . Then the informedness of a supplier j can be written as

$$IS_j = \sum_{k=1}^q y_{jk} \quad (4)$$

where: IS_j = Informedness of supplier j

y_{jk} = Number of consumer segments where supplier j knows the value of attribute k

q = Number of attributes used by all consumers.

If the total number of consumer segments in this particular market is n , then

$$0 \leq IS_j \leq n \cdot q \quad (5)$$

It seems that the informedness of suppliers is partly dependent on the informedness of consumers. If consumers know a lot about the brands in the market and how these relate to their needs, it is easier for them to voice complaints, it is easier for suppliers to do market research, and it is easier for institutions of consumer policy to channel information about unsatisfactory products. Hence we can view the

aggregated consumer informedness (\tilde{IC}) and the supplier's individual propensity to acquire information (AS_j) as the basic determinants of his informedness:

$$IS_j = f(\tilde{IC}, AS_j, n, q) \quad (6)$$

Whenever some consumers do not know how the m brands in the market score on the p_i attributes important to them, or whenever some suppliers do not know which combinations of attribute values are preferred by the n consumer segments, some consumers will experience a welfare loss. This is because they do not realize utility U^* that would be associated with a brand that conformed perfectly to their individual preferences, but some lower utility U . The welfare loss $\Delta U = U^* - U$ can be split up into two parts:

$$\Delta U_i = \delta_i + \nu_i \quad (7)$$

where: δ_i = welfare loss resulting from the fact that a consumer i chooses among the available brands not the one coming closest to his subjective ideal brand
 ν_i = welfare loss resulting from the fact that none of the brands available is completely like the subjective ideal brand.

It is plausible to assume that the first type of welfare loss depends on the consumer's informedness,

$$\delta_i = f(IC_i) \quad (8)$$

and the second on the suppliers' aggregate informedness (\tilde{IS}) and the number of brands in the market (m):

$$\nu_i = f(\tilde{IS}, m) \quad (9)$$

This simple framework can now be used to trace some possible effects of lowering C , the cost of information. We define "cost" in a very broad sense here, so that most measures of information policy, like providing additional information, improving the quality of information, increasing the availability of existing information etc., can be viewed as special cases of "lowering cost".

Direct Information Effects and Market Effects

We distinguish two types of effects: those resulting directly from changes in the information flow (one-party effects), and those resulting from reactions of the market to such changes. For each possible effect we will list its prerequisites, possible obstacles, and ways these can be handled by public policy.

We have two direct information effects. One stems from changes in consumer behavior, the other from changes in supplier behavior. Both can increase consumers' welfare.

Lowering C may induce some consumers to acquire more information, thus raising IC_i for these consumers. This may enable them to make a better choice among the brands available, thus lowering δ_i . This effect has two prerequisites:

- a lower C must induce at least some consumers to increase IC_i . This is obviously not always the case. We know many examples where newly presented information was not used by consumers (e.g. Carman 1972; Monroe & LaPlaca 1972; Lenahan, Thomas, Taylor & Call 1973). Usually, this can be explained by deficiencies either in information content, which does not conform to the consumer's information needs (Grunert & Saile 1978), or in format, which does not conform to the consumer's information processing capabilities (Bettman & Kakkar 1977). Careful design of measures of information policy can circumvent these obstacles;
- a higher IC_i must actually result in a lower δ_i , i.e. the consumer must be able to use the additional information in his decision in an efficient way. This, too, is

not always the case. We have some experimental evidence that with increasing amounts of information consumers resort to simpler decision rules (Payne 1976; Park 1978), thus giving away possible welfare gains from a better choice. Again, careful design of the information, which takes into account how people integrate information and how this is influenced by the information format, may circumvent this obstacle.

Lowering C may also have direct informational effects on suppliers. This occurs if C raises \tilde{IC} , which in turn raises IS_j for some suppliers. That is, due to the increased informedness of consumers some suppliers know more about these consumers' preferences, for the reasons indicated above, causing them to change their products. It is also possible that the mere existence of neutral consumer information detrimental to a supplier's product may cause him to change the product. These changes in the brands on the market may cause a lower ν_i for some consumers and hence increase welfare.

Apart from the constraint that a lowering of C must not necessarily result in a rise of \tilde{IC} , as stated above, we have two more prerequisites for the supplier effect to occur and increase welfare:

- a rise in aggregate consumer informedness must actually result in better informedness of the suppliers about consumers' preferences. Poor market research on part of the supplier may inhibit this. Information policy could help greatly here by institutionalizing communication channels between consumers and suppliers, so that the latter can take advantage of the expertise of the former;
- suppliers must be motivated to use an increase in IS_j to design products which are in better accordance with the preferences of consumers. This is not necessarily so. If a supplier has acquired monopoly power to an extent that disregarding the wishes of consumers is no longer punished by the market, the welfare effect will not occur. This, however, is a problem for competition policy, which can try to enhance competitive pressure in the market and/or facilitate market entry so that unwarranted monopoly power cannot endure.

If these direct informational effects occur, the other members of the market will probably react to the new developments. This is the cause for market effects. To discuss possible market effects, it is useful to start with a general process model of the functioning of a competitive market, which was developed to explain the dynamics of competition resulting for example from a pioneering supplier taking a lead in offering a new or improved product (Herdzina 1973):

- in the action phase some supplier offers a new, improved product;
- in the substitution phase, consumers react by substituting the new product for others, thus creating increased demand and some degree of monopoly power for this supplier;
- in the reaction phase, competing suppliers react by offering improved products as well;
- in the resubstitution phase some consumers go back to their old suppliers now also offering improved products.

As a result, the monopoly power of the innovator has been eroded, but the supply has improved. Applying this general scheme to information policy, we find that both direct informational effects mentioned above trigger this whole process and hence cause market effects: if consumers change their demand due to an increased IC_i this constitutes a substitution phase with subsequent reactions, and if suppliers change their supply due to an increased IS_j this constitutes an action phase. In the course of the process additional welfare effects will result:

- consumers who, in spite of the decrease in C , did not acquire more information and continued to buy the same brand, will find that this brand is no longer available and buy a substitute. Depending on the kind of information deficit of the consumer, his preferences, and the

way products are improved, this will result in decreases in S_i and/or μ_i for these consumers;

- suppliers who did not change their product in response to the informational measure may be forced to do so because of demand reactions of the consumers. This again may lead to a decrease in μ_i for some consumers.

The realization of these effects presupposes, of course, a proper functioning of the competitive process. We tried to outline how this process can be stimulated by information policy measures. However, it is important to consider that the described effects depend on the kind of information strategies involved. Inadequate provision of information could affect the functioning of the competitive process even detrimentally.

Pursuing an "inadequate" information policy could for instance mean either to single out special product attributes, like prices, or to inform on selected groups of suppliers exclusively, as is often done in comparative testing magazines.

The possible negative effects on competition of stressing only a special attribute of a product have been demonstrated on the market for used cars, referred to as a market for "lemons". In the absence of information on further product attributes, as for example low defect rates, price becomes the exclusive indicator for the quality of the product. Then, it is a reasonable choice for sellers to refrain from costly, price-raising pre-purchase repairs and sell lower-priced "lemons" instead (Akerlof 1970, FTC 1979). Corresponding to this result, information policy which even enlarges the consumers' incomplete evaluation of products could mean to contribute to a deteriorating quality of supply and to a paralyzing of sellers' competition on quality.

Comparative testing of products can be undertaken by "neutral" institutions, i.e. by organizations, which are neither organizationally nor financially dependent on suppliers, and the results thus derived can be used for information policy strategies. If however comparative testing selects large suppliers and national brands to the detriment of small and local suppliers, consumers would be informed on parts of the existing supply only. This could induce monopolizing tendencies and bring about results which are similar to those discussed in the context of advertising effects. There, it is stressed that information by advertising can increase either the heterogeneity of supply or the concentration towards large suppliers, both effects entailing possible welfare losses by an increase in monopoly power and/or a decrease of brands supplied (Grunert & Stupening, in press).

"Adequate" consumer information policy which aims at improving the competitive market system has to avoid partial or one-sided information measures like the ones mentioned. It rather should try to alleviate consumers' choice among different products by comparable and understandable messages. Consumer information policy should try to choose an information content and format that:

- facilitates comparison on all attributes of competing products which are deemed relevant by individuals,
- analyzes all brands available, instead of particular market segments,
- adapts the information to the information processing capabilities of consumers,
- enables the consumers to react more rapidly to changes in product prices and qualities, thus spurring efficient competition.

The "Workable" Information System

If the host of effects, constraints, and policy variables described in the preceding section allow any strong conclusion, it is the one that maximal information is not an aim to be pursued by information policy, since the welfare effects are so complex that a simple maximization postulate is inappropriate. Information policy is therefore undergoing

a development similar to the one competition policy underwent a few decades ago, when it abandoned the concept of perfect competition as a policy goal in favor of a "workable" competition (Clark 1940). It seems that the concept of "workability" can be used for information policy as well: not a system of perfect information of consumers and suppliers is sought for, but a "workable information system" that tries to take advantage of the possible benefits of informed consumers and suppliers without neglecting welfare detriments resulting from inefficient or exaggerated use of information measures.

This means that information policy, just as competition policy, has to find ways in which the "workability" of an information system can be measured and evaluated. We want to propose that the workability of an information system can be assessed in three steps:

- by analysing the motivational determinants of information supply and demand. This can be termed information demand analysis;
- by analysing the information presently available in a market. This can be termed information supply analysis;
- by analysing whether consumers and suppliers are actually able to obtain in the market the information they need to make informed choices. This can be called information performance analysis.

These concepts will be elaborated in the remainder of the paper. We will concentrate on information for consumers and only hint at information for suppliers from time to time.

Assessing the Workability on the Basis of Perceived Risk: Concept and Applications

"Perceived Risk" as a Unit of Analysis

As we tried to show in the preceding section, information is not a homogeneous good. Thus, in trying to assess the workability of an information system, the question is not how much information this system does process but rather what this information is about. In order to make this a meaningful question, we must develop a unit of analysis with which to measure information content.

The basic purpose of consumer information is to enable the consumer to compare products, to find out which brand is best for him. Since usually a product has several characteristics that give rise to utility, the information must be about several product attributes. Indeed, this is what most consumer information is about. However, in order to be "information" to a consumer, two additional prerequisites must be fulfilled: the consumer must consider the attribute to be important, and he must be ignorant about the values the various brands have on this attribute. The well known concept of perceived risk comes in handy here: attribute information is helpful for the consumer in making his product choice only if this attribute is a perceived risk for him, i.e. he does not know the value some brands have on this particular attribute, but he knows that he would dislike some of the possible values the attribute can take. We thus use the term perceived risk in a multi-attribute fashion here. There can be many perceived risks associated with a particular product (e.g. Humphreys & Ingene 1980).

Information Demand: Consumer's Motivation

If a consumer perceives risks he is motivated to acquire information. Whether he will actually do so depends on a host of factors, the most prominent one being probably the expected availability of the information and the cost of obtaining it. But while a consumer motivated to acquire some information might actually do so or not depending on situational factors, an information supplied for which the consumer is not motivated will never be acquired and hence is useless and a waste of resources. Therefore, it makes sense to start with the motivational side, and use it later as a criterion in evaluating the supply side.

We propose to assess the motivational side by means of a risk analysis. This includes the following steps:

- a first consumer survey to obtain a list of possible risks consumers associate with a particular product,
- editing this list with the help of consumer experts,
- a second consumer survey to establish the relative importance of the risks for various consumer segments.

In our study of the German automobile market, the first consumer survey used an open-ended questionnaire, including questions about experiences with cars, reasons for buying a car, and main uses. The purpose was to find as many attributes as possible which at least some consumers perceive as "risks" in the sense described above. For this reason, the number of respondents surveyed was not fixed in advance. Instead, the survey was stopped when it was found that additional respondents did not contribute any new "risks" to the list. This was achieved after interviewing 60 respondents, approximately equal halves of which came from urban and rural areas. Since the aim at this stage is only to obtain a list of risks as comprehensive as possible, the sample does not have to be representative.

Experts are needed because automobiles are complex products and the answers of consumers cannot always be taken at face value. Risks perceived by consumers may no longer exist for the brands now on the market, or two risks worded differently by two respondents may actually refer to the same thing. For this reason, the advice of experts was sought and the list of risks edited accordingly. The resulting list contained 84 items.

Another important contribution of the experts results from the fact that for many risks there is no information directly available to reduce them, but rather the consumer has to use certain indicators. For example, there is no way to measure directly whether a car will rust. But if certain parts of the car body are galvanized, we can take this as an indicator that rust will be less probable. Relating indicators to risks and separating valid from invalid indicators requires considerable technical expertise and can be done only with the help of product experts.

The second consumer survey is necessary because the list of risks is a conglomerate of answers from respondents. Probably no single consumer will consider all risks of the list as important. His motivation to acquire information will probably only extend over a few risks. Therefore it is important to know the relative importance of the risks, and to identify segments of consumers according to the risks they stress. The second consumer survey tries to achieve just this. A representative sample of 500 drivers is asked to rate the relative importance of the risks. To obtain maximally differentiated answers, the method of magnitude scaling is used (Wegener 1980). Table 1 shows the results of a pretest, where 19 respondents rated a subset of 40 risks.

The results from such a risk analysis are important in assessing the workability of an information system, because they serve to approximate the total set of information consumers may be motivated to acquire. Of course, an individual consumer will be motivated to acquire only a small subset of this. But this subset is bound to vary as the consumer learns about other risks or puts the product to new uses. Therefore, the totality of possibly experienced risks serves as a useful yardstick. Especially, it makes possible to single out the information that is less likely to be helpful for consumers.

TABLE 1
Average Rating of 40 Selected Risks

RISKS	AVERAGE RATING SCORES
1. BAD TRACK-HOLDING	116.086
2. HIGH FUEL CONSUMPTION	106.221
3. LEAKING BODY	90.805
4. EXPENSIVE PARTS	90.562
5. TRANSMISSION FAILURES	88.963
6. MISPLACED HANDLING INSTRUMENTS	86.152
7. SHIVERING STEERING	83.319
8. FREQUENT ENGINE TROUBLES	79.101
9. BAD ACCELERATION	75.500
10. BAD OVERALL SIGHT	75.095
11. BAD ROAD GRIP IN WINTER	65.948
12. EASY DENTING OF BODY-SIDES	64.833
13. EASY BREAKING OF V-BELT	60.910
14. CROSSWIND-SENSITIVITY	57.634
15. INSUFFICIENT NUMBER PLATE ILLUMINATION	57.023
16. EASY CORRODING	54.881
17. CONDENSING CAR-WINDOWS	54.549
18. SMALL TANK	53.596
19. IRREGULAR WEAR OUT OF TYRES	52.901
20. UNSMOOTH GEAR	51.987
21. UNSMOOTH STEERING	50.288
22. UNSOLID CAR-ROOF	48.650
23. UNCOMFORTABLE SEATS	48.413
24. BAD HILL-CLIMB QUALITIES	44.762
25. BAD LOADING FACILITIES	42.562
26. BAD PAINTING	40.694
27. MISPLACED SPARK PLUGS	40.135
28. SQUEAKING BREAKS	36.948
29. BAD MANOEUVRABILITY	36.395
30. BAD WINDING-UP OF SAFETY BELTS	36.054
31. EASY FREEZING OF DOOR-LOCK	35.656
32. SMALL BAGGAGE BOOT	35.237
33. MISPLACED SPARE TYRE	32.640
34. UNSOLID BUMPER	32.567
35. NOISY ENGINE	31.057
36. BAD SPRING-SUSPENSION	22.858
37. SMALL GLOVE COMPARTMENT	15.668
38. PAINTING DIFFERS FROM SAMPLE	11.943
39. MISPLACED ASH-TRAYS	10.658
40. SLOW REVERSE GEAR	10.000

Information Supply: Business and Neutral Sources

The results of the risk analysis can now be used to analyse the information supply. Inasmuch as it is suitable to reduce the risks experienced by consumers?

Neglecting personal communication, product information is supplied by business and neutral sources, the former being far more prominent. The main business sources are sales brochures, sales talks, and advertisements. While the evaluation of sales talks poses a number of methodological problems (which are currently being investigated), sales brochures and advertisements can be analysed by quantitative content analysis.

In our automobile market study a sample of sales brochures and advertisements was obtained according to the following criteria:

- the analysis referred to the information supply in 1978;
- only information from manufacturers having a market share of more than 0.9 % was included, and only about models not exceeding or falling short of certain thresholds for price and cylinder capacity. This served to delimit the market;
- within these limits, all sales brochures available from

manufacturers (totalling 39 documents) were included in the analysis. However, 5 documents were dropped later for technical reasons;

- as for advertisements, 5 magazines and 4 newspapers were selected by systematic criteria. From these, issues were selected randomly, resulting in a total of 395 advertisements. Discounting duplicates, a sample size of 261 was obtained.

A content analysis involves splitting the documents into units of analysis and sorting these into categories. In this case, the sentence was used as the unit of analysis. The category system reflects two dimensions:

- which risk, if any, is alluded to in the sentence
- how this risk is dealt with.

We distinguish four ways in which a risk can be dealt with:

- the risk itself is referred to in an objectively verifiable way. Example: "this car makes 35 miles per gallon";
- the risk itself is referred to, but in a way not objectively verifiable. Example: "this car has a very good mileage";
- an indicator for the risk in question is referred to in an objectively verifiable way. Example: "the aerodynamics of this car are described by an air resistance value of 0.4";
- an indicator for the risk in question is referred to in a way not objectively verifiable. Example: "the car has a very aerodynamic body".

For 84 risks and 4 ways of dealing with a risk, this results in a total of 336 categories.

A pretest with manual coding for some selected ads and risks showed severe reliability problems, so that we decided to attempt a computer-assisted content analysis using the program TEXTPACK. Computer-assisted content analysis requires that

- all documents be transcribed to machine-readable format;
- all categories are defined in an extensive way, i.e. by a comprehensive list of words or word-combinations which, when they appear in a sentence, constitute evidence that this sentence belongs to a certain category.

Construction of the category system thus becomes a complex interactive process in which word functions are defined, tested with the data, and redefined until the validity of the category system seems satisfactory. This possibility of recoding the data as often as seems necessary to construct good categories is a main advantage of computer-assisted content analysis, along with the fact that problems of reliability are completely eliminated.

Some sample results are shown in tables 2 and 3. They show the average numbers of sentences in sales brochures and advertisements dealing with four of the more important risks in the four different ways described above.

The results show that generally the information situation in sales brochures concerning these four risks is quite satisfactory. Most sales brochures carry information about all four risks. There are, however, differences in how this is done. Objectively verifiable information is readily available for the risks concerning fuel consumption and baggage boot. This is not the case for the risks concerning the crushable zone and corrosion, where objective measurement procedures are not as popular as for the other two risks. As for corrosion, there is at least considerable information about indicators. Concerning the crushable zone, however, quite a number of sales brochures does not carry any objectively verifiable information.

Quite obviously, an advertisement can carry much less information than a sales brochure due to space limitations. On the other hand, this space limitation could serve as an incentive to present in a concise and condensed way some of the information consumers want most urgently. The data in table 3 do not bear this out. In spite of the obvious importance of these four risks, a majority of the ads in the sample does not carry any information about them. The relative coverage of the four risks, however, appears to be similar to the sales brochures.

TABLE 2
Coverage of Selected Risks in a Sample of 34 Sales Brochures

SELECTED RISK	AVERAGE NUMBER OF SENTENCES PER BROCHURE REFERRING TO:					
	RISK			INDICATOR		
	VERIFIABLE			VERIFIABLE		
	NO	YES	BOTH	NO	YES	BOTH
HIGH FUEL CONSUMPTION	3.0	10.6	1.3	1.2	5.2	-
SMALL CRUSHABLE ZONE	1.6	0.2	0.3	2.4	0.3	0.3
SMALL BAGGAGE BOOT	5.0	4.9	0.2	0.1	1.5	-
EASY CORRODING	2.9	-	-	1.2	4.6	0.7

ONE SALES BROCHURE COVERS 509.7 SENTENCES ON AVERAGE

TABLE 3
Coverage of Selected Risks in a Sample of 261 Advertisements

SELECTED RISK	AVERAGE NUMBER OF SENTENCES PER ADVERTISEMENT REFERRING TO:					
	RISK			INDICATOR		
	VERIFIABLE			VERIFIABLE		
	NO	YES	BOTH	NO	YES	BOTH
HIGH FUEL CONSUMPTION	0.14	0.22	0.09	0.07	0.15	-
SMALL CRUSHABLE ZONE	0.01	-	-	0.02	0.01	-
SMALL BAGGAGE BOOT	0.12	0.16	-	-	0.01	-
EASY CORRODING	0.03	-	-	-	0.07	-

ONE ADVERTISEMENT COVERS 13.2 SENTENCES ON AVERAGE

- Analysing the information supply like this on the basis of perceived risks can have a number of applications:
- it can be used to compare the information supply in comparable markets, to see how different information supplies affect the workability of the market;
 - it can be used to obtain hints where the development of new, standardized measurement procedures would be especially helpful for consumers, like a procedure to measure a car's susceptibility to rust;
 - it can be used to guide the development of neutral consumer information, showing where this information is most badly needed.

Of course, existing outlets of neutral consumer information

are also part of the information supply and can be analysed and criticized by the same means. The most important form of neutral consumer information is probably the product test. The results of the risk analysis can be used to assess whether the criteria used in such tests are in accordance with what consumers would like to hear. Even a quick glance at product tests shows that this is not unequivocally the case. Many times, the test engineers seem to measure the criteria most conveniently to measure, which are not necessarily the same that consumers are interested in.

Drawing on the results from a review of automobile tests in the two major German automobile magazines, these tests seem to be wanting in the following ways:

- some of the published "tests" do not even seem to merit this name; they consist merely of some casual observations made while driving the car;
- some results presented are simply copied from the manufacturers' sales brochures;
- some results could more conveniently be obtained by the consumer himself by inspecting the product;
- many risks are either not dealt with at all or are dealt with indirectly by measuring indicators whose relation to the risks is not clear.

Thus, the limited resources available for tests seem to be partly wasted to present results which can be obtained elsewhere, at the cost of information that consumers could obtain only in such tests. Such information concerns mainly:

- long time usage effects, like susceptibility to corrosion;
- technical properties relevant to risks which cannot be assessed by the consumer himself;
- technical properties which could basically be noted by the consumer, but only in special situations (like handling capabilities in winter weather).

Especially for the long time effects, adequate measurement procedures may not even be available at present. This would be an important task for basic research into comparative testing.

Table 4 shows some data concerning the coverage of some selected risks - two for each of the three categories just named - in all "tests" published in 1979 in two major German automobile magazines. The results show that many risks are covered only rarely or not at all.

TABLE 4
Coverage of Selected Risks in a Choice of 133 Automobile Tests

SELECTED TEST CATEGORIES	SELECTED RISKS	NUMBER OF AUTOMOBILE TESTS REFERRING TO SELECTED RISKS IN TWO GERMAN AUTOMOBILE MAGAZINES			
		ONE-PRODUCT TEST (144 IN TOTAL)	COMPARATIVE TEST (7 IN TOTAL)	ENDURANCE TEST (12 IN TOTAL)	Σ (133 IN TOTAL)
LONG TIME USAGE	DIFFICULT MAINTENANCE	2	0	0	2
	HIGH COST OF REPAIR	0	2	12	14
EXPERT USAGE	SPECIAL TOOLS FOR DO-IT-YOURSELF REPAIRS	0	0	0	0
	HIGH EMISSION	6	0	0	6
CASUAL USAGE	CROSSWIND-SENSITIVITY	5	3	1	9
	INSUFFICIENT WIPER-ZONE	0	0	0	0

Information Performance: The Gap between Supply and Demand

Information demand and supply together determine inasmuch the information flow in the market is suitable to allow the actors in the market to make informed decisions. If we compare the totality of risks experienced by consumers with the information supplied by suppliers and neutral sources, the resulting gap can be said to be an indicator of the information performance in this market.

This, however, is only one possible indicator for information performance. Other possible indicators would be information usage or decision quality, which also reflect the gap between supply and demand. The direct measurement of the gap by comparing information supply and information demand is quite cumbersome and can be recommended to obtain a basis assessment of the information situation in a market. If the effect of remedial measures of information policy on information performance is to be measured, indicators like information usage or decision quality, which are less costly to obtain, are more adequate. This, too, can be done on the basis of a risk analysis, as has been shown elsewhere (Grunert 1980).

Towards an Improved Information System

An improved information system, in our view, is one where consumers get more information suitable to reduce risks perceived by them, and where suppliers get more information about risks perceived by consumers, resulting in a long-run net welfare increase for consumers. In the preceding section we have tried to show how the workability of an information system can be measured and how the results obtained can be used to hint at ways for improvement.

The examples we gave all referred to consumer information and ways to improve it. But obviously information for the suppliers is as important. If suppliers do not have adequate knowledge about the wants of consumers, no consumer information program can bring about an efficient allocation of resources. Of course, suppliers try to obtain this kind of information by market research. But market research differs widely in quality, favors the big firms, and deals with questions formulated by suppliers, which might bypass certain wants of consumers. We think that institutions of consumer policy have a vast field for action here, obtaining and canalizing information from consumers, no matter whether complaints about current products or wants for future products, and offering it to all suppliers in the pertinent market. A workable information system involves functional information flows in both directions - a fact that consumer policy has not always realized.

REFERENCES

Akerlof, George A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", Quarterly Journal of Economics, 84, 488-500.

Bettman, James R. and Kakkar, Pradeep (1977), "Effects of Information Presentation Format on Consumer Information Acquisition Strategies", Journal of Consumer Research, 3, 233-240.

Carman, Jones M. (1972), "A Summary of Empirical Research on Unit-Pricing in Supermarkets", Journal of Retailing, 48, 63-71.

Clark, John M. (1940), "Toward a Concept of Workable Competition", American Economic Review, 30, 241-256.

FTC (1979), Consumer Information Remedies, Washington, D.C.: U.S. Government Printing Office.

Grunert, Klaus G. (1980), "Consumer Information by Objectives: An Approach for Defining Goals and Measuring Results", Advances in Consumer Research, Vol. 7, ed. Jerry C. Olson, Chicago: Association for Consumer Research, pp.664-668.

Grunert, Klaus G. and Saile, Heinz D. (1978), "Information Needs, Supply and Demand as a Basis for the Development of Consumer Information Material: The INVORMAT Method", Journal of Consumer Policy, 2, 338-348.

Grunert, Klaus G. and Stupening, Eduard (in press). Die gesellschaftliche Bewertung der Werbung: Argumente und Befunde, Frankfurt: Campus.

Herdzina, Klaus (1973), "Oligopolistische Interdependenz, funktionsfähiger Wettbewerb und Wettbewerbsvoraussetzungen", Jahrbuch für Sozialwissenschaft, 24, 55-84.

Humphreys, Marie Adele and Ingene, Charles A. (1980), Perceptions of Risk: A Theory of Consumer and Producer Behavior, paper presented at AMA Special Educators Conference, February 10-13, 1980.

NEGATIVE INFORMATION: PERSPECTIVES AND RESEARCH DIRECTIONS

Marc G. Weinberger, University of Massachusetts/Amherst
Chris T. Allen, University of Massachusetts/Amherst
William R. Dillon, University of Massachusetts/Amherst

Abstract

The topic of negative information encompasses a variety of research areas which heretofore have been considered as distinct and separate from one another. The literature from each of these unique research perspectives is reviewed here and then related to the broader scope of negative information effects. Finally, possible directions for negative information research are suggested.

Introduction

Though consumer researchers have had a long tradition of interest in the impact of information on consumer decision-making, relatively little effort has been devoted to examining the effect of negative consumer information. This absence of concentrated study prevails despite dramatic growth in the amount of negative information in the marketplace. Today's manager is confronted with a hostile and often uncontrollable external information environment. Negative information offered for public consumption can be traced to consumer and environmental groups, to regulatory agencies, and even to competitors. In other instances the unfavorable information takes the form of untraceable rumor. The strength and longevity of effects of these unfavorable cues on an individual's perceptions, attitudes, and behavior are of obvious potential interest and importance to both the marketing academic and practitioner. Moreover, the impact of negative information in the marketplace would seem salient to those who study the relationships among major societal institutions because there is potentially a cumulative effect of a profusion of negative information cues on consumers' general attitudes towards business practices and institutions.

Definitionally negative information about people, products, issues or companies represents the presentation of input which somehow denigrates the object of the message. Issues such as person perception, personnel evaluation, fear appeals, rumor, product recall, counter advertising, comparison advertising, product recall reports, corrective advertising, affirmative disclosure, and self-denigrating product information all involve some aspect of negative information that may be confronted, subsequently placed in short or long-term memory, and potentially withdrawn to be used in a current or future decision. In all instances an attempt is made to either discourage some behavior or attitude and focus on establishing a new attitude or behavior.

To date research about negative information has been scattered under diverse areas in psychology, sociology and marketing. As a result micro research traditions related to negative information issues have developed within each discipline with little recognition that they relate to a common theme. The major intent of this paper is to first review some of the major paradigms and conceptual foundations which have been applied to the study of negative information. Secondly, this paper will examine empirical findings from a variety of disciplines which relate to negative information effects. Finally, some suggested directions for future research in this area will be outlined. In general, what is labeled here as negative information is not novel but the vision of the related component issues under a common rubric is new.

Explaining the Effect of Negative Information

Different experimental paradigms and theoretical constructs have emerged in generating and testing specific hypotheses concerning negative information. The simplest and most widely used research paradigm has been impression formation. In its most primitive cause and effect form a stimulus triggers some belief, attitude or action in a subject population. The impact of mediational factors such as anxiety, susceptibility to social influence, receptivity and so on are not of primary concern. In most impression formation studies unfavorable adjective descriptions or scenarios have been the stimuli employed with the target object being a known and/or unknown individual or group. Variations in the amount and intensity of negative and positive stimuli have been gauged using simple paper and pencil measures of liking, behavioral intention and so on. The advantage of this approach to studying negative information has been the ability to retain very tight internal control over the experimental setting.

A more complex extension of impression formation is attribution theory where the individual is viewed as a being who perceives a situation, examines it for key elements of information, and subsequently makes an inference about a stimulus object (Kelley 1973). To date attribution theory has served as an explanation for negative informational effects rather than as an empirical paradigm. Using an attributional framework Kanouse and Hanson (1971) suggest that negative information has the strong impact that it does because it stands out more than positive information and in Kelley's terminology would therefore have more distinctiveness. This apparently results from the fact that there are more positive cues in the individual's social environment. As a result, negative cues attract more attention and are therefore more heavily attributable to the stimulus object. This underestimation of important contextual factors is an attributional phenomenon observed in both a laboratory setting (Jones and Nisbett 1971) and at the societal level (Lazer 1980).

A more mediational view of negative information effects might be taken with both assimilation-contrast theory or threat-compliance models. Both are conceptual approaches that help explain the situational effects of an incoming stimuli in relationship to some set of individual predispositions. Assimilation - contrast theory (Sherif, Sherif and Nebergall 1965) posits that prior experience or attitude of the receiver and its relative position to that advocated in the message will determine whether it is drawn closer into the receiver's sphere (assimilated) or expelled further outside the receiver's position (contrasted). Thus, negative information consistent with the receiver's beliefs would be assimilated and that which is incompatible would be contrasted and rejected with subsequent low compliance. Since individuals bring to a setting some set of predispositions which affect their perceptions, the ability to assess these relevant predispositions might help predict the impact of negative information.

In a similar mediational vein McGuire (1968) developed a threat - compliance model where the impact of a message on opinion change is a function of one's predisposition to receive a message (receptivity) and yielding. Without considering receptivity one might expect that greater anxiety arousal would lead to greater yielding to an

influence attempt. When receptivity is factored in as a potential mediator, a negative relationship between anxiety and persuasion is expected. Receptivity adds a non-monotonic dimension to McGuire's threat - compliance analysis and emphasizes the need to consider predispositional factors like receptivity. Since much of the negative information in the consumer context represents a threat of varying magnitude, this conceptual approach might be attractive in dealing with the situational complexities of many negative information circumstances.

Empirical Research

Negative information has been investigated most often using impression formation as its basic paradigm focusing on the subareas of person perception and personnel decisions.

Person Perception

Psychologists have examined the impact of negative and positive information in the context of forming inferences or impressions about people. For example, Goodman (1950) found that trait words such as "cold" were more powerful in influencing impressions than positive trait words such as "warm." Similarly, Osgood, Succi and Tannenbaum (1957) found that, contrary to their proposed congruity principle, equally polarized positive and negative information did not have a balancing effect on impression formation; rather, in every instance of a reported error the direction of influence favored the negative information.

In 1965 Anderson found that negative adjectives seem more powerful than positive adjective sets in affecting overall evaluations. Similarly, Feldman (1966) and Richey, McClelland and Shinkunas (1967) found that the weight given negative adjectives exceeded weight given to positive adjectives when several must be combined into one overall evaluation.

Others have attempted to identify the situational variables that affect person perception. The strong impact of negative information occurred when the target person was a female rather than a male (Richey and Dwyer 1970), and when a variety of different personality variables were tested (Briscoe, Woodyard and Shaw 1967). In an examination of situational message variables Cusumano and Richey (1970) manipulated order and intensity factors and Richey, Koenigs, Richey, and Fortin (1975) varied amounts of negative and positive information. In both studies, negative information was more salient than would be predicted using a simple averaging theory. In the latter study, one piece of negative information effectively neutralized five positive behaviors. The generalizability of negative information effects receives support in a cross-cultural study conducted in Denmark where Bernadette Gray-Little (1978) found that Danes, like Americans, evince a disproportionate negative weighting scheme. Despite the Danish norms of cooperation and tolerance it appears that Danes, like Americans, are prepared to believe the worst about others.

Personnel Decisions

In the more applied context of employment and personnel decisions the impact of unfavorable information in impression formation has also been studied. Bolster and Springbett (1961) indicated that interviewers tended to give more weight to unfavorable information and subsequently these findings were replicated by Webster (1964). With a contrary note Hollman (1972), using a unique methodology, found that rather than overweighting negative information, interviewers tended to underweight positive information. The accumulated research in person perception and personnel evaluation strongly indicates that in evaluating others, negative information is more influential

than positive information.

Rumor

Rumor study represents a special type of impression formation dealing with the diffusion of positive or negative information about some commercial or non-commercial entity. From a marketer's perspective derogatory rumors present a challenge to public relations activities. Here the absence of substantive fact or the inability to establish the original source make it the most insidious type of information. Negative rumor is particularly problematic in a business context where individuals in this age of consumerism often are predisposed to believe the worst about companies and often are ready to absorb unfavorable information.

Allport and Postman (1947) suggested that the dynamics at work in the birth of a rumor were driven toward perceptual simplicity, orderliness, and completion. Researchers believe that the amount of rumor activity is a function of situational ambiguity and thematic importance. In this vein Rosnow and Fine (1976) present results which suggest that a combination of uncertainty and anxiety set the stage for rumor.

Similarly, Shibutani (1966) purports that the level of suggestibility derived from anxiety, tension, frustration, or lack of trusted information sources will play a role in affecting the readiness of a group to accept rumor. In his view rumor must be plausible and fit with persons' beliefs and inclinations. The fact that embarrassing rumors are seriously entertained, then, is an indication that many people are already predisposed to believe them.

The response to negative rumor is typically a denial. Denial by a trusted source of an absurd rumor is effective according to Shibutani, but usually denial is ineffectual because people often confuse the denial with the charge or those who have not heard the charge may find the denial unconvincing. Lasswell (1933) insists that the only really effective defense against propaganda is promotion favoring some alternative program.

Tybout, Calder and Sternthal (1979) examined in an experimental setting strategies that might be employed to combat rumor. Their results indicate as suggested by others that direct refutation is ineffective. Alternatively, the authors suggest two other strategies based on information processing theory which attempt to link rumor with some other object (storage strategy) or highlight features or services not implicated in the rumor (retrieval strategy). These two approaches had a significant impact on subjects' purchasing attitudes and behavior.

Fear Appeals

A negative information message involves an attempt to discourage some belief, attitude, preference, or behavior and therefore encourage some alternative set of ideas or actions. Fear appeals as studied by social psychologists and marketers, though not historically viewed in the context of generalized negative information, present a classic case of an attempt to discourage and in turn encourage some alternative viewpoint or action.

Hovland, Janis and Kelley (1953) viewed fear as anxiety arousal leading to a tension reducing response such as a change in one's attitudes, opinions or behavior. Janis and Feshbadi (1953) concluded from their research that low levels of fear were most effective because a great deal of fear provokes defensive reactions and interferes with yielding. Powell (1965) found that high fear produces more opinion change than mild fear when the communication was focused on a loved one, attributed to either a highly credible source or a source independent of the content. Capon and Hulbert (1973) found that the results

of mild versus high fear messages vary with a delay in post-measurement and Sigall and Helmreich (1969) purport that the context of the fear message can have a marked moderating impact on the effect of fear provoking communication. It has also been shown that the personal variables of self-esteem, perceived vulnerability, and coping style mediate the strength of fear appeals (Highbee 1969; Leventhal 1965; Ray and Wilkie 1970).

The controversy in the fear literature regarding a monotonic versus curvilinear fear - yielding relationship and the evidence of situational influences is potentially a microcosm of the broader negative information setting. As such, the fear literature should offer insight in the design of negative information research programs.

Empirical Research--The Consumer Context

In a general sense negative consumer information focused on products has been examined on a very limited basis. Arndt (1968) studied the effects of negative word-of-mouth information about a new brand of coffee and found that high risk perceivers were influenced more than low risk perceivers no matter what type of information was received. In particular though, the high risk perceivers were most strongly affected by negative word-of-mouth information while the low risk perceivers were not. Thus, it seems that risk perceptions affect receptivity to information generally and especially to negative information.

Wright (1974) focused a study on decision time and the use of information. The results indicated that as decision time compressed subjects placed a greater reliance upon unfavorable product information perhaps as a means of simplifying their task environment.

In a study using Consumers Union as the source and a detergent as the focus Lutz (1975) found that negative information had greater impact than positive information on cognitive structure and attitude. He suggests that the result might be due to the novelty and consequent higher information value of negative information.

In two studies Weinberger and Dillon (1980) and Weinberger, Dillon and Allen (1980) found strong effects of negative information. In the first study housewives were given either positive or negative ratings for a set of unbranded goods and services. Strong main effects for negative versus positive information, and product type (goods/services) were found. Significant interactions for source and rating type were uncovered where the independent rating source was a significantly stronger source of negative information than either a trade source or other housewives.

Weinberger, Dillon and Allen (1980) demonstrated the impact of negative product news information. An actual negative news story about an automobile was imbedded in a network news broadcast and its effect was assessed. Strong effects on opinions, beliefs and intentions were demonstrated in immediate and two week delayed responses. Direct rebuttals by the impacted firm had little neutralizing effect. Notably the laboratory effects conform to a sharp drop in sales that followed the actual airing of this news story.

Most negative information studies in a consumer context have not been of as general a nature as those described in the preceding section. Rather, they have focused on issues of comparative, counter and corrective advertising, self-denigrating information and product recall. The research in each of these areas will be discussed very briefly.

Comparative Advertising

Comparative advertising as negative information is distinguished by the source (competitor), format (advertising) and the typically low key intensity of messages. Lincoln and Samli (1979) reviewed the literature and found that denigrating a competitor has highly equivocal implications. For the advertiser, some consumers felt that such an approach to advertising was offensive, in some cases less believable, and in general were skeptical of claims. High intensity comparisons were found to be less informative.

Self-Denigrating Information

In most circumstances the marketer is concerned about the impact of negative information from external or uncontrollable sources. At times, however, the source of negative information might be the firm itself. This latter situation might arise from Federal Trade Commission requirements for affirmative information disclosure of certain information, corrective advertising to rectify a previous deception, or voluntary recall action. Alternatively, a firm might intentionally give some negative information about itself to enhance its overall believability.

Corrective Advertising. Corrective advertising orders have been issued by the Federal Trade Commission for the past decade in an effort to force advertisers to rectify misimpressions that develop from deceptive product claims.

Research studying the impact of corrective advertising in remedying false claims or deceptions has been mixed. Hunt (1972) and Mazis and Adkinson (1976) found that corrective ads can dissipate false claims whereas Dyer and Kuehl (1978) showed that limited exposure to corrective ads did not resolve misimpressions brought about from false ads.

Source and message intensity have been the major situational variables studied in conjunction with corrective advertising. Mazis and Adkinson (1976) found no source effects between company or the FTC while Dyer and Kuehl (1978) in a series of investigations revealed that the FTC was more effective than a company source for print media though the same effect did not occur in broadcast media.

A number of studies examined whether alternative levels of specificity produce different effects on falsely held impressions. Hunt (1974) found no significant effects of supported and unsupported ads while Dyer and Kuehl (1974) and Hunt (1973) found that high intensity correctives were more effective than low intensity messages in dispelling false impressions.

Clearly the format (advertising context), and the generally low intensity and weak dominance of the corrective message distinguish it from more blatant forms of negative information. These inherent factors might help explain the less dramatic impact of corrective advertising as a form of negative information.

Affirmative Disclosure. One strategy which the FTC used in attempting to prevent deception by advertisers has been to call for presentation of certain information which by its absence might be misleading. "May cause nausea or vomiting", "Batteries not included" and other such shreds of information are for the most part unfavorable from the advertiser's perspective. To date there is no substantive empirical work indicating whether disclaimers of this type have any impact on consumers.

In a study related to the affirmative disclosure concept, Settle and Golden (1974) conducted an experiment in which they varied the amount of negative disclaimer information which an advertiser might voluntarily include in an ad. Though their theoretical conceptualization and interpretation of results have been challenged (Hansen and Scott 1976) their findings do suggest that some negative infor-

mation about oneself in the context of expected positive information might enhance claim and source credibility. Golden and Settle's findings provide evidence that unfavorable information need not have an inherently deleterious effect and that in some contexts to give such information about oneself in mild forms may even enhance credibility.

Counter Advertising

A form of communication which has been suggested as a remedy for deceptive advertising is the counter commercial. Here a spokesperson offers counter arguments to information set forth by an advertiser or group of advertisers. Reactions to commercials about heavily sugared cereals or automobiles with poor safety records might be permitted but to date none of the counter commercials have appeared on network TV. Because the doctrine of counter advertising has not been accepted little empirical investigation of the phenomenon has been undertaken. One notable exception is Hunt (1972) who reports a significant decrease in affective attitude after exposure to such counter advertising. These results occurred whether government, consumer or original sources were employed.

Product Recall

The product recall unlike corrective or comparative advertising involves an infusion of negative information typically separate from the usual format of commercial advertising. The issue of recall results from a company marketing a product deemed to be unacceptably hazardous either by an outside agency or the company itself. Experimental work in a non-product context (Walster 1966; Shaver 1970; and Schroeder and Linder 1976) indicates that the more severe a possible injury, the more responsibility will be attributed to the person potentially at fault.

In a series of studies Mowen (1980; 1980a; 1980b) manipulated severity of injury, voluntary or coercive nature of recall, number of recalls and order of presentation of company information about the recall. In the first study Mowen (1980) found main effects for severity, coerciveness of recall, and number of prior recalls. In a follow-up investigation (Mowen 1980b), however, the company conducting a voluntary recall was surprisingly perceived as having greater responsibility for the problem than companies forced to conduct a recall. The third study in the series (Mowen 1980c) attempted to determine if the order of information released in a recall would differentially affect perceptions of the company. Results indicate that the company was perceived most favorably when they provided the most severe possible outcome first and later released reported effects at moderate levels.

Investigation of product recall is in its early stages and the dimensions which mediate its impact are still largely speculative. Based upon actual market share data presented by Wynne and Hoffer (1976) it appears however, that this form of negative information has a direct relationship to product sales and is thus an important area for future research.

Summary of Empirical Research

The divergent literature dealing with negative information tells us a great deal about its effects and its situational determinants. In circumstances where the negative information is the major input, studies in person perception, rumor, product perception and counter advertising indicate that there are powerful negative effects which can offset disproportionate positive input. In addition, person perception results have been shown to hold up in a cross-cultural environment.

The precise effect of the negative information is moderated by message, source, target, receiver and company response. The literature reviewed in the previous pages can be briefly summarized with respect to these factors:

Message. The message effects of negative information are situational as found in the message intensity studies in the fear literature. Corrective advertising and product recall research reveal that message intensity, dominance and severity of the problem play a role in determining informational impact. In addition, the context of the message including time pressure on the individual to make a decision plays a role in enhancing the deleterious effects of the unfavorable information.

Source. The empirical evidence indicates a clear interaction between negative information and message source. In the fear appeals area a credible source independent of the situation evokes the greatest effects. In a consumer context, the neutral spokesperson has been shown as a powerful source in several settings though with corrective advertising source effects research is more equivocal. In the case of comparative advertising where a competitor is the source, the impact of the unfavorable information is less powerful reflecting a potentially ineffective competitor source inherent in the comparative form of advertising.

Target. The target of the negative message can influence the impact of the information. When a loved one is the focus, when the theme is important, and when the target is a service rather than a tangible product, the impact of negative information is enhanced. One study revealed that when the negative information was about relatively unimportant features of the target and given by the firm itself the negative information could have an opposite effect and enhance the credibility of an advertiser denigrating itself.

Receiver. As suggested by McGuire (1968) the receptivity of the receivers and their predispositions can play a role in predicting informational effects. The fear literature tells us that self-esteem, vulnerability and coping styles are important mediators in accepting unfavorable input; rumor research finds that anxiety and frustration are key factors; and a word-of-mouth marketing communication study reveals that risk perception is critical. By whatever mechanism is operative these personal factors apparently impact on the receiver's receptivity and subsequent willingness to assimilate negative information.

Response. Concern for the proper response to negative information is common to several research areas. The product recall literature reveals contrary to popular belief that a voluntary recall may lead to perceptions of greater fault. Additionally, the order of company press releases of product problems giving the mild, moderate and severe recall scenarios is important. Rumor research asserts that denials tend to be ineffective and this finding has been supported in a marketing context with automobiles and hamburgers. Alternatively, response strategies focusing on storing and retrieving diversionary information have been found effective in offsetting negative rumor information.

In general this review reveals that we have limited insight into the effects of negative information in a consumer context and even less guidance for the marketer in combating negative information. The next section provides some potential research directions which are necessary to more fully explore negative information issues.

Research Directions

While there is considerable evidence for a "negative information effect", in comparison and corrective advertising contexts, the effects are mild, and with self-denigrating messages negative information has a positive effect. The implication is there may be a number of factors which qualify the impact of negative information and research is needed which examines these factors. There has not been

enough experimental research wherein a variety of situational factors have been controlled or manipulated for conclusions about the potency of negative versus positive information to be unequivocally drawn. Studies manipulating message source and content along with information valence are needed. Techniques such as functional measurement where the value of units of information are indirectly determined using efficient within subject designs might be a reasonable approach to such a task.

One of the more compelling arguments for a "negative information effect" in the American marketplace is furnished by Lutz (1975). He argues on the basis of Brock's commodity theory that negative information is more potent than positive because of its relative scarcity. If true, one should be able to demonstrate that a given bit of negative information in a particular setting or cultural environment where there is a more equal balance between negative and positive information, should have less information value and subsequently less impact. This suggests possible cross-cultural or laboratory settings where informational norms or expectations vary and where research could be conducted. Such theory based empirical investigations are essential if a systematic study of negative information effects is to be accomplished.

The fact that in prior research consumers have been exposed to negative information in a confrontational as opposed to a search (Bettman 1978) mode raises some interesting questions concerning consumers' processing of negative information. For instance, do individuals consider the valence of information as they search for it? Are some individuals more motivated to seek out negative information about a product before purchase than others? Are there particular product categories for which negative information is more actively sought prior to purchase?

It might be argued that from the consumer's viewpoint negative information is received at random time intervals and thus the heavy emphasis on the confrontational mode in past research is appropriate. An important question concerning the processing of negative information in this mode involves how and where negative information is stored in memory. Are consumers more prone to retain negative as opposed to positive information? Are individuals more apt to process and store factual information about products when it is negative? This latter question might be examined using thought listing methodologies to establish whether or not consumers are more likely to rehearse negative as opposed to positive factual information. It may be that negative information is stored in the form of a generalized negative affect about a brand. It may also be that the carryover effects of negative information are stored in non-brand-specific dispositional constructs such as consumer alienation from the marketplace (Allison 1978). An emphasis on the confrontational mode may be appropriate in studying negative information effects; however, research dealing with this mode of information processing should certainly stress delayed measurement to establish the strength and nature of the carry-over effects of negative information. In addition, knowing the secondary effect of negative word-of-mouth spread from a person in the confrontational mode to other persons in a similar or more decision oriented mode represents an important research issue.

The effects of negative information on targeted product attributes is certainly of concern to the marketer, but other types of effects are also of potential interest. Kanouse (1971) suggested that negative information might have a kind of general negative halo effect on one's enjoyment of the product as a whole. The extent of such secondary effects would provide additional evidence concerning the potency of negative information.

On a more macro level the cumulative effects of negative information in the market on basic dispositions like

alienation or consumer discontent (Lundstrom and Lamont 1976) raises important public policy issues. In a comparative study of alienation in the U.S. versus Sweden Schewe et al. (1980) found alarmingly strong perceptions of normlessness in business practices among U.S. consumers. If a potential trend toward increased consumer alienation and discontent can be linked to growth in negativity of the consumer's information environment, research regarding methods for neutralizing the cumulative effects of negative information would seem desirable.

Anomalies in the effect of negative information such as the weak effects of comparative advertising, the extent of reverse effects of some self-denigrating information, and boomerang effects of some negative information disclosures all merit investigation. Why might a ban on saccharin as a cancer causing ingredient trigger consumer hoarding of saccharin based products? Is it due to a source effect or the fact that the deleterious effects are long-term? Is it an avoidance reaction? Or is it simply the threat of withdrawal leading to greater desire as predicted by reactance theory?

It is, of course, probable that individuals will differ in their receptivity to negative information. A useful stream of research might attempt to isolate variables that are effective for identifying persons who are more or less receptive to negative information. The theoretical constructs of the assimilation-contrast/threat-compliance paradigms offer a firm basis from which to study receptivity and predispositions to receive and yield to negative messages. It would be valuable for a firm to know what type of person is going to be most influenced by negative information. Subsequent neutralizing strategies might be targeted if groups of such individuals could be isolated and their receptivity better understood.

Finally, there is the very practitioner-oriented issue of what type of communication strategies are effective in neutralizing the effects of negative information. Past studies indicate that strategies which attempt direct refutation are ineffective (Tybout, Calder, and Sternthal 1979; Weinberger, Allen, and Dillon 1980). Empirical research is needed to test other alternatives. Tybout, Calder, and Sternthal (1979) suggest a pair of response strategies for offsetting rumor effects they developed through application of an information processing conceptualization to the problem. This theory-based approach may be employed to derive other candidates for empirical evaluation in this process of identifying strategies for neutralizing the influence of negative information.

References

- Allison, Neil K. (1978), "A Psychometric Development of a Test for Consumer Alienation from the Marketplace," *Journal of Marketing Research*, 15, 565-75.
- Allport, G. W. and Postman, L. J., (1974), *The Psychology of Rumor*, reprint, ed., New York: Russell and Russell, 1965.
- Anderson, Norman H., (1965), "Averaging Versus Adding as a Stimulus-Combination Role in Impression Formation," *Journal of Personality and Social Psychology*, 9, 272-279.
- Arndt, J., (1968), "Word-of-Mouth Advertising and Perceived Risk," in *Perspectives in Consumer Behavior*, H. Kassarian and T. Robertson, eds., Glenview, Ill.: Scott, Foresman and Company.

- Asch, S. E., (1953), "Effects of Group Pressure Upon the Modification and Distortion of Judgements," in Group Dynamics, D. Cartwright and A. Zander, eds., New York: Harper and Row.
- Bettman, James R., (1978), "Consumer Information and Search Strategies," in The Effect of Information on Consumer and Market Behavior, Andrew A. Mitchell, ed., Proceedings, American Marketing Association, 35-48.
- Bolster, B. I. and Springbett, B. M., (1961), "The Reaction of Interviewers to Favorable and Unfavorable Information," Journal of Applied Psychology, 45, 97-103.
- Briscoe, M. E., Woodyard, H. D. and Shaw, M. E., (1967), "Personality Impression Change as a Function of the Favorableness of First Impressions," Journal of Personality, 35, 343-357.
- Capon, H. and Hulbert, J.. (1973), "The Sleeper Effect, an Awakening," Public Opinion Quarterly, Fall, 333-358.
- Cusumano, D. R. and Richey, M. H., (1970), "Negative Salience in Impressions of Character: Effects of Extremeness of Stimulus Information," Psychonomic Science, 20, 81-83.
- Dyer, R. F. and Kuehl, P. G., (1974), "The Corrective Advertising Remedy of the FTC: An Experimental Approach," Journal of Marketing, 38, 48-54.
- _____ and Kuehl, P. G., (1978) "A Longitudinal Study of Corrective Advertising," Journal of Marketing Research, 15, 38-48.
- Feldman, Sheldon, (1966), "Motivational Aspects of Attitudinal Elements and Their Place in Cognitive Interaction," in Cognitive Consistency, Sheldon Feldman, ed., New York: Academic Press.
- Goodman, S. M. (1950), Forming Impressions of Persons From Verbal Reports, unpublished doctoral dissertation, Columbia University.
- Gray-Little, B., (1973), "The Salience of Negative Information in Impression Formation Among Two Danish Samples," Journal of Cross-Culture Psychology, 4, 193-206.
- Hansen, Robert A., and Scott, Carol A. (1976), "Comments on 'Attribution Theory and Advertiser Credibility,'" Journal of Marketing Research, 8, 193-197.
- Heider, F. (1944), "Social Perception and Phenomenal Causality," Psychological Review, 51, 358-374.
- Hewgill, M. and Miller, G., "Source Credibility and Response to Fear-Arousing Communications," Speech Monographs, 32, 95-101.
- Higbee, K. L., (1964), "Fifteen Years of Fear Arousal: Research on Threat Appeals," Psychological Bulletin, 72, 426-444.
- Hollman, T. D., (1972), "Employment Interviewers' Errors in Processing Positive and Negative Information," Journal of Applied Psychology, 56, 130-134.
- Hoveland, C. I., Janis, I. L., and Kelley, H. H. (1953), Communication and Persuasion, New Haven, Conn.: Yale University Press.
- Hunt, H. K., (1972), "Source Effects, Message Effects, and General Effects in Counter-advertising," in Proceedings, 3rd Annual Conference, Association for Consumer Research, M. Venkatesan, ed., 370-381.
- Hunt, H. K., (1973), "Effects of Corrective Advertising," Journal of Advertising Research, 13, 15-24.
- Janis, I. L., (1967), "Effects of Fear Arousal on Attitude Change: Recent Developments in Theory and Experimental Research," in Advances in Experimental Social Psychology, Vol. 3, L. Berkowitz, ed. New York: Academic Press, 167-224.
- _____ and Feshbach, S., (1953). "Effects of Fear-Arousing Communications," Journal of Abnormal Social Psychology, 48, 78-92.
- Jones, E. E., and Nisbett, R. E., (1971). The Actor and the Observer: Divergent Perceptions of the Causes of Behavior. Morristown, N.J.: General Learning Press.
- Kanouse, David E., (1971), Language, Labelling and Attribution, Morristown, N.J.: General Learning Press.
- Kanouse, David E. and Harison, Reid, L., Jr., (1971). "Negativity in Evaluations," in Attribution: Perceiving the Causes of Behavior, Morristown, N.J.: General Learning Press, 47-62.
- Kelley, H. H., (1973), "The Process of Causal Attribution," American Psychologist, 28, 107-128.
- Kuehl, P. G. and Dyer, R. F., (1977), "Applications of the 'Normative Belief' Technique for Measuring the Effectiveness of Deceptive and Corrective Ads," in Advances in Consumer Research, Vol. 4, W. L. Perrault, ed., Proceedings of Association for Consumer Research.
- Lasswell, Harold D.. (1933), "Agitation," Encyclopedia of the Social Sciences, Vol. 1. New York: MacMillan, 487-488.
- Lazer, William, (1980), "Lucrative Marketing Opportunities Will Abound in the 'Upbeat' 1980's," Marketing News, 14 (June 11), 1-14.
- Leventhal, H., (1965), "Fear Communications in the Acceptance of Preventive Health Practices," Bull. N.Y. Acad. Med., 41, 1144-1168.
- _____, Singer, R., and Jones, S., (1965), "Effects of Fear and Specificity of Recommendation Upon Attitudes and Behavior," Journal of Personal and Social Psychology, 2, 20-29.
- Lincoln, Douglas J., and Samli, A. Coskun, (1979), "Empirical Evidence of Comparative Advertising's Effects: A Review and Synthesis," in Proceedings of The American Marketing Association Conference, N. Beckwith, et al., eds., Chicago, Ill.: American Marketing Association, 367-372.
- Lutz, Richard J. (1975), "Changing Brand Attitudes Through Modification of Cognitive Structure," Journal of Consumer Research, 1, 49-59.
- Mazis, M. B. and Adkinson, J. E., (1976), "An Experimental Evaluation of a Proposed Corrective Advertising Remedy," Journal of Marketing Research, 13, 178-183.
- McDougall, Gordon H. G., (1977), "Comparative Advertising: Consumer Issues and Attitudes," in Proceedings of The American Marketing Association Conference, B. Greenberg and D. Bellenger, eds., Chicago, Ill.: American Marketing Association, 286-291.
- McGuire, W. J., (1968), "Personality and Susceptibility to Social Influence," in Handbook of Personality Theory and Research, E. F. Borgatta and W. W. Lambert, eds, Chicago: Rand-McNally.
- Mowen, J. C., (1980), "Consumer Reactions to Product Recalls: An Empirical and Theoretical Examination," unpublished manuscript, Oklahoma State University.

Mowen, J. C., (1980a), "Further Information on Consumer Perceptions of Product Recalls," in Proceedings of The Association for Consumer Research, J. Olson, ed., 519-523.

_____ and Englebrecht, J., (1980b), "Order Effects and Reporting Negative Corporate Information: More Fall-out from Three Mile Island," unpublished manuscript, Oklahoma State University.

Osgood, C. E., Succi, G. J., and Tannenbaum, P. H., (1957), The Measurement of Meaning, Urbana: University of Illinois Press.

Powell, F. A., (1965), "The Effects of Anxiety-Arousing Messages When Related to Personal, Familiar, and Interpersonal Referents," Speech Monograph, 32, 95-101.

Pride, William M., Lamb, Charles W., and Fletcher, Barbara A., (1977), "Are Comparative Advertisements More Informative for Owners of the Mentioned Competing Brand Than for Nonowners?" in Proceedings of The American Marketing Association, B. Greenberg and D. Bellenger, eds., Chicago, Illinois: American Marketing Association, 298-301.

Ray, M. L. and Wilkie, W. L., (1970), "Fear: The Potential of an Appeal Neglected by Marketing," Journal of Marketing, 34, 53-62.

Richey, M. H., Koenigs, R. J., Richey, H. W., and Fortin, R., (1975), "Negative Salience in Impressions of Character: Effects of Unequal Proportions of Positive and Negative Information," The Journal of Social Psychology, 97, 233-241.

_____, and Dwyer, J. D., (1970), "Negative Salience in Impressions of Character: Sex Differences," Psychonomic Science, 20, 77-79.

Richey, M. H., McClelland, L. and Shimkunas, A., (1967), "Relative Influence of Positive and Negative Information in Impression Formation and Persistence," Journal of Personality and Social Psychology, 6, 338-342.

Rosnow, R. L. and Fine, G. A., (1976), Rumor and Gossip: The Social Psychology of Hearsay, New York: Elsevier.

Schewe, Charles D., Chris T. Allen, Bertil Liander, and Gösta Wijk, (1980), "An Exploratory Investigation of Consumer Alienation from the Marketplace: Sweden vs. the United States," in Proceedings of the American Marketing Association, R. Bagozzi, et al., eds., Chicago: American Marketing Association, 444-447.

Schroeder, D. A. and Linder, B. E., (1976), "Effects of Actor's Knowledge of Prior Accidents Upon Attributions of Responsibility," Journal of Experimental Social Psychology, 12, 340-356.

Settle, Robert B., and Golden, Linda L., (1974), "Attribution Theory and Advertiser Credibility," Journal of Marketing Research, 11, 181-185.

Shaver, K. G., (1970), "Defensive Attribution: Effects of Severity and Relevance on the Responsibility Assigned for an Accident," Journal of Personality and Social Psychology, 14, 101-113.

Sherif, C. W., Sherif, M., and Nebergall, R. E., (1965), Attitude and Attitude Change, Philadelphia, Pa.: W. B. Saunders.

Shibutani, Tamotsu, (1966), Improvised News, New York: The Bobbs-Merrill Company, Inc.

Sigall, H. and Helmreich, R., (1969), "Opinion Change as a Function of Stress and Communicator Credibility," Journal of Experimental Social Psychology, 5, 70-78.

Tybout, Alice M., Calder, Bobby J., and Sternthal, Brian, (1979), "Of Worms and Consumers: When the Best Laid Marketing Plans Go Astray," paper presented to Annual Meeting of Association of Consumer Research, San Francisco.

Walster, E., (1966), "Assignment of Responsibility for an Accident," Journal of Personality and Social Psychology, 3, 73-79.

Weinberger, Marc G. and Dillon, William R., (1980), "The Effects of Unfavorable Product Rating Information," in Proceedings of the Association for Consumer Research, J. Olson, ed., Ann Arbor, MI.: Association for Consumer Research.

_____, Allen, Chris T., and Dillon, William R., (1980), "Assessing the Potency of Negative Information in the Marketplace," working paper 80-4, University of Massachusetts.

Wright, P., (1974), "The Harassed Decision Maker: Time Pressures, Distractions, and the Use of Evidence," Journal of Applied Psychology, 7, 1-29.

FACTORS INFLUENCING CONSUMER RESPONSES TO PRODUCT RECALLS:
A REGRESSION ANALYSIS APPROACH

John C. Mowen, Oklahoma State University
David Jolly, Oklahoma State University*
Gary S. Nickell, Oklahoma State University*

Abstract

Two hundred consumers were surveyed to determine their perceptions of four companies which had previously recalled products. The results revealed that impressions of the companies were influenced by: the knowledge that a recall had been made, the perceived danger of the defective product, the perceived corporate responsibility of the company, the knowledge of recalls by other companies, and the perceived responsibility of the company for the defect.

Introduction

Negative or unfavorable public information regarding a company and/or its products is of major concern to business. Exposure to unfavorable information results from such sources as reports of defective products, news media presentations of corporate earnings (e.g., the oil industry), and publicity concerning questionable competitive practices. An excellent case in point is the Three Mile Island nuclear incident. The negative information resulting from the incident has seriously jeopardized the future of the nuclear power industry in the United States. In a recent study regarding the effects of negative product information, Weinberger and Dillon (1979) found that: (1) unfavorable product ratings tended to have a greater impact on purchase intention than did favorable ratings (2) unfavorable product information received from an independent testing agency or peer as a source had a relatively stronger effect on purchase intentions than did similar information communicated by the trade and professional association source, and (3) consumers probably place more reliance on outside sources of information when purchasing a service as opposed to purchasing a product. These findings emphasize the necessity of an organization to react effectively when consumers receive negative information concerning that organization.

Product recalls have become a common avenue by which consumers are exposed to negative corporate information. Approximately 25 percent of all consumer goods firms listed in Fortune's 500 were involved in recall campaigns in 1974 (Kerin 1974). The Conference Board estimates that 25 million product units will be recalled every year (McGuire 1975).

Because of the increasing number of product recalls, many articles have been written on the subject from a "corporate management" perspective. Researchers have written articles concerning the physical tracing and recall of the products (Fisk and Chandran 1975; Warner 1975), contingency plans for recalling products (Kerin 1974), the impact of recalls on market share (Gray 1978; Wynne and Hoffer (1976), assessing the effectiveness of drug recalls in pharmacies (Gumbhir and Jamison 1975), surviving product recalls (Snyder 1974), and the high costs associated with product recalls (Tamarkin 1978).

For a number of reasons it is likely that the number of product recalls will continue to increase. One is the

consumer activist movement. Since the early 1960's, consumers have become more vocal in their criticisms of business organizations in general and those that may have produced a faulty product in particular. The end result is that consumers are becoming much more inclined to report defective products. A second reason is the more activist stance taken by such government agencies as the Consumer Product Safety Commission and the Food and Drug Administration. This increase in intervention is due partly to the increase in voiced demands by the consumers. A final reason for the predicted increase in recalls is the increasing complexity of products. In order for manufacturers to comply with consumer demands and regulatory requirements for safer and better quality products, the products are becoming more complex, thereby, increasing the possibility that defects will occur.

As previously mentioned, a small literature has developed based on the managerial question of "how to" recall a product. It appears, however, that only two empirical studies have been performed to specifically investigate consumer reactions to the negative information resulting from a product recall. Mowen (1979a) manipulated the three independent variables of (1) the extent of injury resulting from the defective product, (2) the length of time taken by the company to make the recall, and (3) the number of previous product recalls made by the company. The results of this study revealed that the extent of injury resulting from a defective product, the number of previous product recalls, and the length of time to recall the defective product each influenced consumer perceptions of the corporation and of a replacement product.

In a subsequent study, Mowen (1979b) manipulated the independent variables of: (1) familiar versus unknown company making the recall, (2) action or inaction by the Consumer Product Safety Commission to force the recall, and (3) whether or not other manufacturers had had similar problems. The results revealed that consumers perceived a familiar company as significantly less responsible for a defect than an unfamiliar company. It was also found, unexpectedly, that consumers perceived the company as more responsible when it acted prior to intervention by the Consumer Product Safety Commission.

In the work by Mowen, studies were conducted in which subjects received fictitious magazine articles describing a product recall. Within the articles, information was varied in order to determine the effects of certain variables on the perceptions of companies. Such an experimental approach has the advantage of establishing high internal validity. However, the external validity of the studies are suspect. In the "real world," consumers simply may not attend to product recall information. Thus, while one may find large effects in an experimental setting, in a naturalistic setting the effects may be minimal.

To overcome the above problems in using an experimental methodology, the present study utilized a survey research approach. Two hundred randomly selected subjects completed surveys asking questions about the public's perceptions of Ford Motor Company, Firestone, Inc., Corning Glass Works, Inc., and Conair, Inc. Each of the companies had previously made product recalls. Responses to the questionnaires were then analyzed by regression analysis. The dependent variable was the perception of the company, and

*David Jolly is an MBA student, and Gary Nickell is a doctoral student in social psychology.

the major predictor variables were knowledge of the recall, impressions of the danger of the product, the company's social responsibility in making the recall, the company's responsibility for the defect, the length of time to make the recall, whether or not other companies had made recalls, and whether or not the company had made previous recalls. (Several additional predictor variables were also included.) If the results of the regression analyses supported the experimental results, one would have impressive collaboration of the previous studies.

Method

Seven research assistants were assigned to obtain respondents from seven areas in a midwestern city (population 42,000). The areas were selected to represent lower middle, middle, and upper middle residential housing areas. Surveyors identified respondents through the following procedure. They would enter the area from any direction they wished. From a random numbers table, they determined which side of the street to sample from and how many houses to count down the street in order to identify the selected house. After completing the survey they would go to the first cross street, randomly select whether to go to the left or the right and then follow the same procedure as before to select the proper house. If the house was vacant, the surveyor would return twice more before selecting a new house. After selecting the house, the surveyor would determine if the respondent answering was a college student or a high school student. Such individuals were excluded from the study.

Respondents cooperated excellently with the surveyors. Only two totally refused to participate, and six others participated reluctantly. Three percent of the respondents were less than 21 years old; 48 percent were 21-25 years, 24 percent were 26 through 39 years old, and 25 percent were over 40 years of age. Ninety-six subjects were male and 105 were female. Forty-one percent of the subjects owned their own homes.

The surveyors were junior and senior students majoring in marketing, who volunteered to work on the research study. Students were carefully selected for their maturity and competency.

On the questionnaire respondents were first asked about their general experience with recalls, their age, and whether or not they owned a home. They were then asked questions concerning Conair, Corning Ware, Firestone, and Ford. Because of the length of the questionnaire, half of the subjects answered questions about Firestone and the other half about Ford. Respondents were given a copy of the questionnaire, and the interviewer would read the questions and place the answers on a code sheet. All rating scales were composed of six points with appropriate anchors designated.

Results

Twenty point four percent of the subjects reported having owned a recalled product. Of the 41 respondents owning a recalled product, 58.5 percent reported taking an action to correct the problem. The majority of the products reported to have been recalled were autos (53.0 percent), tires (10.2 percent), coffee pots (18.4 percent), and hair dryers (14.3 percent). The low rate of return of known defective products matches reports found in the popular literature (Duns Review 1979).

In the first set of data analyses the independent variables of interest (the consumer's knowledge of a recall, the perceived danger of the product, the company's social

responsibility in making the recall, the length of time to recall, the number of previous recalls by the company, the presence of recalls by other companies, and the company's responsibility for the defect) were entered into a stepwise regression and utilized to predict the dependent variable of favorability regarding the company. The analyses were performed separately on data for each of the four companies.

Looking first at the results for Conair, one finds that danger entered first in the stepwise regression ($p < .01$), and responsibility for the defect entered second ($p < .05$). The overall R^2 was .09. Thus, the greater the perceived danger of the product and the greater the perceived responsibility of the company, the less favorable was the impression of the company.

Following the same procedure as above for Corning resulted in no findings of significant effects. This is, consumer favorability toward Corning was not significantly influenced by the independent variables selected.

The analysis for Ford revealed that one significant predictor variable entered the equation. The perceived social responsibility of Ford in making the recall was significantly related to respondent's favorability towards Ford ($R^2 = .19, p < .0001$).

The analysis for Firestone showed substantially divergent results from that found above. Here the company's social responsibility in making the recall ($p < .0001$), previous knowledge of the recall by the respondent ($B = .75, p < .0001$), recalls by other corporations ($p < .05$), and danger of the product ($p < .10$) entered the equation ($R^2 = .32$). Firestone was perceived more favorably when perceived as acting socially responsibly, when respondents were aware that Firestone had recalled the tire, when the danger of the tire was perceived as low, and when other companies were known to have recalled tires. Table 1 summarizes these analyses.

TABLE 1
Significant Effects for Each Company
Using Stepwise Regression Analysis

Dependent Variable	Company			
	Conair	Corning	Ford	Firestone
Consumer's Knowledge of Recall	--	--	--	$p < .0001$
Danger of Product	$p < .01$	--	--	$p < .10$
Company's Social Responsibility	--	--	$p < .0001$	$p < .0001$
Length of Time to Make Recall	--	--	--	
Recalls by Other Companies	--	--	--	$p < .05$
Company Responsibility for Defect	$p < .05$	--	--	
Previous Recalls by Company	--	--	--	
Percentage of Variance Accounted for	9%	--	19%	32%

Discussion

The results of the present research partially support the previous work by Mowen (1979a, 1979b). Mowen previously found, using an experimental methodology, that varying the danger of the defective product, the length of time to recall, and the number of previous recalls by the company influenced consumer favorability regarding the company. Table 1 reveals that for Firestone and Conair the perceived danger of the product was significantly related to the consumers' perceptions of the company. In no cases were

knowledge of previous recalls and the length of time to make the recall significantly related to corporate perceptions.

While the perceived length of time to recall and the recognition of previous recalls failed to directly influence the perceptions of the company, these variables may operate through moderator variables. In particular, the length of time taken to make the recall may act by first influencing the perception of the company's social responsibility. Indeed, this was argued by Mowen (1979a). The argument was based on the hypothesis that consumers will tend to expect that companies will delay recalling a product for as long a time as possible in order to maintain corporate profits. On the other hand, if the company acts rapidly and decisively to recall a product, consumers will perceive it as acting against its own short term interest and in the interest of consumers. Consequently, the belief that the action to recall the product was consumeristically oriented would be augmented (Kelley 1971). For both Ford and Firestone social responsibility did enter the regression equation (both p 's $< .001$). In addition, timeliness was significantly correlated with perceived corporate social responsibility (for Firestone, $R = -.36$, $p < .0001$; for Ford, $R = -.46$, $p < .0001$). Preliminary work using path analysis supports these suggested relationships. Thus, the length of time taken to make a recall may influence perceptions of the company by first influencing the impression of the company's social responsibility. Mowen (1979a) also argued that the number of previous recalls affected perceptions by first influencing the perception of the corporation's responsibility for the defect. For Conair responsibility for the defect did enter the equation $p < .05$. However, responsibility was not significantly correlated with the number of previous recalls by the company. Indeed, the preliminary path analytic work tends to reject responsibility for the defect as a moderating variable for the number of previous recalls made by the company.

As a summary statement, evidence from the regression works partially supports the direct impact of variations in the danger of the product and of the indirect impact of the length of time taken to make the recall on consumer perceptions of a company. No evidence was found indicating that the presence of previous recalls influenced perceptions of the companies.

The results for the Firestone analysis revealed that the mere knowledge of a recall can influence corporate perceptions. A second interesting outcome of the analysis on Firestone was the number of significant predictor variables entering the equation. It is difficult to say just what factors contributed to these differential perceptions of Firestone. In terms of awareness more people were aware of the Ford recall (90.0 percent) than Firestone (71.3 percent), Conair (29.8 percent), or Corning (23.3 percent). One possibility is that direct or indirect knowledge of problems with Firestone tires may have been substantially greater than such knowledge for the other products. Of the 22 respondents who reported owning Firestone 500 tires, 13 indicated that they had had problems with the tires. Only one of the seven individuals who reported owning a Pinto indicated that they had a defective gas tank, and only two of 24 Corning percolator owners reported problems. (The problems with the hairdryers was unrelated to performance.) Based on the high percentage of individuals who had problems with the 500 tires and the large number of individuals owning the tires, it is possible that individuals may have also heard of problems with the tires from friends. Thus, personal knowledge of either a direct or indirect form may have been greater for Firestone than the other companies.

One question arising from the results involves why no significant effects were found in the analysis of the data for Corning Ware. One explanation is that with a relatively small percentage of subjects aware of the recall, it would be difficult to statistically isolate significant effects.

A second explanation is based upon experimental work performed by Mowen (1979b). In this work it was found that the impression of a well-known, highly respected company was not influenced by knowledge of a product recall to the same extent as less favorably viewed companies. Support for the speculation comes from the mean favorability of 1.7 on the six-point scale. Thus, knowledge of the Corning recall may not have influenced the respondent's impressions.

Thus far the published series of investigations performed to determine factors influencing consumer reactions to product recalls includes two experimental studies and the survey research presented in the present article. Future work should involve three separate objectives. The next major step involves the completion of the path analysis work mentioned briefly in this article to test the models developed in the Mowen (1979a, 1979b) studies. The second step will consist of replicating the previous experimental work (Mowen 1979a, Mowen 1979b) across divergent subject populations and product classes. The final stage will involve tracking over time consumer reactions to a series of companies who have made product recalls.

References

- Dunns Review (1979), "A Record Year for Recalls," 113, 28-34.
- Fisk, G and Chandran, R. (1975), "How To Trace and Recall Products," Harvard Business Review, 53, 90-96.
- Gray, R. (1978), "Recalls Haven't Hurt Ford," Advertising Age, 49, 19, 8.
- Gumbhir, A. K. and Jamison, A. C. (1975), "Assessing the Effectiveness of Drug Recalls in Pharmacies," Drug and Cosmetic Industry, 117, 40-3, 113-16.
- Kelley, Harold H. (1971), Attribution in Social Interaction, New York: General Learning Press.
- Kerin, R. A. and Harvey, M. (1975), "Contingency Planning for Product Recall," MSV Business Topics, 23, 5-12.
- McGuire, E. P. (1975), "Product Recall and the Factors of Life," The Conference Board Record, 12, 13-15.
- Mowen, J. C. (1979a), "Consumer Reactions to Product Recalls: An Empirical and Theoretical Examination," unpublished working paper, Department of Marketing, Oklahoma State University.
- _____ (1979b), "Further Information on Consumer Perceptions of Product Recalls," Proceedings of the Association for Consumer Research, Tenth Annual Conference, Advances in Consumer Research: Volume VII, Jerry C. Olson (Ed.), San Francisco, California.
- Snyder, J. D. (1974), "How to Survive a Product Recall," Sales Management, 113, 23-27.
- Tamarkin, B. (1978), "Recalls--The Costs Soar," Forbes, 122, 79-80.
- Warner, H. W. (1977), "Guidelines for Product Recall," Public Relations Journal, 33, 11-13.
- Weinberger, M. C. and Dillon, W. R. (1979), "The Effects of Unfavorable Product Rating Information," Proceedings, Association for Consumer Research, Volume 7, Jerry C. Olson (Ed.), San Francisco.
- Wynne, J., and Hoffer, G. E. (1976), "Auto Recalls: Do They Affect Market Share," Applied Economics, 8, 157-63.

THEORETICAL PERSPECTIVES ON THE IMPACT OF
NEGATIVE INFORMATION: DOES VALENCE MATTER?

Carol A. Scott, University of California, Los Angeles
Alice M. Tybout, Northwestern University

Abstract

Marketing researchers and marketing practitioners have long been interested in the factors that determine which product cues, or pieces of information, are used by consumers in forming judgments. Recent marketplace events such as product recalls, dramatic news stories on product defects, and occasional wild but damaging rumors have focused practitioners' interest more specifically on the utilization of negative product cues and on ways to lessen their impact on evaluations and choice. This special session attests to the fact that some researchers also feel that we should devote some attention to the effects of negative information.

Yet, from a theoretical perspective, a focus on valence *per se* may be wrongheaded and ironically unlikely to provide us with any clear directions for moderating the impact of negative information. Using an information processing theory approach, we contend that research may be better directed toward examining those attributes of cues and of the environment in which the cues are received that make them more salient to the individual and/or more informative,¹ and thus more likely to be differentially attended to and used by consumers. A cursory review of the literature indicates that valence *per se* may not be one of these attributes. Further, we argue that researchers should investigate the effects of various cue properties on the information processing strategies or cue combination rules used by consumers as well as on the ultimate evaluation. Examining the underlying determinants of attention and cue integration will provide direction for designing more effective influence strategies. This abstract briefly sets forth this theoretical view and illustrates its value.

Does Valence Matter?

Several studies have suggested that negative information is given relatively more weight than positive information in judgment formation (e.g., Kanouse and Hanson 1971; Lutz 1975; Wright 1974). It is important to note the emphasis on differential weighting, as certainly no one would be surprised that the addition of negative cues to an information base produced less favorable evaluations. Most attitude models would predict this latter outcome (cf. Fishbein and Ajzen 1975). A strong test of the differential weighting effect is extremely difficult to conduct, though, because one must somehow devise a situation in which two pieces of information are equal on all other plausible dimensions except valence.

In this regard, it is interesting to note that most explanations proposed for the observed differential impact of negative information entail the use of some other descriptive construct that essentially pinpoints a difference other than valence between the information pieces. Thus, negative cues are not given more weight because they are negative, but rather because negative cues happen to be statistically rare (i.e., one encounters nega-

¹Fiske (1980) argues that informative cues are by definition salient or attention attracting. However, as empirical research has shown (Taylor and Fiske 1978), salient cues are not always high in informativeness.

tive information less frequently than positive information in everyday life), because they are rare in context (i.e., one negative cue amidst several positive cues in a particular context), or because they are non-modal, or extreme (i.e., the property of being "bad" on an attribute when almost all other similar objects are "good" on the attribute). As Fiske (1980) argues, cues that are rare or non-modal are more informative than common or modal cues because they discriminate between similar objects. If, as recent work in social cognition suggests, people are "cognitive misers" (Nisbett and Ross 1980; Taylor 1980), they may follow the adaptive strategy of differentially attending to these more informative cues. And, differential attention has been shown to determine the weight given to a cue in forming judgments.

One must be very careful, then, in generalizing the conclusion that negative information is weighted more heavily than non-negative information, especially to product evaluation contexts. The rarity or distinctiveness of a cue can change over time and situations independently of its valence. Many of the studies frequently cited as supporting the negative information effect, for example, have been conducted in the context of evaluating or forming impressions of other people where indeed social customs make negative information unusual or unexpected. Clearly, one can imagine product evaluation contexts where very positive information would be as rare or as distinct (e.g., an honest and reliable car repair shop). Fiske (1980) recently has demonstrated that extremely positive, and thus statistically rare, information is given more attention (as measured by looking time) and more weight in forming judgments than other cues.

A focus on the underlying properties of cues which cause differential attention (e.g., the informativeness produced by rarity or non-modality) and thus differential impact is clearly warranted by extant research. Such a focus has a number of advantages. Obviously, it will allow us to specify the conditions under which negative (or positive) information will have an especially heavy impact on product evaluations. Further, it will open the door to investigation of other properties which may affect attention to cues. A knowledge of important situational qualifiers indicates when a manager should be especially concerned about negative product information, and a knowledge of factors that produce increased attention may suggest ways of decreasing attention to certain cues, or at least enable one to counter with informational cues of equal attention value.

Apart from the relationship between cue attributes and attention, however, the information processing approach suggests that one should also be concerned with the relationship between cue attributes and cue integration processes. Certain types of cues may be processed more easily and thus be used more readily. Other types of cues may be more difficult to interpret, and thus may affect evaluations indirectly by causing the individual to shift to a different cue integration strategy. For example, information about products can be obtained through learning about other people's experiences or opinions. Experiential information may be processed fairly easily, but information from others (external information) may require more cognitive work (see Kelley 1967). This would be especially true if the external

information is presented in abstract, statistical terms (e.g., "80% of the people who tried product x found ...") as opposed to concrete, vivid terms (e.g., "When Joe Smith tried his new lawn mower ..."), as discussed by Borgida and Nisbett (1977). Thus, the source (external vs. experiential) and the format (abstract vs. concrete) of the information may affect the process by cues are combined to form judgments.

This type of reasoning formed the basis for several experiments recently conducted by the authors. Specifically, we hypothesized that receipt of negative abstract external information about a product (i.e., 80% of the people who tried it preferred another brand to it) would produce uncertainty about the true character of the product. If this information was received after actual experience with the product, we postulated that the uncertainty would be resolved by recalling the more easily processed experience cue, and that consumers would follow a simple integration strategy that entails merely adding or averaging informational cues. If the person had no experiential information, however, the uncertainty caused by the external information should motivate the individual to recall any past behaviors related to the product and to combine these cues according to the causal analysis rules described in Bem's (1972) self-perception theory.

This proved to be the case. Individuals who had agreed to try a new soft drink in return for a coupon incentive evaluated the drink less favorably than those who did not receive an incentive for trial (i.e., the typical discounting effect predicted by self-perception theory) only when the negative evaluations of others were received prior to tasting the product. On the other hand, when the negative information was received after the taste experience, all cues were simply added such that the incentive group (with the additional positive reward cue) evaluated the drink more favorably than the no incentive group. This finding was observed across several dependent measures (semantic differential scales, overall evaluation measures, behavioral intentions). Further, the incentive-information after group reported the least desire for more information about the product before using it regularly, recommending it to friends, etc.

The only problem with concluding that negative external information produced the uncertainty and thus the differences in cue integration processes is the fact that the same results were found for positive external information. Even when subjects tasted a good product and they received positive information from other people, the order of the two types of information (experiential and external) had a dramatic effect on subsequent evaluations. Thus, one must conclude at this point that the ambiguity of the external information, and not its valence caused the differential processing of cues. If this is the case, then one might expect that introducing factors to make the external information less ambiguous would extinguish this effect. Providing external information in vivid, concrete terms might be one way to achieve this.

Summary

As the discussion of new and old literature in this paper illustrates, we believe that an examination of the underlying properties of negative information that sometimes make it more informative or attention arousing or that cause individuals to shift to different cue integration strategies is a useful approach. Ecologically, negative information may possess some of these attributes. But it is important from a practical and a theoretical perspective to separate valence from other properties.

Placing the research on negative information in the more general information processing framework provides a richer basis from which to design strategies to counter negative information.²

References

- Bem, D. J. (1972), "Self-Perception Theory" in Advances in Experimental Social Psychology, Vol. 6, ed. Leonard Berkowitz, New York: Academic Press, pp. 1-62.
- Borgida, E. and Nisbett, R. E. (1977), "The Differential Impact of Abstract vs. Concrete Information on Decisions," Journal of Applied Social Psychology, 7, 258-271.
- Fishbein, M. and Ajzen, I. (1975), Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading, Mass.: Addison-Wesley.
- Fiske, Susan T. (1980), "Attention and Weight in Person Perception: The Impact of Negative and Extreme Behavior," Journal of Personality and Social Psychology, 38, 889-906.
- Kanouse, David E. and Hanson, L. Reid, Jr. (1972), "Negativity in Evaluations," in Attribution: Perceiving the Causes of Behavior, ed. E. E. Jones et al., General Learning Press.
- Kelley, H. H. (1967), "Attribution Theory in Social Psychology," in Nebraska Symposium on Motivation ed. D. Levine, Lincoln, Nebraska: University of Nebraska Press.
- Lutz, Richard J. (1975), "Changing Brand Attitudes Through Modification of Cognitive Structure," Journal of Consumer Research, 1, 49-59.
- Nisbett, R. E. and Ross, L. (1980), Human Inference Strategies and Shortcomings of Social Judgment, Englewood Cliffs, N. J.: Prentice-Hall.
- Taylor, Shelley E. (1980), "The Interface of Cognitive and Social Psychology," in Cognition, Social Behavior, and the Environment, ed. J. Harvey, Hillsdale, N. J.: Erlbaum.
- Taylor, Shelley E. and Fiske, Susan T. (1978), "Salience, Attention, and Attribution: Top of Mind Phenomena," in Advances in Experimental Social Psychology, Vol. 11, ed. Leonard Berkowitz, New York: Academic Press.
- Tybout, Alice, M., Calder, Bobby J., and Sternthal, B. (in press), "Using Information Processing Theory to Design Marketing Strategies," Journal of Marketing Research.
- Wright, Peter (1974), "The Harrassed Decision Maker: Time Pressures, Distractions, and the Use of Evidence," Journal of Applied Psychology, 59, 555-561.

²An excellent example of how counter-strategies can be developed is provided by Tybout, Calder, and Sternthal (in press). Using available knowledge concerning the mechanisms of cue-storage and retrieval, these authors were able to demonstrate why the refutational strategy used by McDonald's to counter the recent rumor about worms in their hamburgers was ineffective. Further, they demonstrated that other strategies, based on information processing principles, were effective in neutralizing the rumor.

VIEWER MISCOMPREHENSION OF TELEVISED COMMUNICATION:
A BRIEF REPORT OF FINDINGS¹

Jacob Jacoby, Purdue University
Wayne D. Hoyer, Purdue University²
David A. Sheluga, Purdue University²

Abstract

Some of the major findings stemming from a large scale investigation of viewer miscomprehension of televised communication are presented. These findings reveal that virtually all viewers miscomprehend to some extent, and that no communication is immune from being miscomprehended. Further, the average amount to which the core content of each communication was miscomprehended was a surprisingly high 30%. Finally, commercials seem to be slightly better comprehended than program content. A few of the implications of these data are also discussed.

Introduction

Considerable interest has recently been focussed on issues related to consumer miscomprehension of advertising. Consumer advocates, regulatory agencies, and academic researchers have been devoting an increasing amount of attention to the issues of deceptive and/or misleading advertising and their remedies (e.g., corrective advertising, affirmative disclosure statements). They have not, however, addressed the logically prior set of questions concerning the existence, degree, correlates, and causes of miscomprehension. The evidence that is available in these regards tends to be fragmentary, anecdotal, and impressionistic.

Implicit in the arguments advanced by many consumer advocates is the assumption that commercial advertising is prone to causing miscomprehension whereas other forms of communication are generally not so prone. A problem with this view, however, is that it ignores our contemporary understanding of the communication process. In contrast, an alternative assumption is virtually all forms of communication are subject to being miscomprehended. Stated somewhat differently, there is not necessarily any correspondence between a given communication as conveyed by the source and that same communication as interpreted and remembered by the receiver. This is because the meaning a receiver extracts from a given communication consists

of both directly asserted meaning (as expressed in the message) and meaning inferred by that individual receiver. These latter meanings are a unique function of each receiver's total sum of prior experiences and the set of expectations he brings to the situation (cf. Harris and Monaco 1978).

If the latter assumption is correct, then the fundamental question to be resolved is not "Does advertising cause miscomprehension?" but, rather "Is there any greater level of miscomprehension associated with advertising than is associated with other comparable forms of mass media communication?" In other words, it is conceivable that miscomprehension of advertising may not be uniquely attributable to advertising per se, but may reflect a natural error rate common to all comparable forms of one-way mass communication.

Given the lack of hard benchmark evidence, the Educational Foundation of the American Association of Advertising Agencies decided to fund three investigations on the questions surrounding miscomprehension. This report represents a highly condensed summary of the basic findings generated by the first of these investigations. The primary objectives of the present investigation were:

1. To determine whether viewers do, in fact, miscomprehend televised communication.³
2. Assuming that some degree of miscomprehension is detected, to determine whether there is a "normative range" of miscomprehension associated with televised communications.
3. To determine whether commercial advertising tends to have a rate of miscomprehension which differs from other forms of televised communication.
4. To determine whether viewers possessing certain demographic characteristics are more prone to miscomprehending televised communications than are others.

Method

Given these objectives, three core concepts needed to be operationalized. These were: (1) communications, (2) viewers, and (3) miscomprehension.

Communications

Sixty different communications which had been broadcast over TV during the six months spanning the last quarter of 1978 and first quarter of 1979 constituted the set of test communications. These communications were broadly representative of most material being broadcast over commercial television. The set of sixty communications can be grouped into three categories: commercial advertising (n=25), noncommercial advertising (n=13), and program excerpts (n=22). The commercial ads consisted of 18 ads for brands and services and 4 "image" or "corporate"

¹This paper is a highly condensed version of a monograph describing a recently conducted investigation sponsored by the American Association of Advertising Agencies. The authors acknowledge with gratitude the many constructive contributions made by the members of the Academic and Industry Review Committees established by the AAAA to monitor this project from inception to completion. The members of the Academic Review Committee were Professors Jacob Cohen (New York University), Stephen Greyser (Harvard University), and William McGuire (Yale University). The Industry Review Committee members were Mrs. Rena Bartos (J. Walter Thompson), Dr. Seymour Banks (Leo Burnett), Dr. Theodore Dunn (Benton and Bowles) and Dr. Benjamin Lipstein (originally at SSC&B, now at New York University). The full description is provided in Jacoby, Hoyer, and Sheluga (1980).

²Wayne Hoyer is now with the Department of Marketing, University of Texas at Austin, Austin, Texas. David Sheluga is now with the Quaker Oats Company, Barrington, Illinois.

³Given that the major share of national advertising is devoted to television and regulator interest has been most evident in regard to this medium, attention was confined to televised communications.

ads. The non commercial ad group included 8 public service announcements and 5 ads for social causes. Finally, the group of program excerpts consisted of 18 excerpts from popular TV programs (including national and local news) and 4 excerpts from speeches and editorials.

Viewers

The sample consisted of 2,700 people, 13 years old and older. Sampling quotas insured that there were equal numbers of males and females tested at each of 12 testing sites and that there were 386 respondents subsumed under each of the following age brackets: 13-17; 18-24; 25-34; 35-44; 45-54; 55-64; and 65+. Compared to U.S. Census Bureau data, the sample was found to be broadly representative in terms of years of formal schooling completed, household income claimed for the preceeding year, marital status, and race. As a whole, the sample had a slight upward bias in terms of both education and income. That is, it claimed to be more highly educated and reported a higher level of prior year's income than did the population at large.

Miscomprehension

A six item miscomprehension quiz (patterned after Preston and Scharbach 1971) was administered after exposure to each communication. Each quiz consisted of six brief statements which sampled the universe of important information content regarding the object of that communication. Viewers were asked to indicate whether each statement was "True" or "False" based upon what was stated or implied in the communication. Two of the six statements on each quiz were accurate (i.e., "True"); four were inaccurate (i.e., "False"). Half related to objective facts (i.e., they were accurate statements of what had been directly asserted in the communication) and half to inferences which could be drawn from what was stated in the communication regarding the product (or product equivalent).

The procedure involved intercepting potential volunteers in the twelve shopping malls and inviting them to participate in the investigation. If the individual agreed and conformed to pre-established sampling quotas, s/he was invited into the testing room. Shortly thereafter s/he viewed the first of the two communications to which s/he had been randomly assigned and then responded to the miscomprehension quiz and a questionnaire regarding that communication. This was followed by viewing the second test communication assigned and responding to the set of corresponding questionnaires.

Several points should be made regarding the assessment procedures. First, interest focussed specifically and exclusively on miscomprehension and not on any of the prior or subsequent effects usually theorized to result from communication. Thus, the forced exposure procedure -- to insure that exposure, awareness, and attention had actually occurred -- was considered appropriate.

Second, miscomprehension was assessed immediately after exposure to each communication (rather than after exposure to both communications) because of the rapid decay of memory traces found with such communications. As one example, Bogart (1967 p. 109-110) cites research to indicate that fewer than 20% of a group of 5275 respondents could recall the identity of the TV commercial they had viewed just a few minutes earlier.

Third, the procedures involved two sets of counterbalancing. On the one hand, the order of presentation of test stimuli was counterbalanced for main effects due to order. Each communication was presented a total of 90 times -- first in 45 cases and last in the other 45 cases. On the

other hand, so as to eliminate geographic biases, each of the 60 communications was tested at each site. Since each communication was involved in 90 separate tests, no less than 7 nor more than 8 of these tests took place at any one of the 12 testing sites.

Selected Findings

A series of multiple regression analyses were first conducted to determine whether variations in communication complexity,⁴ communication familiarity, product awareness-trial-usage, order of testing, or testing site exerted any appreciable effect on the miscomprehension scores obtained. In none of these analyses did these covariates cumulatively explain more than 8.4% of the variation in the scores. This suggests that miscomprehension is unaffected by variations in any of these factors.

Objective 1: Does Miscomprehension Occur?

The question "Does miscomprehension occur?" can be re-phrased as three separate questions: (a) What proportion of viewers miscomprehend at least some portion of the communications which they viewed? (b) What proportion of the meanings in the test communications were miscomprehended? (c) What proportion of the set of 60 test communications were miscomprehended?

Proportion of Viewers who Miscomprehended: Only 16.8% of the viewers "fully comprehended" (i.e., were able to correctly answer all six quiz items for) either the first or second communication which they viewed. This means that in 83.2% of the 5400 viewings (2,700 viewers X 2 test communications each), viewers miscomprehended at least some portion of the communications which they viewed.

Correctly comprehending the content of one communication did not mean that that person was necessarily able to correctly comprehend another communication. In fact, of the 16.8% who fully comprehended one of the test communications (in the sense of answering all six quiz items correctly); most (13.3%) miscomprehended some portion of the second communication which they saw. That is, only 3.5% of the viewers fully comprehended both test communications to which they were exposed (i.e., answered all 12 items correctly).

Proportions of Meanings Miscomprehended: One may also ask: To what extent was the content of the 60 test communications miscomprehended? The findings reveal that 29.6% of the (2700 respondents x 2 communications per respondent x 6 quiz items per communication =) 32,400 meanings that were tested were answered incorrectly. Both the median and model rates of Overall Miscomprehension were 28%. These figures probably underestimate the actual amount of miscomprehension, since they do not adjust for instances where viewers who did not know the correct answer managed to guess correctly.

Proportion of Communications Miscomprehended: It is noteworthy that at least some degree of miscomprehension was associated with each and every one of the 60 test communications. Across the set of 60 test communications, the level of miscomprehension ranged from a low of 11% to a high of 50%.

⁴ Complexity was assessed via a battery of eight separate indices consisting of a copy point index, assessments of both visual and auditory complexity (see Watt and Krull 1974), and an assessment of the difficulty level of the verbal copy (using the Farr, Jenkins, and Paterson 1951, version of the Flesch count).

In sum, regardless of whether considered in terms of viewers, meanings, or communications, the answer to the first research question is a resounding "yes." Miscomprehension of televised communications does occur and seems to be far more prevalent than might have been anticipated. Perfect comprehension rarely occurs and may not be generally attainable.

Objective 2: What is the Typical Range of Miscomprehension?

The interquartile range (that is, the range encompassed by the middlemost 50% of the test communications) was used to estimate the typical range of miscomprehension.⁵ Considering the entire set of 60 test communications, the interquartile range extended from 23% to 36%. Generalizing from these data suggests that one might expect anywhere from one-quarter to one-third of the material information content contained in communications that are broadcast over commercial television to be miscomprehended.

Objective 3: Is Advertising More Miscomprehended than Non-Advertising?

Program excerpts vs. commercial advertising vs. non-commercial advertising. Based on the sample of 60 test communications, the 22 program excerpts were associated with a significantly⁶ higher rate of miscomprehension (\bar{X} =32.2%) than were either the 25 commercial (\bar{X} =28.3%) or 13 non-commercial (\bar{X} =27.6%) advertisements. However, this difference is quite small and, though statistically significant, may be considered immaterial in practical terms.

Product-service advertising vs. entertainment information program excerpts. The broad three-category classification system has the potential to obscure what is perhaps the most interesting comparison of all, namely, the contrast between product-service advertising on the one hand, and entertainment-information program excerpts on the other. These are the two most frequently broadcast kinds of communication appearing on network television. This investigation examined 22 product-service advertisements and 18 entertainment and information program excerpts. Are these two kinds of televised communication associated with different levels of miscomprehension?

The answer is "yes." Entertainment-information program excerpts were associated with a small, but significantly higher rate of miscomprehension than were product-service advertisements. The mean levels of miscomprehension were 31.0% and 28.8%, and the interquartile ranges extended from 27% to 37% and 22% to 34%, respectively.

Objective 4: Are There Demographic Differences Associated with Miscomprehension?

A number of socio-demographic characteristics were assessed. These included the respondents' sex, age, marital status, employment status, race, years of formal education, and household income for the preceeding year. Only two of these factors -- age and education -- were significantly

⁵The interquartile range was employed in preference to "plus-or-minus one standard deviation from the mean" for two main reasons. First, the distribution of miscomprehension scores was decidedly not normal. Second, the interquartile range lends itself more easily to layman comprehension.

⁶All statements of "significance" refer to being statistically significant at the $P=.01$ level or better. These data have been independently checked and confirmed by our statistical consultant, Professor Jacob Cohen.

related to miscomprehension. Both younger and older viewers were more likely to miscomprehend the communications, and miscomprehension appears to decrease (although not appreciably) as amount of formal education increases. However, while statistically significant, for all intents and purposes, these relationships are so weak as to have little practical applicability.

Discussion

Based upon the findings, six specific conclusions appear justified. First, a large proportion of the American television viewing audience tends to miscomprehend communications broadcast over commercial television. The vast majority (96.5%) of the 2700 respondents in this investigation exhibited some degree of miscomprehension. Second, it would appear that no communication is immune from being miscomprehended. Each and every one of the 60 test communications was miscomprehended at least some of the time by some of the viewers.

Third, the average amount of miscomprehension associated with each of the 60 test communications was 30% (actually 29.61%). In other words, approximately 30% of the relevant informational content contained within each communication was miscomprehended. Fourth, as a preliminary estimate, the typical range of miscomprehension is 29.5% + 6.5%. That is, typical TV communications seem to have their content miscomprehended at anywhere from 23% to 36%.

Fifth, non-advertising communications were associated with higher levels of miscomprehension than were advertising communications. Of particular interest, excerpts of TV programs were miscomprehended at higher levels than were commercial advertisements (for products-brands-services). However, though statistically significant, these differences are practically trivial. Finally, for all practical purposes, no major demographic variables appear to be meaningfully associated with miscomprehension. Miscomprehension seems to be widespread throughout the populace, occurring at all age, income, and education levels in our society and to the same degree.

Selected Implications of Our Findings

Among other things, this investigation reveals that, once again, the mere provision of information does not automatically translate into this information having any effect, much less the intended effect (cf. Jacoby 1974; Jacoby, Chestnut, and Silberman 1977). Rather, the receiver/viewer brings a storehouse of past experience and an ongoing mental set to each communication transaction and tends to interpret, and misinterpret, communications in terms of these mental phenomenon. Given that it is not possible to eradicate either the influence of past experience or the individual's current mental set, it may well be impossible to eradicate miscomprehension. This leads directly to a set of implications for public policy makers, educators, and researchers alike.

Perhaps the most fundamental implication for policy makers is that, just because there is a demonstrable degree of miscomprehension associated with a particular advertisement does not necessarily mean that that particular advertising contains something out of the ordinary to provoke said miscomprehension. Such an assumption may be totally unwarranted. Indeed, the findings of the present investigation revealed that the rates of miscomprehension associated with advertising were significantly lower than the rates associated with typical program content. In other words, a certain proportion of miscomprehension may simply reflect a natural error rate associated with all types of televised communication. The ramifications that this implication has for such regulatory actions as cease and desist orders, corrective advertising, and affirmative

disclosure orders are substantial.

Of course, a reasonable argument is that the consequences of miscomprehending most TV programs tend to be trivial when compared to the consequences of miscomprehending, and then acting upon, advertising. Hence, the rate for advertising should be lower than that for other forms of televised communication. While the authors might be persuaded that advertising miscomprehension rates are probably unacceptably high, they would find completely unreasonable any attempt to use "zero-based miscomprehension" as the criterion for evaluating advertising and as a basis for formulating public policy.

Perhaps the data from the present investigation might serve a useful function in this regard. Regulatory interest in correcting misleading and/or deceptive advertising might use the boundary separating the third and fourth quartile of communications as a triggering mechanism, so that advertising which exceeds this level, while not automatically considered deceptive or misleading, would be targeted for further scrutiny. Of course, any attempt to employ such a cutoff level would have to accommodate the fact that the data adduced here probably underestimate the extent of miscomprehension present under normal viewing conditions. Further, it needs to be recognized that there may be instances (as in the case of over-the-counter and prescription pharmaceutical products) where miscomprehension rates of 10% or even 5% might be deemed totally unacceptable (see Jacoby and Small 1975).

There are parallel implications for advertisers. Given recognition of the fact that some degree of miscomprehension is likely associated with any ad and may, if detected at sufficiently high levels, become the focus of regulatory attention, advertisers might be encouraged to pressure their advertising departments and advertising agencies for lower rates of miscomprehension. Aside from concern over possible regulatory activity, there is a very practical reason for engaging in such research. Namely, by lowering the rate of miscomprehension, the advertiser would insure that more of his advertising message is being accurately comprehended. Perhaps the syndicated advertising evaluation services could be encouraged to devise and provide a "miscomprehension index" along with the other standard communication impact indices which they currently do provide.

Another implication -- one which is consistent with much previous research -- is that broadcast advertising may not be a suitable medium for communicating substantial amounts of product information, particularly complex product information. In terms of the classical hierarchy-of-effects conceptualization (Lavidge and Steiner 1961) and more contemporary models of consumer decision making (e.g., Engel, Blackwell, and Kollat 1978), perhaps the basic function of advertising is simply to stimulate awareness and generate problem recognition. Labelling and other information media would seem to be better vehicles for providing the detailed supporting information.

Conclusion

Some will no doubt quarrel with the procedural details of this investigation or with the specific findings obtained arguing, as examples, that the 30% miscomprehension rate is too high to be correct, or that commercial advertising is more readily miscomprehended than is TV programming. It may well be that these arguments are correct. No claim is made of providing definitive answers to the research questions which stimulated this investigation. The authors merely claim that the findings represent preliminary ballpark estimates which are in need of further confirmation and refinement. However, the fact remains that, even with a procedure -- involving unclutter forced

exposure and volunteer participants who were likely more attentive to the test communications than they normally would otherwise have been -- which should have reduced the rate of miscomprehension, nearly every single one of the 2700 respondents managed to miscomprehend at least some portion of the communications which they viewed, each and every one of the 60 test communications was miscomprehended, and the average amount of miscomprehension associated with any given communication was an unexpectedly high 30%. Further, these findings were quite robust, holding true for respondents of different ages, incomes, education levels, sexes, and marital status.

The implications of these findings are substantial. Perhaps foremost among these is that to require advertising -- or any other comparable form of communication, for that matter -- to conform to a "zero base" miscomprehension rate appears totally unreasonable. Further, to contend that advertising is necessarily more prone to being miscomprehended than are other forms of televised communication is similarly unreasonable. If nothing else, the present findings strongly suggest the need for a programmatic series of research investigations designed to shed considerably greater light on the various questions surrounding viewer miscomprehension of televised communications.

References

- Bogart, L. (1967), Strategy in Advertising. New York: Harcourt Brace, Chapter 5.
- Engel, J. F., Blackwell, R. D., & Kollat, D. T. (1978), Consumer Behavior (3rd ed.). Hinsdale, Illinois: Dryden Press.
- Farr, J. N., Jenkins, J. J., & Paterson, D. G. (1951), "Simplification of Flesch Reading Ease Formula." Journal of Applied Psychology, 35(5), 333-337.
- Harris, R. J. and Monaco, G. E. (1978), "Psychology of Pragmatic Implication: Information Processing Between the Lines." Journal of Experimental Psychology: General, 107(1), 1-22.
- Jacoby, J. (1974), "Consumer Reaction to Information Displays: Packaging and Advertising." In S. R. Divita (Ed.), Advertising and the Public Interest. Chicago: American Marketing Association, pp. 101-118.
- _____, Chestnut, R. W., and Silberman, W. (1977), "Consumer Use and Comprehension of Nutrition Information." Journal of Consumer Research, 4(2), 119-128.
- _____, Hoyer, W. D., & Sheluga, D. A. (1980), The Miscomprehension of Televised Communication. N. Y. American Association of Advertising Agencies.
- _____, & Small, C. B. (1975), "The FDA Approach to Defining Misleading Advertising." Journal of Marketing, 39(4), 65-68.
- Lavidge, R. J. & Steiner, G. A. (1961), "A Model for Predictive Measurements of Advertising Effectiveness." Journal of Marketing, 25(4), 59-62.
- Preston, I. L. & Scharbach, S. E. (1971), "Advertising: More Than Meets the Eye?" Journal of Advertising Research, 11(3), 19-24.
- Watt, J. & Krull, R. (1974), "An Information Theory Measure for Television Programming." Communication Research, 1(1), 44-68.

IMPACT OF PUBLICITY ON CORRECTIVE
ADVERTISING EFFECTS

Kenneth L. Bernhardt, Georgia State University
Thomas C. Kinnear, University of Michigan
Michael B. Mazis, American University
Bonnie B. Reece, University of Michigan

Abstract

Previous research on corrective advertising has generally involved a forced exposure in a laboratory setting using student subjects. Treatment and criterion variables have also usually been developed by the researcher independent of the real interests of regulators. This paper presents an overview of a national field study of the STP corrective advertising campaign, as it was actually undertaken by the Federal Trade Commission. The details of this study are currently under review for publication elsewhere.

Background

There have been several important corrective advertising studies conducted in a laboratory setting, but frequently subjects have viewed ads under conditions of forced exposure thereby limiting generalizability (Armstrong, Gurol and Russ 1979, Dyer and Kuehl 1974, 1976a, 1976b, 1978, Gurol 1977, Hunt 1972, 1973, Kassarian et. al. 1975). In addition, while most of the corrective campaigns have appeared in the print medium, many of the studies have concerned broadcast advertising for Listerine (Armstrong et. al. 1979, Dyer and Kuehl 1978, Gurol 1977, Mazis and Adkinson 1978, Mizerski et. al. 1979, Sawyer and Seminick 1978). Further almost all reported studies have used students as subjects. The impact on actual consumers in the marketplace is essentially unresearched.

The current research is the first to examine corrective advertising as it was actually undertaken by the Federal Trade Commission in the field. It is designed to measure, though a "before and after" design (Campbell and Stanley 1963), the effectiveness of the 1978 STP corrective advertising campaign which appeared in a number of periodicals and received substantial publicity in the news media.

In February 1978, the STP Corporation signed a consent decree under which it agreed to discontinue certain advertising claims about its motor oil additive and to make a \$700,000 settlement, including \$200,000 to place a "public notice" advertisement in 14 periodicals.

During February 5-9, 1978, several days before initiation of the "public notice" campaign, 823 persons who were at least 18 years of age, who were regular readers (read three of every four issues) of one of the target periodicals were interviewed on the telephone. While a stratified random sampling procedure was used to contact potential interviewees (as described below), a quota procedure was used to obtain a 60:40 ratio between "general public" respondents and respondents with "business related occupations."

Subsequent to these interviews, the "public notice" advertisement was placed as follows:

February 10 - Wall Street Journal
February 10 - New York Times (Financial Section)
February 13 - Washington Post
February 13 - Barron's

February 20 - Newsweek
February 20 - U.S. News & World Report
February 20 - People
February 20 - Business Week

The advertisement also appeared in the Time Business during February. However, since it was impossible to determine which Time readers received the business edition, Time was excluded from the list of relevant publications. Additionally, during March and April the advertisement appeared in Esquire, Forbes, Guns and Ammo, Harvard Business Review, and National Geographic. Since these placements appeared after Wave II was completed, readership of these periodicals was not relevant for inclusion in the samples. A second wave of 845 interviews with a separate group of respondents was conducted from February 27 through March 2 using the identical procedure followed for the first wave. Respondents were contacted shortly after initiation of the campaign in order to reduce the impact of "history effects" as a source of invalidity and to more accurately measure consumer perceptions of the advertisement. As such, the study focuses solely on short-term effects generated by the campaign and publicity. Additionally, a special sample of 197 advertising industry managers, selected from a proprietary list, was interviewed during Wave II only.

The sample, based on the Louis Harris national sample selection procedure, was selected based upon the national population of the United States, excluding Alaska and Hawaii, and those in prisons, hospitals or religious and educational institutions. Figures are updated annually by intercensal estimates provided by the Bureau of Census, with sample locations selected biannually to reflect demographic changes.

Results

The results of the campaign were measured on a number of variables: knowledge of problems about STP advertising, intention to purchase STP, attitudes toward the STP Corporation, and credibility of oil additive claims in general. Additional recall measures about the campaign were taken on Wave II participants.

The campaign and the attendant publicity had a marked effect on awareness of a problem with STP advertising, and resulted in a decrease level of intention to purchase STP. The effects of the campaign did not spill over into oil additive claims in general, nor did it negatively impact attitudes held about the STP Corporation. The effects appear to be very specific to STP.

Examination of recall results raises questions as to how these effects occurred. Both unaided and aided recall levels were extremely low. How then could the observed effects have occurred? It is our speculation that the substantial attendant publicity for the campaign (features on network television news for example) must be a significant contributor to the observed results. Unfortunately, no evidence is available to support or refute this conclusion.

References

- Armstrong, Gary M., Gurol, Metin N. and Russ, Frederick A. (1979), "Detecting and Correcting Deceptive Advertising," Journal of Consumer Research, 6 (December), 237-246.
- Campbell, Donald T. and Stanley, Julian C.(1963), "Experimental and Quasi-Experimental Designs for Research on Teaching," in Handbook of Research on Teaching, N. L. Gage, ed., Washington: American Educational Research Association, 171-246.
- Chadduck, H. W. (1972), "In Brief Summary: Prescription Drug Advertising, 1962-71," FDA Papers (February).
- Dyer, Robert F. and Kuehl, Philip G.(1974), "The Corrective Advertising Remedy of the FTC: An Experimental Approach," Journal of Marketing, 33 (January), 40-44.
- _____ (1976a), "Applications of the 'Normative Belief' Technique for Measuring the Effectiveness of Deceptive and Corrective Advertisements," in Advances in Consumer Research, Volume IV, W. D. Perreault, ed., Atlanta: Association for Consumer Research, 204-209.
- _____ (1976b), "Brand Belief Measures in Deceptive-Corrective Advertising: An Experimental Assessment," in Marketing: 1776-1976 and Beyond, K. L. Bernhardt, ed., Chicago: American Marketing Association, 373-379.
- _____ (1978), "A Longitudinal Study of Corrective Advertising," Journal of Marketing Research, 15 (February), 39-48.
- Gurol, Metin Nazmi (1977), "Deception in Advertising: A Review of Past and Current FTC Practice and an Experimental Evaluation of a New Approach for Detecting Deception and a New Approach for Eradicating its Effects," unpublished Ph.D. dissertation, University of North Carolina at Chapel Hill.
- Hauck, Mathew and Cox, Michael (1974), "Locating a Sample by Random Digit Dialing," Public Opinion Quarterly, 38 (Summer), 253-260.
- Hunt, H. Keith (1972), "Source Effects, Message Effects, and General Effects in Counteradvertising," in Proceedings, M. Venkatesan, ed., Iowa City, IA: Association for Consumer Research, 370-381.
- _____ (1973), "Effects of Corrective Advertising," Journal of Advertising Research, 13 (October), 15-24.
- Kassarjian, Harold H., Carlson, Cynthia J. and Rusin, Paula E. (1975), "A Corrective Advertising Study," in Advances in Consumer Research, Volume II, Ann Arbor, MI: Association for Consumer Research, 631-642.
- Landon, E. Laird and Banks, Sharon K.(1977), "Relative Efficiency and Bias of Plus-one Telephone Sampling," Journal of Marketing Research, 14 (August), 294-299.
- Mazis, Michael B. and Adkinson, Janice E.(1976), "An Experimental Evaluation of a Proposed Corrective Advertising Remedy," Journal of Marketing Research, 13 (May), 178-183.
- Mizerski, Richard W., Allison, Neil K. and Calvert, Stephen (1979), "A Controlled Field Study of Corrective Advertising Using Radio," unpublished working paper, University of Cincinnati, Department of Marketing.
- Sawyer, Alan G. and Semenik, Richard J.(1978), "Carryover Effects of Corrective Advertising," in Advances in Consumer Research, Volume V., H. Keith Hunt, ed., Ann Arbor, MI: Association for Consumer Research, 343-351.

CORRECTING CORRECTIVE ADVERTISING¹

Jacob Jacoby, Purdue University
Margaret C. Nelson, Purdue University
Wayne D. Hoyer, Purdue University

Abstract

Prior discussion and research on remedial (i.e., affirmative disclosure and corrective advertising) statements has addressed the impact of such statements on beliefs, attitudes, and purchase intentions. It has been implicitly assumed that such statements are satisfactorily comprehended by the general public. The present investigation tested this assumption using three heterogeneous samples (n=451) and six remedial statements proposed by the Federal Trade Commission. The findings revealed that remedial advertising statements may be miscomprehended as much as -- or even more often than -- the advertising they are supposed to remedy.

Introduction

The subject of deceptive-misleading advertising has experienced increasing regulatory and academic interest over the past decade. The remedies that can be employed in the event an ad (or ad campaign) is judged to be deceptive-misleading have also been the focus of considerable discussion. Two of the more controversial of these remedies are "corrective advertising" and "affirmative disclosure."

Corrective advertising is advertising mandated by regulators which is designed to modify an erroneous belief purportedly caused by previous advertising for the same product, brand, or service. Affirmative disclosure refers to the requirement that an advertiser provide new or additional information so that the consumer will be better able to make a more informed choice. The intent of the regulators is the same in either case: to somehow affect the consumer's belief structure regarding the advertised product, brand or service. In the case of corrective advertising, this means changing or "correcting" a purportedly erroneous belief. In the case of affirmative disclosure, this means generating a new belief. In both cases, the implicit objectives are to make consumers better decision makers and to redress sales gains as a result of what are considered to be unfair advertising practices.

Early empirical research focused on the impact that corrective advertising had on such factors as: brand awareness; attitudes toward the brand, the company (i.e., company image), and the advertisement; and purchase intentions. Later research zeroed in on the impact that such remedial advertising had on consumer beliefs.

As is readily recognized, awareness, beliefs, attitudes, and intentions are among the dependent variables traditionally employed to assess communication impact. These variables are usually major components in most widely accepted conceptualizations of the communication process. A consideration of these models -- from the hierarchy-of-effects model (Lavidge and Steiner 1961) through more contemporary conceptualizations (e.g., McGuire 1976;

Engel, Blackwell and Kollat 1978, especially Chapters 1 and 13) -- reveals in every case that some degree of comprehension is assumed to precede belief formation and change.

Hence, regulator efforts to use affirmative disclosure and corrective advertising statements to change an existing belief or to provide a new belief, and researcher efforts to assess the impact of such advertising, necessarily rest on a critical assumption -- that the logically prior step of consumer comprehension of the message has been accomplished. This assumption is as yet completely unverified. It stands to reason that before a corrective advertising or affirmative disclosure statement can exert the intended impact upon consumer beliefs, it must first be understood by the consumers toward whom it is targeted. If the meaning contained in the remedial message has not been satisfactorily comprehended, then the logically subsequent step of belief formation or change cannot be achieved.

In other words, there exists the potential for remedial statements to be at least as confusing and misleading as the original advertising messages which they were designed to counteract. The present investigation examined this possibility using the actual language proposed by the Federal Trade Commission (FTC) for implementation in a case which, as of the present writing, is still under active judicial consideration.

This case revolves around two familiar analgesic products manufactured by the Bristol-Myers Company, namely, Bufferin and Excedrin. The FTC contends that an integral part of Bufferin advertising has been the claim that Bufferin is a faster pain reliever and gentler to the stomach than plain aspirin. Similarly, it is asserted that Excedrin advertising contains the claim that the product is a more effective pain reliever than is plain aspirin. The FTC further contends that a significant proportion of American consumers now hold these beliefs as a result of exposure to television advertising.

The problem, according to the FTC staff, is that the degree of empirical proof available to support the claims is not as conclusive as the FTC staff believes it should be. Hence, the FTC staff would like consumers to know that, while a certain amount of proof is available to show that the claims may be true, the degree of proof is insufficient according to some authorities.

With the expectation that the FTC Administrative Law Judge hearing the case would rule in its favor, the FTC Complaint Counsel sent a letter to the Judge. The letter provided two sets of remedial statements, one for each product, and proposed that one of the statements in each set be mandated for the respective products.

"For inclusion in Excedrin advertising, the disclosure that:

1. Excedrin has not been proven to be a more effective pain reliever than aspirin;
- or; 2. It is not known whether Excedrin is a more effective pain reliever than aspirin;

¹This report is a highly condensed version of a longer manuscript (Jacoby, Nelson, and Hoyer, submitted).

- or; 3. There is a real question whether Excedrin is a more effective pain reliever than aspirin.
- For inclusion in Bufferin advertising, the disclosure that:
4. Bufferin has not been proven to be a faster pain reliever or gentler to the stomach than aspirin;
- or; 5. It is not known whether Bufferin is a faster pain reliever or gentler to the stomach than aspirin;
- or; 6. There is a real question whether Bufferin is a faster pain reliever or gentler to the stomach than aspirin."²

In serving as an expert witness for the defendant, the senior author expressed several opinions regarding these proposed remedial statements. Principal among these was the belief that a substantial number of consumers would either be confused as to their intended meaning, or would extract some non-intended meaning. The investigation described below is a direct outgrowth of this testimony and represents an empirical attempt to confirm (or disconfirm) these opinions.

Method

Instrument

Comprehension was assessed via a closed-ended questionnaire. A set of 8 "derived meanings" was generated for each of the six proposed remedial statements. For each such set, one (or two) of the derivations were judged roughly equivalent in meaning to the corrective statement under consideration; the other six (or seven) derivations were judged to be incorrect. Figure 1 provides an example for one of the six proposed statements. Each of the remedial statements, followed by its set of derivations, constituted a separate page in a questionnaire. The opening instructions requested that the subjects read the statements typed at the top of each page, then check as many or as few of the interpretations as they felt accurately paraphrased the meaning which they had derived. These pages were randomized across statements and respondents. In addition to the set of eight derivations, each page contained a ninth response option for "none of the choices presented above reflect the meaning I get from this statement" and a tenth option for "I am uncertain as to what the statement actually means and therefore cannot check any of the above meanings as being either correct or incorrect."

FIGURE 1

"Excedrin Has Not Been Proven To Be A More Effective Pain Reliever Than Aspirin."

1. Excedrin is a more effective pain reliever than aspirin.
2. Excedrin is not a more effective pain reliever than aspirin.
3. Excedrin is neither more nor less an effective pain reliever than aspirin.

² Fisherow, W. Benjamin. Complaint Counsel's statement concerning corrective advertising for Excedrin and Bufferin. Letter dated January 31, 1979 and sent to Montgomery K. Hyun, Administrative Law Judge. In the matter of: Bristol-Myers Company, Ted Bates and Company, Inc., and Young and Rubicam, Inc. Docket No. 8917, Federal Trade Commission.

4. Excedrin is a less effective pain reliever than aspirin.
5. Excedrin may be more effective than aspirin, but no proof whatsoever exists for this statement.
- *6. Excedrin may be more effective than aspirin, but no conclusive proof exists for this statement.
7. Excedrin may be more effective than aspirin, but no studies have yet been done which would show this to be true.
- *8. Excedrin may be more effective than aspirin, but there is at present insufficient evidence to show that this is true.
9. None of the choices presented above (1 through 8) reflect the meaning I get from this statement.
10. I am uncertain as to what the statement actually means and therefore cannot check any of the above as being either correct or incorrect.

For any given page, an individual's response was scored "correct" if he checked the one accurate derivation in the case of Bufferin, or one of the two or both accurate derivations in the case of Excedrin, and marked no other response options. An answer was scored "incorrect" either if the individual checked any of the inaccurate derivations on the page and failed to also check an accurate response, or if the individual marked the ninth response option ("None of the choices . . ."). A response was scored "confused" either if the individual marked a correct derivation as well as one or more incorrect derivations, or if the individual chose the tenth response option ("I am uncertain . . .").

Samples

Three separate samples were employed. The first consisted of 92 Purdue University undergraduates enrolled in an introductory course on consumer behavior. The second consisted of 281 members of social and fraternal organizations (Breakfast Optimists; Rotary Club; League of Women Voters, and Parent Teachers Association) in Lafayette, Indiana. The third sample consisted of 78 middle-class adult blacks living in south suburban Chicago. All subjects were volunteers and were unaware of the purposes of the investigation.

Results

Table 1 summarizes the results obtained with each of the samples. As can be seen from these data, extracting either an incorrect or confused meaning was much more prevalent than extracting the intended meaning. In other words, despite the considerable amount of thought and well-intentioned effort which underlie the development of these messages, remedial statements have the potential to generate at least as much miscomprehension as the advertising messages they were designed to counteract. However, if one were forced to make a choice from among the three versions proposed, it is clear that the "There is a real question. . ." version is best understood and least confusing.

*Note: Options 6 and 8 were judged to be accurate paraphrases of the meaning structure intended by the remedial statement.

TABLE 1

Table 1. Proportion of respondents who extracted correct, incorrect, or confused meanings from the proposed remedial statements.

	Remedial Statements					
	Excedrin			Bufferin		
	1	2	3	4	5	6
Correct						
Sample 1	.21	.23	.46	.13	.15	.33
Sample 2	.25	.17	.40	.17	.13	.25
Sample 3	.19	.15	.48	.15	.23	.23
Weighted mean	.23	.18	.43	.16	.15	.26
Incorrect						
Sample 1	.26	.20	.06	.33	.33	.33
Sample 2	.32	.40	.13	.56	.52	.41
Sample 3	.35	.33	.20	.58	.46	.35
Weighted mean	.31	.35	.13	.52	.47	.39
Confused						
Sample 1	.53	.57	.48	.54	.52	.35
Sample 2	.43	.43	.47	.27	.35	.33
Sample 3	.46	.52	.32	.27	.31	.42
Weighted mean	.46	.47	.44	.32	.38	.35

Sample 1 = 92 Purdue University undergraduates
 Sample 2 = 281 post-college age adults residing in Lafayette, Indiana.
 Sample 3 = 78 black adults residing in S. Sub. Chicago
 Total N = 451

Discussion

The principal objective of the present investigation was to examine the question: Are corrective advertising or affirmative disclosure statements as subject to being miscomprehended as are the advertising messages they are designed to remedy? From the data obtained, the answer would appear to be a resounding "yes". The six remedial statements proposed by the FTC staff for inclusion in Bufferin and Excedrin advertising appear to be more often associated with confused or incorrect comprehension than with correct comprehension. In other words, the implicit assumption that remedial advertising statements will automatically be accurately comprehended by the public at large seems both unwarranted and untenable.

The difficulty involved in accurately communicating meaning is often underestimated, and regulators would seem to be no exception in this regard. As Rumelhart (1977, p. 102) notes: "Of all the complex tasks the human information system carries out, perhaps the most sophisticated is language processing . . . Only when we stop and analyze the intricate reasoning that appears to underlie even the simplest cases of language comprehension can we begin to appreciate the complexity of the language understanding process."

Given that the receivers of communication messages differ substantially in terms of their past experiences and present "semantic expectations" (see Rumelhart 1977, p. 148), it is not surprising to find substantial levels of miscomprehension. The fact that high levels of miscomprehension were found for remedial statements generated and proposed by the FTC suggests caution in being quick to blame the advertiser when some consumers miscomprehend his message.

It could be argued, of course, that the relatively high rates of confusion obtained (that is, extracting the correct meaning along with some additional, but incorrect meaning) should not be of major concern since they suggest that at least some of the correct meaning was being received. However, consider Cohen's (1974, p. 9) description of the criteria used by the FTC to arrive at a finding of deception in advertising:

On the basis of FTC cases and court interpretations, an advertisement will be considered false, misleading, or deceptive when the following [applies]:

.....

An advertisement that is capable of two meanings, one of which is false, is misleading.

One might therefore reasonably ask: Should not the same criteria apply to remedial statements as well?

Perhaps the most important implication of the present investigation is that, just as is the case for advertising proper, the likely impact of remedial statements needs to be thoroughly researched before such statements are inserted in advertising. Regulatory agencies need to recognize that the language of the remedial statements which they devise may be equally (and in some cases even more) confusing to the consumer than the advertising message it is designed to correct. It cannot be safely assumed that language decided upon in good faith and with the consumer's best interests at heart will be sufficient for accomplishing its objective. There is no substitute for empirically assessing the perceived meanings of corrective advertising or affirmative disclosure statements before such remedial statements are mandated for use in advertising.

References

- Cohen, D. (1974), "The concept of unfairness as it relates to advertising legislation." *Journal of Marketing*, 38, 8-13.
- Engel, J. F., Blackwell, R. D., & Kollat, D. T. (1978), *Consumer Behavior* (3rd ed.), Hinsdale, Illinois: The Dryden Press.
- Jacoby, J., Nelson, M. C., & Hoyer, W. D. "Corrective advertising and affirmative disclosure statements: Their potential for confusing and misleading the consumer." Submitted.
- Lavidge, R. J. and Steiner, G. A. (1961), "A model of predictive measurements of advertising effectiveness," *Journal of Marketing*, 25, 59-62.
- McGuire, W. J. (1976), "Internal psychological factors influencing consumer choice." *Journal of Consumer Research*, 2 (4), 302-319.
- Rumelhart, D. E. (1977), *Introduction to Human Information Processing*, New York: J. Wiley and Sons.

THE EFFECTS OF RECALL ON BELIEF CHANGE:
THE CORRECTIVE ADVERTISING CASE

Neil K. Allison, University of South Carolina,¹
Richard W. Mizerski, Federal Trade Commission

Abstract

This study reports the association of product or brand recall with various beliefs consumers hold about a product subject to the corrective remedy. Respondents recalling exposure to a test product's corrective ad exhibited belief change for only the target belief of the corrective message. Those subjects who were unable to recall exposure had overall beliefs similar to a control group not exposed to corrective advertising.

Introduction

The 1970's have seen the Federal Trade Commission (FTC) expand its role in consumer protection. An area of FTC jurisdiction concerns advertising that the Commission has defined to be false or misleading. This type of advertising is considered to be an unfair method of competition for which the Commission is able to enforce several remedies, the most severe and controversial being an order for corrective advertising. Such a ruling not only requires the advertiser to stop the false or deceptive communication but, in addition, to engage in a corrective advertising campaign that seeks to "... eliminate the residual effects of inaccurate or misleading information" (Wilkes and Wilcox 1974, p. 58) presented to the consumer in the previous advertising.

The most important action involving corrective advertising, in terms of both the size of the settlement and the legal precedent, was levied against Warner-Lambert's Listerine Antiseptic Mouthwash. The case against Warner-Lambert began in 1971 with an FTC complaint that Listerine ads falsely promoted the product as an "aid in preventing and ameliorating colds," because medical evidence failed to support this claim. Unlike previous corrective advertising cases, Warner-Lambert refused to sign a consent decree. In response, the FTC ordered Warner-Lambert to judicial review. This action set the legal testing ground for the FTC's power to issue and enforce the corrective advertising remedy. The order was upheld with slight modifications by the Court of Appeals for the District of Columbia in 1977, and in 1978 the U.S. Supreme Court refused to hear the case, thus establishing the legality of corrective advertising.

The modified corrective advertising order required Warner-Lambert to include the corrective message in all Listerine advertising until it had spent \$10 million, an amount equal to the average annual advertising expenditures for the previous ten-year period. The content of the message was limited to "Listerine does not cure or prevent colds, or lessen their severity." The purpose of this and all other corrective messages was to change only consumers' falsely formed beliefs about the product and not to favorably or unfavorably affect other beliefs about the product, the manufacturer, or the competition. This additional impact would be considered punitive and beyond the legal bounds of the FTC.

¹The authors would like to acknowledge the financial assistance of the University of Cincinnati, College of Business, for support of this research.

Since corrective advertising was first proposed in 1969 as a remedy for false or deceptive advertising, the FTC, and the courts in the Listerine case, have been continually developing guidelines for its use and implementation. The following are a sample of significant decisions that have had an effect on the corrective advertising rule as it is presently enforced (Maddox and Zanot 1979).

-Corrective advertising can be justified in situations where there are residual effects of an advertising campaign that is no longer running (Firestone Tire and Rubber Company Case).

-Corrective advertising must at least reach the same target with similar impact as the original deceptive advertising (Sugar Association Case).

-Corrective advertising must appear in the same media and in the same manner as the original deceptive advertising (Lens Craft Research and Development Company Case).

-Corrective advertising need not include the phrase "contrary to past advertising" (Listerine Case).

-Background sounds or music during the corrective disclosure are not allowed (Listerine Case).

The evolution of the application of corrective advertising still leaves many issues unanswered. In general, the FTC lacks objective criteria for its decision making in the corrective advertising area. For example, expenditures for a corrective campaign have been mandated in different cases to be either 25% of all media expenditures, one in four ads for corrective advertising, or the \$10 million figure in the Listerine case. Another issue that is a concern of both industry and government involves the impact of corrective advertising on the consumer. There have been a number of empirical studies with conflicting findings that address this issue (Hunt 1973a, 1973b, Dyer and Kuehl 1978, Mazis and Adkinson 1974). A recent study (Mizerski, Allison and Calvert 1980) concluded that multiple exposure to an actual corrective radio commercial had a significant influence on lowering the subjects' target belief for which that remedy was developed. At this point, it appears that corrective advertising, as it is presently implemented, is capable of removing a false belief without harming the offending advertiser or its competition.

As with any persuasive communication, standards for establishing the effectiveness of corrective advertising must be defined. A common industry technique for advertising effectiveness is recall testing. This includes aided and unaided recall performed on a day-after-exposure basis or with a more substantial time delay. (For a discussion of recall techniques, see Darrell Blaine Lucas and Stuart Henderson Britt, Measuring Advertising Effectiveness, Chapter 4). There have been many questions raised as to the validity of recall testing as a measure of advertising effectiveness. There is evidence that, over time, advertising may induce attitude change even though the specifics of the verbal content of the message itself are not recalled (Percy 1978). If this phenomena is true, then a measure of general advertising recall may be a necessary and sufficient criteria for establishing the persuasive impact of a communication, rather than a measure of

detailed information recall. This study investigates the value of a measure of general advertising recall for establishing the effectiveness of corrective advertising.

Based on previous studies of corrective advertising and research on advertising recall, the following hypotheses are proposed:

1. Experimental subjects who recall hearing a corrective advertisement will have a weaker belief that Listerine "fights colds and sore throats" (target belief) than either experimental subjects who do not recall exposure to a corrective advertisement or control subjects.
2. Experimental subjects who recall hearing a corrective advertisement will not have different levels of beliefs concerning non-target attributes of Listerine than either experimental subjects who do not recall exposure to a corrective advertisement or control subjects.
3. Experimental subjects who recall hearing a corrective advertisement will not have a different belief concerning the honesty of Listerine's manufacturer than either experimental subjects who do not recall exposure to a corrective advertisement or control subjects.
4. Experimental subjects who recall hearing a corrective advertisement will not have a different affect/attitude toward Listerine than either experimental subjects who do not recall exposure to a corrective advertisement or control subjects.

Methodology²

Sampling Plan

In order to more realistically investigate the effects of corrective advertising on consumers' cognitive structure than in previous studies (Hunt 1973a, 1973b, Dyer and Kuehl 1978, Mazis and Adkinson 1974) the natural setting of an on-campus radio station was used as the medium of the corrective advertising message. The station reaches only dormitory residents of a large mid-Western university by direct transmission via carrier current and does not broadcast to the general public, thus allowing for a controlled field experimental environment. The programming of the station follows the format of a typical contemporary rock commercial station with music, disc jockeys, news, commercials, games and give aways.

A systematic random sampling plan from a sample frame of all dormitory residents was used to make initial contact with students. Potential respondents were telephoned and asked to participate in a study to help improve the programming and format of the on-campus radio station. This guise for the corrective advertising experiment was supported by an article in the campus newspaper discussing the need for a radio station survey that was to be conducted by university personnel.

All potential subjects were first asked whether or not they were able to receive the station on the radio in their own dormitory room. Because of the nature of the carrier current transmission and how it is affected by a building's structural design, individuals next to each other in the same dorm differed in ability to receive the station. Those students who reported they were able to receive the station were considered potential experimental group subjects and those who could not were considered potential

control group subjects.

Potential experimental subjects were informed that participation would require four one-half hour listening periods over the course of a week. In order to increase participation, they were to be paid \$5 for their efforts. As a manipulation check, during each of the four experimental sessions, the disc jockey announced a "listener survey password" that had to be presented to the researchers at the end of the study in order for the subject to receive his/her pay. To make participation in the survey even more attractive, students were given a chance to win their choice of record albums in the course of the four listening sessions. Twice during each half-hour period a student's social security number was randomly drawn and read out over the air. The respondent had five minutes to telephone the radio station and claim his/her prize, thus providing an additional manipulation check. After attrition due to extensive missing data, response bias or incomplete participation, there were 49 experimental subjects available for data analysis.

Those students who could not receive the radio station and who agreed to participate were told they would need to fill out a questionnaire concerning the on-campus radio station. The four one-half hour test sessions were not mentioned to this group of control subjects. For their participation, they received 50¢ coupons to a local fast-food restaurant. There were 51 control group subjects after removing those with obvious response bias or extensive missing data. Upon conclusion of the entire study, debriefing sessions showed that no control subjects were exposed to the experimental treatments.

Experimental Treatments and Measurements

The study employed a Solomon Four-group experimental design incorporating a pre-test measurement for half of the experimental subjects and half of the control subjects. No data collected from the pre-test will be used in the analyses presented in this paper and, therefore, it will not be discussed.

The experimental subjects were required to listen to the on-campus station in their own dorm rooms for at least the four designated one-half hour periods that ran from Monday to Thursday of one week. They were requested to listen and behave during these listening periods as they normally do while listening to the radio. This was done to prevent an inordinate amount of attention on involvement in the radio broadcast, thus keeping the deception and the subjects' behavior as realistic as possible.

The corrective ad was a professionally prepared adaptation from a current Listerine TV commercial that was neither corrective nor deceptive in nature. Although the theme of this ad was not being used on radio at that time, it was eventually extended to this medium after the experiment was completed. The actual Listerine corrective advertising campaign was not aired until data collection had been completed. The corrective message was placed near the middle of three of the four half-hour test periods. This was done to help disguise the true nature of the study. Commercials for other products and services were also aired during the test sessions.

Upon completion of the experimental treatments, the subjects were requested to go to the radio station's office to fill out a questionnaire concerning their feelings and attitudes toward their listening experiences. All respondents were required to report to the station within 11 to 13 days of their last experimental exposure. This spacing of the exposure from the data gathering was done to reduce demand and to lower sensitization to the experimental treatment as well as provide a more realistic measure of advertising recall.

²The data reported in this study are part of a more detailed investigation into the effect of corrective advertising. For a more complete discussion of the methodology see Mizerski, Allison and Calvert 1980.

The post-test questionnaire consisted of demographic data, opinions and attitudes toward the radio station's programming, and attitudes and beliefs toward commercial products. There were twenty-four belief statements, four each for a brand of beer, a local audio store, the University bookstore, a brand of toothpaste, a brand of soap, and Listerine. All belief statements were presented with a seven-point horizontal scale ranging from how "unlikely" to "likely" the belief statement represented the respondent's feelings about the product or service. The belief statements for Listerine were:

1. Listerine is effective for killing germs.
2. Listerine leaves mouth feeling refreshed.
3. Listerine fights colds and sore throats (target belief of the corrective advertisement).
4. Listerine has long lasting effects.

Perceptions of honesty for Listerine and the other product/service manufacturers were measured on a seven-point horizontal belief statement ranging from "likely" to "unlikely" while the seven-point affect/attitude measure for each product/service had scale labels of "like" to "dislike." Purchase intention as based on a 0% to 100% probability the respondent would buy the particular brand of product/service the next time the need to purchase that type of product/service arose.

Upon completion of the questionnaire, each experimental subject was interviewed by one of the investigators. They were first asked for any general comments and suggestions about the radio station. Unaided recall for all commercials and advertisements was then elicited followed by aided recall of any ads not mentioned under unaided recall. Finally, the subjects were asked what they thought to be the true nature and intent of the study. All respondents thought the sole intention was a programming study for the campus radio station without any underlying commercial or other research purpose.

Control subjects received a questionnaire similar to the experimental group questionnaire in both content and structure. All questions pertaining to product/service and demographic information remained the same. Information elicited about the radio station was phrased in terms of their "favorite local radio station."

Upon completion of the data collection, a de-briefing letter was sent to all experimental and control subjects. They were informed of the true nature of the study and that the Listerine ads were test ads, although the corrective statement would be similar once the company began its corrective advertising campaign.

Analyses and Results

Based on the experimental subjects' response to the Listerine ad during the exit interview, respondents were classified as either no recall (N=36), unaided recall (N=11), or aided recall (N=2). For purpose of analysis, unaided and aided recall subjects were grouped together into one recall category (N=13). No subjects who reported recall of the Listerine ad under the aided or unaided condition were able to remember any specific corrective or non-corrective information from the advertisement thus providing a measure of only source or general recall. Due to the self-selective nature of the recall variable, a series of chi-square analyses were computed in order to check for any inherent bias between people who could recall the Listerine ad and those who could not recall the ad. No differences between recall groups were found for product usage (three levels:

nonuser, regular Listerine user, and regular other brand user), sex, class standing, age, ethnic origin or experimental grouping (pre-test vs. no pre-test group of the Solomon Four group design). These data suggest that there was no previous bias in the two groups that would affect their processing of a corrective advertisement (Table I).

TABLE I
Chi Square Analyses

Variable	χ^2	p.
Product Usage	0.77	0.68
Sex	0.00	1.00
Class Standing	2.49	0.65
Age	2.25	0.69
Ethnic Origin	0.01	0.96
Experimental Grouping	2.25	0.14

TABLE II
Analyses of Variance for Belief Variables

Variable	Group Means			F	p.
	Recall	No Recall	Control		
Belief 1 - Kills germs	5.17	4.69	5.10	.747	.48
Belief 2 - Leaves mouth refreshed	5.67	4.83	4.66	1.75	.18
Belief 3 - Fights colds and sore throats	2.29	3.86	4.61	4.17	.02
Belief 4 - Long lasting effects	4.67	4.22	4.44	.412	.66
Honesty of Manufacturer	4.83	4.77	4.50	.49	.61
Overall effect/attitude	4.75	4.00	4.00	.90	.41

Weighted one-way analyses of variance to adjust for unequal cell sizes were computed to test all hypotheses (Winer 1971, Nie, et. al, 1975). These data are presented in Table II. The only variable for which significant differences existed between recall, no recall, and control subjects was the target belief of "fights colds and sore throats," thus supporting all four hypotheses. In order to investigate whether significant differences existed between groups, a Scheffe test was performed on the third belief variable. The Scheffe test, which is an appropriate statistic to use for unequal cell sizes, is a strict or conservative post-hoc test. In order to report statistical significance, differences have to be rather substantial (Nie, et. al. 1975, Kerlinger 1973). At the $P < .05$ alpha level, the Scheffe procedure found the recall group subjects to be different from both the no recall and the control group subjects, and the no recall group to be equal to the control group.

Summary, Conclusions and Implications

Multiple exposures to an actual corrective radio advertisement had significant influence on lowering recall subjects' target belief of "fights colds and sore throats" while those subjects who did not recall the corrective ad were not effected by the experimental manipulation. No

differences between recall, no recall, and control subjects were found for any non-target beliefs, perception of honesty, or affect/attitude toward Listerine. It appears that corrective advertising can be effective for removing or negating false beliefs formed and developed by deceptive advertising without affecting other beliefs and attitudes about the product. This appears true when the consumer is able to report or recall the ad, but can occur without awareness or recall of the specific corrective information presented in the communication.

Limited external validity of these data exists due to the product and brand studies, the student sample, and the procedure and the incentives used to induce cooperation and to elicit cognitive responses. Nonetheless, the validity of these findings and significance are enhanced by the field experimental nature of this study as opposed to the laboratory settings of all other empirical research assessing the impact and effectiveness of corrective advertising. At this point, it appears that corrective advertising, as it is implemented in the Listerine case, is capable of removing a false belief without harming the offending advertiser or its competition if the consumer has had sufficient exposure and cognitive activity to develop recall of the message. Such data should be useful to policy makers in guiding future enforcement of corrective advertising in terms of amount and endurance of the corrective advertising campaign.

References

- Dyer, R. F. and Kuehl, P.G. (1978), "A Longitudinal Study of Corrective Advertising," Journal of Marketing Research, 15, 39-41.
- Hunt, H.K. (1973a), "Source Effects, Message Effects, and General Effects in Counteradvertising," Proceedings, Association for Consumer Research, 370-381.
- _____ (1973b), "Effects of Corrective Advertising," Journal of Advertising Research, 13, 15-24.
- Kerlinger, F. N. (1973), Foundations of Behavioral Research, 2nd ed. New York: Holt, Rinehart and Winston, Inc.
- Lucas, D.B. and Britt, S.H. (1963), Measuring Advertising Effectiveness. New York: McGraw-Hill Book Company.
- Maddox, L.M. and Zanot, E.J. (1979), "Corrective Advertising: Review and Prognosis," Current Issues and Research in Advertising, Division of Research, Graduate School of Business Administration, University of Michigan, 53-63.
- Mazis, M.B. and Adkinson, J.E. (1976), "An Experimental Evaluation of a Proposed Corrective Advertising Remedy," Journal of Marketing Research, 13, 178-183.
- Mizerski, R.D., Allison, N.K., and Calvert, S. (1980), "A Controlled Field Study of Corrective Advertising Using Multiple Exposures and a Commercial Medium," Journal of Marketing Research, 17:3 (August).
- Nie, N. H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. (1975), Statistical Package for the Social Sciences, 2nd ed., New York: McGraw-Hill Book Company.
- Percy, L. (1978), "Some Questions on the Validity of Recall Testing as a Measure of Advertising Effectiveness," Current Issues and Research in Advertising, Division of Research, Graduate School of Business, University of Michigan, 121-130.
- Wilkes, R.E. and Wilcox, J.B. (1975), "Recent FTC Actions: Implications for the Advertising Strategist," Journal of Marketing, 38, 55-61
- Winer, B.J. (1971), Statistical Principles in Experimental Design, 2nd ed., New York: McGraw-Hill Book Company.

AN EXPERIMENTAL STUDY OF THE EFFECTS
OF COMMERCIAL TV ADVERTISING AND PRO-CONSUMER
PRODUCT TEST RESULTS ON TV

Emmanuel J. Chéron, Université du Québec, Rimouski
Jean Perrien, Université de Sherbrooke, Québec

Abstract

The relative effect of commercial TV advertising was experimentally tested in comparison with pro-consumer TV program giving product test results. The experiment involved 125 consumers who were on a shopping trip to a local store. Subjects who volunteered were asked to view a 10 minute videotape arrangement and to complete a short questionnaire. Product brand evaluation and intention to buy were the two main dependent variables. Results indicated that pro-consumer TV program had greater influence than commercial advertising. Significant interaction effects were also observed between the two sources of communication.

Introduction

With the ever increasing influence of consumerism, the Canadian Broadcasting Corporation (C.B.C.) recently sponsored a new type of TV program designed to present test results on various products and inform consumers about their rights. This new form of communication to consumers has been an immediate success as audience figures indicated as high ratings for this program as for the most popular series. The present research is designed to test the effects of this new source of communication on consumers in contrast with traditional commercial advertising.

Background

Previous research on persuasive communication has focused on influence of source credibility (Hovland 1952), learning without involvement (Krugman 1965) and various types of effects (message, presentation, communicator, audience and risk level) as studied by Levitt (1965) in the context of industrial selling. The nature of these studies was to test identical messages from different sources with similar audiences. Although the Levitt study manipulated the quality of message presentation and type of audience, sources of communication were in fact identical in message content and all under control of the marketer. As stated by Levitt: "In commercial communications particularly, no two sources selling the same product to the same audience are ever likely to employ the same message".

A research by Hempel (1966) tested the effects of printed pro-consumer information such as Consumer Reports against a tape recorded commercial sales presentation. The product involved in the study were two brands of men's white dress shirts. Respondents were college students and results indicated that Consumer Reports effects were greater than the commercial sales presentation. Contrary to expectation, no significant interaction was observed between the two sources of communication. The nature of the respondents and product (low ego-involvement) imposed limitations to the validity of the experiment, further, sources of communication differed as to their form (printed and verbal tape recorded). More validity could have been attained with a printed commercial advertising rather than a tape recording of a salesman.

Recent studies (Settle and Golden 1974, Smith and Hunt 1978) have linked the effects of commercial communication

and attribution theory. These studies tended to focus on the processes by which individual perceive and explain causal relationships and give meaning to events. Hence these studies indicated that the effect of commercial communication was greater when product claims were not all favorable. The notion of causal schemata (Kelley 1971) introduced the idea of causal inference by which individuals explain their own behavior or that of others. Individuals tend to interpret the situation (the communication) in terms of intrinsic or extrinsic factors. This process has a direct effect on inferences made by the individual about communicators' intentions. When a behavior is perceived as being caused by an external factor such as a proper payment from an advertising agency, it is most likely that there will be less confidence in the communication. However if the behavior is perceived under the influence of internal factors controlled by the actor, there is probably more confidence in the presentation.

The objective of the present research was to assess the relationships among the effects of two impersonal communication sources on consumer product brand evaluation and intention to buy. From the consumer perspective, information on products is available through personal (sales presentations, personal contacts, friends) and impersonal sources (advertising, product test results, editorials). The study was intended to experimentally compare a pro-consumer TV broadcast and commercial TV advertising both prepared on video-tapes. The design of the experiment was adapted from Hempel (1966).

Problems

This study concerns the relative effect on consumers of commercial TV information as compared with product test results such as given by consumer protection broadcasts.

More precisely, a controlled experiment was designed to answer such questions as: How do consumers react to information sources? How information sources influence product evaluation and brand choice? Does the effect of information sources vary with demographic and socioeconomic differences?

For the business community, the impact of product test results on TV may be of much consequence. Previous research by Perrien and Chéron (1979) indicated a large audience familiarity with such a program (75.3% of the respondents on an unaided recall). Further, the audience was not restricted to a socially advantaged group of consumers. Hence, if present, the effect of such information will involve more consumers than was the case with pro-consumer printed information. Following a test comparison on TV, businesses could have a strong incentive to adjust their advertising strategies and marketing mix. Facing a favorable test result, the seller could anticipate increasing demand and should adjust inventories, physical distribution arrangements and/or other marketing-mix elements. In presence of unfavorable test result, the seller could consider various advertising and/or other marketing-mix strategies to counterbalance a probable negative effect on demand.

Hypotheses

Formally, the following hypotheses were tested:

- H1 - Pro-consumer TV broadcast information on product test results will have greater influence on product evaluation and intention to buy than a similar commercial TV advertising.

The study by Hempel (1966) supported this hypothesis, further, attribution theory suggested that when behavior is perceived under control of the actor, there is more confidence in the communication. Results of a previous survey (Perrien and Chéron 1979) indicated high confidence of respondents (mean score of 5.32 on a seven point scale) in the independent nature of information as given by pro-consumer TV programs.

- H2 - There will be significant interaction between the effects of the two sources of information. The influence of one source of information will depend upon the other source of information to which the respondents is exposed.

Although the previous study (Hempel 1966) expected interaction to be significant, analysis of variance results indicated no significant departure from additivity of effects. With the expected greater influence of product test results over commercial advertising it is likely that product evaluation and intention to buy will be mainly determined by this source of communication and thus presents significant interaction with commercial advertising.

- H3 - The effect of information sources will vary with demographic and socio-economic variables. Since previous research has dealt with homogeneous student subjects, this hypothesis is proposed to explore the issue of different responses with consumers of varying age, marital status, sex and education.

Method

The relative effect of the two sources of information were experimentally tested involving the evaluation of two brands of microwave ovens (high ego-involvement products). Subjects were one hundred and twenty five real consumers who agreed to take part in an experiment on consumption while on a shopping trip to a local food and hardware store. Arrangements were made with the store manager to reserve an enclosed area with twenty seats where videotape viewing could be organized. About one out of four consumers accepted to participate, coffee and cookies were served until about fifteen consumers were recruited for each of the nine parts of the experiment.

The experiment consisted of videotape arrangements showing typically:

- 1 - A commercial TV advertising for one or two fictive brands of microwave oven (TORIKO or CGX).
- 2 - A short cartoon unrelated with the topic.
- 3 - A TV product-test as given by pro-consumer TV broadcast comparing the two fictive brands of microwave oven and favoring brand TORIKO or CGX.

Special care was involved in the preparation of the communication messages. The commercial TV advertising was of the testimonial type showing two "slices of life" scenes involving a career woman, and a secretary. Motive to use microwave oven were time, money saving and safety. Form and contents of this commercial were based on an analysis of various advertising materials for different brands of

microwave ovens. Production of the commercial was done by the audio-visual department at a major university and TV product test results were prepared with participation of the host of the C.B.C. TV broadcast.

Comparison of the two microwave ovens were presented on criteria such as protection against radiation, efficiency, oven design, price and overall test result was given favoring one of the two brands. This set of criteria was based on real tests presented in an issue of the "Canadian Consumer". Therefore, both types of messages would be regarded as externally valid.

Communication messages were tested with fifteen adult consumers. Subjects were recruited through the local boy-scout organization in exchange for a donation. Comparison of two brands of microwave ovens was presented as the object of the experiment and participants were given a questionnaire to fill in at the end of the viewing. Questions were asked on product brand evaluation, intention to buy, demographic and socioeconomic characteristics and overall impression and comments about the objective of the experiment. Comments indicated that respondents felt they were in a real situation and two out of fifteen indicated that the objective of the experiment was to measure the effect of the C.B.C. TV broadcast. One respondent suspected the objective of the experiment was to measure the influence of advertising.

In the store, consumers were randomly assigned to one of the nine combination treatments shown on figure 1.

FIGURE 1

Experimental design

Product-test results Commercial advertising	Product-test results		
	No-message	TORIKO brandname	CGX brandname
No-message	1 T_0CGX_0	4 T_0T_{TR}	7 T_0CGX_{TR}
TORIKO brandname	2 $T_{AD}CGX_0$	5 $T_{AD}T_{TR}$	8 $T_{AD}CGX_{TR}$
CGX brandname	3 $CGX_{AD}CGX_0$	6 $CGX_{AD}T_{TR}$	9 $CGX_{AD}CGX_{TR}$

As indicated, consumers might be exposed to each of nine combinations of no-message, test results and commercial advertising supporting either one or the other of the two brands of microwave ovens. In the "no-message" situation, subjects were only shown a slide presenting a microwave oven of the appropriate brand (T_0 or CGX_0 or both). Commercial advertising was of the testimonial type as mentioned above, contents of the two commercials were identical except for the brand name of the microwave oven. Product test results, as presented by the host of the program, were also identical in contents except for the winning brand name. For example, subjects assigned to cell $CGX_{AD}T_{TR}$ would be exposed to a commercial advertising favoring brand CGX and a product test result supporting brand TORIKO.

Presentation of a short cartoon between the advertising exposure and the test result contributed to add realism to the situation. All nine treatment cells were prepared beforehand on videotape so that presentation to subjects could be made on color TV screen without loss of continuity. The fact that some respondents were accompanied by

children added a source of distraction more conform to the real TV viewing situation.

Following the exposure, each respondent completed a questionnaire designed to measure product brand evaluation and intention to buy. Other information requested were: ownership of a microwave oven, viewing and frequency of viewing test results on TV and demographic and socio-economic data. The dependent variables of the experiment were the subject's brand evaluation and intention to buy recorded on a seven-point bipolar rating scale with one of the two brands at each end of the scale.

Results

Frequency data on respondents revealed that 117 out of 125 or 93.6% did not own a microwave oven and that 119 or 95.2% had seen product test result broadcasts on TV. These figures indicated a fairly homogeneous group with respect to these variables. Average age of respondents was 30, 60% were married and 57.6% were females.

Results on product evaluation and intention to buy indicated that the experimental treatment produced significant differences among the nine exposure combinations.

The two-way analysis of variance for the effects of the two types of communications on product evaluation and intention to buy is shown in table 1.

TABLE 1

Two-way ANOVA for the effects of the two types of communication messages

	PRODUCT EVALUATION			INTENTION TO BUY	
	D.F.	Mean square	F ratio	Mean square	F ratio
Product test	2	49.210	51.98**	41.90*	37.45**
Commercial	2	3.766	3.95*	3.268	4.81**
Interaction	4	5.525	5.79**	6.354	5.78**
Error	116	.954		1.095	

** Significant at .01 level: $F_{.99}(2,116) = 4.79$; $F_{.99}(4,116) = 3.43$

* Significant at .05 level: $F_{.99}(2,116) = 3.07$

Hypothesis one suggested that pro-consumer test results will have greater influence on product evaluation and intention to buy than commercial advertising. As both message effects were statistically significant a measure called "omega-square" was computed: (Hays cited by Green 1978). This measure allows to estimate the relative importance of each type of message in contributing to variation in product evaluation and intention to buy.

TABLE 2

Hay's omega-square measures

	PRODUCT EVALUATION	INTENTION TO BUY
Product test	0.400*	0.325*
Commercial	0.025*	0.034*
Interaction	0.076*	0.085*
TOTAL	0.499	0.442

* Computed according to: $\omega^2 = \frac{SS_{\text{among}} - (DF_{\text{among}} \times MSE)}{SS_{\text{total}} + MSE}$

Results in table 2 show the relative importance of the message variables on product evaluation and intention to buy. These measures can be roughly interpreted as R^2 in multiple regression. Product test messages account for 40% of the variation in product evaluation while commercial messages and interaction account respectively for 2.3% and 7.6% of the variation in product evaluation.

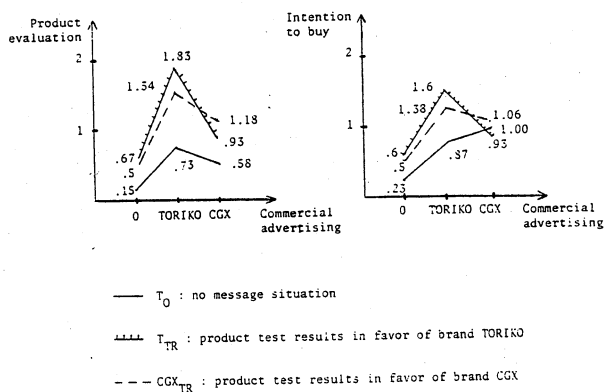
Overall, 50% of the variation in product evaluation is accounted for by the two message variables. Results for intention to buy indicate a lower relative importance of product test messages than for product evaluation. However relative importance of commercial messages and interaction is higher for intention to buy than for product evaluation.

Overall, relative importance measures for product evaluation and intention to buy support the hypothesis that product test results has greater influence than commercial advertising.

Hypothesis two proposed that there will be significant interaction between the effects of the two sources of information. The analysis of variance in table 1 confirm the presence of statistically significant interaction. Graphically this can be seen on figure 2.

FIGURE 2

Interaction patterns



Diagrams in figure 2 can be interpreted as follow; exposure to a commercial advertising in favor of brand TORIKO and viewing product test results in favor of brand TORIKO yielded a mean score of 1.83 for product evaluation and 1.6 for intention to buy. These scores were derived by taking the absolute difference between cell means and the mid-point on the seven-point scale featuring one brand at each end of the scale. (Overall grand mean of observations in the experiment was equal to mid-point score). Interaction effect is apparent as response functions depart from parallelism (beyond what might be expected by chance).

These results clearly confirm hypothesis two indicating that responses to one source of information will depend upon the other source of information to which the respondent is exposed.

Hypothesis three was exploratory in nature and proposed to test the effect of variations in demographic and socioeconomic variables on product brand evaluation and intention to buy. The effects of marital status, education and age indicated that although the effect of product-test results remained significant, the effect of TV commercial adverti-

sing tended to be less significant (confirming hypothesis one) and interaction was slightly less significant (disconfirming hypothesis two).

An analysis of covariance controlling for age as a concomitant variable resulted in an F ratio of .368 with associated probability of .546 for product evaluation, and an F ratio of .039 with associated probability of .843 for intention to buy. Clearly age was not a significant variable for measuring the effect of the two sources of communication on product evaluation and intention to buy.

The effect of sex as a discriminating variable revealed differences of responses to commercial advertising. Results of the two-way analysis of variance for males and females is shown on table 3 for product brand evaluation.

TABLE 3

Two-way ANOVA for male and female product brand evaluations in response to the two sources of information

	FEMALE RESPONDENTS			MALE RESPONDENTS		
	D.F.	Mean square	F ratio	D.F.	Mean square	F ratio
Product test	2	27.959	27.9**	2	22.750	24.95**
Commercial	2	6.513	6.52**	2	0.051	0.05 ^{NS}
Interaction	4	2.585	2.59*	4	3.291	3.60*
Error	63	.999		43	.912	

** Significant at .01 level: $F_{.99}(2,63) = 4.98$; $F_{.99}(2,43) = 5.13$

* Significant at .05 level: $F_{.95}(4,63) = 2.53$; $F_{.95}(4,43) = 2.61$

NS : Not significant

As can be seen in table 3, the main effect of the commercial advertising is not significant for male respondents. Similar results were obtained with intention to buy, commercial advertising although significant for female respondents was not significant for male respondents. This result confirms the marginal effect of commercial advertising as compared with product test results. As regards interaction, this effect is still significant for both male and female respondents.

Discussion

The overwhelming influence of product test results confirmed hypothesis one with the implications that businesses would be well advised to take this source of information into account for commercial strategy decisions. Relative importance of product test results tended to decrease when consumer decision was closer to the buying act (intention to buy vs product brand evaluation) as shown by omega-square measures in table 2. Previous research (Levitt 1965) indicated similar results. In relatively higher risk situations, the audience tended to discount the highly credible source of information.

Independent variables and interaction tended overall to account for less of the variation in intention to buy (44%) than of the variation in product evaluation (50%). This is consistent with results of a previous survey (Perrien and Chéron 1979) where the effects of product test result programs on TV were measured on the three components of an attitude. Scores on the conative component were lower than on the cognitive and affective components indicating decreasing effects of product test results when consumer decision was closer to action.

The presence of interaction confirming hypothesis two indicated unanticipated patterns of changes. Effects on pro-

duct evaluation and intention to buy tended to be higher when the two sources of information were in favor of two different brands rather than both in favor of the same brand (see figure 2). This situation may be accounted for by a defensive reaction in presence of intense persuasive messages such as observed in previous research (Hempel 1966) and by consumer attributional processes and effects in presence of messages with only positive claims (Smith and Hunt 1978).

Exploration of demographic and socio-economic variables revealed that sex was a discriminating variable for responses to commercial advertising. For male respondents the main effect of commercial advertising was not significant. This difference could be accounted for by the type of product chosen. Previous research (Capon and Burke 1977) indicated that microwave ovens scored high on a perceived risk scale (compared to non-durables lower priced items). These products were thought to be rather high ego-involving and implying a substantial money outlay with likely joint husband and wife buying decision. However microwave ovens are likely to be used more often by women and it is suspected that in the experiment, men had more difficulty to identify to the commercial advertising situation showing scenes involving women. Further research will have to be conducted to clarify this issue.

With respect to interaction, hypothesis three was still confirmed for both men and women. However, unanticipated results of lower product evaluation in presence of the two messages favoring the same brand rather than different brands was the results of male rather than female respondents.

Female respondents tended to show the anticipated pattern, scoring higher on product evaluation when both messages favored the same brand rather than different brands. This result although speculative would be in contradiction with previous research (Krugman 1965). Female respondents were probably more involved, and they were expected to have higher defence reaction than male respondents. However response delays may be too short and real choice situations too distant for such a tentative comment.

Some of this unexpected variation in product evaluation may be accounted for by different connotations associated with the two fictive brand names. In fact, mean cell results indicated larger variations in product evaluation and intention to buy when brand TORIKO rather than brand CGX was involved (see figure 2). In the French Canadian environment, brand name CGX was possibly less appealing to respondents than brand TORIKO.

Conclusion

The effects of commercial TV advertising and pro-consumer product test results on Canadian TV program were experimentally compared.

Findings indicated the great influence of this pro-consumer source of information. Presence of significant interaction with commercial advertising implied that the net effect of the two sources of information will be different from a simple addition of the separate effects of each source of communication.

Not only sources of communication, but message contents, seemed to have different influence on male and female respondents. Further research should investigate whether the extent of ego-involvement with the product is the same for men and women and whether fictive brand names do have different connotations. In spite of these limitations it is felt that the experiment was closer to the real situation than research conducted with student subjects within university walls.

From the seller's point of view, pro-consumer product test results on TV programs introduces a new kind of uncontrollable source of mass communication influencing product evaluation by consumers. Marketers should consider this new variable for the necessary adjustments to their marketing-mix.

From the public policy maker point of view, pro-consumer TV programs giving product test results can be viewed as an efficient means to counter-balance the effect of commercial TV advertising among the wide segment of the population less likely to be reached with printed media.

However, our findings imply that magnitude of communication effects is not easily predictable. Another problem is the relative short-term of these types of communication aggravated by the rapid turnover of brands and products on the market. Communication policy makers with consumer welfare in mind should try to reach an adequate balance between concrete, highly involving product test results and less attractive general consumer education with longer term effect.

References

Capon, No#1 and Burke, Marian (1977), "Information Seeking Behavior in Consumer Durable Purchases", Proceedings of the American Marketing Association, 110-115.

Golden, Linda L. (1977), "Attribution Theory Implications for Advertisement Claim Credibility", Journal of Marketing Research, 115-118.

Green, Paul E. and Tull, Donald S. (1978), Research for Marketing Decisions, Fourth Edition, Prentice Hall, 369-372.

Hempel, Donald J. (1966), "An Experimental Study of the Effects of Information on Consumer Product Evaluations", in Science Technology and Marketing, edited by Raymond Haas, American Marketing Association, 589-597.

Hovland, Carl I. and Weiss, Walter (1952), "The Influence of Source Credibility on Communication Effectiveness", Public Opinion Quarterly, Vol. 15, no. 4, 635-650.

Kelley, Harold H. (1971), "Causal Schemata and the Attribution Process", in Edward E. Jones et al., eds, Attribution: Perceiving the Causes of Behavior, General Learning Press, 151-174.

Krugman, H.E. (1965), "The Impact of Television Advertising: Learning Without Involvement", Public Opinion Quarterly, Vol. 29, no. 2, 349-356.

Levitt, Theodore (1965), Industrial Purchasing Behavior: a Study of Communications Effects, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 5-11.

Perrien, Jean and Chéron, Emmanuel J. (1979), "Pro-Consumer Information on TV: Results and Implications of the Canadian Experience", 8th Annual Meeting of the European Academy for Advanced Research in Marketing, D1-D23.

Smith, Robert E. and Hunt, Shelby D. (1978), "The Effectiveness of Personal Selling Over Advertising: an Attributional Analysis", Educator's Proceedings, American Marketing Association, 158-163.

Smith, Robert E. and Hunt, Shelby (1978), "Attributional Processes and Effects in Promotional Situations", Journal of Consumer Research, vol. 5, 149-158.

INQUIRY RESPONSE RATES, COST AND REVENUE
PER INQUIRY OF REPETITIVE PRINT ADVERTISING

Arch G. Woodside, University of South Carolina

Abstract

How many exposures are enough? Theoretical arguments by Krugman (1972) propose that three exposures may be optimal for remembering broadcast advertising messages. Evidence presented here suggests that two exposures may be best for producing the highest direct inquiry response rates and lowest cost per inquiry from print advertisements.

Introduction

How are costs and revenues affected by repetitive advertising aimed at producing direct response inquiries and sales? Unfortunately, published reports appear to be nonexistent on the costs, number of inquiries, and revenues related to each of repeated insertions of the same advertisement in the same magazine or newspaper.

Simon (1979) has pointed out that recall and attitude measures of advertising are simply proxies for sales, and ultimately they must be related to sales. His analysis of Pomerance and Zielske's (1958; Zielske, 1959) field laboratory study comparing two 13-week media schedules indicated that "the amount of advertising impact (as measured by recall-weeks) is much higher for the every-four-weeks schedule than for the weekly schedule" (Simon 1979, p. 418). Simon noted some "severe" qualifications and criticism of this study including the following two: 1) The data were generated by just one advertising medium, direct mail, and direct mail would not be a likely medium for a staple food product (the product advertised in the study); and 2) A sales measure might show a different sort of response pattern than the recall measure used. Also, a recall measure has a built-in saturation limit (100%) that is not relevant for actual sales purposes.

Ray and his associates (Ray and Sawyer 1971, Swinyard and Ray 1977) report increases in purchase intentions of later versus first and second insertions. In the Swinyard and Ray (1977) study, female household subjects were asked the following question after receiving four direct mailings of appeals for blood donors and Red Cross volunteer workers: "If you were going to volunteer work for one of these organizations, which one would be your first choice? Second choice?" Behavioral intentions were coded as the proportion of respondents rating the Red Cross as first choice among five name social health organizations. Intention to volunteer for the Red Cross increased from 5% to 17.4% to 22.7% from the first to the second to the fourth exposures for the subjects receiving the direct mail appeals.

Information on costs, number of inquiries, and estimated revenues from a print advertising campaign which included multiple insertions of the same advertisement in the same magazines and newspaper is reported in the present article. The immediate objective of the campaign was to maximize the number of direct inquiries with the lowest possible costs. The ultimate objective of the campaign was to produce the greatest possible sales given a fixed advertising budget.

The main purpose of the research was to measure the changes in costs per inquiry (CPI) of several insertions of the same ad in each of several magazines. The same ad was repeated 3 to 13 times in 7 magazines and 1 newspaper.


Assuming a constant cost per advertisement insertion over a campaign, decreases in CPI for the second and third inser-

tions compared to the first insertion were hypothesized. Then CPI was expected to increase for later insertions. Thus, a "U" shaped relationship between CPI and the number of insertions was hypothesized.

What rationales can be provided for the expected increases in the number of inquiries for the second and third insertions compared to the first (and thus decreases in CPI given a constant cost per insertion)? Increases in recall (Swinyard and Ray 1977), and brand awareness (Stewart 1964) for second, third, and later advertisements versus an initial advertisement found in prior research provide one rationale for the hypothesis. The theoretical work of Krugman (1962; 1968; 1972) provides another rationale. Krugman develops a cognitive theory to indicate "why three exposures may be enough" to produce behavior and thus sales.

Method

The media costs, inquiry response rates, and estimated revenues (for repeated insertions in two magazines) were studied for 7 magazines and 1 newspaper which included three or more insertions of the same advertisement. The advertisement was designed to promote a direct response inquiry to receive a "free South Carolina Trip Kit." The advertisement is shown in the exhibit:



South Carolina.
It's a lot of great vacations.

Whatever kind of vacation you're looking for, you'll find it in South Carolina.

And to help you plan your vacation, we'll send you the new South Carolina Trip Kit—free! It'll tell you about things to do and places to stay. And there are maps to show you how to get there once you get here.

Wherever you go in South Carolina, you'll find a lot of great vacations. Just clip the coupon to find out how they can all be yours.

FREE SOUTH CAROLINA TRIP KIT

Name _____
Address _____
City _____
State _____ Zip _____

Mail to South Carolina Division of Tourism,
Room 000, Box 78, Columbia, S. C. 29202

Each advertisement included a unique room number in the address on the coupon to identify the magazine and specific insertion. The number of inquiries was counted per insertion for each magazine. The cost of the advertisement for each insertion was divided by the number of inquiries received, resulting in a CPI.

CPI, similar to CPM (cost per 1,000's of circulation), is often used as a measure for comparing the performances of competing magazines. A recent ad for Sport magazine compares its CPM with the CPM's for Playboy, Popular Mechanics, Popular Science, and five other magazines (Advertising Age 1980, p. 11). Data on CPI are collected by individual advertisers and are often used to answer the following question: Did the magazine or newspaper ("book") offering a low CPM also provide a low CPI in comparison to other books? The implication is that a book promoted as a "good buy" for an ad placement because of its comparatively low CPM may be, in fact, a poor buy because it delivers a comparatively high CPI for the specific ad placed. Also, a book having a comparatively high CPM may be, in fact, a good buy because its CPI is also comparatively low (cf. Woodside and Reid 1974).

Revenue per inquiry (RPI) was estimated for each of multiple insertions for two magazines in the study. A mail questionnaire was sent to samples of inquirers for each insertion six months after nearly all the inquiries were received (in October). The samples were selected systematically from each "room" for each insertion. The questionnaire had 18 questions including questions on visiting South Carolina and expenditures in South Carolina. RPI was computed by first multiplying the total number of inquiries, e.g., 500, by the percent of those survey respondents who visited South Carolina, e.g., 20 percent: $500 \times .20 = 100$. Second, the estimated number of visitors is multiplied by the average expenditure per visitor, e.g., \$30. The result is the estimated total revenue: $100 \times \$30 = \$3,000$. RPI is estimated by dividing the estimated total revenue by the total number of inquiries: $\$3,000 / 500 = \6 .

The average rate of response of questionnaires sent was 50 percent for the eight insertions studied following a second and third mailing of the questionnaire to nonrespondents. (Percentages of visits to South Carolina were not statistically different between respondents from the three mailings for each magazine.)

Results

The hypothesis was partially supported. The CPI declined from the first to the second insertions for 7 of the 8 books as shown in Figures 1 and 2. However, the relationship between CPI and the number of insertions was more "v" shaped than "u" shaped. The CPI was greater for the third insertion compared to the second and to the first insertions for 6 of the 7 books, exhibiting a "v" shaped relationship between CPI and the number of insertions.

The inquiry response rates as percents of circulation are listed in Table 1 for the three insertions in Popular Mechanics. The difference (.02%) in response rates between the January and February insertions is small but meaningful given the large circulation of the magazine. Thus, CPI drops \$0.17 or 5% with a .02% increase in inquiry response rate.

Note that RPI changes in the same direction as CPI for three insertions in Popular Mechanics. The poor evaluation implied by the relatively high CPI of \$4.27 of the March insertion is offset by the relatively high estimated RPI of \$217 for the same insertion.

FIGURE 1
COST PER INQUIRY FOR SEVEN BOOKS

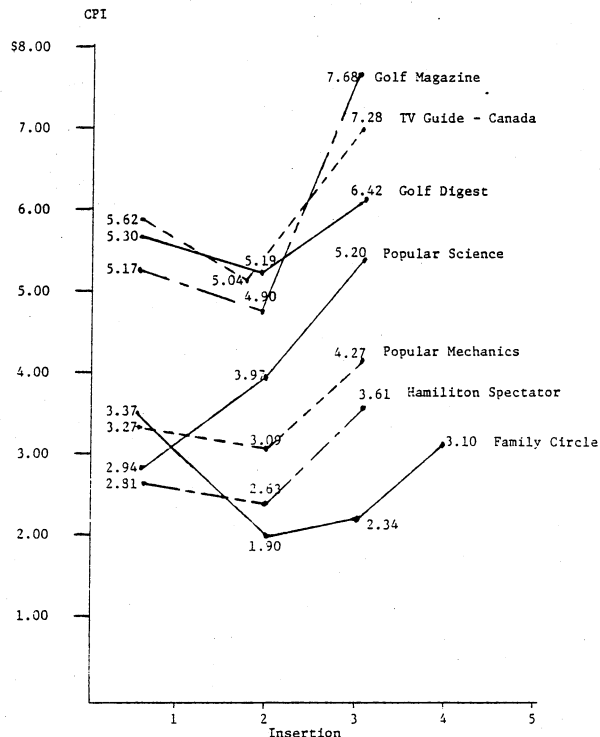
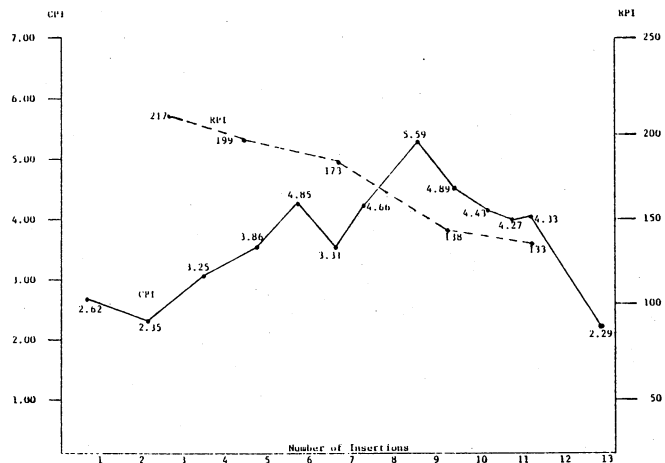


FIGURE 2
PLOT OF CPI AND RPI FOR THIRTEEN INSERTIONS
IN TV GUIDE^a



^aCPI is plotted on the left vertical axis and RPI is plotted on the right vertical axis.

TABLE 1
INQUIRY RESPONSE RATES, COSTS AND REVENUES
FOR THREE INSERTIONS IN POPULAR MECHANICS

Date	Response Rate ^a	CPI	RPI
January	.46%	\$3.26	\$166
February	.48	3.09	157
March	.37	4.27	217

^aCirculation was 1,671,000 for each month.

A total of 13 weekly insertions of the ad was placed in TV-Guide starting on January 6, 1979. The resulting CPI values are shown in Figure 2. Note that a "v" shaped relationship exists for CPI and the first three insertions. CPI increases peaking at \$5.59 for the eighth insertion, and then declines to its lowest point (\$2.29 for the last insertion). These findings suggest a substantial seasonal effect of insertions on CPI. The early January ad may have attracted many inquiries intending to visit South Carolina for the spring season (and famous coastal gardens). The March 31 insertion (the last insertion) may have attracted many inquiries planning summer vacation trips. (State traced data indicate that June, July, and August followed by April are the four months having the greatest percent on nonresident travelers in South Carolina.) Thus, insertions in late February may be expected to have the highest CPI's. Unfortunately, only the limited information shown in Figure 2 is available to support this hypothesis.

Details on the circulation, costs for each ad insertion, and the percent inquiry response rate per insertion are presented in Table 2 for seven books. The percent inquiry response rates for the second compared to the first or third insertions were higher for all the books except Popular Science. The response rate declined for Popular Science from .68% to .50% to .38%. Thus, CPI increased for the second and third insertions in Popular Science since the cost for each ad insertion was constant.

In Table 3, the response rates, CPI, and available RPI are listed for the ad insertions in TV-Guide. Note that RPI continually declines substantially from \$217 to \$133 for the five insertions included in the survey research study. This may suggest that early ad insertions in the campaign may produce greater RPI than later insertions. Unfortunately, the RPI for the last insertion was not estimated to learn if the substantial drop in CPI between insertion 12 and 13 was coupled with a substantial increase in RPI.

TABLE 3
INQUIRY RESPONSE RATES, COSTS AND REVENUES
FOR THIRTEEN INSERTIONS IN TV-GUIDE

Date	Response Rate ^a	CPI	RPI
January 6	.55%	\$2.62	--
January 13	.62	2.35	\$217
January 20	.45	3.25	--
January 27	.37	3.86	--
February 3	.35	4.15	199
February 10	.44	3.31	--
February 17	.31	4.66	173
February 24	.26	5.59	--
March 3	.30	4.89	--
March 10	.33	4.43	138
March 17	.34	4.27	--
March 24	.33	4.33	133
March 31	.63	2.29	--

^aCirculation was 4,267,000 for each date.

Discussion

Krugman offers the following thesis as why three exposures may be enough, "campaign effects based on, say, 20 to 30 exposures are only multiples or combinations of what happens in the first few exposures." For television advertising, he believes that the cognitive task presented by the first exposure is dominated by a "what is it?"

TABLE 2
CIRCULATION, AD COST, AND PERCENT INQUIRY RESPONSE RATE FOR REPEATED INSERTIONS

Magazine	Circulation (000)	Cost for Each Ad Insertion	Percent Inquiry Response Rate Per Insertion			
			1	2	3	4
Golf Digest	946	\$3,915	.78	.80	.64	--
Golf Magazine	724	2,840	.76	.80	.51	--
Popular Mechanics	1,671	2,485	.46	.48	.37	--
Popular Science	1,862	3,724	.68	.50	.38	--
Family Circle ^a	2,350	6,384	.685	1.43	1.16	.88
TV-Guide, Canada	687	1,950	.51	.56	.39	--
Hamilton Spectator	143	484	1.20	1.29	.94	--

^aThe cost of the first insertion was \$5,426 while the cost for each of the second, third, and fourth insertions was \$6,384.

response - an attempt to define or understand the ad stimulus. The first exposure is likely to cause high recognition scores for the second exposure. The second exposure is likely to be dominated by the evaluative response of "what of it?" where the consumer judges the personal relevance of the now somewhat familiar ad. The third exposure becomes a reminder but also the beginning of disengagement.

Assuming that print advertising requires more consumer involvement than television commercials (cf. Krugman, 1965; Robertson, 1976) might suggest that such learning with involvement requires fewer exposures for optimal effect. Readers have the opportunity to linger and ask what is it?, what of it?, and what should I do about it? In a review of the literature Sawyer and Ward (1979) concluded that "the existing evidence suggests that the audience-controlled input of print advertising may generate more immediate learning and cognitive responses than low involvement "broadcast exposures" (p. 289). Thus, higher involvement in print versus broadcast advertisements is one rationale for the finding that CPI is lowest for the second of three print exposures.

Clearly, more unobtrusive field research is needed than available currently to learn the behavioral effects of repetitive advertising exposures. The results of the reported study support an inverted V-curve relationship between inquiry response rate and exposures, and thus, a V-curve relationship between CPI and exposures given a constant cost per exposure. Further evidence with other products and services (and in tourism advertising) is needed to support or refute this prediction before the conclusion is reached that two exposures may be enough in print media vehicles.

Do the results from the present study really suggest that two exposures are enough? This implication may be inaccurate. Managers may discover that multiple insertions of related but different ads produce the greatest sales returns per advertising dollar. Varying the size of the ad and the use of color might increase the sales impact of multiple exposures. Unfortunately, the investment in research to learn the sales impact of repetitive advertising is no higher than the low levels reported by Stewart in 1964. Stewart's field experiment on repetitive newspaper advertising effects on awareness, attitude, and sales remains the best work available.

References

Advertising Age (1980), 51 (April 7).

Krugman, Herbert E. (1965), "The Impact of Television Advertising: Learning Without Involvement," Public Opinion Quarterly, 29, 349-356.

_____, (1968), "Processes Underlying Exposure to Advertising," American Psychologist, 23, 245-253.

_____, (1972), "Why Three Exposures May Be Enough," Journal of Advertising Research, 12 (December), 11-14.

Pomerance, Eugene and Zielske, Hubert (Hugh) (1958), "How Frequently Should You Advertise?" Media/Scope (September).

Ray, Michael L. and Sawyer, Alan G. (1971), "Repetition in Media Models: A Laboratory Technique," Journal of Marketing, 8 (February), 20-29.

Robertson, Thomas S. (1976), "Low-Commitment Consumer Behavior," Journal of Advertising Research, 16 (April), 19-24.

Sawyer, Alan and Ward, Scott (1979), "Carry-over Effects in

Advertising Communication," in Research in Marketing, Volume 2, Greenwich, CT: JAI Press, 259-314.

Simon, Julian L. (1979), "What do Zielske's Real Data Really Show About Pulsing?" Journal of Marketing Research, 16 (August), 415-420.

Stewart, John B. (1964), Repetitive Advertising in Newspapers, Boston, Division of Research, Graduate School of Business Administration, Harvard University.

Swinyard, William R. and Ray, Michael L. (1977), "Advertising-Selling Interactions: An Attribution Theory Experiment," Journal of Marketing Research, 14 (November), 509-516.

Woodside, Arch G. and Reid, David M. (1974), "Tourism Profiles Versus Audience Profiles: Are Upscale Magazines Really Upscale?" Journal of Travel Research, 12 (Spring), 17-23.

Zielske, Hubert (Hugh) (1959), "The Remembering and Forgetting of Advertising," Journal of Marketing, 23 (January), 239-243.

CONSUMER RESPONSE TO ADVERTISING:
IMPLICATIONS FOR COPY TESTING AND COPY

Mary Jane Schlinger, University of Illinois, Chicago

Abstract and Introduction

In a 1976 "state of the art" review, Jacoby identified several trends in research on consumer behavior. The first and most important was the attempt to apply and, more significantly, develop theory. Another was the move toward searching for cause-effect relationships. The three papers included in the session titled "Consumer Response to Advertising" exemplify these trends; they are theoretically oriented and search for explanations. There also have practical implications for the development of copytesting research, and it is those implications that are discussed in this overview.

Guilt Arousal

Chingold's paper titled "Guilt Arousing Marketing Communications: An Unexplored Variable" claims that guilt appeals are widely used in advertising and suggests that guilt arousal as persuasive device should be studied. On a broad level, this paper raises the question of how emotions in general relate to consumer responses to advertising.

"Emotion" is one of those hypothetical constructs that fall through the definitional and theoretical cracks in the jigsaw puzzle of psychology. "All experimenters know that emotion is a topic about which there is no agreement...." stated Murray in 1938. Nearly forty years later, Zimbardo and Ruch (1977) point out that psychologists still have failed to develop a precise definition of emotions, which variously are seen as motives, traits, bodily changes, and subjective feelings experienced and reported by the individual.

Lists of emotions, like lists of motives, differ both in length and content. Among those mentioned in the literature are admiration, affection, amusement, anger, anxiety, confidence, disgust, enjoyment, fear, friendliness, frustration, gratitude, guilt, hate, irritation, jealousy, love, passion, pity, pride, shame and surprise (Davitz 1969).

Although psychologists do not agree about what emotions are, there is a substantial body of research on the topic that primarily focuses on: (1) the physiology of emotional arousal; (2) the judgment of emotion in others; and, (3) the negative effects of emotion on behavior and health.

In contrast to the psychologists, consumer behavior researchers tend to have ignored the construct, except for the research on fear appeals. Chingold's paper is interesting because it suggests the possibility and potential usefulness of measuring not only guilt arousal but also a wide range of emotional, nonrational, impressionistic responses to advertising. At the present time, such reactions either are not measured, or the range is limited, e.g., to liking, amusement, empathy and irritation.

With regard to guilt, Chingold's paper prompts three observations. First, it is important to distinguish guilt as an appeal from shame. Guilt arises when the individual goes against his conscience. Shame, on the other hand, occurs when a cultural norm is violated in the presence of a real or imaginary audience. Advertising about ring-around-the collar, bad breath, underarm odor, spots on dishes, smelly carpets, and household dirt seems more likely to threaten the viewer with shame than guilt.

Guilt may be more appropriate and more often used for relatively involving products, such as low tar cigarettes, or for nonprofit and social advertising, such as anti-alcohol abuse campaigns and fund solicitations.

Second, it is a fairly common strategy for advertising to try to allay guilt about using products such as convenience goods. The purpose is to help the consumer avoid or reduce guilt arousal. One typical approach is to suggest or imply that the time saved by using the advertised brand can be devoted to more important matters -- family activities, exercise, and so on. One wonders if the guilt reduction effect of such messages is assessed.

Finally, any discussion of arousing guilt or other negative emotions must consider the moral and ethical implications. Advertising is criticized for manipulating emotions, both positive and negative, when the appeal is seen as exploitive, debasing or insulting. Rokeach (1970) has urged advertisers to avoid promoting products via destructive emotions or beliefs, i.e., negative self-conceptions and primitive fears about rejection, self-worth, self-identity and self-competence.

Celebrity Endorsers

Nearly every advertising textbook includes a typology of commercial executional styles or structural designs such as slice of life, humor, product demonstrations, celebrity endorsers, etc. (Hefzallah and Maloney 1979). The latter design is the subject of the paper by Mowen and Brown: "Effectiveness Measures of Celebrity Endorsers."

It is estimated that from five to ten percent of television commercials feature celebrities (Shimp 1979, McCollum and Spielman 1980). This is a substantial proportion, considering the disadvantages: stars can be expensive and risky for the advertiser both in terms of money and loss of control, i.e., live stars, unlike Tony the Tiger, may get involved in scandal, lose their star status, die, etc. On the other hand, the advantages that celebrities potentially can offer include credibility for the claims, attention value and memorability for the commercials, and prestige and positive affect that may be generalized to the brand.

Recently McCollum and Spielman (1980) analyzed data from hundreds of celebrity commercial tests conducted by the company over the past twelve years. Among their findings: fewer than half of the celebrity endorser commercials scored above product category norms on either clutter awareness or attitude change measures. The implication is that there is another serious risk in using celebrities: the risk that they may not "work."

Given that the use of stars is both common and chancy, it is not surprising that advertisers are keenly interested in finding ways to evaluate their potential effectiveness. Mowen and Brown approach this problem via a study that is theoretically and operationally complicated, i.e., it is based on an integration of Heider's balance theory with Kelley's attribution theory, and the resulting experiment utilizes multiple manipulations and several dependent variables. The fact that the predicted results were not obtained in the experiment may reflect inadequacies or misinterpretation of the theories (cf.

Kiesler, et. al. 1969, Eagly and Himmelfarb 1978, Kelley and Michela 1980) or inadequacies in the research design and execution.

With regard to implications for evaluating and selecting celebrity presenters, Mowen and Brown's study and the few other studies that have been published on the subject suggest that a multifaceted and multiple criteria approach will be required. Among the factors that need to be considered are the celebrity's likeableness, credibility (expertise and trustworthiness), attention value, compatibility with the brand and message, and distinctiveness with respect to endorsing one versus multiple products and brands. There definitely is a need for additional theoretically oriented research on celebrity endorsers, but the research may need to aim at theory building rather than the testing of theories borrowed from other disciplines.

Cognitive Processes

"Linear Effects of Cognitive Response to Advertising" by Percy and Lautman examines reactions to television commercials in an information processing framework. More specifically, it illustrates the practical application to copytesting of a theory, i.e., that cognitive processes mediate affective and conative reactions to advertising, and a measurement technique, i.e., the quantification of cognitive reactions by content analysis of verbatim protocols (cf. Wright 1973, Olson and Muderrisoglu 1979, and Mitchell 1980).

It is common for advertising agencies to utilize nondirected, free elicitation procedures to obtain verbatim protocols from subjects who have viewed an advertisement of commercial. Typical open-ended questions are "what went through your mind as you watched the commercial?" (or looked at the ad) and "Tell me in your own words what the advertisement said and showed." It also is common for the coding or content analysis of such protocols to be simplistic and atheoretical. For example, an analyst might count the number of respondents who: mentioned various executional elements, actions or characteristics; played back messages, stated the brand name, and made evaluative comments about the brand or execution. In copytesting research, the codes and coding procedures usually are not carefully defined, and the reliability is not measured.

In short, the content analysis of open ended verbatims generated from copytests could be, should be and probably will be improved in the future. The paper by Percy and Lautman suggests one approach, i.e., the coding of support arguments. There are, of course, many other analytic possibilities not only for coding of different types of mediating cognitions but also for linguistic analysis and the measurement of nonverbal factors such as voice pitch and response time.

There is an unanswered question about the Percy and Lautman data that is relevant to the implications of the study for copytesting. Respondents were shown a commercial for a well known brand which they had been told to watch very carefully and then asked to relate all of their thoughts while viewing the commercial. The number of support arguments elicited for each respondent was counted, and it was found that as the number of support arguments offered by a respondent increased so did her positive attitude toward the brand and her intention to buy. It is not clear what influence a viewer's prior exposure to advertising, product experience, and pre-existing knowledge, attitude and behavioral intentions with regard to the brand had on her cognitive processing of the commercial and the number of support arguments that she produced. For example, pre-existing favorable intentions to buy may have led both to more support arguments and to a high post-viewing intention to buy rating. If copy testers are to evaluate and

compare advertisements for established brands in a cognitive processing framework, then the influence of prior experience will need to be controlled or sorted out from the influence of the advertisements.

References

- Davits, Joel R. (1969), The Language of Emotion, New York: Academic Press.
- Eagly, Alice H. and Himmelfarb, Samuel (1978), "Attitudes and Opinions," in Annual Review of Psychology, Vol. 29, eds. M. Rosenzweig and L. Porter, Palo Alto, California: Annual Reviews Inc., 517-574.
- Hefzallah, Ibrahim M. and Maloney, Paul (1979), "Are There Only Six Kinds of TV Commercials?" Journal of Advertising Research, 19 (August), 57-62.
- Jacoby, Jacob (1976), "Consumer Psychology: An Octennium," in Annual Review of Psychology, Vol. 27, eds. M. Rosenzweig and L. Porter, Palo Alto, California: Annual Reviews, Inc., 331-358.
- Kelley, Harold H. and Michela, John (1980), "Attribution Theory and Research," in Annual Review of Psychology, Vol. 31, eds. M. Rosenzweig and L. Porter, Palo Alto, California: Annual Reviews, Inc., 457-501.
- Kiesler, Charles, Collins, B.E. and Miller, N. (1969), Attitude Change, New York: John Wiley & Sons.
- McCollum/Spielman & Company (1980), "Starpower," Topline, 2 (August).
- Mitchell, Andrew A. (1980), "The Use of an Information Processing Approach to Understand Advertising Effects," in Advances in Consumer Research, Vol. 7, ed. J. Olson, Ann Arbor: Association for Consumer Research, 171-177.
- Murray, Henry A. (1938), Explorations in Personality, New York: Oxford University Press, 88.
- Olson, J.C. and Muderrisoglu, A. (1979), "The Reliability of Responses Obtained by Free Elicitation: Implications for Measuring Attribute Salience and Memory Structure," in Advances in Consumer Research, Vol. 6, ed. W. Wilkie, Ann Arbor: Association for Consumer Research, 269-275.
- Rokeach, Milton (1970), Beliefs Attitudes and Values, San Francisco: Jossey-Bass, Inc.
- Sampson, Edward E. (1971), Social Psychology and Contemporary Society, New York: John Wiley & Sons.
- Sheth, Jagdish N. (undated manuscript), "The Surpluses and Shortages in Consumer Behavior Theory and Research."
- Shimp, Terence A. (1976), "Methods of Commercial Presentation Employed by National Television Advertisers," Journal of Advertising, 5 (Fall), 30-36.
- Wright, Peter L. (1973), "The Cognitive Processes Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-62.
- Zimbardo, Philip G. and Ruch, Floyd (1977), Psychology and Life, 9th edition, Glenview, Illinois: Scott, Foresman and Company.

LINEAR EFFECTS IN COGNITIVE RESPONSE TO ADVERTISING

Larry Percy, Creamer Inc.
Martin R. Lautman, Associates for Research in Behavior, Inc.¹

Abstract

This paper presents the results of an exploration of linear effects in cognitive response on cognitive structure. Specifically, elicited cognitive response to advertising stimuli were related to beliefs and purchase intent associated with the advertised product. A strong linear relationship was found between supportive arguing responses and related cognitive structure measures.

Introduction

Interest in the joint cognitive response-cognitive structure model for studying the impact of persuasive communication has grown since its early presentation by Lutz and Swasy (1977). A number of studies have been reported which attempt to link the number of cognitive responses to a communication stimulus (usually in the form of advertising) with a corresponding effect upon elements of cognitive structure. Olson, Toy, and Dover (1978), for example, studying, specifically, five levels of price information, demonstrated a strong relationship between the number of counterarguments and support arguments as related to the amount of information contained in several advertisements. Wilson and Muderrisoglu (1979, 1980) have utilized much the same procedure, applying the model to an analysis of comparative advertising. In each case, strong support for the notion of the number of cognitive responses elicited from an advertising communication stimuli mediating cognitive structure was shown.

All of this research, however, has been concerned with the total or average number of cognitive responses elicited by the communication stimuli. Accepting the basic hypothesis that there is a relationship on the macro level between the number of cognitive responses and the effect upon cognitive structure measures, it is the purpose of this paper to explore beyond the total effect and consider the functional relationship between different magnitudes of cognitive response upon cognitive structure. More specifically, what will be explored is whether respondents exhibiting greater cognitive activity, operationally defined as offering a greater number of cognitive responses, also exhibit a corresponding change in cognitive structure.

Methodology

Selection of Experimental Stimuli and Respondents

The methodology was designed to test the theory of individual intensity in cognitive response by utilizing both specific communication test stimuli as well as the general experience of respondents. The specific experimental stimuli used consisted of two television commercials for a well-known brand of frozen food, with a high positive image. Each commercial featured a variety of products: one with several people featured; one only a couple. In addition, one group of respondents was not exposed to a commercial and was asked only to think about advertising for the product with which they were familiar.

To ensure a realistic response, subjects were carefully screened to be certain only those people who belong to the product's target market, and who had experience with the product category, were included. A total of 300 respondents were interviewed (female household heads, 18+, with family incomes over \$15,000), 100 in each of three widely distributed geographical markets. One-third of the respondents in each market received each of the treatment conditions. Respondents in each city were randomly assigned to treatment cells. Those assigned to the first two cells were exposed to one of two television commercials. Each respondent was exposed independently, and asked to watch the commercial very carefully, learning everything from it that they felt the advertiser wanted them to know. Next, following the general procedure advanced by Greenwald (1968) and discussed by others (cf. Calder, Insko, and Yandell 1974; Wright 1973 and 1974), respondents were asked to relate all of their thoughts while viewing the commercial. Respondents in the third cell were not shown any specific advertising; rather, they were asked for all of the things that came to mind when thinking of looking at or reading advertisements for the product. In all treatment cells thoughts were elicited and recorded one at a time.

After all thoughts were recorded, a series of questions were asked relating to selected cognitive structure measures. Each respondent was asked how likely she would be to buy the product advertised, measured over a five-point Likert type scale; then how strongly she felt the product reflected specific beliefs associated with the advertising, using a seven point balanced attitude scale.

Analytic Plan

Initially, each of the cognitive responses elicited from a respondent were classified in the manner suggested by Wright (1973). Supportive arguments were classified as those thoughts, beyond a simple copy-point playback, which evidenced a favorable reaction to the product, its use, or the belief messages communicated. Counterarguments were those thoughts that were generally directed against the product or its use, and those that refuted message claims. In addition to these critical responses, three other categories were recorded: positive and negative ad-related thoughts which addressed only the executional elements, and neutral statements. Then, rather than averaging the total number of responses elicited for each classification, those respondents offering only one thought within a classification, two thoughts, etc. were grouped. In this way it is possible to preserve the frequency of individual cognitive responses by type of response.

Each of these distributions was compared over the three treatment cells. It was hypothesized that there should be no significant difference in the general distribution of cognitive response between cells, at least for the important supportive and counterarguments, since the two commercial stimuli used had been pre-tested and both communicated the same general message beliefs; and, importantly, these were the same message beliefs generally communicated through the advertising for this brand over the last several years (a campaign which included one of the two commercials tested here). As a result, those respondents not directly stimulated by a specific commercial should nonetheless have exhibited the same, or at least similar general cognitive structures to those exposed to the commercials.

¹¹The authors would like to thank Ms. Carol Farquhar, Northwestern University, for her assistance in designing and analyzing this study.

Incidence of responses were then compared with cognitive structure measures of belief and intention to buy in order to establish whether a linear relationship might mediate cognitive structure.

Results

Table 1 shows the mean number of cognitive responses offered by classification for each treatment cell studied. The high number of supportive arguments vs. other types of cognitive responses is, no doubt, a function of the existing highly positive image of the product.

TABLE 1
SUMMARY OF MEAN NUMBER OF COGNITIVE RESPONSES ELICITED

Classification	Commercial Cell #1	Commercial Cell #2	Non-cued Cell #3	% None
Supportive arguments	2.27	2.12	2.11	11
Counterarguments	0.09	0.09	0.30	86
Positive ad-related	0.50	0.25	0.07	82
Negative ad-related	0.27	0.20	0.01	89
Neutral	1.55	2.11	1.04	27

It is generally accepted (cf. Olson, Toy, and Dover 1978) that positive and negative ad-related responses, as well as neutral responses, do not mediate cognitive structure. For this reason, along with the fact that very few respondents provided positive or negative ad-related responses, only supportive arguments and counterarguments will be analyzed. Furthermore, because of the very low total number of counterarguments elicited, and the fact that 34 of the 41 respondents offering a counterargument gave only one, the statistical analysis will further be restricted to only the supportive arguments.

The distribution of supportive arguments between treatment cells is shown in Table 2. Little difference in either the overall distribution or average number of arguments elicited appears to be in evidence. A chi-square analysis of the distribution indicated no significant difference between the distributions ($X^2 = 4.90$, d.f. = 8, $p > .05$).

TABLE 2
DISTRIBUTION OF SUPPORTIVE ARGUMENTS

Number of Supportive Arguments	Commercial Cell #1	Commercial Cell #2	Non-cued Cell #3
One	19	24	21
Two	33	27	24
Three	23	19	17
Four	8	6	12
Five or more	8	9	9

These results would suggest that there is no significant difference in the general difference in the general distribution of cognitive responses (or at least, in this case, supportive arguments) among treatment cells. Therefore, the subsequent analysis permits us to look at total response over cells where appropriate.

Belief measures for six constructs reflected in the advertising for this product were measured via a seven point attitude differential scale. Mean beliefs averaged over the six constructs for respondents offering one through five or more supportive arguments are shown in Table 3. The one inversion, between three and four supportive arguments was primarily the result of one belief construct--the only one showing low overall belief.

TABLE 3
MEAN BELIEF BY INCIDENCE OF SUPPORTIVE ARGUMENTS

Belief	Incidence of Supportive Arguments				
	1	2	3	4	5+
Belief	2.15 ^a	2.26	2.44	2.38	2.63

^aMean values reflect seven-point differential scale where -3 = very unlikely to believe and +3=very likely to believe. A 3 x 5 ANOVA of this attitudinal data (see Table 4) indicated only Support Arguments to be statistically significant--this the result of a significantly increasing linear trend as shown in Table 3. As the number of support arguments offered by a respondent increased so did his positive attitudinal evaluation of the product.

TABLE 4
ANALYSIS OF VARIANCE FOR ATTITUDINAL DATA

SOURCE	SS	d.f.	MS	F	Sig.
Treatment Cells (T)	17.33	2	8.66	0.56	n.s.
Support Arguments (S)	199.75	4	49.94	3.22	0.04
linear	179.49	1	174.49	11.59	0.01
deviations from linear	20.26	3	6.75	0.44	n.s.
T x S	158.95	8	19.87	1.28	n.s.
Residual	3,995.58	258	15.49		

Finally, the intention to purchase also appears to be strongly mediated by the number of support arguments offered by a subject. Intention did increase linearly, as shown in Table 5.

TABLE 5
INTENTION TO PURCHASE

	Incidence of Supportive Arguments				
	1	2	3	4	5+
	3.86	4.02	4.08	4.15	4.42

^aMean values reflect 5-point Likert scale where 5.0 = Definitely buy and 1.0 = Definitely would not buy

An ANOVA of this data indicated that there was a significant difference of this particular measure between treatment cells, this was occasioned by a mean purchase intent of 3.84 in Cell 1 (those who saw one of the commercials) vs. 4.04 and 4.06 (among those seeing the second commercial and those seeing no commercial, respectively). One commercial was less effective than either the other commercial or the control in generating purchase interest. However, there was no interaction between this difference and the number of support arguments elicited; and, there was a significant linear trend in support arguments, as indicated in Table 6. As the number of support arguments offered by an individual respondent increased so did her interest in buying the product.

TABLE 6
ANALYSIS OF VARIANCE FOR PURCHASE INTEREST

SOURCE	SS	d.f.	MS	F	Sig.
Treatment Cells (T)	4.99	2	2.49	3.05	0.05
Support Arguments (S)	6.31	4	1.58	1.93	0.10
linear	5.81	1	5.81	7.09	0.01
deviations from linear	0.05	3	0.17	0.21	n.s.
T x S	6.51	8	0.81	0.99	n.s.
Residual	211.24	258	0.82		

Discussion

The results here indicate that there is a strong reason to suspect a linear effect mediating the relationship between cognitive response and cognitive structure. Overall belief increased with the number of support arguments offered, as did intention to purchase. Because of the high positive image nature of the product involved, one is tempted to believe these results are even more significant. Since both belief and interest are already high, to stimulate even greater interest as a function of advertising based cognitive response is surprising. Unfortunately, owing to that same high positive image, linear effects among counterarguments could not be tested because of sample size limitations. However, the trends, at least for purchase intent, suggested the predicted linear functional relationship. That is, more counterarguments resulted in lower purchase interest.

It is reasonable to predict that a similar linear relationship might be observed for products with less positive images. Here, however, one is equally likely to suspect that the quality of support arguments, not only their existence and frequency, might be of significance.

The elicitation measure used in this study differs considerably from that commonly employed in pretesting and on-air testing of advertising. Most typically, tabulations are obtained of the percentage of respondents recalling various messages. Occasionally, responses are tabulated by type of response (ad-related, sales point related, etc.). Rarely, if ever, is a coding scheme employing cognitive response categories used. Yet, the data presented here strongly suggest that the mean number of responses within cognitive response categories (a "recall-type" measure) may be a significant correlate of a motivational construct -- purchase interest. Whether these responses must be precursors of shifts in purchase interest, however, remains a critical question.

The findings of this study suggest that segmentation of groups by number of responses might reveal if there is a critical number of support arguments correlated with purchase interest; and/or, if any specific arguments or combinations of support arguments are integral to promoting purchase interest, given that the critical number of responses have been elicited. Additional commercial tests (with larger samples) would be necessary to explore these questions.

References

- Calder, Bobby J., Insko, Chester A., and Yandell, Ben. (1975), "The Relation of Cognitive and Memorial Processes to Persuasion in a Simulated Jury Trial," Journal of Applied Social Psychology, 4, 62-93.
- Greenwald, A.G. (1968), "Cognitive Learning, Cognitive Responses to Persuasion, and Attitude Change," in A.G. Greenwald, T.C. Brock, and T.M. Ostrom (Eds.), Psychological Foundations of Attitude, New York: Academic Press.
- Lutz, Richard J. and Swasy, John L. (1977), "Integrating Cognitive Structure and Cognitive Response Approaches to Monitoring Communications Effects," in W. Perreault, Jr. (Ed.), Advances in Consumer Research Vol. IV, 363-371.
- Olson, Jerry C., Toy, Daniel R., and Dover, Philip A. (1978), "Mediating Effects of Cognitive Responses to Advertising on Cognitive Structure," in H. Keith Hunt (Ed.), Advances in Consumer Research Vol. V, Ann Arbor, Michigan: Association for Consumer Research.
- Toy, Daniel R., Dover, Philip A. and Olson, Jerry C. (1977), "Communication Effects on Cognitive Responses, Cognitive Structure, and Attitude," Paper presented at the American Psychological Association 85th Annual Convention, San Francisco.
- Wilson, R. Dale, and Muderrisoglu, Aydin (1979), "An Analysis of Cognitive Responses to Comparative Advertising," Pennsylvania State University Working Series in Marketing Research, Number 84.
- _____, and Muderrisoglu, Aydin (1980), "An Analysis of Cognitive Responses to Comparative Advertising," in Jerry C. Olson (Ed.), Advances in Consumer Research Vol. VII, Ann Arbor, Michigan: Association for Consumer Research.
- Wright, Peter L. (1973), "The Cognitive Process Mediating Acceptance of Advertising," Journal of Marketing Research, 10, 53-62.
- _____, (1974), "On the Direct Monitoring of Cognitive Response to Advertising," in G.C. Hughes and M.L. Ray (Eds.), Buyer/Consumer Information Processing, Chapel Hill: University of North Carolina Press, 200-248.

ON EXPLAINING AND PREDICTING THE
EFFECTIVENESS OF CELEBRITY ENDORSERS

John C. Mowen, Oklahoma State University
Stephen W. Brown, Arizona State University

Abstract

Consumer behavior scholars and advertising practitioners are devoting increased attention to promotional endorsements. Building upon two theories from social psychology, balance theory and attribution theory, this paper examines a new approach to explaining and predicting endorser effectiveness. More specifically, the results are presented of an experiment investigating consumers' perceptions of endorsers who promote multiple versus single products and of products promoted by multiple versus single endorsers.

Introduction

Celebrity endorsers are regularly used to promote an endless list of products, including credit cards, automobiles, headache remedies, beer, and diet programs. A number of factors have been mentioned as important in using endorsers, such as the endorser's personal qualities--reputation, believability, likeability, etc. (Nelson 1974). With the large number of products endorsed by celebrities, a natural outcome is that some celebrities endorse a number of products, e.g., Jack Nicklaus, Joe Namath. However, as noted in a monograph published by the advertising firm, Foote, Cone, and Belding (1978), care should be exercised to select an endorser not tarnished by the association with other products. Of course, "virgin" endorsers tend to be highly expensive, leading to one practical question. Does the endorsement of multiple products indeed tarnish a celebrity's effectiveness? A second practical consideration involves whether or not to use multiple endorsers, such as in advertisements for Miller's Lite Beer and American Express Credit Cards. Are there advantages in having a number of different endorsers recommend a product to offset the cost of such an approach?

As proposed previously by Mowen, Brown, and Schulman (1979) a general communications model, based upon the integration of balance theory (Heider 1958) and attribution theory (Kelley 1967) can be used to analyze factors affecting the effectiveness of product endorsers. In the Mowen, et al. study, balance theory was used to define the relationships among the triad of cognitive elements of consumer (C), endorser (E), and product (P). According to the analysis, the consumer's perceptions of an endorser and a product represent sentiment or affective relations. The relationship between the endorser and the product represents a unit connection, or a perception by the consumer of the extent with which the endorser is associated or bonded with the product. Based upon the cognitive consistency analysis, an endorser will be maximally effective when both a strong sentiment relationship exists between the consumer and the endorser, and a strong unit relationship exists between the endorser and the product. With the development of strong positive sentiment and unit relations, the consistency forces are hypothesized to cause the consumer to cognitively reorganize the weaker consumer-product (C-P) relation so as to perceive the product more favorably (Abelson and Rosenberg 1958).

The extensive literature on communicator effects (Cohen 1964, Simons, Berkowitz, and Moyer 1979, Sternthal, Phillips, and Dholakia 1978) adequately details the factors influencing the consumer-endorser (C-E) sentiment relation.

However, surprisingly little attention has been devoted to delineating the nature of unit relations in general and the relation between an endorser and a product in particular. Following the suggestions of Mowen (In Press), Mowen, et al. (1979) investigated the application of attribution theory (Kelley, 1967) as a method of analyzing the unit connection between endorser and product.

From an attribution theory perspective, consumers may be conceptualized as seeking to determine the causal reasons for a celebrity endorsing a product. The question is, does the celebrity endorse the product because of his or her belief in its positive characteristics or because of external factors, such as monetary incentives? To the extent that consumers perceive that the endorser really believes in the product, the unit connection between endorser and product (the E-P relation) is hypothesized to be strengthened.

The use of attribution theory as an approach for understanding the nature of the unit connection is new. Despite the proposition's novelty, however, several factors suggest its application. First, the concept of association, delineated by Heider (1958) to describe the nature of the unit connection, is generally considered to be one of the fundamental elements in determining causality. Attribution theory was developed to predict individuals' perceptions of social causality. Therefore, it makes conceptual sense to use attribution theory as an approach to specify the nature of the unit connection. Second, that attribution theory and balance theory may be related should not be surprising for both were analyzed first by Heider in his seminal (1958) work.

To analyze the factors potentially influencing consumer perceptions of the E-P relation, one may utilize Kelley's (1967) attributional approach for individuals having multiple observations of an endorser's actions. Kelley argued that in circumstances in which an observer has multiple observations, three sets of attributional cues are utilized to establish causality--distinctiveness, consistency, and consensus. In an advertising context, distinctiveness refers to the extent that the endorsement occurs uniquely in the presence of the product. If a celebrity endorses several products, the relation between himself and a particular product is not distinctive, leading to an inference that the nature of the particular product was not the reason for endorsement. Thus, attribution theorists tend to agree with some professional advertisers, i.e., that endorsers can become tarnished by endorsing multiple products (Foote, Cone, Belding 1978). Consistency deals with the relationship between endorser and product over time and modality. High consistency will strengthen the perceived E-P bond. For this to occur, the endorser should advocate the product over a variety of media in a long running campaign. The third criteria, consensus, concerns the consumer's perception of whether other individuals view the product similarly to the endorser. To the extent which friends, reference others, or possibly other endorsers support the product, the consumer will perceive the message to result from the nature of the product rather than from situational factors such as a payment for services. Thus, by using multiple endorsers, advertisers may be effectively utilizing the concept of consensus.

Utilizing George C. Scott, Steve McQueen, and Paul Newman as endorsers of a fictitious television set, Mowen et al. (1979) manipulated distinctiveness and consensus information available to subjects. Distinctiveness was varied by informing subjects that the endorsers had signed contracts to endorse only one product--a new television set (high distinctiveness) or the television and four additional products (low distinctiveness). Consensus was varied by depicting the fictitious company as signing only one individual to endorse the product (low consensus) or signing six celebrities to endorse the television set (high consensus).

The results of the study supported neither the balance nor the attribution theory analyses. First, the manipulation of consensus and distinctiveness information failed to influence consumer perceptions of the causality for the endorsement (the endorser-product connection), thereby, not supporting the attributional analysis. With the lack of variations in the unit connections, according to the balance analysis, one would not expect the manipulations to influence the perception of the product. In addition, several complex interactions were found involving consensus information, distinctiveness information as well as the age of the consumer. Such interactions were also not supportive of the balance analysis.

The research presented in this paper conceptually replicates the Mowen et al. (1979) research. However, two important modifications were made. First, a new product was selected for endorsement. Consumers saw a storyboard of an advertisement for a fictitious new ball point pen--the "Easy Writer." Utilizing a television set as the product in the previous experiment could have resulted in consumers perceiving risk in allowing an advertisement to influence their perceptions of a relatively costly consumer good. Furthermore, in the earlier study, the actors (Newman, McQueen, and Scott) may not have been perceived by consumers to have much relevancy to a television set. By developing advertising copy in which the endorser discussed using the pen to sign autographs, it was hoped that greater relevancy would be established. A second modification in the present study was the addition of conditions in which the endorser was said to have signed contracts to endorse products normally used by higher social status individuals. In the earlier study, the products were relatively low status--motorcycles, cigars, cologne, and rums. In the present study, higher status products were utilized (i.e., Hilton Hotels, Chivas Regal Scotch, and Mercedes Benz Automobiles) as well as the lower social status products.

The experiment emerging from these considerations was developed as a 3x3, full factorial, between subjects design. The three levels of distinctiveness were: high distinctiveness (the celebrity endorsed only the pen), low distinctiveness-low social status product, and low distinctiveness-high social status product. The three levels of consensus information were: low (no others endorsed the pen), high consensus-celebrity (Steve McQueen also endorsed the pen), and high consensus-student (a student also endorsed the pen).

Method

Procedure

Ninety-nine individuals of both sexes served as subjects. Subjects were students interviewed in their dorm rooms at a large midwestern university. Students were selected as subjects for two reasons. First, students are a primary target market for producers of ball point pens. Second, the use of a relatively young, homogeneous population allowed a more precise test of our hypotheses. In previous research, Mowen et al. (1979) found interactions between age and the variables of consensus and distinctiveness.

In order to test for other possible explanations for the results of the Mowen et al. (1979) study, it was important to hold age relatively constant in order to avoid a larger, more cumbersome experimental design.

Subjects were approached in their dorm rooms by a same sexed experimenter. The experimenter would briefly state that a new advertising campaign was being tested, and ask if they could help out. All subjects agreed to help and were then handed an experimental booklet. On the first page subjects read:

We are interested in obtaining your opinions concerning a particular test advertisement. You will be shown an advertisement and then asked several questions concerning your reaction to the ad and to the particular product described in the ad.

One important point to remember: since this is a test advertisement, it is presently in rough draft form. We ask that you keep in mind the fact that if this ad were to appear in a magazine, it would be in color and professionally laid out. Everything else, however, would remain exactly as you see it. Do you have any questions?

On the second page of the experimental booklet, subjects read material which was labeled background information. The experimenter explained that since the subject was receiving only a draft of the proposed magazine ad, he or she should get additional information. All subjects read the same first paragraph:

This advertising campaign has a total budget of \$1,200,000. The "Easy Writer" will be distributed in all 50 states and in ten foreign countries through better stationery and department stores.

Subjects then read that Paul Newman had signed a two-year contract to endorse the pen. Next, the manipulation of distinctiveness was made. In the high distinctiveness condition subjects read that "Paul Newman has refused to endorse any products other than the "Easy Writer". In the low distinctiveness, low status conditions, subjects read: "Paul Newman has also signed contracts to endorse Puerto Rican Rums, Grendiers Cigars, Yamaha X5650 motorcycles, and Brut Aftershave Cologne." In the low distinctiveness, high status conditions, subjects read: "Paul Newman has also signed contracts to endorse Hilton Hotels, Mercedes Benz Automobiles, Chivas Regal Scotch, and Botany 500 suits."

Subjects were next shown the 11 x 14 inch storyboards. The manner in which the storyboards were shown constituted the manipulation of consumer information. In low consensus conditions, subjects saw only one storyboard having a large photograph of Paul Newman on it. To the side of the photograph, the below standard endorsement was printed:

PAUL NEWMAN does. Why?

"I've signed a lot of autographs during my career. But whether I'm signing autographs or writing a check or letter, I require a superior writing instrument. And Easy Writer is very definitely a superior writing instrument."

Your Easy Writer will never skip, never stick. It is so well engineered, that it can't leak or blot, either on your fingers or on your paper. The Easy Writer flows with such ease and comfort that it actually lessens writing fatigue.

And since it's refillable, you can expect to sign your name again and again and again.

Ask for the Easy Writer in any good art, stationery or college bookstore.

In the "high consensus-celebrity" conditions, subjects saw the Paul Newman storyboard and in addition saw storyboards with pictures of Steve McQueen and George C. Scott. The same standard endorsement was printed for each of the celebrities. In the high consensus, average person conditions, subjects saw the Paul Newman storyboard plus two storyboards, each having a different picture of two male college students. To the right of the photo of the student with pen in his hand, the copy was placed.

STEVE SMITH Does. Why?

"Being a college student requires a lot of writing everyday. But whether I'm taking notes, an exam, or writing a check, I require a superior writing instrument. And Easy Writer is very definitely a superior writing instrument."

Your Easy Writer will never skip, never stick. It is so well engineered, that it can't leak or blot, either on your fingers or on your paper. The Easy Writer flows with such ease and comfort that it actually lessens writing fatigue.

And since it's refillable, you can expect to sign your name again and again and again.

Ask for the Easy Writer in any good art, stationery or college bookstore.

Dependent Variables

Eleven dependent variables were taken assessing subjects' impressions of the product, the advertisement, and the major celebrity endorser--Paul Newman. Concerning the product, subjects were asked: "How favorable an opinion do you have of the Easy Writer?", and "If you were in the market for a pen, would you buy the Easy Writer?" Concerning the advertisement, subjects were asked, "What is your overall reaction to the advertisement?" In addition, subjects rated the advertisement on a series of five semantic differential scales: interesting-dull, appealing-unappealing, believable-unbelievable, informative-uninformative, and eye-catching-not eye-catching. Concerning their impression of Paul Newman, subjects were asked: "Is Paul Newman doing the advertisement solely for the money or because he really believes in the Easy Writer?", "To what extent does Paul Newman believe in the Easy Writer?", "How likeable is Paul Newman?" and "How much do you trust Paul Newman?" All dependent measures were assessed on seven point scales anchored with appropriate descriptors. (Lower scale values indicate a more positive impression.)

Results

Subjects' perceptions of Paul Newman were analyzed first. For the question asking subjects to indicate their liking for the actor, one significant main effect was found for the consensus variable, $F(2,90) = 3.03, p < .055$. The pattern of means revealed that subjects liked Newman most when the student endorsers were also shown ($M = 1.45$) and liked Newman less when his endorsement appeared alone or in conjunction with Scott and McQueen (M , alone = 2.45; M , celebrity endorsers = 2.64).

On the two questions asking for subjects' reactions to the "Easy Writer Pen", similar effects were found. When asked for their opinion of the pen, $F(2,90) = 4.3, p < .02$ and whether they would buy the pen, $F(2,90) = 3.7, p < .05$, a main effect for distinctiveness occurred. In each case, one finds that subjects reacted more favorably to the pen in high distinctiveness ($M = 2.09$ for opinion and $M = 2.18$

for buying likelihood) than in the two low distinctiveness conditions (Opinion-HSS $M = 2.75$, Opinion-LSS = 2.56; Buy-HSS = 2.85, Buy-LSS = 2.99).

Of the six questions asking subjects to give their impressions of the advertisement, two significant effects were found, each paralleling those found directly above. On the question asking for the subjects' overall reaction to the advertisement, a main effect was found for distinctiveness $F(2,90) = 4.12, p < .02$, revealing that impressions were most favorable in the high distinctiveness conditions. (M high distinctiveness = 2.42, M HSS = 3.12, M LSS = 3.03). Similarly, subjects found the advertisement more interesting in the high distinctiveness conditions (M high distinctiveness = 2.60, M HSS = 3.15, M LSS = 3.15).

On the question asking subjects to indicate the extent which Newman believes in the product, a significant consensus by distinctiveness interaction occurred, $F(4,90) = 2.87, p < .03$. Figure 1 reveals the nature of the interaction. (Table 1 presents the means for all of the dependent variables.) As shown by Figure 1, the interaction resulted from the pattern of the "low distinctiveness-high SS product" conditions diverging substantially from the two remaining distinctiveness conditions. Thus, one finds that in the low distinctiveness-high SS conditions, subjects viewed Paul Newman as believing in the pen in "low consensus" conditions. However, in the "high consensus-celebrity" condition, belief decreases and in the "high consensus-student" conditions, perceptions of his belief drop substantially. This pattern diverges from that of the "low distinctiveness" conditions and "low distinctiveness-low SS" conditions, in which the perceptions of belief increased in the "high consensus-student" conditions. No significant effects were found for the variables assessing subjects trust in Paul Newman. On the question asking whether Newman was doing the advertisement for the money or for the product, a marginally significant interaction occurred, $F(4,90) = 2.13, p < .10$, which generally followed the same pattern as the belief variable.

Discussion

The results of the experiment generally supported the predictions for the effects of distinctiveness information, as derived from attribution theory. When the celebrity endorsed only one product, the product was seen more favorably, individuals indicated a greater interest in buying the product, subjects viewed the advertisement more favorably, and subjects responded more positively to the promotional message. These important distinctiveness effects show the value to an advertiser of establishing an exclusive agreement with a celebrity.

While generally supporting attribution theory, the results for distinctiveness information failed to support the balance theory analysis. According to the analysis, perceptions of the product are hypothesized to change as a result of a highly positive endorser being placed in a unit connection with the product. As such, high distinctiveness information should first influence subjects' perceptions of the connection between endorser and product and their view of the product. (Perception of the unit connection were operationalized as the perception that the endorser believes in the product and that the endorser is doing the endorsement for the product, not for the money). However, these perceptions were not influenced in the same manner by variations in distinctiveness information as were variables assessing perceptions of the product. Therefore, as with the Mowen et al. (1979) study, the balance theory analysis was not supported.

Turning to the results for consensus, significant main effect was found in which subjects liked Newman most when a student endorser also recommended the product.

The student subjects could apparently relate more to Newman when people like themselves (i.e. student actors) were also favorable toward the same product he was endorsing.

A complex pattern emerged in the consensus by distinctiveness interaction for the dependent variable of Newman's belief in the pen. (See Figure 1.) For high status products, Newman's association with the student endorser resulted in lower believability than in other conditions. This result suggests that advertisers should be careful not to associate a celebrity, who typically endorses high status products, in the same advertising campaign in which the "masses" also endorse the same product. In the high distinctiveness condition, a similar pattern of results occurred for the believability measure in consensus and in high consensus, celebrity endorser conditions. Higher believability resulted when Newman endorsed only the pen, and the pen was also endorsed by students. This finding supports both the distinctiveness and consensus predictions and suggests that an "exclusive" celebrity can be effectively used with lay endorsers whose characteristics (e.g. age, occupation) are similar to the target audience.

In the low distinctiveness-low social status conditions, the association of Newman and the students with the pen produced more believability for the celebrity, even when Newman was also endorsing other products. This suggests that even multiple product celebrity endorsers associated with low status products can have relatively high believability if linked with "lay endorsers" similar to the target audience.

The results from the present research differ from the earlier Mowen, et. al. (1979) study. The 1979 study supported neither the balance nor attribution theory analyses. The study reported in this paper, however, provides evidence that distinctiveness information does provide relevant input which consumers utilize in forming or not forming unit connections between the endorser and the product. Nevertheless, the complex interactions involving the types of consensus and distinctiveness information were found. In general, these interactions were not supportive of the theoretical propositions.

Conclusions

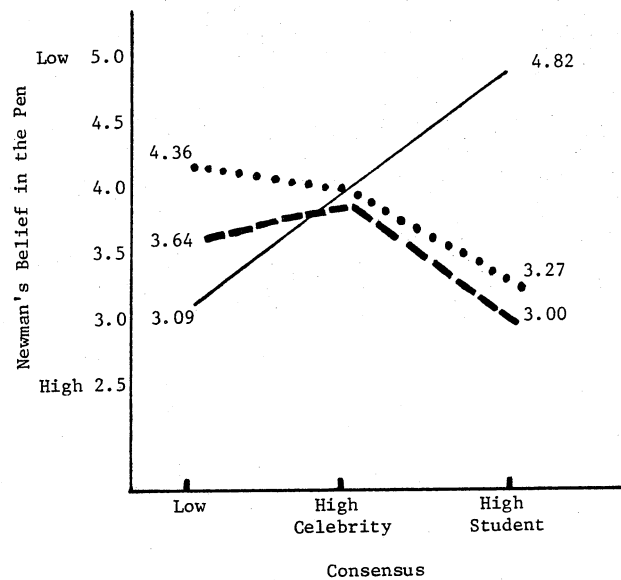
The results of this study provide some guidance for marketers, advertisers, and public policy makers. Each group should conduct careful pretesting of promotional messages to assess, a) whether the audience perceives the product more favorable when one versus multiple endorsers are used, b) audience perceptions of both the endorser and of the other products he or she endorses, and c) whether alternative market segments perceive distinctiveness and consensus information differently.

This study's findings provide a clearer answer to the question of whether an integration of balance theory and attribution theory adequately explain the effects of celebrity endorsers than the earlier study by Mowen, Brown, and Schulman (1979). This is not startling since the more recent study benefited from the earlier experience. At this point in time, little evidence exists that balance theory contributes significantly to our understanding of celebrity endorsers. According to the balance analysis, distinctiveness and consensus information are hypothesized to affect first the unit connection between endorser and product. Secondly, the distinctiveness and consensus information affect perceptions of the product. The results in the present study and in the Mowen et al. (1979) research reveal, however, that one consistently finds that distinctiveness and consensus interact to influence the unit connection measure but influence the product ratings in a non-interactive manner. Such results cannot be predicted from the balance theory approach.

The research does indicate greater applicability of concepts derived from attribution theory and in particular the distinctiveness concept. A lack of support exists, however, for predictions based upon the consensus formulation. It may be that in advertising contexts, attempts to create the impression of consensus are doomed to failure. In order for consumers to gain an impression of consensus, it is probably necessary that endorsements appear independent to consumers. That is, the endorsers need to have the appearance that each acted independently of the other and were not "bought off" by the company. However, with consumers increasingly becoming wary of advertisements, the ability to create the impression of the independence of endorsers may be impossible, even if it in fact exists.

Despite the completion of two related studies, additional work is certainly needed on the topic of endorser effectiveness. Such studies could introduce additional product types, new endorsers, non-student subjects, and different operationalizations of distinctiveness, consensus, and consistency information. Further research, including replications with extensions, will help consumer research scholars and communication practitioners better understand and measure endorser effectiveness.

FIGURE 1
3-Way Interaction Depicting
Newman's Belief in the "Easy Writer" Pen



Distinctiveness*

- - - High distinctiveness
- . . . Low distinctiveness, LSS product
- Low distinctiveness, HSS product

*LSS = low social status products, HSS = high social status products.

TABLE 1
Mean Responses for the Major Dependent Variables

Dependent Variables ²	Independent Variables ¹								
	A1 B1 ³	A1 B2	A1 B3	A2 B1	A2 B2	A2 B3	A3 B1	A3 B2	A3 B3
Overall Ad Reaction	3.36	3.18	2.45	3.27	3.00	2.64	2.73	3.00	2.18
Interesting - Dull	3.27	3.54	2.36	3.36	3.09	2.45	2.82	3.27	3.00
Appealing - Unappealing	3.18	3.27	2.45	3.27	2.91	2.36	2.73	2.90	2.64
Believable - Unbelievable	2.90	3.64	3.00	3.45	3.18	3.00	3.27	2.82	2.45
Informative - Uninformative	2.45	3.09	2.82	3.27	2.82	2.18	2.91	2.27	2.45
Eye Catching - Not Eye Catching	3.00	3.36	2.09	2.82	2.82	2.27	2.18	2.91	2.73
Opinion of Easy Writer	2.64	2.36	2.90	3.00	2.91	2.36	2.64	2.45	1.82
Buy Easy Writer	2.45	2.82	2.00	3.09	3.36	2.27	3.00	2.82	2.27
Newman Doing For Money or Belief In Pen	4.09	5.36	4.36	5.09	4.36	4.90	4.82	3.73	3.54
Newman Likeability	2.45	3.09	1.73	2.64	2.27	2.27	1.45	1.90	2.00
Newman's Belief in Easy Writer	3.09	4.36	3.64	3.91	3.91	3.73	4.82	3.27	3.00

¹A1 = Newman Alone, A2 = Newman Plus Celebrities, A3 = Newman Plus Students

B1 = Low Distinctiveness - High Social Status, B2 = Low Distinctiveness - Low Social Status,

B3 = High Distinctiveness

²Low scale values indicate a more positive impression.

³N = 11 in each cell.

References

- Abelson, R.P. and Rosenberg, M.J. (1958), "Symbolic Psychology: A Model of Attitudinal Cognition," Behavioral Science, 3, 1-13.
- Cohen, A.R. (1964), Attitude Change and Social Influence, New York: Basic Books.
- Fireworker, R.B. and Friedman, H.H. (1977), "The Effects of Endorsements on Product Evaluation," Decision Sciences, 8 (July), 576-583.
- Foote, Cone, and Belding (1978), Profiting from Celebrity Endorsers, New York: Foote, Cone, and Belding Communications, Inc.
- Heider, F. (1958), The Psychology of Interpersonal Relations, New York: John Wiley and Sons, Inc.
- Kaplan, G. (1978), Profiting from Celebrity Endorsers, New York: Foote, Cone, and Belding Communications, Inc.
- Kelley, Harold H. (1967), "Attribution Theory in Social Psychology," in Nebraska Symposium on Motivation, David Levine, ed., Lincoln: University of Nebraska Press, 192-238.
- Mowen, John C. (In Press), "On Product Endorser Effectiveness: A Balance Model Approach," Current Issues and Research in Advertising--1980, James H. Leigh and Claude R. Martin, Jr. (eds).
- Mowen, John C., Brown, Stephen W., and Schulman, Meg (1979), "Theoretical and Empirical Extensions of Endorser Effectiveness," 1979 Educators' Conference Proceedings, Neil Beckwith et al., editors, Chicago: American Marketing Association, 258-262.
- Nelson, Alan R. (1974), "Can The Glamour and Excitement of Sports Really Carry the Ball for Your Product?" Marketing Review, 29 (February), 21.
- Simons, H., Berkowitz, N. and Moyer, R. (1970), "Similarity, Credibility, and Attitude Change: A Review and Theory," Psychological Bulletin, 73, 1-16.
- Sternthal, B., Philips, L.W. and Dholakia, R. (1978), "The Persuasive Effect of Source Credibility: A Situational Analysis," Public Opinion Quarterly, (Fall), 285-314.
- Winer, B.J. (1971), Statistical Principles in Experimental Design, New York: McGraw Hill Book Company.

GUILT AROUSING MARKETING COMMUNICATIONS: AN UNEXPLORED VARIABLE¹

Morry Chingold, Pennsylvania State University

Abstract

The relevance of guilt arousing marketing communications is developed and a rationale for investigation proposed. A brief review of the literature and a discussion of the possible inadequacy of related paradigms follows. A "working draft" conceptual model and several individual difference variables believed to be dominant are then presented. Several research questions in the form of broad hypotheses and possible approaches to measurement are offered to guide preliminary research.

Introduction

Advertising and promotion continue to be central topics within the field of marketing for practitioners and researchers alike. One avenue of inquiry has focused upon the use of emotional appeals in persuasion. Past research dealing with emotional appeals in persuasion has been multidisciplinary, attracting the interests not only of the scientific community but of special interest and regulatory agencies in the public sector as well.

As has been the case in most other fields, the bulk of research within marketing has centered upon fear arousing communications and the relationship between the level of fear and subsequent shifts in attitude or behavior. Notwithstanding a stream of research which has spanned three decades, the evidence to date has remained equivocal. This may be due in part to disparities in conceptualization or operationalization (Higbee 1969, Ray and Wilkie 1970, Sternthal and Craig 1974).

Surprisingly, however, much less attention has been afforded other plausible emotional appeals. Roseman (1979) has identified 13 distinct emotions believed to span the affective spectrum and many of these (e.g. anger, hope, guilt) seem well suited to persuasive communications. Within a marketing context, a cursory inspection of advertising themes will probably disclose advertising appeals designed to arouse or exploit a variety of emotional responses. One such emotion may be guilt.

The concept of guilt has long been dominant in such diverse fields as theology, philosophy, and counseling and psychiatry, disciplines in which guilt and guilt arousal are viewed as effective agents of attitude and behavior change (see Schneiders 1967, 1968; Stein 1968). However, it would seem that only marginal effort has been expended in examining the role of guilt in persuasive processes, a field of inquiry central to marketers and marketing theorists.

Research scarcity alone, however, is insufficient cause to merit investigation of guilt arousing communications. Preliminary interest could focus on differences existing between guilt arousing communications and other persuasive communications which also use emotional appeals but may differ theoretically, as may be the case with fear appeals.

Another incentive for investigation of guilt arousing communications, albeit less compelling, might be the sheer number of advertisements which appear to have the intention of arousing some degree of guilt in one form or another

(for some possible examples of guilt arousing communications in print media see Time 1980a, 1980b, 1980c). Wheatley and Oshikawa (1970) suggest that advertising practitioners tend to prefer positive appeals over negative or anxiety arousing appeals such as fear or guilt. In light of this preference, an impartial observer might cautiously surmise that at least some advertisers believe negative appeals to be effective for their own purposes considering the vast resources allocated for production and dissemination of these messages.

Although perceived (by the author) to be in fairly widespread use in marketing communications, guilt appeals have thus far not been investigated within the discipline. The lack of conceptualization, theory, and research has prevented greater understanding of this enigmatic marketing variable. This paper, therefore, has three purposes: 1) to introduce a new area of interest, 2) to offer a review of the literature dealing with guilt arousing persuasive communications, and 3) to attempt to conceptualize this construct and suggest research on guilt arousing marketing communications.

Research Overview

The purpose of this section is to follow the course of research which has dealt with guilt arousing communications, a particular form of negative-emotional/anxiety arousing communications. Following sections will attempt to differentiate between guilt and other negative-emotional appeals, of which fear has been singularly predominant, and to begin development of a "first draft" conceptual model.

A major obstacle which has hampered the development of a clear linkage between guilt appeals and research in the area of negative-emotional or anxiety arousing communications has been the concentration of effort in the domain of fear research. Fear appeals research prospered for more than two decades, and clearly has been the preferred negative emotion or anxiety for investigation. This state of affairs has best been exemplified by Leventhal and Tremblay (1968). In their article "Negative Emotions and Persuasion," the authors restricted their experimentation and discussion to fear alone, rather than other plausible negative emotions which may have persuasive potential such as guilt, anger or sorrow. Whether these authors believed that a fear paradigm was sufficient to explain the relationship between all negative emotions and persuasion or that all negative emotions were either interchangeably similar with or a subset of fear remains unclear. Perhaps in light of the absence of competing paradigms for other emotions or anxieties their approach was totally appropriate. It should, however, have signaled the need for conceptualization and research in new areas.

Wheatley and Oshikawa (1970) reported on the relationship between anxiety and positive and negative advertising appeals. The authors stated that researchers in the social sciences had compiled evidence sufficient to attest to the power of negative appeals and suggest that advertisers may be overlooking or minimizing an important persuasive tool. Citing past research the authors proposed that "negative communication can be effective in inducing the behavior advocated by the communicator (pg. 85)." Applying a general anxiety reduction drive model, Wheatley and Oshikawa expected that reduction of an aroused anxiety would constitute an effective reinforcement and could lead to conformity with

¹Special thanks to Dr. R. G. Roden and Peter H. Roden, University of Texas Medical Branch, and all my colleagues at Penn State for their help in the preparation of this paper.

a communicator's recommendation. This general model would seem appropriate for a variety of negative emotions or anxieties but as has often been the case only a fear manipulation constituted the negative appeal.

In contrast to many other streams of research, investigation of guilt arousing communications has not evolved through a tradition of theory building. Rather, the emphasis on fear related approaches to research in the areas of anxiety or negative appeals is also evident in the literature on guilt arousal and persuasion. Of the few investigators who have approached the issue of guilt arousing communications, most have seen fit, rightly or wrongly, to graft their research onto existing fear appeal paradigms (Haefner 1956, Zemach 1966, Yinon et al. 1976). So engaged, these researchers could hope for no more than a general level hybrid theory of guilt arousing communications juxtaposed with fear research.

More to the point, attempts at hybrid guilt theory building involved optimally matching empirical findings to existing fear models. The earliest work, by Haefner (1956), involved a manipulation of guilt in two levels, high and low, perhaps in an effort to replicate the "inverse relationship" in fear research reported by Janis and Feshback (1953, 1954). One may interpret Haefner's approach as viewing guilt as a subset or alternate form of fear. Haefner was exploring negative emotional arousal, operationalized as fear and guilt (both of which were significant), then comparing his findings to the earlier Janis and Feshback (1953, 1954) results.

Subsequent investigation of guilt arousing communications reflected changes in the dominant fear paradigm. Zemach (1966) undertook an experimental investigation of the effects of guilt arousing communications, however, the thrust of her research "piggybacked" on fear research. "In view of the paucity of observations and systematic research on guilt arousal, this study was undertaken to investigate whether principles derived from the literature on fear arousing communications might be fruitfully employed in predicting acceptance of recommendations motivated by guilt arousal (pg. 4)." Thus, Zemach's two fundamental hypotheses (the curvilinear "inverse U" relationship between level of guilt arousal and adherence to recommendations, and the dissipation of the differential effects of difference levels of guilt arousal with the lapse of time) were derived from fear research rather than a priori conceptualization. Although both hypotheses were supported, it remains unclear which, if any, theory was supported, the existing fear model or a unique guilt model.

More recently, Yinon et al. (1976) reported an experimental investigation of the effects of guilt arousing communications on acceptance of behavioral recommendations. Citing the competing fear paradigms of Janis (1967) and Leventhal (1970) as evoking "theoretical curiosity of the social psychologists," Yinon et al. sought to test Janis' (1967) contention that the curvilinear "inverse U" model also applied to guilt arousing appeals. Although the curvilinear model was supported, Yinon et al. failed to measure any individual difference variables, as Zemach (1966) had done. As well, Yinon et al. reported operational difficulties in effectively manipulating three levels of guilt, leading to a posteriori collapsing of the design, and casting at least some shadow of doubt on the experimental findings.

Kelman (1979) discussed the relationship between various emotional states and attitude change. Although Kelman (1979) did not specifically address guilt arousing communications, he did propose that guilt, as an emotional response, was among the most likely moral dilemmas to lead to attitude change. Changes in attitude may occur toward the object or issue involved in an action or toward the action itself. Kelman proposed that when justifying an action or making it more acceptable were not available as possible changes in attitude, individuals must confront the dis-

crepancy between their actions and role expectations and "may actually change their attitudes in the direction of more favorable evaluation of the object or issue involved in the action (pg. 74)." Kelman stated that these notions were derived from dissonance theory.

In summary, the paucity of guilt appeal research reported by Zemach (1966) has continued to the present. Clearly a stream of research founded upon three studies conducted over a period of twenty years could not be expected to have nurtured much theory. Rather, that research which has been reported has really been in the domain of fear research. Although one cannot dismiss this approach out of hand, the fact that guilt arousal has not received attention as a unique construct unto itself is disappointing. In light of the differences between guilt and fear to be developed below, this paper seeks to argue in favor of viewing guilt arousing communications as a distinct persuasive mechanism employing an emotional appeal theoretically different than fear. As such, a "first draft" conceptual model specific to guilt arousing communications will be proposed.

Toward a Conceptualization of Guilt Arousing Persuasion

Guilt as a Psychological Construct

As a psychological term, English and English (1976) define guilt as "a realization that one has violated ethical or moral or religious principles, together with a regretful feeling of lessened personal worth on that account." Wolman (1973) added that with the lowering of self esteem or feeling of lessened personal worth also comes a need to make retribution for the transgression. So defined, guilt may be viewed as an a posteriori emotional response which follows a particular action or thought, what Wolman (1973) called the transgression. The need to make retribution may be considered as one possible outcome resulting from a desire to reduce the level of guilt feelings to a more tolerable level, assuming some idiosyncratic threshold had been surpassed.

Unlike most other negative emotions guilt may apply equally to positive and negative outcomes (Roseman 1979). Thus, someone who shoplifts in public but is not apprehended may experience guilt even though the outcome of the act was essentially positive. Roseman maintains that both forms of guilt are phenomenologically, physiologically, and behaviorally alike inasmuch as all "guilty" people have done something "wrong" and can expect to be sanctioned by others.

Fear, in contrast, is more an anticipatory emotional response. Fear is the anxiety brought about by anticipated consequences to particular actions or cognitions, in other words an a priori emotional response. Fear results from the perception of some danger and brings about the goal of avoiding or preventing the threatened outcome. Guilt, on the other hand, arises from the perception that an individual has him/herself committed some injustice and prompts behavior designed to undo or atone for the offense (Roseman 1979). Since guilt is brought about by the actual consequences of behavior and fear is caused by the anticipated consequences of behavior, theoretically the two forms of anxiety should be different. Thus, guilt should be considered as a persuasive construct unto itself.

There is, however, a "fuzzy" area of overlap apparent between these and other forms of anxiety. Situations may arise in which it is unclear which emotion is operative or central when an individual senses "fear of guilt" or "guilt from fear" for example. Whether the response is a priori or a posteriori depends on the interpretation of the situation. "Ring around the collar" may be viewed as a fear appeal, causing one to seek to avoid or prevent this unwanted outcome, or alternatively as a guilt appeal, prompting the launderer to undo or atone for his/her error. The interrelatedness of anxieties or negative emotions is the

cause of the overlap. Roseman (1979) has identified 5 "dimensions" that give rise to emotions and relate discrete emotions to one another in a structural model of affect. Roseman's model of affect can be used to account for perceived similarities among various emotions and predict patterns of co-occurrence and sequential occurrence in complex emotional reactions.

Guilt as Cognitive Dissonance

As noted earlier, Kelman (1979) argued that guilt was a powerful agent in attitude change. Kelman referred to guilt as one of a number of moral dilemmas. In the case of guilt, attitude change would be in the form of retrospective justification for an action in the same manner postulated by dissonance theory (Festinger 1957).

Dissonance theory (Festinger 1957) can easily be adapted to the problem of guilt. Feelings of guilt are, for most individuals, clearly dissonant cognitions creating inconsistency in some cognitive and/or perceptual elements. The two central hypotheses of dissonance theory are 1) the presence of dissonance motivates the individual to reduce it, thereby achieving consonance, and 2) when dissonance is present the individual will avoid situations and behavior which will increase it.

There are some problems, however, in applying dissonance theory to the problem of guilt arousing marketing communications. Firstly, dissonance theory is very broad in nature and has been modified by a number of researchers who have specified particular conditions under which dissonance will or will not occur (Calder 1973). "The real problem has been in tying down the theory. The very looseness which invites creative derivations renders the theory all but impossible to disprove (Calder 1973, pg. 259)." Thus the broad nature of dissonance theory may render it too general to effectively deal with guilt arousing communications at all levels of analysis. Rather, individual responses to guilt appeals represent but a particular form of cognitive dissonance, in much the same way that other anxiety arousing communications and theories of cognitive consistency are related to the general theory of cognitive dissonance.

A second difficulty rests in measurement and operationalization issues. Given the voluminous research in the area of cognitive dissonance, it would seem that there has not been a generally practiced or accepted way of effectively manipulating and measuring varying levels of "aroused" cognitive dissonance, due in part to the abstractness of the construct and in part to its general nature and vagueness since cognitive dissonance can deal with any cognitive elements. In dealing with a particular form of dissonance, such as guilt arousing communications, it may prove easier to variably manipulate the construct and obtain measures of arousal. Clearly measurements in the domain of an anxiety such as guilt cannot achieve perfection, but they may be easier to fashion because of the narrower and better defined boundaries of the construct of interest. (Several possible methods for measuring aroused guilt will be discussed briefly later in the paper.)

If guilt and cognitive dissonance are but similar constructs, and not in fact one and the same, there may be marked differences between an individual's threshold for cognitive dissonance and his/her threshold for aroused guilt, or in the intensity of the individual's drive to return to the desired mental state. Similarly, one might anticipate different resolution strategies for cognitive dissonance and aroused guilt. Although the essential notion is the same, resolving the undesired mental state (dissonance or guilt), the methods available and selected may not be interchangeable or even compatible.

In general, although guilt arousal is a mental state which behaves similarly to the stipulates of dissonance theory, it is unique at the same time. More to the point,

dissonance theory is so broad and malleable that it may be adapted to encompass many diverse phenomena since it is concerned with any and all cognitions. Understanding dissonance theory may be viewed as necessary but not sufficient for understanding guilt. The specificity and particularities involved require that guilt be studied as a separate entity even though clear linkages may be drawn to dissonance theory.

Guilt as Information Processing Conflict

Investigation of guilt appeals may also view aroused guilt as a form of conflict in information processing. Berlyne (1957) described conflict as the result of competing and incompatible response tendencies. Thus, in the case of choosing between a response tendency of self gratification versus altruism, (i.e., dining out instead of renewing one's donation to the Red Cross) the conflict which may subsequently occur may be analogous to guilt. Often, the guilt or conflict may not even be known to exist until aroused, perhaps as by a communication in a marketing context.

Bettman (1979) cites Hansen (1972) and Howard and Sheth (1969) in suggesting that conflict be viewed as an equilibrium model. Individuals would be viewed as having desired optimum levels of conflict, above or below which undesired disequilibrium would occur. As with cognitive dissonance, the notion of conflict represents a form of inconsistency which would lead to a response aimed at its resolution. Kelman and Baron (1968) proposed a framework of alternative modes for resolving inconsistencies, as those created by conflicts, as well as hypotheses concerning the conditions under which the various strategies might be used.

Although guilt is clearly a form of conflict arising from competing response tendencies, a conflict resolution model would seem not to be generalizable to all guilt arousal situations. However, there is some similarity between certain forms of guilt and the special cases of avoidance-avoidance conflicts and approach-avoidance conflicts. The overlaps between guilt and information processing conflict occur, however, only with special cases of each and not as a general principle and the two are not interchangeable since approach-approach conflict is not analogous to guilt. As well, the notion of an equilibrium model for conflict seems somewhat counterintuitive when generalized to guilt. As a rule it would probably be safe to say that few individuals seek to increase their feelings of guilt because their current mental state is too free to guilt. Rather, a threshold effect appears more logical.

As was the case with cognitive dissonance theory, there are clear linkages between conflict in information processing and a conceptualization of guilt arousal in persuasion. However, there also appear to be several clear distinctions, such as the equilibrium effect and inapplicability of positive response tendencies, and it would be inappropriate to assume that information processing conflict and guilt arousal are one and the same phenomenon. Rather, in specific situations and under particular conditions they are quite similar.

In light of the dubious piggybacking on fear research which has occurred in the past, and the apparent inappropriateness of cognitive dissonance and information processing conflict as fully viable theoretical explanations, there is a need for a unique conceptualization concerned with guilt arousal in a persuasive context. A "first draft" of such a paradigm will now be proposed.

A "First Draft" Conceptualization of Guilt Arousing Communications

Given the definition of guilt as a psychological construct noted earlier, several constructs are implied which must be incorporated into a model of guilt appeals. Firstly, a hypothetical mental state of aroused guilt must be assumed

to exist. Implicit in such a construct would be the notion of an idiosyncratic threshold, that level beyond which guilt could no longer be tolerated. Since it is likely that a single communication would elicit varying degrees of aroused guilt in different individuals, the construct mental state of aroused guilt should be mediated by the associated construct susceptibility to guilt arousal, although the abstractness of this construct would make it difficult to distinguish and measure.

Assuming a communication has aroused guilt in an individual beyond the tolerance threshold, but below an intolerance threshold, the next logical construct would refer to the drive mechanism of reducing or resolving the feelings of guilt, an effort to restore moral equilibrium between the individual and his/her environment (Roseman 1979). The drive to reduce undesirable feelings of guilt manifests itself in other more abstract constructs, such as beliefs and/or attitudes in cognitive structure, a desire for additional information (possibly for purposes of bolstering or transcendence), or behavioral intent and ultimate behavior for example. It is likely, however, that another mediating variable will impact upon these abstract constructs, that being the individual's susceptibility to persuasion (which in this instance would be related to susceptibility to guilt arousal).

Given that a marketing communication had been judged to arouse guilt in a message recipient, and the message had been duly attended to, the directionality of the proposed paradigm would be as follows. Dependent upon the individual's susceptibility to guilt arousal, a mental state of aroused guilt would be expected to occur. Should the threshold for aroused guilt be surpassed in the mental state the individual would be expected to enter a drive state intent upon reduction or resolution of the evoked feelings of guilt. This drive might then bring about movement, cognitive or behavioral, to resolve the undesired mental state. A reasonable expectation would be that both the mental state of aroused guilt and subsequent resolution tendencies would be related to the individual's susceptibility to aroused guilt since those more susceptible would likely surpass the hypothesized threshold point sooner than individuals with a similar threshold but less susceptibility to guilt arousal.

Although it may be impossible to compile an exhaustive list of resolution strategies, several dominant ones with direct marketing implications bear consideration. Of interest in a marketing context would be resultant changes in beliefs, attitude or more generally, cognitive structure in the direction advocated by the communication. Such changes would depend on the individual's susceptibility to persuasion in the particular context of the communication. For example, an individual with a large latitude of rejection would likely contrast the communication as being too extreme, possibly demonstrating a polarization of attitude rather than attitude change (Sherif and Hovland 1961, Sherif 1979). Behavioral intent and ultimate behavior are also plausible vehicles for manifestation of resolution strategies, although as above, changes would depend upon susceptibility to persuasion, which in the case of guilt appeals would in turn depend upon susceptibility to guilt arousal. A further resolution strategy of consequence for marketers could be a desire for additional information, perhaps as a prelude to reevaluation and subsequent attitudinal or behavioral change.

Similar to models of cognitive dissonance, the proposed paradigm is essentially a threshold model for the hypothetical construct representing the mental state of aroused guilt. When the level of guilt surpasses the individual specific threshold, resolution of the guilt is desired. However, should the level of guilt be too high, surpassing the threshold of intolerance, it may well be that the persuasive attempt will be without impact. This may be due to possible denial or distortion of the message content to avoid intolerably high levels of guilt, rather than attempting to resolve them. Once the transgression has been

resolved (rather than denied or distorted) as with changes in cognitive structure, behavioral intent or behavior, the individual would cease efforts to reduce the feelings of guilt, returning to a mental state in which the level of aroused guilt was below the threshold.

In sum, by arousing feelings of guilt which refer to a particular context and that surpass a given threshold (but not the threshold of absolute intolerance), an individual would be expected to attempt to alleviate such feelings, possibly through changes in cognitive structure, behavioral intent or behavior. If the feelings of guilt were aroused by a marketing communication which included recommendations, implicit or explicit, leading to a reduction of guilt, the recommendations would more likely be adhered to as mediated by susceptibility to aroused guilt and/or persuasion in terms of attitude, behavioral intent or behavior than if the recommendation were posited without the accompanying guilt arousing communication. (A graphic representation of the "first draft" model can be seen in Figure 1.)

Some Possible Individual Difference Variables

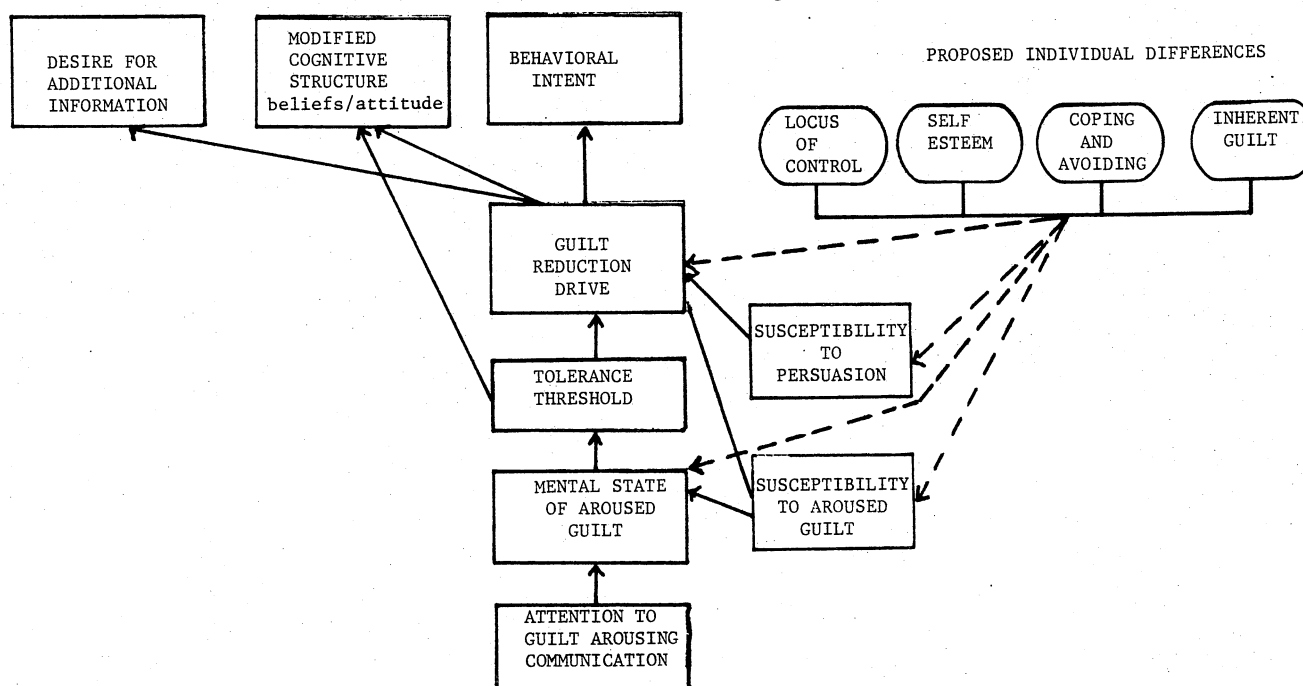
A number of individual difference variables may be hypothesized to mediate the efficacy of guilt arousing communications in a given population. Although many such variables are generalizable to other forms of persuasion, some are more closely linked to the paradigm proposed herein. Since one may conceptualize a great number of relevant individual difference variables, only those considered of significant impact will be discussed at this early stage of model building.

One trait related to many forms of persuasion is locus of control (Rotter 1966). Rotter differentiated between individuals believing in an internal locus of control or an external locus of control. In broad terms, internals perceive themselves to control their own destinies, whereas externals expect control from outside forces. Rotter argued, with the support of empirical evidence, that externals would be more prone to subtle suggestion and persuasion. With respect to guilt arousing marketing communications, one would expect that an external would more likely adhere to the recommendations advocated in the communication as a means of resolving feelings of aroused guilt. An internal might possibly seek his own solution to the moral dilemma. Thus, locus of control may mediate resolution strategies to aroused guilt.

Another general individual difference variable in persuasibility is self esteem. Leventhal and Perloe (1962) found that for certain message types subjects high in self esteem were influenced more by positive communications than by negative communications. Subjects low in self esteem showed the opposite pattern. With respect to guilt arousing communications, Zemach (1966) reported that self esteem was a powerful explanator in her investigation, although it remains unclear whether self esteem mediated overall susceptibility to persuasion alone or susceptibility to guilt arousal as well, since the two are probably correlated and it would be difficult to separate and independently measure the constructs without confounding one with the other.

Goldstein (1959) hypothesized that the acceptance or nonacceptance of recommendations contained in a persuasive appeal would be related to the subject's characteristic reaction to tension-producing stimuli based on Mainord's (1956) distinction between copers and avoiders. Coping and avoiding behavior parallels the later formulation of Kelman and Baron (1968) for handling inconsistency. Copers would be expected to exhibit greater impact of the tension-producing persuasive appeal than would avoiders. Thus, with respect to guilt arousing communications, copers should be more susceptible to high levels of aroused guilt and persuasion in general. Following Goldstein's findings, low tension-producing (i.e., guilt arousing) communications

FIGURE 1
A "First Draft" Model of Guilt Arousing Communications



would be more effective for avoiders than for copers.

A final variable which may be postulated to mediate guilt appeal effects would be a trait of self blame or inherent guilt in an individual. This psychological construct has not been applied to the area of persuasion to date, although it appears relevant when considering guilt arousing communications. Storm et al. (1958) developed the construct as well as a scale for its measurement. It would seem logical that individuals with a tendency for self blame may be more susceptible to communication induced guilt, and as such, more likely to react in accordance with the recommendations advocated in a guilt arousing communication. In contrast, it may be postulated that individuals high in inherent guilt possess higher thresholds for aroused guilt, delimiting the persuasive efficacy of guilt arousing communications. Thus, although it would seem that this construct intervenes in the persuasion process, arguments as to the directionality of the influence may support either suppression or facilitation of the guilt appeal effect.

In short, four factors have been proposed as likely individual difference variables which may mediate the effects of guilt arousing communications; locus of control, self esteem, coping and avoidance behavior, and self blame and inherent guilt. Although other variables may come into play, those noted above are likely to be prominent and are therefore worthy of consideration at this preliminary stage of the theory building process.

A Research Agenda: Some Preliminary Hypotheses

Although the graphic representation of the model (Figure 1) appears rather elaborate, a single experimental investigation could be sufficient as a preliminary test of the proposed paradigm. The research would be of an exploratory nature, centering upon a phenomenon of which little is known. Of particular interest in the study to be proposed are the following broad hypotheses:

H₁: The mental state of aroused guilt will be greater

under manipulations of moderate guilt than under minimal or maximal guilt.

- H₂: The greater the susceptibility to guilt arousal, the greater the level of aroused guilt in the mental state.
- H₃: The greater the susceptibility to aroused guilt, the greater the likelihood of a drive to resolve feelings of aroused guilt and subsequent susceptibility to the persuasive appeal.
- H₄: The greater the level of aroused guilt in the mental state, the greater the drive to resolve the feelings of guilt.
- H₅: The greater the drive to resolve feelings of guilt, the greater the changes in cognitive structure, behavioral intent, or desire for additional relevant information.

Thus, the hypotheses infer a curvilinear relationship between levels of guilt in the communication and the mental state of guilt. This reflects the double threshold effect described above. Susceptibility to aroused guilt is also hypothesized to affect the mental state of guilt, as well as the drive to resolve the feelings of guilt and subsequent susceptibility to the persuasive appeal. The drive to reduce feelings of guilt is hypothesized to be manifested in one of several ways, through changes in cognitive structure (e.g., changes in beliefs or attitudes in the advocated direction or bolstering of existing attitudes), changes in behavioral intent relevant to the message recommendations, or the desire for more information relevant to the message content, although other means may likely exist.

Measurement of Aroused Guilt

Before closing this discussion it may be useful to briefly outline some approaches to the measurement of aroused guilt since the contention of this paper has been that guilt

arousing persuasion is an unexplored and enigmatic variable in the marketing discipline. Several methods will be proposed since empirical tests should not rely on a single measure.

Given that an individual has been presented with a guilt arousing communication, a cognitive response format would appear to be the best way to begin. The first question might ask the respondent to elicit (verbally or in writing) any thoughts that the communication brought to mind. A second question might direct the respondent more towards affective responses, asking that he/she elicit any feelings or emotions that the communication made him/her feel. Conceivably these questions could "create" a great deal of data of which only a small fraction would be of interest. Coding and coder reliability would also be problems and the analyst would have to specify a priori which responses were indicative of guilt feelings and/or to what degree. Nonetheless, cognitive responses immediately after exposure might offer the best measures of aroused guilt.

Another method of measurement could be the use of a scaling technique which has respondents indicate the degree to which they feel or experience a particular emotion. Haefner (1956) used a cluster of nine adjectives as an index of guilt (ashamed, blameworthy, conscience-stricken, contrite, guilty, regretful, remorseful, repented, and sorry) judged by clinical psychologists to be relevant to that form of emotional arousal. Haefner found that his treatment groups showed substantial shifts on this cluster although he did not factor analyze them. This is essentially the method used by Zemach (1966, pg. D1, D2) to evaluate the affective responses to the communications used in her study, although she did not use all the adjectives in Haefner's cluster. Although open to demand characteristics, this method could be expected to realize useful measures of aroused guilt.

Similar to the above method, semantic differential scales could also be used to allow the respondent a means of expressing how the communication makes him/her feel. Care should be taken to avoid redundancy with any adjectives used in other scales or measures.

Lastly, attitudinal measures could be used to gauge the arousal of guilt relevant to the communication. If an advertisement was used which was intended to arouse guilt feelings over the plight of the world's underprivileged and garner charitable donations, respondents could be asked the extent to which they feel partly to blame for the plight of the poor, or the extent to which they regret not having done more to help the underprivileged. These measures are more specific to the message content but should also yield some indication of the level of guilt aroused by the communication.

In sum, several approaches to measuring aroused guilt may be used; cognitive responses, affect scales, semantic differentials, and attitudinal measures. It would be advisable to incorporate multiple measures in an experimental design to offer at least some opportunity for tests of validity.

In Conclusion

This paper has proposed that guilt arousing communications are frequently occurring phenomena in a marketing context and worthy of investigation by consumer researchers. The insufficiency of past explorations into guilt appeals and the presumed differences between guilt and related paradigms of fear, cognitive dissonance and information processing conflict provided a rationale for more focused research. Constructs were explicated and an initial "first draft" model was presented. Several individual difference variables were proposed and a research agenda was outlined in the form of broad preliminary hypotheses and approaches to measurement issues. No doubt this paper has raised more questions than it has answered but it is hoped that this first attempt at

conceptualization of the phenomenon will arouse interest and stimulate additional efforts to add increased understanding and control of this exciting marketing variable.

References

- Berlyne, D. E. (1957), "Uncertainty and Conflict: A Point of Contact Between Information-Theory and Behavior-Theory Concepts," Psychological Review, 64 (November), 329-339.
- Bettman, J. R. (1979), An Information Processing Theory of Consumer Choice: Addison Wesley Publishing Company.
- Calder, B. J. (1973), "Cognitive Consistency and Consumer Behavior," in H. H. Kassarjian and T. S. Robertson, (eds.), Perspectives in Consumer Behavior, Glenview, IL: Scott, Foresman and Co., 247-263.
- English, H. B. and English, A. C. (1976), A Comprehensive Dictionary of Psychological and Psychoanalytical Terms, New York: David McKay Co.
- Festinger, L. (1957), A Theory of Cognitive Dissonance, Stanford: Stanford University Press.
- Fishbein, M. and Ajzen, I. (1975), Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading: Addison-Wesley Publishing Company.
- Goldstein, M. J. (1959), "The Relationship Between Coping and Avoiding Behavior and Response to Fear-Arousing Propaganda," Journal of Abnormal and Social Psychology, 58 (March), 247-252.
- Haefner, D. P. (1956), Some Effects of Guilt Arousing and Fear Arousing Persuasive Communications on Opinion Change, unpublished technical report.
- Hansen, F. (1972), Consumer Choice Behavior: A Cognitive Theory, New York: Free Press.
- Higbee, K. (1969), "Fifteen Years of Fear Arousal: Research on Threat Appeals, 1953-1968," Psychological Bulletin, 72 (December), 426-444.
- Howard, J. A. and Sheth, J. N. (1969), The Theory of Buyer Behavior, New York: John Wiley and Sons.
- Janis, I. L. (1967), "Effects of Fear Arousal on Attitude Change: Recent Developments in Theory and Experimental Research," in L. Berkowitz (ed.), Advances in Experimental Social Psychology, New York: Academic Press, 167-225.
- _____ and Feshback, S. (1953), "Effects of Fear-Arousing Communications," Journal of Abnormal and Social Psychology, 48 (January), 78-92.
- _____ and _____ (1954), "Personality Differences Associated with Responsiveness to Fear-Arousing Communications," Journal of Personality, 23 (December), 154-166.
- Kelman, H. C. (1979), "The Role of Action in Attitude Change," paper prepared for the 1978-79 Nebraska Symposium on Motivation, Department of Psychology, University of Nebraska, Lincoln.
- _____ and Baron, R. M. (1968), "Determinants of Modes of Resolving Inconsistency Dilemmas: A Functional Analysis," in R. P. Abelson et al. (eds.), Theories of Cognitive Consistency: A Sourcebook, Chicago: Rand McNally, 670-683.
- Leventhal, H. (1970), "Findings and Theory in the Study of Fear Communications," in L. Berkowitz (ed.), Advances in Social Psychology, New York: Academic Press, 119-186.

_____ and Perloe, S. (1962), "A Relationship Between Self Esteem and Persuasibility," Journal of Abnormal and Social Psychology, 64 (June), 385-388.

_____ and Trembly, G. (1968), "Negative Emotions and Persuasion," Journal of Personality, 36 (March), 154-168.

Mainord, W. A. (1956), Experimental Repression Related to Coping and Avoidance Behavior in the Recall and Relearning of Nonsense Syllable, unpublished doctoral dissertation, University of Washington.

Ray, M. L. and Wilkie, W. L. (1970), "Fear: The Potential of an Appeal Neglected by Marketing," Journal of Marketing, 34 (January), 54-62.

Roseman, I. (1979), "Cognitive Aspects of Emotion and Emotional Behavior," paper presented at the 87th Annual Convention of the American Psychological Association, New York City.

Rotter, J. B. (1966), "Generalized Expectancies for Internal Versus External Control of Reinforcement," Psychological Monographs (Whole no. 609), 1-28.

Schneiders, A. A. (1967), "Catholics and Psychology," Catholic Psychological Record, 5 (2), 102-110.

_____ (1968), "The Nature and Origins of Guilt," Transactions of the New York Academy of Sciences, 30 (5), 705-713.

Sherif, C. W. (1979), "Ego Involvement and Attitude Change," paper prepared for the 1978-79 Nebraska Symposium on Motivation, Department of Psychology, University of Nebraska, Lincoln.

Sherif, M. and Hovland, C. I. (1961), Social Judgment Assimilation and Contrast Effects in Communication and Attitude Change, New Haven: Yale University Press.

Stein, E. V. (1968), Guilt: Theory and Therapy, Philadelphia, PA: The Westminster Press.

Sternthal, B. and Craig, C. S. (1974), "Fear Appeals: Revisited and Revised," Journal of Consumer Research, 1 (December), 22-34.

Storm, T., Rosenwald, G. C. and Child, I. L. (1958), "A Factor Analysis of Self-Ratings on Social Behavior," Journal of Social Psychology, 48, 45-49.

Time (1980a), advertisement appearing on page 9, Vol. 115 (5).

_____ (1980b), advertisements appearing on pages C2 and 58, Vol. 115 (7).

_____ (1980c), advertisement appearing on page 17, Vol. 116 (1).

Wheatley, J. J. and Oshikawa, S. (1970), "The Relationship Between Anxiety and Positive and Negative Advertising Appeals," Journal of Marketing Research, 7 (February), 85-89.

Wolman, B. B. (1973), Dictionary of Behavioral Science, New York: Van Nostrand Reinhold Company.

Yinon, Y., Bizman, A., Cohen, S. and Seger, A. (1976), "Effects of Guilt-Arousal Communications on Volunteering to the Civil Guard: A Field Experiment," Bulletin of the Psychological Society, 7 (6), 493-494.

Zemach, M. (1966), The Effects of Guilt-Arousing Communications on Acceptance of Recommendations, unpublished doctoral dissertation, Yale University.

CONSUMER RESPONSE TO ADVERTISING: A DISCUSSION OF THREE STUDIES

Robert L. King, Virginia Polytechnic Institute and State University

Abstract

Three papers related to consumer response to advertising are discussed. Two papers report results of field investigations, one concerning linear effects in cognitive response to advertising, and the other concerning the effectiveness of celebrity endorsers. The third paper proposes a conceptualization of guilt-arousing communications. Some questions are raised concerning methodology and design of the first two studies, and some suggestions are made for extending the scope of the third paper.

Percy and Lautman's Study of Linear Effects in Cognitive Response to Advertising

The authors' report of their empirical investigation of linear effects in cognitive response to advertising is both interesting and relevant to the concerns of a substantial part of its Association for Consumer Research audience. Further, the study's design is simple, appropriate for the present investigation, and essentially well executed. Finally, the authors' make good use of recent relevant research literature, and they provide support for existing theory in this field of investigation.

If the paper contains disappointments for this reader, they center on two points. First, early in the design of their study the authors made some critical decisions which had potential for influencing their results significantly, but they did not share the rationale for these decisions with their audience. Second, the reader who seeks a relatively full discussion of the applications and implications of this research remains disappointed. While the value of parsimony may be exceeded only by its rarity in professional dialogues such as ours, in the present situation neither the authors nor the audience seems well served by the brevity of the report in these regards.

Several examples may serve to illustrate the first concern, i.e. inadequate explanation of some basic decisions related to the universe which was studied, and to the testing devices which were developed for data collection purposes in this study. A fuller explanation of issues such as the following would be helpful, particularly to prospective users who have a marketing practitioner orientation:

First, why did the authors select frozen food as the product subject for this investigation? Is it conceivable that their findings might have been substantially different had they chosen a consumer durable product such as a videotape recorder? Might the degree of product diffusion among the general population alter the nature of the responses? What are the implications of selecting a well-known brand of the product, as opposed to a lesser known brand or even a newly introduced brand of the same product type?

Second, what was the basis of the authors' decision to feature "a couple" in one experimental commercial, while using "several people" in the other? More specifically, do the people in these two advertisements represent similar or different age groups, marital statuses, social classes, and life styles, or do they reflect essentially different populations? What are the potential consequences of these similarities and differences to the responses which are forthcoming?

Third, is the definition which is suggested for the product's target market acceptable (i.e., female household heads, aged eighteen or older, with family incomes over \$15,000, who have had experience with the product category)? Are such women actually indicative of the product's target market? Or are there substantial regional, urban/rural, racial, religious, and other demographic differences in the actual target market which may not be accounted for in either the study's definition of the target market, or in the three "widely distributed" geographic areas from which respondents were selected, or both of these factors? More specifically, what does "experience with the product category" imply? Is there a measure of brand usage experience? And perhaps more important from the marketing practitioner's perspective, is attention given to differences in responses from heavy and light users of the product type?

This reader's second concern relates to the paucity of the authors' remarks about applications and implications of their findings. Specifically, how projectable are their findings across various product categories? What are the implications for promotional program development and evaluation generally, and for advertising pre-testing specifically? On the other hand, have the authors possibly generalized excessively when they refer to the broad area of "cognitive response"? Their data refer essentially to "support arguments," which are in fact a subset of "cognitive response."

In conclusion, the paper is a sound one which adds credence to existing theoretical and empirical work. However, a fuller discussion of some of the decisions underlying the study's design, and of possible applications of the findings, would have proved extremely useful.

Mowen and Brown's Study on Explaining and Predicting the Effectiveness of Celebrity Endorsers

According to legend, the first recorded instance of a modern athlete leasing his name to a commercial sponsor occurred in 1905 when Honus Wagner, shortstop for the Pittsburgh Pirates, gave the J. F. Hillerich and Son Company permission to use his name on its Louisville Slugger bats for a consideration of \$75. However, the non-smoking Wagner sued the American Tobacco Trust, which put his name on a Sweet Caporal baseball card without his permission. In 1907 Wagner appeared with Ty Cobb (who had a cigarette and a candy bar named for him) in Coca-Cola's "The Great National Drink of the Great National Game" advertising campaign ("Tinker to Evers to JWT," 1978).

Two issues which were central to the situation of 75 years ago are addressed by Mowen and Brown, who properly remind us that definitive answers have not yet been found to the questions:

(1) Does endorsement of multiple products tarnish a celebrity's effectiveness?

(2) Is it advantageous to the sponsor to use a number of different endorsers to recommend a given product?

These questions take on a new urgency in view of contemporary promotional practice, which is characterized by increasing use of celebrity endorsers, especially in tele-

vision commercials. The substantial fees paid to celebrity endorsers also give these questions a new urgency. For example, shortly before his death John Wayne filmed several commercials for Great Western Savings and Loan, a regional advertiser, under a three-year contract which paid Wayne an estimated \$1.5 million. Recent trade estimates suggest that some athletes are paid \$25,000 or more for making a single national commercial (Forkan 1979, and Morris 1979).

The authors are due appreciation, then, for raising relevant and interesting issues, and for attempting to extend previous related research. Similarly, they have provided us with a useful review of the limited literature in this area of research. In retrospect, however, one may wonder whether the cause would not have been better served had they chosen to investigate only one of these related but difficult questions, rather than both, and had they structured the design of that more limited inquiry much more tightly. More specifically, it appears that the authors have relied unduly on assumptions produced by logic as the basis of much of their study's design, when a "factual" basis could have been obtained relatively painlessly. The latter approach would have given the report's audience greater confidence in its results. Several examples are offered to illustrate this point.

First, reference is made to the possibility of respondents having perceived risk when allowing advertising to influence their perceptions of relatively costly television sets in their earlier study. In fact, was consequential risk perceived by those respondents? Why was "a new ballpoint pen" selected as the product-subject in the present study? Is its comparatively low price sufficient evidence that respondents would not perceive consequential risk if they allowed advertising to influence their perceptions of it? In short, in conducting this replication of their earlier study, did the researchers objectively measure the risk levels associated with letting advertising influence one's perceptions of various product-subjects?

Second, products are classified by status level in the report. But, are the products which are characterized as relatively "low status" (motorcycles, cigars, cologne, and rums) and "higher status" (Hilton hotels, Chivas Regal scotch, and Mercedes-Benz automobiles) in fact significantly different in their status identifications among respondents? Were these differences tested? And in any event, how does product status level relate to the research problem, as it is defined in this study?

Third, the researchers' operational measure of "relevancy" seems open to question. Does the use of a script in which the endorser of the ballpoint pen discusses its performance in signing autographs establish "greater relevancy"? Or does the authors' concern, which was expressed about respondents' perceptions of the three actors' relevancy in the television set study, call for a more objective measure of relevancy in the present study?

Fourth, similar questions can be raised concerning the operational measures of "concensus" (endorsement by multiple sources) and "distinctiveness" (endorsement by the endorser of a single product only). Also, it seems that "who" the endorser(s) is (are) must be examined. That is, did Steve McQueen represent essentially the same image to all of the respondents? Did George C. Scott? What is the impact on responses of this particular mix of endorsers? Does this mix possibly introduce additional complexities into the study?

Fifth, direct attention needs to be given to the issue of "credibility" of individual celebrity endorsers in order to understand the phenomena under investigation here. A substantial professional and trade literature exists to help us in this regard. (See Middleton 1974; Eagly and Chai-ken 1975; Dholakia and Sternthal 1977; and Morris 1979,

for example.) Unfortunately, this literature is not well represented in this study's "References" section.

Similar questions could be raised concerning other uncontrolled sources of variation, the absence of reported manipulation checks, and the rather thinly disguised experimental conditions in the research. One would hope that the authors will continue their investigation of this timely and interesting subject after reconsideration of some aspects of their operationalizations and manipulations. More rigorous testing would yield much needed guidance for business academicians and practitioners alike.

Chingold's Study of Guilt Arousing Marketing Communications

Whereas the previous papers reported results of field investigations, the present paper is essentially concerned with conceptualization of a construct for exploring an infrequently investigated phenomenon, guilt-arousing marketing communications. The author has very adequately met his three stated objectives. He introduces this largely new area of research interest for marketing communicators, acknowledging the distinction as well as the interrelationship which exists between "fear" and "guilt." His review and synthesis of the literature dealing with guilt-arousing persuasive communications, along with the accompanying bibliography, should be particularly welcomed by concerned marketing communicators. Finally, his "first draft" conceptualization of guilt-arousing communications is a fresh, thought provoking statement, although it is subject to possibly extensive revision.

Even so, some concerned academicians and practitioners in the field of marketing communications will be dissatisfied with, and less well served than was necessary by, several characteristics of the paper. First, the paper may appear to be too truncated: the first one-third of an empirical paper, omitting the methodology and analysis sections. Certainly, the paper would have been much stronger if it presented empirical findings, however tentative they might be. Failing this, it would be useful to know the author's more detailed thoughts concerning measurement approaches. Probably the major shortcoming of the paper in the eyes of its critics is its lack of a suggested research design for testing the recommended hypotheses.

Second, the reader is offered insufficient guidance in identifying specific guilt-arousing marketing communications, and in measuring the incidence of their use by marketing communicators. The author perceives guilt approaches to be in "fairly widespread use" by marketing communicators, but his perception seems to be untested in the market place. An expanded listing of contemporary examples of guilt-arousing commercials, according to the author's definition, would be quite helpful. In fact, how common are such appeals in marketing communications? We would be assisted in identifying examples if we had an improved or expanded definition of guilt-arousing communications which specifically reflected a marketing communications context. Presently we must ask whether "lowering of self esteem, or feeling of lessened personal worth" and "a need to make retribution for the transgression" (Wolman 1973) result from exposure to advertisements such as the following:

- (1) Wisk's "ring around the collar" advertisements, suggesting the housewife's distress upon finding her laundry efforts publicly wanting;
- (2) Shower-to-Shower powder advertisements, showing passengers piling into the rear seat of an automobile in order to avoid unpleasant body odor of the person in front;
- (3) Church of the Latter Day Saints advertisements, asking whether one has hugged his children today;
- (4) Crest toothpaste advertisements directed toward parents, reporting reduced incidence of tooth decay among using children of "cavity-prone" age; and
- (5) CARE and Save the Children Foundation advertisements,

showing the lean faces and hungry eyes of undernourished children.

Third, in view of the author's necessarily heavy reliance on non-marketing literature, reinforced by the paucity of references to the conceptualization's specific marketing communications implications and applications, one may wonder why he elected to develop and present his work within the context of marketing communications. With a minimum of editing, the report would seem equally applicable to a variety of essentially non-marketing communications, such as those between candidate and voter, office holder and citizen, supervisor and worker, pastor and parishioner, teacher and student, and parent and child. While it is not surprising that applications are more generalized than area-specific in the "first draft" conceptualization in a less explored field, the paper does seem unduly brief in setting forth specific marketing communications examples, implications, and applications.

The paper raises a wide range of additional issues, each of which is worthy of fuller discussion. For example, is guilt research's "piggybacking" on fear research really so "dubious"? Is the inappropriateness of cognitive dissonance and information processing for understanding guilt possibly more apparent than real? What is the relevance to the marketing communicator of such distinctions as high vs. low self-esteem, copers vs. avoiders, and high vs. low inherent guilt? Might not the suggested hypotheses take on increased relevance and be more finely tuned to the paper's title if they were expressed specifically in a marketing communications context? Might not an additional hypothesis, reflecting the possible dissipation of guilt arousal with the lapse of time, be appropriate? Could/should tolerance and intolerance thresholds of aroused guilt be incorporated into the "model"?

In conclusion, the author has developed a thought-provoking conceptualization. But one recalls his earlier statement that "a single experimental investigation could be sufficient as a preliminary test of the proposed paradigm." One would agree, while regretting that the author did not include the results of such an investigation as part of his paper, but hoping that he will continue his interesting efforts and share those results with us in the near future.

References

- Dholakia, Ruby Roy and Sternthal, Brian (1977), "Highly Credible Sources: Persuasion Facilitators or Persuasive Liabilities?," Journal of Consumer Research, 3, 223-232.
- Eagly, A. and Chaiken, S. (1975), "An Attribution Analysis of the Effect of Communication Attractiveness," Journal of Personality and Social Psychology, 32, 136-144.
- Forkan, James P. (1979), "You Can Reach for the Stars, Talent Agents Say," Advertising Age, April 23, 1979, 52-53.
- Middleton, Patricia N. (1974), Social Psychology and Modern Life, New York: Knopf.
- Morris, Bourne (1979), "Will a Personality Sell a Product Better? Pros & Cons," Advertising Age, February 5, 1979, 43-44.
- "Tinker to Evers to JWT" (1978), Media Decisions, November 1978, 124 ff.
- Wolman, B. B. (1973), Dictionary of Behavioral Science, New York: Van Nostrand Reinhold Company.

PUBLIC POLICY UPDATE: PERSPECTIVES ON THE FEDERAL TRADE COMMISSION¹

Kenneth L. Bernhardt, Georgia State University
Ronald Stiff, University of Baltimore

Abstract

This paper presents a summary of a special session devoted to a briefing on current and future developments at the FTC and the role of consumer research at the Commission. Examples of topics likely to be important for consumer research in the future are presented, together with a brief discussion of each of a number of recent research projects.

Purpose

This paper presents the highlights from a special session concerning current and future perspectives on the Federal Trade Commission. Participating in the session were Michael Pertschuk, Chairman of the FTC; Robert Reich, Director of the Agency's Office of Policy Planning; and Ronald Stiff, who was Consumer Research Advisor at the time of the session. The session was moderated by Kenneth L. Bernhardt, who had served as Consumer Research Advisor at the FTC the previous year.

The session focused on three subjects: 1) what topics are likely to become increasingly important at the FTC in the future; 2) what is the role of research at the FTC, what types of research are conducted, and what have been the successes and failures of consumer research at the Commission; 3) what are some examples of research projects that have been conducted by the FTC.

In sum, the goal of the session was to provide those attending with a briefing on not only what types of research are currently being conducted at the agency, but also to alert researchers to topics that are likely to be requiring consumer research inputs in the next few years. The rest of this summary presents the conclusions offered by the speakers at the session.

Future Areas of Interest

Six areas were identified as being hot topics for the future. It was emphasized that the FTC was not abandoning its current programs, advertising substantiation for example, but that these are not likely to be the growth areas of the future. The future growth areas which were identified are briefly described below:

1. Credit Programs. Because of the continued importance of credit in inflationary and recessionary times, this should be an increasingly important area for the FTC in coming years. Specific areas mentioned included debt collection practices, balloon payments, deficiency judgments, and the provisions included in the proposed Credit Practices Rule.
2. Comparative Disclosures. The purpose of the comparative disclosures program is to help consumers make better comparative judgments in evaluating product alternatives. Current programs in this area include energy savings comparative disclosures such as the appliance energy

- labeling program, and the insulation R-value program. The FTC is currently looking at other areas where comparative performance information would be valuable to consumers.
3. Product Quality/Defects. As consumers become more concerned with product quality and as products become more complex, this program should increase in importance. The FTC currently has been investigating product quality concerning housing, automobiles, and heat pumps.
 4. Professional Services. For the past four years, the FTC has been studying the marketing and advertising of professional services, and this program is likely to increase in importance in the next couple of years. Specifically, the agency has studied advertising and marketing by opticians, dentists, nurse practitioners, dental hygienists, lawyers, and doctors. The U.S. Appeals Court has recently affirmed the FTC decision in the American Medical Association case which removes the constraints imposed by the AMA, inhibiting doctors from advertising. The FTC is also investigating the business aspects of these professions, including such questions as whether dental clinics can be opened up in shopping centers and what qualifications should be required for ownership of clinics.
 5. New Media and Marketing. The FTC is currently investigating the "communications revolution." Such things as videodiscs, videotext, and cable communications should open up whole new areas for FTC investigation. The agency has recently completed a Media Policy Session and report on new media technology and legal change.
 6. Energy, Health Care, Transportation, Shelter, and Food. There has been a pronounced shift in the priorities the FTC assigns in evaluating potential new programs. The bulk of the programs in the future will be closely matched to their importance in terms of consumer spending. In other words, the agency will concentrate its efforts on increasing consumer savings in those areas where the potential net benefit is the greatest.

Consumer Research at the FTC

The atmosphere for research has changed considerably since the first consumer research study, costing \$1200, was funded seven years ago. The current budget contains \$1.2 million for research, almost all of which is for consumer research. This amount does not include any funds spent internally for personnel, supplies, travel, etc., but is only the amount spent on external contract research.

The evolving nature of the authorizing legislation has insured that the FTC will continue to engage in consumer research. For example, the 1980 FTC Improvements Act requires considerable economic analysis on proposed and existing regulations, and much of this analysis will be consumer research based. In addition, the political environment is contributing to an increase in consumer research--the agency must be able to defend its actions. Institutionally the agency is committed to do consumer research to help pay attention to the real world effects of its actions. Temperamentally, the agency is ready to accept and use consumer research findings.

Several of the commissioners are now asking consumer research questions of the attorneys when new regulatory programs are being proposed. For example, the questions being asked include: a) what empirical evidence do you have that

¹The views reflected in this summary statement are the views of the individuals participating in the session and do not represent the official views or policies of the Federal Trade Commission.

there is a problem requiring FTC action?; b) what empirical evidence do you have that would indicate that your recommended remedy would actually work?; and c) what empirical evidence are you gathering or have you gathered that will enable us to later evaluate whether or not the remedy is working as intended?

FTC attorneys have become very sophisticated with respect to consumer research. It's not unusual to have discussions at the agency concerning statistical significance, sample size requirements and procedures, regression and other multivariate analysis techniques. Less and less the discussions concern whether or not research is needed. Instead, the discussions center on what research would be most helpful and how to do it.

Consumer research has been helpful in halting programs that research indicated had a low probability of success. For example, a study of detergent buying behavior was done by Joel Cohen and Bill Wilkie showed that even if consumers were given product information indicating parity of performance between major brands and private labels that people would still buy the major brands. The program was eventually halted, at least in part because of the evidence that it would not work.

Consumer research has been used to help formulate programs. For example, focus groups and quantitative followup studies were conducted as part of the development of the appliance energy labeling program. The purpose of the research was to measure if consumers could understand the energy information, and to determine which formats among several alternatives were most effective in communicating the information. The research showed that expressing energy usage in dollars was far more effective than using an "energy efficiency ratio" and the final format of the label actually being used was heavily influenced by consumer research.

The identification of needed FTC action has been another result of consumer research. For example, a study of the quality of new homes purchased and experiences of the buyers indicated that defects in new housing were an important problem and that there were some builders who were not resolving problems which had a major financial impact on these consumers. The study also helped identify what the problems were, the magnitude of the consumer injury, and what types of houses had problems. This study thus helped define the nature of the problem, provided direction for possible remedies, and served as a benchmark against which the success of any government action could be later measured.

Consumer research has also been used by the FTC to help identify and prioritize advertising substantiation cases. In several instances, copy tests have been used to determine consumer perceptions of potentially deceptive and unfair advertisements to determine if a case should be initiated against the advertiser.

In spite of the successes described above, there also have been some failures in the consumer research area. The biggest failure has probably been the inability of consumer researchers to adequately explain to Presiding Officers, Administrative Law Judges, and others involved in a proceeding exactly what they have done and why it is appropriate. This inability to adequately explain what was done and why has diminished the value of the research to the decision makers involved.

Another failure of consumer research concerns lapses in the quality of individual studies done for various proceedings. Many consumer researchers have failed to recognize the higher quality standards necessary for studies done in the public arena for public policy purposes. Thus, many studies, with small sample sizes, selected inappropriately, analyzed improperly or with sloppy controls, have been thrown out or given very little weight by the decision

makers.

Examples of FTC Research Projects

Many different types of consumer research techniques have been used, including copy tests, experimental design studies, mall intercept studies, consumer mail panel surveys, and random-digit-dial telephone surveys. It has not been possible to engage in test marketing of proposed programs, but in several studies individual states, with regulations similar to those proposed at the national level, have been used as control groups to measure the impact of the programs. Listed below are examples of 16 studies that have recently been completed or are in progress at the agency. The studies represent a diversity of problems and research methodologies, but show the spectrum of the kinds of research being conducted by the FTC.

1. Unavailability of Advertised Specials. A survey of consumers' values of high levels of availability of advertised specials in grocery stores and consumer experiences with unavailability. (Report Released)
2. STP Corrective Advertising. Measures of consumer and business attitudes toward the company and the product were taken before and after the corrective message. In addition, consumer and business perceptions of the corrective ad were determined after the ad appeared. (Report Released)
3. Warranties Consumer Baseline Study. A survey of a national sample of 4,300 consumers who purchased products with warranties in 1976 was conducted to determine their knowledge of warranties and the use of this information in the purchase decision. (Report Released)
4. Automobile Warranties Content Analysis. This study analyzed the changes from 1967 through 1977 in the content of automobile warranties obtained from automobile manufacturers who had a significant share of the market. (Report Released)
5. Appliance Warranties Content Study. This study analyzed the changes in the content of warranties offered for 40 consumer products in 1978 compared to 1974. (Report Released)
6. Antacid Prospective Study. This project focused on the potential impact of warning messages in antacid television commercials on subsequent consumer behavior with regard to antacids. The study measured the impact of warnings on the target population (e.g., antacid users on a salt-free diet) to see if warnings can be effectively communicated via 30-second TV commercials. In addition, the test measured whether legitimate antacid users (not on salt-free diets) are dissuaded from using the product. Various warning messages, ranging from very specific warnings about antacids to more general warnings (e.g., read the label) were tested to measure their impact. (Report a part of OTC Drug Advertising Staff Report)
7. Appliance Energy Labeling. This is a first part of a study to assess the impact of appliance energy labels on (1) consumer purchasing patterns for appliances, (2) shifting purchases to products which use less energy, and (3) promotional practices of appliance marketers. The data will be used in conjunction with a later study to provide pre-to-post analysis of the effects of the labeling program. (Report Released)
8. Affirmative Disclosures. This project involves assessing the effectiveness of recent (since 1970) affirmative disclosure remedies. Past orders will be categorized and analyzed, and the contractor will make recommendations about future affirmative disclosure policy. (Report expected to be released in December 1980)
9. New Housing Construction. This study examines the nature, severity, and extent of defects or problems occurring in new housing. Telephone interviews were conducted with 1,812 home owners who have owned their home for 1 or 2 years. Actual home inspections were conducted for a subsample of 299 of the homes to verify consumer reported defects and costs. (Report Released)

10. Listerine Corrective Advertising. This is an analysis of consumer awareness, beliefs, and attitudes towards Listerine and the corrective advertising for the product. Initial measures were taken following the introduction of the corrective advertising campaign. The study will enable an assessment of the effectiveness of the corrective advertising over time. (Report expected to be released in December 1980)
11. Insulation R-Value. This is a baseline study measuring the extent to which consumers understand and utilize R-value information provided as a result of the rule. (Report to be released in December 1980)
12. Eyeglasses Advertising. The Bureau of Economics (BE) conducted a cross-sectional study of the effect of state eyeglass advertising restrictions on the cost and quality of eyeglasses and eye exams. (Report Released)
13. Cooling-Off Rule. This study is a post-rule evaluation of the Three Day Cooling-Off Period for Direct Sales Rule. The costs and benefits of the rule will be examined. A survey of consumers will be conducted to determine their understanding and use of the rule's provisions. A survey of direct selling companies will be conducted to determine the costs of complying with the rule. (Report expected to be released in 1981)
14. Used Car Study. This study is designed to gather baseline data on the used car market so comparisons may be made with data to be gathered in the future. Private buyers and sellers were interviewed via telephone. (Report expected to be released in 1981)
15. Advertising Information Levels. The purpose of this study is to determine how much information and what kind of information television advertising can communicate. A split-cable television system will be used to allow different amounts of information. Recall of information will be tested. (Report expected to be released in 1981)
16. Care Labeling Study. This study will gather baseline data on the extent to which consumers are being harmed by the insufficiency of information related to clothing apparel and textiles. Data will be gathered on the incidence of incorrect washing, drying, ironing and bleaching. This study will also measure the extent to which care and maintenance instructions are overly restrictive. (Report expected to be released in 1981)

THE FUTURE OF CONSUMER PROTECTION REGULATION

Michael B. Mazis, The American University

Abstract

The purpose of this paper is to provide a perspective on consumer protection regulation in the 1980's. In the course of this discussion, seven allegedly deadly sins of consumer protection are discussed. While there is some truth to each of these "deadly sins", they are not true to the degree suggested by critics of consumer regulation.

Attention About What To Regulate

There has been considerable criticism that consumer protection regulation has followed a scattergun approach in determining the areas selected for regulation. Often, the logic underlying selection of target sectors for government involvement is unknown to even well-informed observers. There appears to be over-regulation in some areas and, contrary to the belief of many business persons, there is likely under-regulation in other areas. The 1980's will see several important shifts in consumer protection regulatory activity.

First, there will be a substantial shift away from advertising regulation toward regulation in other areas. Recent FTC budgets show a substantial decrease in expenditures for advertising activities, with little emphasis on new rule-making activities. The FTC, known as the Tyrannosaurus Rex of regulatory agencies, is likely to show much more discretion in promulgating new investigations and enacting additional rules over the next decade. The agency is quite aware that many of its advertising initiatives have been ill-advised both on substantive and on political grounds.

Most of the food advertising rule has been scrapped and it can be anticipated that very little of this ambitious undertaking will be enacted into final regulation. The children's advertising rule is currently in hibernation and it is unlikely to appear in a form resembling the original proposal. The FTC's over-the-counter drug rule is likely to be discarded.

Second, there will be a substantial increase in de-regulatory activities both from the FTC and HHS. Various restrictions against professional service proliferation, such as not allowing franchises and advertising, will be under serious attack by consumer protection regulators. These efforts will be met with resistance on the part of the Supreme Court, which has supported state restrictions, which bar professional service expansion. The Court has followed a "states' rights" position in overturning a major portion of the FTC's eyeglasses rule. It is unclear at this time whether the Court will allow federal agencies to preempt state law. As a result, there is likely to be a shift to additional effort at the state and local level to overturn regulatory activities which further the profession and harm the consumer. This emphasis on state and local regulation follows the shift in media patterns resulting from cable T.V. expansion and more localized T.V. programming which coincides with more careful targeting by advertisers. As a result, much of the battleground of advertising regulation may shift from the national arena to more localized efforts.

It should be noted that this emphasis on deregulatory activity or free choice may face a counter-cyclical trend. At this point, the pendulum is swinging against free choice in certain areas. For example, motorcycle helmet law repeal has resulted in an enormous increase in accidental deaths. Therefore, in the latter part of the decade there is likely to be additional concern for the consumer's health and safety.

Third, there is likely to be an emphasis on cost or pocketbook issues. In housing, there will be increased pressures by homeowners for warranties. Pressures will build on insurance companies for cost disclosures and there will be serious attempts to provide greater opportunities for alternative providers, such as banks, to sell insurance at a lower cost.

Fourth, there will continue to be a major emphasis on health and safety regulation, especially involving uncontrollable risks. Consumers believe that where information can be supplied through labeling, they are able to control risks and, therefore, product bans are unnecessary. For example, nitrate and saccharin bans are unacceptable to many consumers since they choose to decide whether they wish to be risk-takers or risk-avoiders. On the other hand, when risk cannot be controlled by the consumer, there will continue to be significant regulatory activity. For example, new laws have been passed regulating infant formula and FDA has been given additional regulatory authority in this area. FDA recently enacted patient package insert regulations which provide for additional information to be given to consumers about prescription drugs. This regulation has met with little public controversy. Efforts to protect the water supply, which involves an uncontrollable risk, will come under increased scrutiny by consumer protection regulators in the coming decade.

Finally, there will be increased concern about product liability. For example, the Supreme Court has permitted victims of the cancer-causing drug DES to sue all major DES manufacturers for damages even if the victims cannot identify the specific company that sold them the drug. Under this ruling, a company may be made to pay for damages it never caused to someone who may never have consumed its product. To escape liability the firm must prove that it could not have made the product that caused the injury. This approach departs dramatically from conventional legal practice, which generally requires a showing of direct responsibility before damages can be awarded.

This increase in product liability awards has two important consequences. First, it may reduce the need for the Consumer Product Safety Commission (CPSC) to take action since products liability cases can serve as an important deterrent to manufacturers producing unsafe products. This does not mean that CPSC is likely to be eliminated in the near future; however, the agency will find it difficult to undertake new initiatives. Second, these large awards may cause increased pressures to restrict product liability awards. Similar pressures have been noticed in the medical malpractice area.

Choice of How to Regulate

In the past, there has been an emphasis by consumer protection regulators on product bans. However, public resistance has encouraged regulators to move away from this form of regulatory activity. One major turning point involved public resistance against seat belt interlock systems, which was followed by a public outcry against proposed Laetrille and saccharin bans. A few years ago, FDA routinely passed "death sentences" against entire product categories. The cranberry crop was doomed one year because some of it was contaminated by a dangerous insecticide. Cyclamates were banned on the basis of a handful of reports indicating that it might be a carcinogen when used in small amounts over a long time period. New processes for manufacturing high protein fish flour at greatly reduced costs were rejected because they ground up the entire fish, complete with waste matter.

There is likely to be a shift towards informational approaches as opposed to bans. However, during the decade, questions will be raised about whether consumers can absorb all the information presented, understand the information and use the information in decisionmaking. For example, can consumers understand DOT's complex tire rating system? Several studies have shown the futility of FTC's detailed food advertising disclosures in advertising. Can consumers understand complicated risk information such as is available based on animal testing? Questions will be raised about the processibility of the information disseminated through government regulation.

There will also be a shift from rule making to case-by-case adjudication to avoid the political pressure imposed by legislative vetoes over proposed regulation. The battle over legislative vetoes is likely to persist over the next several years. As Elliot Levitas one of the leaders of the legislative vetoes has suggested, "We are returning to the people the control over these regulations and at the same time we have sensitized the "bureaucracy". However Stanley Cohen from Advertising Age has suggested, "It will enable powerful interests to affect the regulatory process in secret, behind the closed doors of individual committees and in individual members' offices." As the decade begins, the proponents of legislative vetoes appear to be winning the war. However the increased burden on Congress to cope with the extensive lobbying will eliminate additional pressures for legislative veto efforts towards the end of the decade.

Finally, there will be increased emphasis on enforcing current regulations rather than on initiating new regulations. While this is a welcome change for many people, there may be negative effects. Regulators will be content to develop a scorecard of politically safe enforcement actions, which may represent little consumer gain, but may create the illusion of regulator victories.

Sensitivity to Political Pressures and to the Political Process

There is a serious question as to whether regulators' alleged insensitivity to political pressures is a sin or virtue. As the decade begins we are witnessing an end to truly independent regulatory agencies, since they are becoming increasingly politicized. We note the specter of the FTC Chairman Michael Pertschuk being "roasted" on the

Hill rather than serving in his role as an independent commissioner. During the decade, regulatory agencies will develop increased sensitivity to political pressures. On the other hand, these pressures may create management difficulties, such as have been observed at the Department of Energy, since little time will exist for long-range planning. Agencies will engage in more sophisticated "lobbying," although formal lobbying efforts on the Hill are prohibited. Congressional staffs will be given information about key issues and the regulatory agency will be "positioned" as merely acting as the servant of Congressional committees who will demand that the agency initiate an investigation. Regulatory agencies will play the role of feeding ideas to legislative staffs who will hold hearings and "pressure" agencies to respond to public needs.

In addition, regulatory agencies will become more innovative in generating public pressure for needed reforms. Investigative television shows, such as Sixty Minutes or 20/20 are always seeking to expose a consumer fraud or report on an issue of general consumer interest. Regulatory agencies will become more sophisticated in using the media to further regulatory activity by generating public pressure for regulatory action.

Position of Business in Regulatory Proceedings

Business interests have reacted to what they believe is the inherent unfairness of the regulatory process. Many business persons feel that they are convicted through publicity even if they are found innocent of any wrongdoing. They charge the system is run by people who are unfamiliar and unsympathetic to the business point of view. They frequently note that the regulatory process is not conducive to outside views since agencies appear to have made a decision before the process starts.

In the 1980's, regulators will be more open to the views of business groups, due to the increased power of the business point of view in Congress. There will be attempts at joint research among business and government, although many of these efforts will fail to produce an acceptable product. Business and government legal staffs counsel against undertaking joint research activity. However, the courts will more carefully scrutinize agency rule-making, which will result in government regulators paying increased attention to evaluating low-cost alternatives proposed by business groups.

Study of How Markets Operate

Regulatory agencies will show increased understanding of business activities and evidence less myopia in promulgating new rules. There will be an increased use of academics and MBA's in regulatory agencies. However, there will continue to be domination by a single perspective at regulatory agencies. For example, lawyers will continue to dominate FTC proceedings; physicians will control FDA activities; and engineers will dominate the activities of DOE, CPSP and DOT.

There will be increased contact between business and regulatory groups, which will facilitate a greater understanding of market activity. However, there will be great difficulty arriving at mutual understandings because both business and government groups will continue to have an incentive for conflict rather than cooperation. The Washington representatives on the business side and the legal staff on the government side both have an interest in maintaining conflict.

The real breakthrough in cooperative activity will come from pressures by U.S. industry to be competitive with foreign competition. As a result, regulators will be forced to work with business interests in order to meet the threat of foreign competition. We have already seen changes in the automobile industry, with EPA and DOE working more closely with the automobile companies. In the anti-trust area there will be increased pressures to allow large firms, especially in the communications field, to meet world-wide competition.

The greatest area of cooperation will be in consumer education. Industry quality standards, such as the homeowner warranty program, will proliferate. These efforts, which have been tried by Japan and Sweden, will be tried in the United States as consumers seek more value from products.

There will be less detailed and more flexible rule-making by regulatory agencies, which reflects a greater confidence and understanding of business. There will be much more emphasis on performance standards, which monitor actual output, rather than on design standards.

Economic Analysis

In the past, there has been little economic analysis to determine if a rule is needed and to discover if a regulation is effective. There have been many proposals for increased economic analysis of proposed regulations. At present, these regulatory analysis proposals are not taken seriously by regulatory agencies. These efforts are viewed as a chore to be undertaken after regulation is developed and a regulatory analysis is developed to justify the regulation. However, as additional economists are hired by government agencies to perform regulatory analyses, economists will develop increased power. They will be able to veto regulations because it can be claimed that a regulation cannot be justified economically. However, economists will have great difficulty in undertaking new initiatives, except in the deregulation area. The economist will serve as a brake on regulatory activity and will tend to be reactive as opposed to proactive.

Use of Consumer Behavior Theory and Research

There will be great increases in regulatory agency research budgets, primarily in the evaluation and regulatory analysis areas. However, these funds can be used to support a variety of consumer projects. Overall, there will be more sophisticated government research. However, there will be little basic research into

information processing and other important areas, since most of the research will continue to concern applied efforts.

Summary

Overall, regulators will be more prudent in their choice of what to regulate. Greatest attention will be devoted to situations where markets have failed or aren't working properly. Deregulation will be emphasized as opposed to regulation. Services will come under greater scrutiny. Information remedies will be used rather than product or advertising bans. However, the problem of too much detailed information on product labels and in advertising will be an increasing problem as the decade of the 1980's concludes.

There will be greater sensitivity to political realities; however, this will be accompanied by a loss of independence by regulatory agencies who will be under increasing Congressional scrutiny. There will be increased cooperation between government and business groups, resulting largely from enhanced power of business groups with Congress. There will be increased use of economic analysis and consumer research. However, these fields will tend to be reactive rather than proactive. Economics, as a result of its greater theoretical grounding, will exert far more impact than consumer behavior research. However, consumer researchers will continue to exercise an increasingly important role. There is a difficult road ahead for consumer protection regulation, but regulators will become more prudent and perhaps more successful in their regulatory efforts.

CONSUMER POLICY ISSUES: GLOBAL TRENDS FOR THE 1980's

Graham T.T. Molitor, Public Policy Forecasting, Inc.

Abstract

Consumer laws move in cycles, the current cycle (projected for 1978-1998) including over 4,000 issues not yet adopted in America. U.S. response lags Western Europe's precursor-Sweden - typically by some 6-8 years. Implementing dates for consumer laws can be accurately forecast years ahead of time using some of 160 indicators of socio-political change.

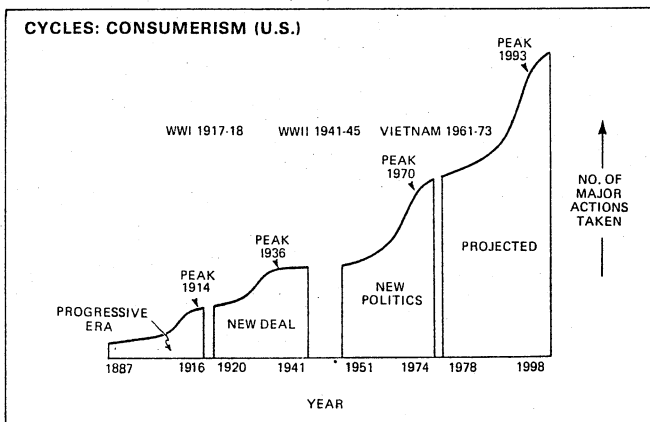
Onslaught of Government Change

How can we tell where we are going? What's in store for the 1980's?

What lies ahead for consumer policy? The current cycle of consumer issues which began in 1978 is growing (Exhibit 1). Presently Public Policy Forecasting, Inc., is following over 4,000 separate and specific consumer issues that have not been enacted by the Federal Government.

EXHIBIT 1

PATTERNS OF CHANGE



© 1979, PUBLIC POLICY FORECASTING, INC.

New laws steadily increase. The U.S. Congress, during a two-year term, receives an average of 20,000 bills and enacts some 1,000 new laws (Exhibit 2). Added to this are an estimated 9,000 federal regulations issued annually. Compound this domestically by the 50 states which entertain some 100,000-125,000 bills yearly, of which 15-25,000 are enacted (Exhibit 3). Add inputs from some 78,218 special districts (counties, school districts, municipal authorities). Next, multiply these results by the 155 or so nations the world over and the overwhelming global magnitude of government-imposed change begins to be perceived. Steadily, and with an increasing rate of growth, this myriad of law and regulation concerns the consumer.

The impact of regulation and accompanying changes in social attitudes also has been seen in the courts. For example, in 1960, there were 50,000 product liability

claims filed; alarmingly the number had swollen to over one million by 1970. In just 3 years between 1974 and 1977, the number of product liability lawsuits filed in federal courts rose from 1,579 to 4,077. Judge and judge-and-jury awards in such cases have mushroomed as well. The record-setting 100 million dollar decision against the Ford Motor Company in a case brought as the result of alleged safety defects of a Pinto automobile is an incredible landmark award. Despite reduction of the award on appeal, the impact of just how far things could go sent waves of apprehension throughout industry. In related judicial trends verdicts against physicians in malpractice suits also dramatically raised their liability insurance costs. Ultimately, of course, it is the consumer who winds up paying.

Paradoxically, consumer legislation and regulation, especially when viewed from the citizen's, as well as from judicial and legislative viewpoints, can be counterproductive to the consumer. Witness the costs passed along to the consumer, such as inflated insurance premiums, added cost of mandatory automobile safety and environmental protection gear, overhead expended for legally mandated paperwork, and so forth.

The tide is toward more and more regulation. More and more change is in store. Large as well as small businesses stand in the growing shadow of government regulation. Products can be banned from the marketplace by the stroke of the lawmaker's pen or by a regulator's arbitrary edict. Limited government is a thing of the past; government-imposed change cannot be ignored (Exhibit 4). Wrenching dislocations may engulf the unwary or the unprepared.

What can be done to get out in front of impending public policy change?

Change. That is the key word.

Changes in public policy seldom come as a bolt out of the blue. Changes evolve.

Change is prompted by social frictions which expose abuses or wrongs. When such "social dissonance" gets out of hand and can no longer be ignored, when the momentum of forces clamoring for reform swell in numbers, government response and reform are not far behind.

The Six Indicators of Change

Six key indicators are most useful for detecting, tracking, measuring, and predicting the trend and direction of public policy change (Exhibit 5).

The process invariably begins with aberrant and unique events. Novel or bizarre, at first, such happenings go largely unnoticed. Their accumulation over time eventually leads to aggregation which helps to reveal meaningful patterns. Event patterns emerge in many different ways. Among them: innovation from introduction of new technologies and social inventions, increases in magnitude, and practical experience.

As events unfold and become the object of attention various authorities and advocates (often the "victimized") observe, analyze and begin to comment on such emerging phenomena. Authorities include the gifted few who can be found in any discipline and the geniuses who propose a

EXHIBIT 2

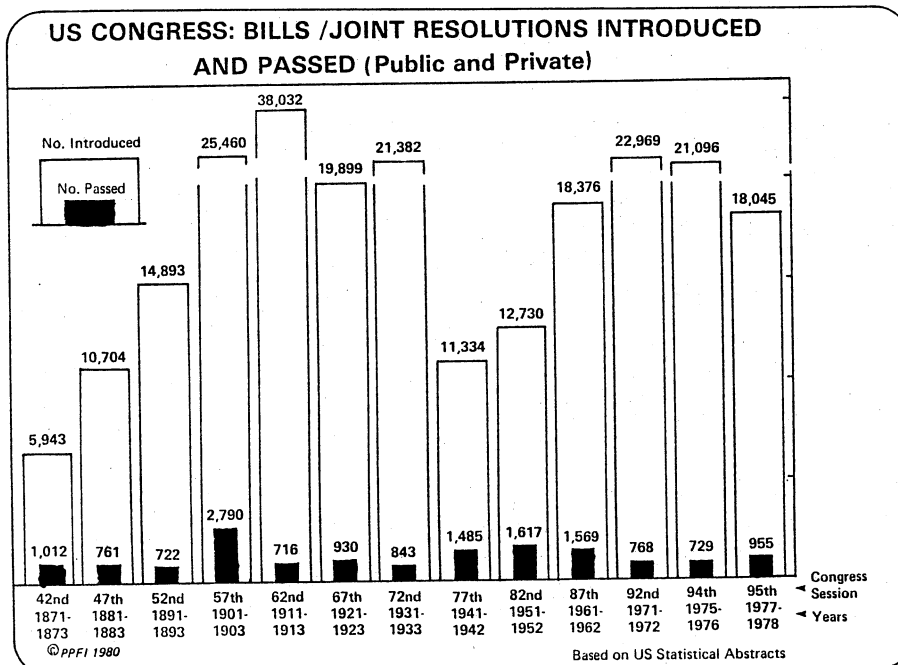


EXHIBIT 3

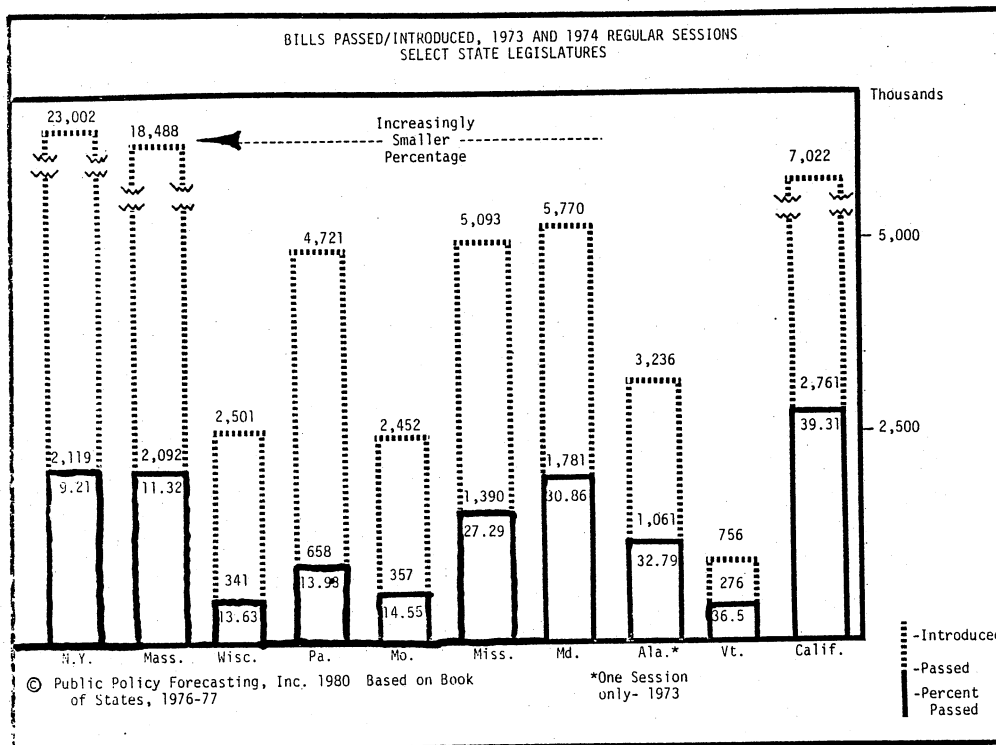


EXHIBIT 4

POLITICAL INVOLVEMENT IN SOLVING SOCIAL PROBLEMS

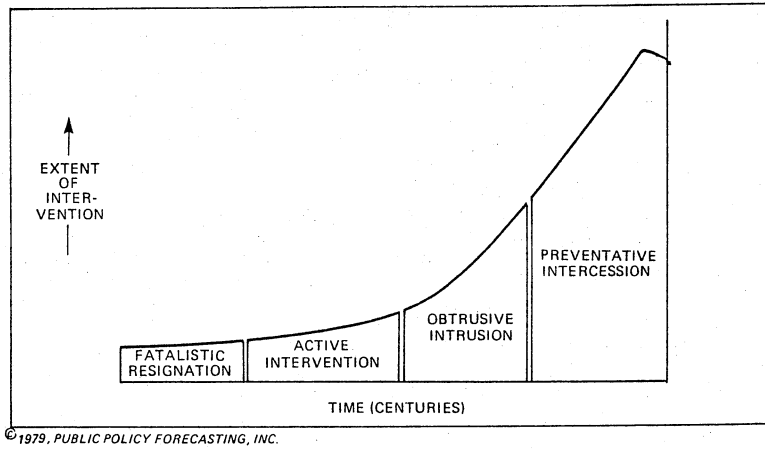
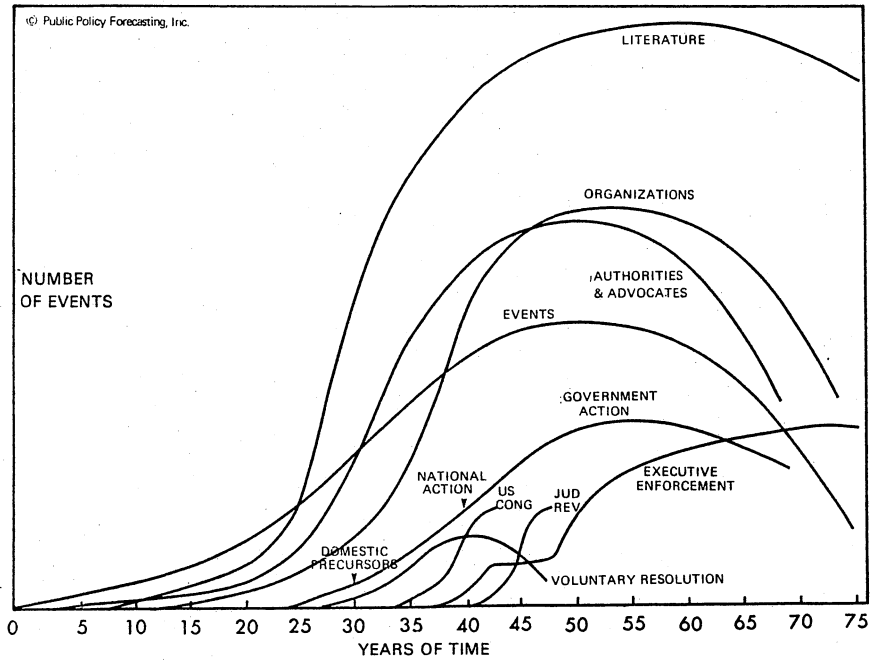


EXHIBIT 5

PATTERN OF CHANGES



theory which may take years to prove. Albert Einstein was one of these men. Less intellectually capable, but extremely vocal, are the victims of an event who can evoke such sympathy that their sufferings begin to focus widespread attention and help bring about change. Galvanizing tragedies often goad public policy change. The sinking of the Titanic on April 14, 1912, led to many reforms and some innovations in the safety of ships which carried into the age of aviation. By 1912, Congress had passed amendments to the Ship Act mandating minimum radio communications/procedures for seagoing vessels. (Later this law was amended by the Radio Act of 1927 which, in turn, was supplanted by the Federal Communications Act of 1934.) The revolutionary fringe, the deviants, cranks, and oddballs of yesterday often start a process which, though ignored at first, may have a profound impact later -- such figures as Alexander Graham Bell, the Wright Brothers, and Benjamin Franklin were not taken seriously initially. Scientific discoveries and technological breakthroughs can be, and often are, major events, even though their impact may not be felt for years.

After the events occur, they are disseminated in what may be termed leading literature, often bringing with it institutionalization in the form of organizations which incorporate the new message and provide a sustained base for advocating change. Popularizers of public policy opinion, the public-spirited crusaders -- the Ralph Naders, the Howard Jarvis types -- make their contribution. So do the prophets of gloom -- the Club of Rome pessimists. Usually such efforts are cast into public interest groups and organizations which institutionalize the cause and provide a sustained base for advancing change.

Politicians and precursor jurisdictions -- domestic and international -- sensing the winds, pick up the trends. Finally, solutions to the excesses and abuses which have been generated are enconced in legislation, regulation, judicial decision, or custom and voluntary undertakings.

The process -- leading events, leading authorities/advocates, leading literature, leading organizations, and implementation by leading political jurisdictions (domestic precursors and leading nation-states) -- tends to follow an S-shaped curve. At the outset, the intensity takes off slowly, follows a steep slope of increase, then tapers off. The curve, when accurately coordinated with time and number or frequency of phenomena indicates past and present position and gives a sense of the momentum for change continuing into the future.

Lead-Lag and Precursor Jurisdictions.

Lead-lag, as used here, describes time lapse between an event and its general adoption. As man's knowledge and capabilities have grown, lead-lag time has shortened. For example, it took hundreds of years to perfect an engine capable of flying the heavier than air ornithopter (1910) which had earlier been designed by Leonardo da Vinci (c. 1500). Only some 15 years elapsed from the time that Alexander Fleming discovered penicillin (1928) to its commercial availability (1940's). A mere 3 years is alleged to have elapsed between invention of the transistor (1948) by Schockley, et. al, and its commercial introduction.

In terms of socio-political lead-lag, it has long been recognized that certain jurisdictions are -- at least for a season in history -- precursors of change. Precursor nations are the adopters, the leaders of innovation. What occurs in these jurisdictions today often is what happens elsewhere tomorrow.

Nations may be broadly classified into three groups according to their willingness and capacity to respond and adapt to change. The first group, bold/innovators and

adopters, include (in typical chronological order) Sweden, Norway, Denmark, The Netherlands, West Germany, the United Kingdom, the United States, and Canada. Diffusion time among these nations varies as to issue. Some 6-10 years for consumer issues has been typical over the last few decades. There are few nations in this group (20-26 countries). Next come the great majority of other nations, which take from 10 to 50 years to follow the leader of the bold early innovators. They tend to range from the industrialized and advanced nations not included in the first group (Spain, Portugal, Italy, and Greece) to some of the progressive emerging nations. A minority of very poorly developed nations comprise the third group, the laggards which typically require many decades to respond to change. Frankly, they are the poorest and the least socio-economically developed nations -- Upper Volta and Zambia, for example. One must remember that the patterns, while normally consistent, are not invariant. At best, such diffusion patterns are a guide to the pace and direction of socio-political change. (Exhibit 6)

Sweden as a Precursor Nation

Over the last few decades, Sweden and the other Scandinavian nations have achieved a remarkably consistent record of "firsts" in consumer affairs. Swedish government officials are realists; indeed, in these nations, nothing succeeds like success and success depends upon tangible accomplishment. Perhaps more important they are motivated by a keen sense of and commitment to social responsibility.

Sweden has many tangible accomplishments to its credit. In the most basic terms -- terms of longevity -- the Swedish people live longer on the average than anyone else. This may be more than an accident of geography or genes. Swedish public policy places a high premium on preventive medicine and related matters. Sweden has a very low death rate from cirrhosis of the liver and a low rate of alcoholism. This is no mere coincidence. To accomplish this the nation placed many restrictions on alcohol, including high taxes based on the content of alcoholic beverages. Distilled spirits have a 90% tax as a percentage of retail price, and the tax graduates down to 50% on light foreign wines, 45% on strong beer, and 10% on very mild beer. Seven percent of Swedish inland revenue came from alcoholic beverage taxes in 1974. Furthermore, the state has a monopoly on the sale of all products with more than 3.6% alcohol, and it operates the package stores -- only 300 in the entire nation in 1976.

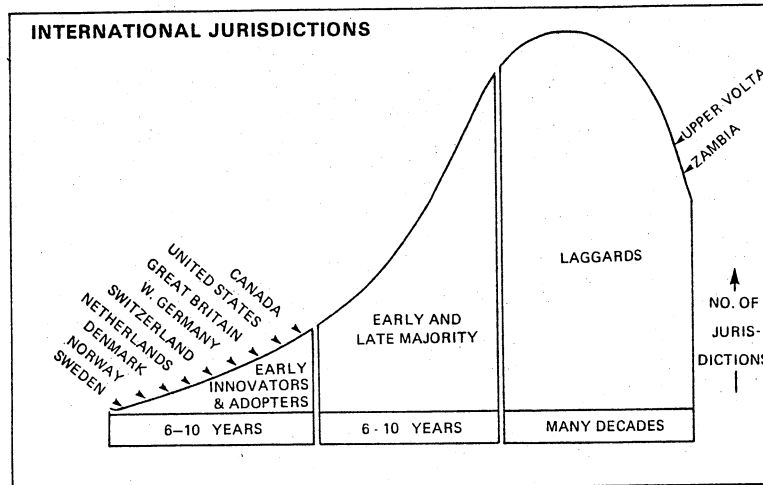
The hand of the Swedish government is felt in many other ways. Extremely stringent controls on advertising over alcohol and tobacco have been enacted. Included have been restrictions in outdoor, movie, and direct mail media to mention a few. Cigarettes must be sold with a "negative marketing approach" in which one of several regularly rotated warnings are placed on the packages. These warnings are much stronger and more persuasive than the U.S. warning. All media and information channels are used to discourage smoking.

The results? Swedes smoked 1,620 cigarettes per capita in 1970 versus 3,971 in the U.S. Part of the result of the state controls can be seen in the voluntary self-regulatory measures adopted by the Swedish Tobacco Branch Association in 1975 which included media limitations (ad bans and restrictions), advertising content restraints (e.g., limitation of pictorial displays, prohibition of pictures of humans or natural scenery unless the picture is part of the trademark), and other promotional restrictions.

Swedish action concerning alcohol, tobacco, and many other products is just a part of a large framework of preventative health measures. Swedish planners long ago

EXHIBIT 6

PATTERNS OF CHANGE



committed themselves to the proverbial ounce of prevention being worth a pound of cure. National health policy, legislated and voluntary, has fallen into line.

The United States as a Precursor Nation: Myth and Reality

Many Americans mistakenly assume that this nation's leadership is all-pervasive, that America always -- or even usually -- leads the world. However, a careful study of history reveals that America has tended to follow the legislative lead of advanced European nations. (Exhibit 7)

During the 1700's and 1800's, the lead-lag was from 30 to 60 years. In more recent times, lead-lag relationships have been contracting for many reasons (the communications and transportation revolutions, for example), and international diffusion occurs at an ever-increasing rate. During the twentieth century, the lead-lag pattern has been reduced to some 5 to 20 years.

The U.S. has often been among the very last of the great nations to legislate a number of major programs:

- The U.S. was one of the last nations to abolish slavery.
- The nation is still a hold-out on compulsory, universal health insurance.
- With the exception of a very few small nations, the U.S. is the last nation to commit itself to metric conversion.

In terms of social and consumer issues, the U.S. lead-lag has been painfully conspicuous. The U.S. adopted:

- pure food and drug laws 46 years after the United Kingdom;
- workman's compensation 37 years after Germany;
- old age assistance 46 years after Germany;
- female suffrage 27 years after New Zealand;

- unemployment insurance 30 years after France;
- national nutritional (macro-nutrient) goals 6 years after Sweden.

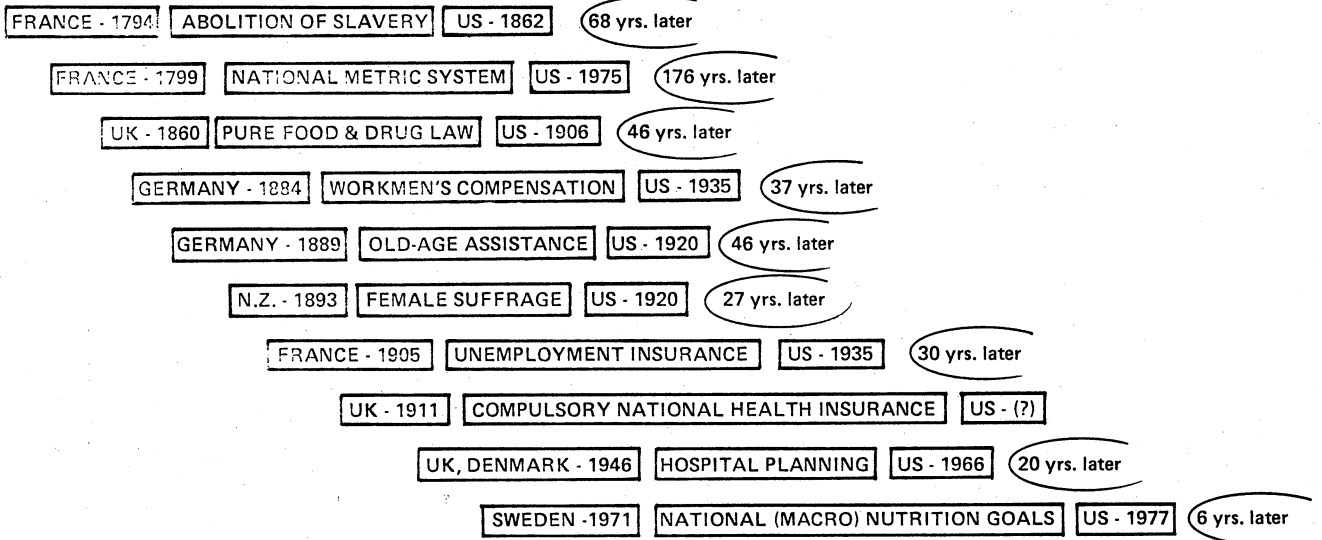
Yet, a cautionary note is in order: just because Sweden or other precursor nations have adopted innovative policy changes does not mean that, *ipso facto*, the changes will be adopted elsewhere. Innovative change and dissemination is not an inexorable process. Fatalistic attitudes are not indicated. Some changes, for example, may not get too far:

- Norway, land of the ski, banned the manufacture, importation and advertising of skateboards in 1979.
- Sweden also during 1979 banned war toys, not an unusual move for a nation that has not been at war for over 150 years.
- Sweden also has taken steps to gird children's rights by restricting and limiting the spanking or humiliation of children.
- Finnish libraries (perhaps with tongue-in-cheek) have banned a corrosive symbol of Western decadence -- a literary character who has nephews of uncertain parentage, who has protracted a romance into a 50-year long love affair without benefit of matrimony, and who runs around with his rear-end exposed. Thus, one should not be so brash as to ask a Finnish librarian for any publication featuring Donald Duck!

Domestic Precursor Jurisdictions in the United States

With minor exceptions historically, certain jurisdictions in the U.S. have been consistent innovators in public policy: New York City and State, Boston and Massachusetts, and, more recently, California, are the leaders. New Jersey often follows the policy lead of its dominant neighbor. Dade County, Florida, is a newcomer that also has moved early in adopting new legislative concepts. Illinois, once a bellweather jurisdiction, has been

INTERNATIONAL DIFFUSION OF PUBLIC POLICY Leads & Lags: Early or 1st Adopters and US



© PUBLIC POLICY FORECASTING INC., 1980

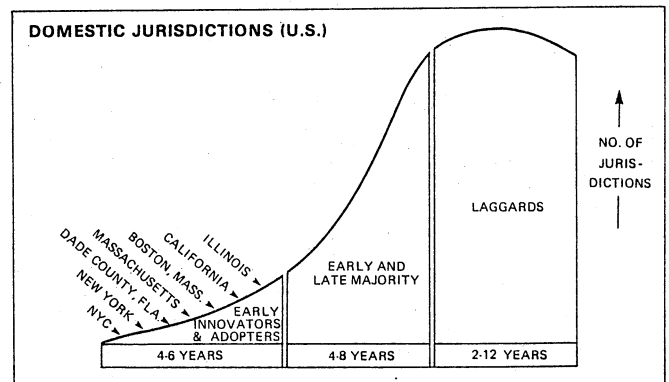
slipping from a vanguard position in recent years. Some states are notable for innovation in one or two areas rather than on a broad-spectrum basis: Oregon with respect to environmental matters, or Minnesota on consumer and environmental issues. (Exhibit 8)

EXHIBIT 8

Once four to six precursor jurisdictions have acted upon an issue, the die often is cast. Taking the cue from experience gained by the early innovators and adopters, most other states are likely to follow suit. The less affluent, less literate, less culturally advantaged places -- Mississippi, Alabama, South Dakota -- usually bring up the rear. The lead-lag time from precursor to laggard states ranges from 5 to 10 years -- sometimes longer; a similar time span has been noted between state adoption of major legislation and federal action.

Federal action is probable when a number of factors -- singly or in combination -- are present. Diverse state approaches can hinder interstate commerce, and almost invite Washington's interest to provide uniformity. State-by-state enactments can be insufficient or unable to deal effectively with problems, particularly those transcending geographic borders. Federal action can cover the interstices. In times of great urgency caused by a spectacular event (airplane crashes in the early 1920's)

PATTERNS OF CHANGE



© 1979, PUBLIC POLICY FORECASTING, INC.

or widespread condition (the Great Depression), the states may be unable to deal with the problem by themselves because of logistical, funding, or leadership problems. Again, a cautionary note -- federal action sometimes precedes state action.

Identifiable factors predispose precursor jurisdictions in the U.S. as they do in Europe:

A large amount of activity in an area (large population, as in New York, Boston, California, intrudes new problems associated with size); large expenditures (as for medicaid in these huge jurisdictions where an astounding 40% of total U.S. expenditures occurred during the mid-1970's);

Special patterns of use (sheer size giving rise to traffic strangulation control in New York and Los Angeles);

Terrain and associated meteorological conditions (bringing on confrontations with air pollution in California and Kentucky).

Philosophical predispositions are also a factor. Liberal states are generally more innovative and progressive, and thus are likely to arrogate power to government. As time goes by, reform breeds further reform. However, excess reform can turn the political climate toward conservatism. The balance between these two extremes constantly is shifting over time -- like the swing of a pendulum.

No jurisdiction has a monopoly on any ideas or issues. Thus, comprehensive analysis of legislation/regulation requires monitoring of both precursor and other jurisdictions. Public policy trend analysis can provide knowledge as to precisely where an issue stands, what issues are flashed in the pan, and what issues are most likely to run their logical course.

In addition to legislative harbingers, it is also wise to monitor court decisions. In this century, the courts, especially on the highest state and federal levels, have exercised a sense of social awareness which has often been translated into reinterpretations of the common law. U.S. courts can, in effect, make law by restating, refining, expanding or contracting its meaning. Judicial precedent often reveals shortcomings that lead to across-the-board legislative standards.

Monitoring landmark decisions, counting cases on a specific issue, following groups which test the limits of the law (e.g., Americans for Democratic Action, the American Civil Liberties Union), and measuring the time that certain cases take to reach final appeal are just some of the many methods for sensing and predicting this aspect of change. The courts have been a major change agent in a day and age when they have assumed an increasing role in legal interpretation, becoming what some observers feel is a shadow legislature.

Finally, the actions of the executive branch should be followed with care. The actions can take several forms:

1. Legislation in effect proposed by the President.
2. Legislation vetoed by the President.
3. The extent and vigor of executive enforcement of law and regulation -- or the lack thereof.
4. Occasional but weighty Presidential actions. Broadcasts to the nation are one of these. Nomination or dismissal of key appointments is another.

The Issue Universe for Consumerism

Analyzing and forecasting a particular problem area begins by charting the total issue universe. The process begins by -

- (1) identifying the broad generic issue -- consumerism in this case,
- (2) delineating the major issue sectors or groups (e.g., product safety, advertising regulation, macro-nutrient content, labeling disclosure),
- (3) sub-categorizing each sector/group into discrete units.

Some of the major issue sectors Public Policy Forecasting, Inc., has charted are diet and health; food safety and quality assurance; advertising regulation; telecommunications; sugar, alcohol and tobacco regulation; organizational and institutional response to consumerism; health care cost containment; and others.

Issues such as these are broken into subcategories in the third stage. For example, under food safety and quality assurance, Public Policy Forecasting, Inc., has identified over 275 impending issues such as food colors, other additives, additive disclosure systems, allergens (primary and secondary reactions), synergisms, Delaney Clause revision prospects, natural foods, testing protocols, procedures and standards, and so forth.

As technology advances, the list grows, often logarithmically. One food safety drive aimed at reducing sugar consumption included over 125 individual issues aggregated into 20 sub-categories. New methods of measurement allow for the detection of one part per trillion of some food ingredients -- the equivalent of one drop of vermouth in 520 railroad cars each containing 30,000 gallons of gin. Coffee has been found to contain an estimated 2,000 individual chemicals. Diesel exhaust contains over 30,000 chemical components. At these rates, almost any product can be indicted as posing a risk to health.

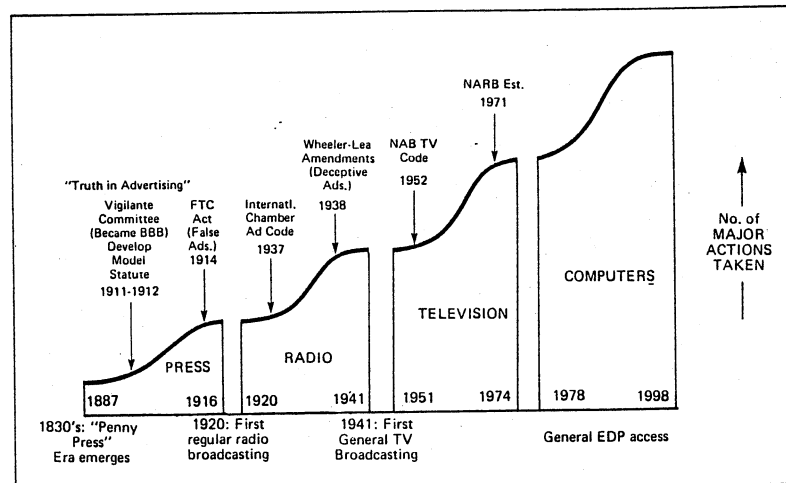
Increasingly, decisions will be premised on such microscopic detail. Chemical reactions, right down to the individual basic atomic elements reactions and relationships, increasingly are becoming the new frontiers of government investigation and policy making. With each potential posing risks (or benefits), the magnitude of the issue universe we are attempting to describe can hardly be underestimated.

Cyclical Change

Though consumerism does not hold the headlines as much as it did in the late '60's and early '70's -- the Ralph Nader years -- and although consumerism may seem to be receding in light of current anti-regulation and tax-cutting (Proposition 13) moods, these are but a temporary pall, a breathing space, before the next large-scale expansion of consumer issues. A growing backlog of such issues continues to build up like the pressure in a knotted house. Remove the restriction and the stream will shoot forth once more.

Consumerism tends to move in spurts or bursts, a phenomenon which is one of the major techniques of investigating public policy matters. Historic cumulative plots of major consumer laws and regulations (Exhibit 1) reveal three major cycles, each lasting about 10 to 30 years. The first, during the Progressive Era, was 20 years in duration -- 1887 to 1917, and emphasized social justice espoused by social reformers such as Lincoln Steffens and Theodore Roosevelt. A second cycle started after the war and the recovery period lasted from 1920 to the beginning

CYCLES : ADVERTISING REGULATION



© 1979, PUBLIC POLICY FORECASTING, INC.

of World War II, and climaxed vigorously during the New Deal.

Following the end of World War II, the recovery phase, and the Korean War, the third cycle, emphasizing post-war and post-industrial transition and the New Politics, lasted from the early 1950's to 1974. This period did not have much movement until 1958 with the enactment of the Automobile Disclosure Act. Following this were Senator Kefauver's drug investigations, the truth-in-packaging drive, and the Kennedy years. The fourth cycle started in 1978. Public Policy Forecasting has predicted this new cycle will last until the turn of the century. It should be noted that issues other than consumerism follow similar cycles. We have found patterns for social welfare laws, women's rights, environmentalism, and other issues.

Each of the four consumer cycles noted also included the appearance and dominance of new communication technologies (Exhibit 9): a largely literate audience reading inexpensive newspapers and magazines at the turn of the century, especially in the cities; regular radio broadcasting beginning in the mid-1920's; television starting in the early 1950's; and computers and other electronic media by the late 1970's. Each major communication mode was accompanied by advertising. In each case overzealous advertising practices introduced abuses and excesses. Such social frictions became unbearable and efforts of voluntary rectification were followed by government-mandated standards.

Each cycle occurred during a period of major socioeconomic change, accompanied by relative prosperity. Conversely, during times of military confrontation, national priorities are altered to a wartime footing, requiring tremendous economic, monetary, and manpower commitments. These displacements relegate social interest issues to a low priority. Following cessation of hostilities a two to three year period of economic readjustment also deters commitment to social policy programs of all kinds, including consumerism. Postponement does not mean abandonment, however.

Some Corporate Advantages Gained from Anticipating and Responding to Impending Consumer Issues: The Role of Forecasting

Consumer issues imposed by governments increasingly intrude on private sector activities, management prerogatives, and market decisions. Often politically, rather than rationally inspired, some of these changes can be difficult to foresee and assess. Nevertheless, short-term trends are generally clear down to the level of specific issues; longer term trends are harder to predict, but the basic directions are usually distinct. What is important in this era of rampant and pervasive change is to keep one step ahead of regulators and to stay abreast of consumer demands, which are invariably mirrored by government action later.

Some advantages business can gain from foresight are:

- Identify or corroborate new marketing opportunities and problems.
- Gain the lead time required to develop, re-design, or reformulate products and procedures. Check design soundness, establish alternate supplies, conduct test marketing surveys, reposition existing products, and so forth.
- Avoid, or at least minimize, the often costly impacts of government mandates which catch unwary managers by surprise (e.g., costly and embarrassing inventory and recall/disposal problems following bans).
- Consider undertaking coordinated industry-wide lobbying to champion, redirect, or defeat impending change, or to cooperate in writing laws that all can live with comfortably.
- In the same vein, avoid the cost and bad public relations in lobbying for lost causes.

- Exercise industry leadership by moving first and setting trends, rather than just responding in a purely reactive mode which often results in a negative public relations image. In the long run, such actions can develop positive public relations potentials by building the corporate image in ways paid advertising cannot.

- Respond positively to humanitarian causes and socially desirable goals. When there is a low ebb of consumer confidence -- as is now being experienced -- responding positively to moral and ethical issues builds corporate stature.

If one can foresee the future, one can help manage it. The past is prologue.

Short shrift must not be given to historical and global perspectives. Much can be gained from past experiences and actions taken elsewhere. Short of this we may be doomed to repeat the mistakes of history.

SWEDISH CONSUMER POLICY, ITS TRANSFERABILITY AND
RELATED RESEARCH IMPLICATIONS¹

Hans B. Thorelli, Indiana University

Abstract

An overview of recent Swedish consumer policy, public and private, constitutes the first part of this paper. The one-round Delphi study which follows is focused on future developments in Sweden. A third part briefly discusses some fascinating issues involved in technology transfer in the public policy area, with examples from consumer policy. Finally a catalog is offered of research implications suggested by the paper.

Recent Swedish Consumer Policy

A dynamic analysis of Swedish consumer policy in the 1940-1975 period was presented in Thorelli and Thorelli (1977). A more detailed, institutionally oriented survey of the pre-1970 years is found in Thorelli and Thorelli (1974). Our purpose here is to update these materials, essentially covering the years 1976-80.

The Setting

The consumerism debate in Sweden has calmed down in the last five years.² It has become less passionate, less personal and there is generally less exaggeration in argumentation. This statement is valid also for Råd och Rön, the consumer journal published by the State Consumer Board. It has almost doubled its circulation from 110,000 subscribers in 1974-75 to some 200,000 in 1979-80, despite a doubling in price to a modest \$5.50 per year. The journal now has a penetration rate about twice that of Consumer Reports. Consumer activists are given less play by the media, although there is a popularly viewed, regularly scheduled TV program.

This does not mean that either consumerism or public debate is dead. The consumerist "movement"—Sweden has no general-purpose consumerist groups—has reached a stage of maturity where it is likely to remain for a while. At the same time the climate of opinion has become more nuanced, generally a healthy development. An example is the series of eight or more articles in the Stockholm Svenska Dagbladet, March-June 1978. Another is the conference "Consumer Policy Faces the 80's" of government, academic, business and media opinion leaders arranged by the Ministry of Commerce in May 1979 and also the excellent

collection published by the Swedish Committee for Economic Development (SNS) in 1979.³

A sign of the times is some serious soul-searching going on in the State Consumer Board. This self-examination is not in any direct way related to the change from social democrat to coalition government by liberal parties. It does seem symptomatic of a readiness to accept or consider some new signals. There is also some unrest among the technical and economic professionals about what a fair number of them consider the "lawyer oligarchy" absorbed by the State Consumer Board (KOV) when it merged with the Office of the Consumer Ombudsman (KO) in 1976 and when KO also became Director General of the Board. In part the spirit of reevaluation is no doubt also related to a certain amount of soul-searching in and about the welfare state as such—also a fairly recent Swedish development.

A recurring theme in the expressions of opinion cited earlier is the need for a gradual shift in emphasis of policy from protection towards emancipation of consumers. In this context the need for a fresh look at consumer information programs is often stressed. Sweden pioneered voluntary informative labeling of consumer products, but this program was abandoned by the State Consumer Board during the years of the Left Twist. The desire for "help towards self-help" more than additional regulations was clearly manifested in a recent major academic survey of Swedish consumers (Wikstrom 1979). On the other hand, another large survey sponsored by the State Consumer Board seemed to call for more protection.⁴ Few such large-scale surveys of consumer opinions of consumer policy issues had been made before in Sweden (Thorelli, Becker, Engledow 1975, pp. 194 ff.).

New Laws and Proposals

The momentum of consumer policy proposals enacted into law or subject to scrutiny by the perennial government investigating committees characteristic of the country has been maintained in the last few years. Sweden took the ultimate step towards centralization of general consumer policy by the merger of the Office of KO and the State Consumer Board (KOV).⁵ At that time the 1971 Marketing Practices Act had been succeeded by a new Marketing Act, which strengthened the old law in two important ways.⁶

¹The author expresses his appreciation to The Swedish Bicentennial Fund for the fellowship which made this research in Sweden possible. The Fund was established in 1976 by the Government of Sweden to promote scientific and cultural relations with the United States. Dr. Sarah V. Thorelli participated in all aspects of the research. We are most grateful for the outstanding cooperation of Sven Heurgren, the Consumer Ombudsman and Director General of the State Consumer Board, and his associates at the Board, and to the Dean and staff of the Stockholm School of Economics, and literally scores of other Swedish friends.

²This is recognized by the Director General and the editors of the State Consumer Board; see Konsumenträtt och Ekonomi (1979:2), pp. 2, 30.

³Bengt Ryden, ed., Mot en ny Konsumentpolitik (Towards a New Consumer Policy, Stockholm: SNS, 1979).

⁴Konsumentundersökningen 1978-79 (Stockholm: State Consumer Board), Swedish mimeograph summary, "Konsumentpolitiken inför 80-talet," by L. Lundvall of the Board. Pending the final report some aspects of the survey have been highlighted in Råd och Rön. A few questions of this survey seemed less professionally worded than might be desired.

⁵The law chartering the reconstituted State Consumer Board is Forordning med Instruktion för Konsumentverket, SFS 1976: 429.

⁶Marknadsföringslag, 15 December 1975 (SFS 1975: 1418).

A firm (or trade association) may now be required by KO to supply product information deemed of special importance to consumers. His decisions may be appealed to the Market Court (T. and T. 1977, p. 223), in which case he has the burden of proof. The second innovation deals with the sale and rental of hazardous commodities. The marketing of such products may be prohibited.⁷ Special product control legislation was already in place for food products and goods which entail danger to the environment.

Another new law of general importance is the Consumer Credit Act of 1977.⁸ A seller or financing agency in marketing credit must give information about the effective rate of interest (including administrative charges, etc.). The same data shall again be given any individual customer before credit is granted. As a rule down payments in instalment sales should not be less than 20 percent. The new law also restricts the possibility of sellers and third parties repossessing goods in case of default.

Sweden is primed to adopt strict liability as a general rule in the sale of consumer goods through the pending law on product liability in such sales. We may note, too, that in 1981 the Central Consumer Complaints Board (ARN)--hitherto regarded as an "experimental activity" within the KOV--will be established as a fully separate activity. It would be quite premature, however, to see in this a return to decentralization of consumer policy.

The committee to study the desirability of giving KO certain powers to subpoena information from individual firms presented its first report in 1979, including a draft law providing for sweeping powers of that kind.⁹ KO in an official commentary to the report saw the need for even broader powers as far as past and present activities of firms are concerned. However, he suggested on both practical and principal grounds that his powers to require data about product planning and other future-oriented company activities be limited to cases where issues of product safety would likely be involved.

Another major report presented a draft Consumer Services Act in 1979.¹⁰ The principal investigator has discussed the proposal in detail in Journal of Consumer Policy (Bernitz 1980). Briefly, the proposed act will cover most services involving work on tangible objects, houses and land. The law governing such services is poorly developed in most countries. The act provides for freedom of contract as to the manner of performance of a service as well as to its price. However, once a contract is made, many of the proposed rules are mandatory, i.e., they cannot be set aside by contract to the detriment of the consumer. An interesting example is that the entrepreneur is liable for information furnished when marketing the service: "acts of marketing, such as advertising, should be considered as belonging to the contracting process. . . . Accordingly, a deviation from statements in

⁷The same applies to commodities which are "obviously unfit for their primary purposes." After a test case involving a notoriously poor washing machine, in which KO won only a Pyrrhic victory at the Market Court, that provision has fallen into disuse.

⁸Konsumentkreditlagen, SFS 1977: 981, amended by SFS 1978: 598.

⁹Konsumentinflytande genom Insyn?, SOU 1979: 5.

¹⁰Konsumenttjänstlag, SOU 1979: 36.

advertisements and similar communications aimed at the general public should be considered to constitute a defect in the service performed." (Bernitz 1980). Two investigating committees recently appointed should be mentioned. One is concerned with "market surveys, comparative testing and other types of product research," as well as the various uses, and the distribution of, the information resulting from such activities. The purpose of the other is an evaluation of KOV activity in the area of marketing guidelines (below).

Public Policy Enforcement

Given an ecologic view of institutions it is only natural that trends of enforcement in Swedish consumer policy have responded to environmental developments as well as to the leadership of Sven Heurgren, KO as well as Director General of KOV since the merger. It is important to remember that Mr. Heurgren--who was KO before his present appointment--is a lawyer, and that he brought much of his legal staff with him to KOV. No one may doubt that the top priority of KOV is the enforcement of the Marketing Act and related legislation. Inevitably other aspects of consumer policy--such as education, information, product testing--have suffered somewhat, at least by comparison. To a fair extent technical and economic staff at KOV have been engaged in the professional backup work necessary in effective law enforcement. Indeed, malice would have it that in effect the KOV has become an enlarged Office of the KO.

The most dramatic shift in policy enforcement has been from the brash jawboning of producers and anti-business tirades of the former Director General to the development of guidelines governing particular industries and trades. This change should be seen against the background that Swedish politics is no longer dominated by the Left Twist phalanx and that Mr. Heurgren, while aggressive and publicity-conscious, comes equipped with a sober legal temperament. He also stands for a sense of fairness, which somehow got lost in the missionary zeal of the predecessor administration.

The day-to-day handling of malpractices by particular firms has been described in Thorelli and Thorelli (1977). Of greater principal interest are the guidelines somewhat similar to the trade regulation rules of the FTC. Such guidelines are not new to the Swedish scene (T. and T. 1977, p. 219). What is new is the emphasis placed on them and the institutionalization of the system. KO himself regards the guidelines as the most important single means of influencing producers. Guidelines are typically worked out in negotiations between KOV and a trade association. Thus, the activity reintroduces in consumer policy enforcement an element of pluralism which had been on the wane during the Left Twist years. KOV has the authority to issue guidelines in the absence of an accord. However, it strives to achieve a "voluntary" agreement. If an industry member abridges the agreement, KO can take him to the Market Court, which alone can make the final decision as to the legality of a guideline.

Most of the guidelines are concerned with product safety, and many of them also prescribe certain minimum standards of product information. Advertising of alcohol and tobacco products is severely restricted. Several guides deal with energy labeling of such products as cars and major appliances. Some 30 guidelines are now published in the official KOV collection ranging from grocery store advertising and in-store information concerning "specials" on food products to safety standards for ski bindings. A similar number of proposed guidelines are in various stages of negotiation and processing, and at least another

score figure in KOV internal planning. The guideline for skateboards is significant in indicating the new trend in policy: sale of the boards is permitted with the proviso 1) that they conform to specified safety standards, 2) that they carry elaborate warnings, and 3) that the seller indicate that skateboards are unsuitable for children under 12. Norway banned skateboards outright, something Sweden could have been expected to do only six years ago.

The development of guidelines is a slow and arduous process--even in cases where the trade is eager to cooperate. It must not be forgotten that there are tens of thousands of products in the modern marketplace. The system is also not free from controversy. The Chairman of the Market Court has pointed to the risk that guidelines might thwart constructive competition, and stated that many firms are sufficiently shy of bad publicity to agree to undesirable restrictions on their freedom of action.¹¹ This same argument was also voiced by a spokesman for the Federation of Swedish Industries, which is trying to stimulate its members to request KOV to demonstrate the need for its proposals more thoroughly than heretofore. As KOV apparently feels it represents consumer interests sufficiently well, there has been little or no effort to engage individual consumers or groups thereof in its work. As KOV is well aware, it is also true that resources do not permit a systematic surveillance of compliance with the guidelines; enforcement effort is typically contingent on consumer or competitor complaint. These types of concerns provide some of the impetus behind the appointment of the investigating committee to evaluate the guidelines. The system will almost surely survive this examination, though likely in somewhat modified form.

KOV nowadays does little by way of comparative testing for consumer information (CI) purposes. Some tests are contracted for CI in Råd och Ron (occasionally even paid for out of the magazine's budget). Most of the tests handled by the technical bureau now are either on contract from individual firms or to provide backup data for guideline negotiations. KOV does participate in the Nordic Council comparative tests which are reported in some form in Råd och Ron. KOV also publishes a quarterly journal on consumer law and economics read mainly by marketing executives and lawyers, home economics teachers and other professionals. Several buying guides are also published. Occasionally a major comparative test will be undertaken, or contracted for, to provide material for a special monograph, such as the excellent ones on washing machines and dishwashers and the economics of different models of cars (acquisition, performance, maintenance, repair, taxes, insurance, etc.). A fair amount of consumer education materials for schools (involving issues of methodology as well as subject matter) is also produced.

Little is still being done about CI in the more restricted sense. The relative decline of comparative testing has been mentioned. The only novelty worth mentioning is what may be called "negotiated CI," i.e., information rules in the guidelines. Most negotiated CI is, however, product-specific, i.e., there is no common scheme or even logo, and the information required is typically confined to safety aspects. The few energy labels among the guidelines provide no information beyond the consumption of fuel or electricity under specified conditions. KO has no interest in allocating resources to a broad-spectrum voluntary

¹¹Justice Peter Westerlind in address before the consumer policy conference of the Ministry of Commerce, May 3, 1979.

informative labeling and/or quality certification program,¹² although it would be within his power to do so. He feels it is the responsibility of individual producers to decide what information they should provide beyond that which may have been mandated in the 30-odd guidelines. For several years KO also refused to support research and development of computerized CI banks, expressing the fear that they might easily become subject to abuse by advertisers or other commercial interests.

In the last five years KOV has supported a good deal of academic research more or less related to consumer policy. Some of it has been of high quality. As mentioned, it has also sponsored a major consumer survey under its own auspices. An interesting initiative is the program currently being launched establishing minimum standards for auto repair shops.

KOV has devoted some special attention to the problems of consumers living in sparsely populated areas where the local general stores and/or gas stations find it increasingly difficult to survive. Such outlets may get modest subsidies for special-purpose investments, such as freezers and mobile markets. KOV has recommended to such state bodies as the Apothecary, the Lottery, the Post Office, the Alcohol Monopoly that they remunerate these distant stores somewhat more generously than before for services performed for these agencies. These authorities were not amused. Support for home delivery services organized by local government may be obtained, and so on. Experiments have also been made with special bussing of distantly located consumers to the nearest store or community. In 1975 a law was passed creating a system of consumer advisory services (budgeting, product information, complaint advice, etc.) to be administered by local and regional government on a voluntary basis. By Fall 1979 some 178 local government units of a total of 277 had decided at least in principle to establish such services. However, only 34 of them had one or more full-time consumer advisers.¹³ Most of these plus some 30 other communities had established consumer councils within the local government structure. KO has been recommending that local consumer services be made mandatory, but there is a real question in our mind whether KOV at this time has enough educational and CI materials to back up such a system effectively.

It may be added that some special areas of public consumer policies are handled by other agencies than KOV, such as the State Food Administration and the Environmental Protection Board.

Consumer Policy at the Enterprise Level

Too often one gets the impression in academic as well as public debate that consumer policy is an activity reserved for governments. But the term reasonably applies to any organized measures to promote consumer interests (Thorelli, 1972). Thus, consumer policymakers include consumer groups, the media, educational institutions, business firms, trade and standardization associations, etc. In

¹²KO stated this in a meeting of the Stockholm group of SNS on May 12, 1980.

¹³The Stockholm office is the largest and very centrally located. However, we found that on two Fridays in May 1980 the office was closed--at a time when the city was full of shoppers. The preceding Thursdays were national holidays and all personnel had been allowed vacation days on these busy Fridays. Though stores are open the office is closed on Saturdays.

the case of individual firms the border between consumer policy and general business policy is none too clear. We shall assume merely that consumer policy refers to measures going beyond "providing a good product at a good price."

No systematic survey has been made of consumer policy at the enterprise level in Sweden. If we include--as one probably should--voluntary cooperation with KO-KOV and the Central Consumer Complaints Board here, it is clear that thousands of firms are engaged in consumer policy to varying degree. What firms do on their own, however, generally appears less impressive. Students of Professor Solveig Wikström surveyed the situation in a dozen firms, and found surprisingly little activity beyond routine complaint handling. There were remarkably few consumer affairs professionals in these firms. We got a similar impression, amplified by a well-connected spokesman of organized industry who said that "this country does not have much, and what there is, is in large part cosmetic." We interviewed the coops and a few companies and found only one truly impressive case: ICA AB, the company serving the Independent Grocers Alliance, the largest retail organization in Sweden. Its weekly magazine with a paid circulation of about 600,000 (three times that of Råd och Ron) is thoroughly consumer-oriented in terms of articles on home economics and household management. Almost every issue carries a comparative test; there is a special issue summarizing scores of tests.¹⁴ Indeed, ICA for years has sponsored a great deal more of comparative testing than has KOV. ICA plans an accelerated effort to increase interest in consumer policy at the member store level.

Government enterprises, being semi-monopolies, are not setting noble examples in this area. The State Telephone Administration, for example, has no consumer affairs department, although we understand that KOV has repeatedly recommended that one be created. At the trade association level, the Furniture Research Institute has sponsored an informative labeling program, "Furniture Facts," which has been moderately successful. This is more due to the largest dealer chains (IKEA and the cooperative department stores) wishing to carry labels on part of their assortment than to manufacturer enthusiasm. A trade group of insurance companies has established a special Insurance Consumers' Advice Bureau as a joint venture with KOV. Speaking generally, the federations of industry, wholesalers and retailers have been more willing to accept--indeed, often endorse--public consumer policy than to launch initiatives of their own.¹⁵ It would seem that thus far Swedish business at the enterprise level has been less eager to engage in independent consumer policy than have American firms as revealed in the 1975 study by Fornell. The latter in 1978 estimated the number of consumer affairs units in Swedish firms at about 20, while in the U.S. there might well be more than 1,000 (Fornell 1978). It may well be that Swedish firms have become so conditioned by welfare state thinking and KOV activity that they simply tend to feel no independent responsibility for consumer policy. We may note, however, that both the Swedish CED and the Federation of Industries have recently given the stimulation of consumer policy at the firm level a prominent place on their agendas.

¹⁴Testat av ICA-kuriren (ICA AB, 1979).

¹⁵For sample policy statements by spokesmen of these organizations, see Henell, Olof, Foretagens Reaktionen På Konsumentpolitiken (Stockholm: Raber Sjögren, 1976).

Swedish Policy: A Futurist View

To get an idea of what the future may hold in store for Swedish consumer policy we undertook what may be called a one-round Delphi. A standardized questionnaire was used to obtain from a limited number of experts their probability estimates of 22 different consumer policy "events" taking place in the 1980-85 period. All events have been mentioned as at least plausible in Swedish debate. The 18 experts consulted were government officials, academics and business executives prominently associated with public and/or private consumer policy. No less than 14 returned filled-in questionnaires. There was no difference in occupational composition of respondents and nonrespondents. Our excellent personal contacts with respondents probably had something to do with the 78 percent response rate. We list the respondents to give the reader an idea of their background and to thank them for their cooperation:

Mr. Tore Adler, The Swedish Furniture Research Institute
Mr. Jan Gillberg, Executive Director, The Federation of Swedish Marketing Associations
Mr. Sven Heurgren, KO and Director General of KOV
Dr. Ernst Johnson, Associate Professor, University of Stockholm
Dr. Claes-Robert Julander, Research Associate, Stockholm School of Economics
Dr. Lars Nieckels, Associate Professor, University of Stockholm
Dr. Tomas Ohlin, Executive Secretary, Government Commission on New Information Technology
Mr. Åke Safbeck, Central Marketing Department, Electrolux, Inc.
Dr. Sten Tengelin, Executive Director, Swedish Industry Delegation on Marketing Law and Rapporteur on Consumer Affairs, International Chamber of Commerce, Paris
Mr. Ulf Tidics, Consumer Affairs Executive, Marketing Department, Philips of Sweden, Inc.
Ms. Margareta Wernevi, Director of Consumer Affairs, Sunlight of Sweden, Inc. (a Unilever affiliate)
Dr. Solveig Wikström, Professor, University of Lund
Mr. Nils-Erik Wirsäll, Adjunct Professor, Stockholm School of Economics and Former President, ICA AB (Independent Grocers' Alliance)
Dr. Karl-Erik Wärneryd, Professor, Stockholm School of Economics

The results of the survey are indicated by Table 1. We note the large standard deviations around the means, indicating that the numbers should be taken with a grain of salt. Interestingly, only five events were given a more than even chance of occurring in the next five years. Event number 1 (better complaint handling and internal communication about complaints) anticipates a stronger commitment on the part of individual firms to consumer affairs. (So do events 7 and 14.) Events 2-5 are all concerned with consumer information and education, lending additional support to the view that a certain shift from protection-regulation towards information-education may be the direction of change in Swedish consumer policy. Event 5 should be compared with event 15. If we include as serious possibilities all events with over 40 percent mean probability assigned by the experts, the number trebles to 15 events. Note that computerized consumer information figures in both events 12 and 17, the more ambitious form being given the greater probability. The experts were also invited to volunteer other likely events. An international informative labeling and/or quality marking system with minimum performance thresholds, expanded product liability, more dialog between public and private policymakers, and more "reality-oriented" education of marketing personnel were mentioned. It should be stressed that these experts are all extremely busy people. There may well be other events which they would assign greater probability than some of those specified in the questionnaire.

TABLE 1
 PERCENT MEAN AND MODAL PROBABILITIES OF 22 CONSUMER POLICY EVENTS IN
 SWEDEN IN 1980-85, BASED ON 14 SEPARATE EXPERT ESTIMATES

Events in order of probability assigned	Mean	Standard Deviation of Mean	Mode
1. Firms strengthen complaints handling, communicate the info to product planning, quality control, etc.	68	29	85
2. Some form of standardized, multi-product consumer information system (e.g., a modernized VDN)	64	29	80
3. Obligatory consumer education in the schools	63	25	100
4. Local government consumer affairs activity becomes mandatory	54	26	100
5. Computerized checkout counters in supermarkets give receipts with price data for each individual product	54	31	50
6. Service firms flagrantly advising consumers may lose their license to do business	49	27	50
7. Internal consumer affairs departments in 100 firms	48	29	50
8. Growing number of import restrictions	44	33	50
9. Consumer representation in KOV operations	44	30	40
10. Price control made permanent	44	33	a
11. Cars banned from the core of five largest cities	43	29	50
12. Computerized consumer information one-way (via cassette, T.V., etc.)	43	23	50
13. New measures for consumers in sparsely settled areas	42	22	60
14. Additional joint business-KOV consumer advice bureaus of the insurance consumer service bureau type	42	20	60
15. Individual cans and packaged goods to be price marked even with computerized checkout system	41	25	50
16. Regret period for all consumer goods purchases	40	25	50
17. Computerized consumer information in dialog mode	38	35	35
18. Identical goods from a certain manufacturer may only use one brand name (a few exceptions possible)	38	26	35
19. Store hours restricted by law	36	24	50
20. Milder credit restrictions for young families with children	35	35	20
21. Consumer class actions okayed	23	24	0
22. Control of marketing course literature at the universities	23	19	a

^a multi-modal

Consumer Policy Precurse and Transfer

Public policy forecasting is increasingly vital to the strategic planning of both public and private policymakers. We just illustrated one of the techniques applicable. Another is to look across national borders for a country which may be a precursor of things to come in our own. Public policymakers would also be much interested in the somewhat related question as to whether, and to what extent, the precursor policy is transferable into their own environment. Given the extreme complexity of this area, and the paucity of work in it, we have only a few tentative thoughts upon it. They are essentially derived from an ecologic view of policy and institutions (T. and T. 1977).

At least three types of precurse are relevant. The first is environmental developments in the precursor country creating a latent need or problem. Latency is hard to recognize--especially in a foreign country's environment. To do this one needs an environmental scanning system with a distant early warning attachment (Ansoff 1977). The second type of precurse involves problem or need identification in public debate, often precipitated by a signal event or, more often, by a sequence of such events. The third type involves the adoption of new policies and/or institutions to deal with the problem (need) in the precursor country. Clearly, the three types of precurse often occur sequentially. However, a country may be a precursor at one of these stages but not at another. Furthermore, sometimes new policies are adopted before the need has become acute, bypassing one or both of the earlier stages.

Without getting lost in detail, it is easy to see that there are some major measurement problems here. For instance, how much public awareness of the problem (need) must there be before we have a precurse of type two? Is one a precursor at all before somebody else recognizes it? When do we have a "new" policy or institution as compared to a mere amendment of an old one?

The precursor concept is rooted squarely in the idea of transferability. If there be no transferability, why then bother to look abroad? For one to speak of transfer the policy must presumably be imported by the follower country from the precursor country, not just adopted in some kind of spontaneous diffusion process. Typically, one would expect varying degrees of adaptation in the transfer process, due to such varying environmental conditions as economic development, political values and legal system. An excellent example of adaptation in the transfer-diffusion process is the adoption of restrictive practices legislation in a dozen West European countries after World War II on the impetus of U.S. antitrust policy (Thorelli 1959). When little adaptation is called for we may speak of robustness of transferability. Robustness tends to be great at a very general level of policy, often declining with increasing specificity. For example, we have claimed that in industrially advanced countries the relative emphasis in consumer policy should be information, education, and protection (in that order), while in the LDC it should be exactly the reverse. Proceeding at a very broad level, we believe this statement is robust. Nevertheless, presumably there is a "gray area" where the switch in policy rankings takes place. An interesting example of robustness even in a rather specific situation

in spite of strong environmental variation was given by KO. In the area of product liability Sweden is likely to adopt the U.S. idea of strict liability, but the Swedes dislike the huge and often seemingly arbitrary damages meted out in U.S. cases. KO feels that there would be little risk of "sensational" settlements in Sweden, as damages traditionally have been more modest (there are no juries in Sweden) and as the local social welfare system would automatically absorb the cost of many ailments due to product failure. An illustration of considerable adaptation, on the other hand, was the strong "Swedish accent" initially given to the guideline system largely inspired by the trade regulation rules of the FTC. Transferability is a relative thing. It is easy to exaggerate both the uniqueness and sameness of countries.¹⁶ The ecologic view cautions against both extremes.

Precurse and Transfer: Sweden and the United States

In discussing precurse and transfer of environmental trends, cultural values and shifts in public opinion, it may be quite meaningful to focus on the broad national scene in both Sweden and the U.S. However, as one gets into specifics it would be unrealistic to overlook that while Sweden is a small, homogeneous and in some respects (such as public consumer policy) fairly centralized nation; the U.S. is of continental dimensions, exceedingly heterogeneous, and in many respects a highly decentralized country. In the area of consumer policy a great number of autonomous initiatives are taken at state rather than federal level.

Precurse types one and two involve changes in conditioning variables and consequent problem (or need) identification. There is relatively great transferability here due to the similarity in environmental circumstances: we are dealing with two of the handful of "post-industrial" societies in the world. As Graham T. T. Molitor (1973) and others have observed, Sweden has typically served as a precursor here. This is due to her having embraced the welfare state concept at an early stage and being a good example of the drift towards the left in economic policy, which is possibly the most significant likely global trend for the remainder of the twentieth century. The U.S. has wished to retain a greater element of individual responsibility for personal welfare, and has not veered left at anywhere near the speed of Sweden. Thus it is but natural that Sweden has been a precursor as regards such "movements" as sexual freedom, women's liberation, environmental protection (and in this area sequentially environmental pollution, resource conservation, noise reduction) and egalitarianism in the sense of making everybody equal (as distinct from equal opportunity). As regards broad consumerist trends, she was also a precursor in the early 1970s push to promote producer jawboning as a pre-sale consumer policy approach and in establishing the Central Consumer Complaints Board as a semi-independent agency in 1968 (done within the U.S. OCA in 1980). Currently, Sweden may well have a precursor role as regards both the practice of cohabitation without marriage and public concern about cohabitants' rights.

¹⁶A precisely analogous situation obtains with regard to the PIMS strategic planning data pool, as member companies agonize over to what extent policies pursued by, say the electrical machinery industry, may be applicable to, say, the farm equipment industry.

Local government activity notwithstanding, Sweden has gone further than any other nation in centralization of consumer policy. In this respect she was a precursor type three (specific policy measures) to the attempts to create a central Consumer Protection Agency in the U.S., a proposal at least temporarily abandoned in favor of a more decentralized approach not remarkably different from that initiated by the Ford administration.

In the past, there has been an unfortunate failure to distinguish precurse and transfer. At the level of general ideation it would seem that there has been some transfer from Sweden to the U.S. in the field of social welfare policy (epitomized by Marquis Childs' Sweden--The Middle Way). The notion of KO appears to have played a role in the creation of Consumer Advocates in a number of American states and major cities. However, only few specifics of U.S. consumer policy would appear to have their origin in transfer from Sweden. A big study focused on Sweden's alleged pace-setter role vis-à-vis the United States could only point to legislation establishing maximum levels of mercury in food, banning DDT and regulating the use of pesticides. Volvo also pioneered three-point safety belts as standard in 1959, before the promotion of such seat-belts became official policy in the U.S. (Forecasting International 1976). On balance, it seems clear that most of the transfer traffic actually has been in the other direction. The influential role of the FTC trade regulation rules on the Swedish guidelines system has already been mentioned. The U.S.--or individual states--were ahead on unit pricing, truth-in-lending and cyclamates. The accident reporting system adopted by the Nordic Council is directly patterned on NEISS. The Swedish coops borrowed the idea of generic foods from Jewel Tea. If there was no direct transfer, the U.S. was at least the precursor regarding labeling of automobile gasoline and appliance energy consumption as well as cigarette health warnings. Sweden is about to introduce an automotive repair service standards and control system which in its essentials is directly patterned on the California legislation.

Some Consumer Policy Research Implications

Several ideas for consumer policy research in the 1980s derive directly from the foregoing. We hear a lot about evaluation of marketing these days, although less is being done than said. Evaluation of public and private consumer policies affecting marketing is equally desirable. A comparative study of the impact of KOV guidelines and FTC TRR's (including possible dysfunctional side effects on competition, small business, etc., as well as the primary, hopefully salutary effects) would be of great interest, for example. Indeed, so would the development of more satisfactory methodology for doing such studies. The costs and benefits of product bans and recalls also needs more research.

The effectiveness of local government consumer affairs departments is a fruitful area of research. In Sweden 170+ local governments are at least nominally active and over 100 have a major commitment. Such numbers are large enough to permit statistical analysis of the determinants of effectiveness in terms of goal attainment. An important related question is what types of consumers these agencies actually reach, and whether these consumers redistribute advice and information. (If, as we suspect, many of them are Information Seekers, they will, although their actual impact remains to be evaluated.) Our Delphi study pointed to obligatory consumer education as a likely

development in Sweden. Educators know that their impact is related to both what is being taught and how it is taught. It would be fascinating to know how Illinois (a state with obligatory consumer education) high school graduates do on a functional consumer test as compared with graduates from a state with little or no consumer education. Or how students do who have used J.C. Penney Company's seemingly excellent Consumer Education Modules as compared to students taught in a more conventional way.

Consumer policy at the enterprise and trade association levels will clearly be a major thrust of the future. Much more research is needed as to what is being done in Europe and the U.S. An excellent point of departure is Fornell (1976). Such studies should be supplemented by more research aimed at how consumers actually experience company consumer policy at the marketplace level. All too frequently, there is a yawning gap between formal company policy and its actual implementation.

Consumer policy research also needs to get involved in technology assessment. From both a consumer and public policy viewpoint we need accelerated effort to evaluate private and public experimentation with the new information technology, such as interactive consumer information programs provided by Viewdata, QUBE, etc., and computerized consumer information systems. How much price information does the consumer actually need in the supermarket shopping process? And what are the measurable tradeoffs between costs and benefits here? There has been enough experimentation in both Sweden and the U.S. now with scanner-based checkout systems to do the research needed for sound policy development (e.g., Hammarkvist 1980).

As these examples and other items in the Delphi table indicate, consumer information will be a high priority area in the future. We have pointed to several desirable avenues of research in earlier publications (T. and T. 1977; Thorelli, Becker and Engledow 1975). Of increasing interest here is the Specialty Seeker, that is, the consumer who is an Information Seeker only for certain items of truly special interest to her/him--be it fur coats, cars, cameras, dishwashers or suntan lotions. To know the cumulative impact of the general and Specialty Seekers in various consumer markets would be of great interest to both private and public policymakers.

The evaluation of the total consumer policy system of a country--comprising all private as well as public policymakers and their interactive effects on the marketplace--is a topic barely scratched (Thorelli 1979). This article has also stressed the importance of research on precure, transfer, and transferability--all specialty aspects of the broader topic which is the ecology of consumer policy. Remarkably little has been accomplished here, even in the limited context of Swedish-American interaction. We need better operational definitions of the various types of precure and transfer. We need to know much more about the determinants of the relative degrees of transferability of policy between nations and cultures. And we certainly need more nitty-gritty research on actual cases of precure and transfer. Beyond transfer is the international cooperative development of consumer policy measures. Probably the best on-going example is that of the Nordic Council, which has been actively engaged in consumer policy research and development for over 20 years. The experience of the Council is valuable also as a guide to variables constraining such joint efforts.

It would be an insult to talk about consumer policy research in the 1980s without mentioning the Third World. As the decade opens the U.N. itself and various affiliates are finally beginning to take an interest in LDC consumers

and their protection, in large part stimulated to do so by the International Organization of Consumers Unions (IOCU). Unfortunately, and incredibly, there is as yet almost no solid research of direct relevance (Thorelli and Sentell 1979 and forthcoming). It would be a great pity if the U.N. and individual LDC were to become heavily engaged in policy formulation without an adequate research base, as there is every reason to believe that blind transfer of consumer policy technology from the industrialized countries is not what the LDC really need. The international consumer research community has a strong moral obligation to build that research base--and there is no time to lose.

References

- Ansoff, Igor (1977), "Managing Surprise and Discontinuity: Strategic Response to Weak Signals," in H.B. Thorelli, ed., Strategy + Structure = Performance: The Strategic Planning Imperative, 53-82.
- Bernitz, Ulf (1980), "Services and Consumer Protection: The Proposed Swedish Consumer Services Act," Journal of Consumer Policy, 4 44-56.
- Forecasting International, Limited (1977), Forecasting the Effects of Events on Consumer Affairs, Phase I Final Report, Vol. 2, Appendices.
- Fornell, Claes (1976), Consumer Impact for Marketing Decisions--A Study of Corporate Departments of Consumer Affairs, New York: Praeger.
- Hammarkvist, Karl-Olov (1980), Detaljhandelns Datorisering (The Computerization of Retailing), Stockholm: Marketing Techniques Center.
- Molitor, Graham T.T. (1973), "A Look at Government -- Forecasting Public Policy Development," Technology Assessment, 2:4.
- Thorelli, Hans B. (1979), "Är Sverige Verldsbäst i Konsumentpolitik?" (Is Sweden Number One in Consumer Policy?), in Rydén, ed., 143-52.
- _____, (1959), "Lectures on European Antitrust Policy," University of Chicago Law Review, 25, 222-36 and University of Chicago Law School Record, 8.
- _____, Becker, Helmut and Engledow, Jack L. (1975), The Information Seekers: An International Study of Consumer Information and Advertising Image, Cambridge, MA: Ballinger.
- _____, and Engledow, Jack L. (1980), "Information Seekers and Information Systems: A Policy Perspective," Journal of Marketing, 44, 9-27.
- _____, and Sentell, Gerald D. (1979), "Consumer Ecology in the LDC: The Case of Thailand," in Proceedings of the Academy of International Business, Honolulu, Hawaii (College of Business Administration, University of Hawaii), 324-33.
- _____, and Thorelli, Sarah V. (1974), Consumer Information Handbook: Europe and North America, New York: Praeger.
- _____, and _____ (1977), Consumer Information Systems and Consumer Policy, Cambridge, MA: Ballinger.
- Wikstrom, Solveig (1979), Swedish Consumerism at the Crossroad: A National Opinion Research Survey of Public, Business and Regulator Attitudes, Lund: Department of Business Administration, University of Lund.

DEFINING THE PROBLEM: THE SCOPE OF
CONSUMER CONCERN WITH FOOD LABELING

James T. Heimbach, Food & Drug Administration

Abstract

Some of the major data-gathering approaches used by the Federal Government in its examination of consumer perceptions of food labeling are described, as well as some of the overall findings. These primary findings are discussed in terms of their implications for redesigning food labels to make them more appropriate to consumer needs and more efficient as communication tools.

Food is a uniquely important commodity. It is an absolutely necessary one: no one can avoid consuming it and very few people in this country can avoid buying at least some of their food through the manufacturing/distribution network.

At the same time, food is highly complex and includes a vast variety beyond that which we can directly grow, pick or kill. Foods are composed of a plethora of substances. Some of these we need; some we should avoid; others we should limit. Different people have different needs for these substances, and different people have different tolerances for them.

Buying food is a complex activity, because the appearance, the texture, even the taste will not usually tell us much about its important characteristics. How can we help people to know the nature of the foods they buy?

The primary medium, of course, is the food label. Food labels carry a truly incredible amount of information about the contents of the container: what it is, how much is there, how much it costs, who made it, what ingredients are in it, what nutrition it provides, when it was packed or when it should be consumed, etc.

But if this panoply of information cannot be understood, if it cannot be used by people who want to use it to make the decisions they want to make, then it is worthless.

A problem, however, is that we often do not know what information consumers really want or why they want it. There has been extensive speculation, for example, that people are becoming increasingly interested in the nutritional values of their foods as one aspect of a more general involvement with physical fitness. Possibly related to this, there appears to be a sort of "back to nature" movement, involving both an outright fear of technology and a desire for more "natural" products. Assuming these trends do exist, to what extent do they translate to need for and use of label information and of what kinds of information?

Another factor is the consumer's "right to know." Virtually all consumer advocates, and possibly many consumers, feel that a person buying a product has a right to know all the characteristics of that product, whether or not he or she has any specific use for the information or will make any decisions based upon it. To what extent does this philosophy affect labeling concerns?

What about the effects of medical successes which have drastically increased life expectancy? Has this trend increased awareness and concern about low-level long-term threats to health -- the kind of thing that does no immediate harm, but may cause problems after sixty or seventy years' exposure? Are there implications for food labeling from this phenomenon?

These factors, and others, mean that we cannot rely on the old food label system to provide in any adequate form the

information consumers want and need in ways they want and need. What is needed is to look at the food label as a holistic entity, existing and functioning in a dynamic way against the backdrop of the concerns and ambitions of our consumer society.

In 1978, the Food and Drug Administration (FDA), along with the U.S. Department of Agriculture (USDA) and the Federal Trade Commission's Bureau of Consumer Protection (FTC) embarked upon a comprehensive program to evaluate the effectiveness of current food labeling and to revise food labels as necessary. The subjects of primary interest are the listing of ingredients and the nutrition label, although the entire label - as a whole - comes under scrutiny.

This paper addresses the first phase of the program: the evaluation of the nature and scope of consumer concern with food labeling and identification of problems. Later papers will discuss the remaining phases of the research and development program.

Data have been incorporated in this paper from a number of sources. The primary source is the FDA 1978 Consumer Food Labeling Survey (Heimbach and Stokes 1979), which included personal interviews with 1374 "primary food shoppers" (the person in each household who does most of the food shopping). The design was a three-stage national probability cluster sample, and the field-work was performed by Response Analysis Corporation of Princeton, N.J., in the fall of 1978.

The second major source of data is a series of public hearings held in five cities across the country in the fall of 1978, along with associated written comments from consumer groups, food and nutrition professionals, and food companies and trade associations. Over 450 individuals testified at the FDA/USDA/FTC-sponsored hearings, and the agencies' widely-publicized request for comments attracted an additional 9,400 written comments. Several hundred comments were subjected to content analysis, from which a coding system was developed and applied to the total of 9,864 oral and written comments by a team from the three agencies (U.S. Department of Health, Education and Welfare 1979).

Two additional sources of data were also used. In the fall of 1978, the then-Commissioner of FDA appeared on QUBE, the interactive cable TV system in Columbus, Ohio. In the course of his discussion of food labeling, he asked a number of multiple-choice questions about viewers' opinions. Over three thousand viewers responded, "touching in" on their home consoles, and QUBE's computer produced instantaneous tabulations (Hecht 1979). Finally, another FDA-sponsored national consumer survey completed in February, 1980, by Market Facts, Inc., also included a number of questions dealing with food labeling (Heimbach 1980).

A few words would be in order about the data-collection techniques. There is nothing new, unusual, or controversial about using survey-research data generated from national probability samples. The QUBE data and those from the hearings and written comments, on the other hand, are known to be biased: there is a large degree of respondent self-selection in these data sources. Those who chose to watch the QUBE program on food labeling and to respond to questions, and even more those who took the time to testify at public hearings or to write letters to the Government, are different: they are more interested in food labeling than the "average" consumer, they are more

concerned and dissatisfied with labeling, and they tend to be more literate and more highly educated. What then is the value of these forms of data collection?

There are two answers to this question. The first is political: these people are the "squeaky wheels" who make the system work, and to whom the Government must be responsive. It is no news to anyone that governments are more influenced by a minority who care devoutly about an issue than by a majority who do not care at all or who care very little.

The second point is that these are often the consumers with real needs. Most of us can - and do - walk into a grocery store and pull packages off the shelf with little regard for anything but the price. Many people, however, are not so fortunate: there are diabetics and prediabetics who must control their intake of sugars, hypertensives who must limit - sometimes drastically - their sodium consumption, people with cardiovascular disease who must watch cholesterol and saturated fats, pregnant and lactating women who must increase their intake of iron, elderly people who suffer calcium deficiency, people with allergies or intolerances for almost any substance imaginable, and so on. There are also numerous people who believe that it is healthier for them to eat more "natural" foods and to avoid artificial flavors and colors and preservatives; who are we to tell them they cannot do this?

Some of these needs can be extremely rare statistically and yet still be important for national policy. There are over 200,000,000 people in this country: a need which exists for only 0.01% of the public still represents over 20,000 people. Yet what is the probability that even one person with a given such low-frequency need will appear in a survey of 1,500 people? It is easy to calculate: the answer is 14% - about one chance in seven. (Even if it is assumed that each respondent can report needs of anyone in the household, the probability is still under 50%.)

Consequently, survey research using any reasonable sample sizes simply will not help us to identify many food problems or their associated labeling needs. At the same time, there is an obvious problem with relying upon needs expressed by a self-selected sample: these writers not only possess needs, they also possess literacy and the ability to process information. Many of the things they propose would not be comprehensible to other consumers who may share the problem but not the education or intelligence.

A good example is percentages. A large proportion of those who wrote in with suggestions advocated percentage ingredient labeling. But many people cannot cope with percentages. For example, in our 1979 survey we showed respondents a square divided into four quadrants, one black and three white, and asked what percentage of the square was white. Only 46% said 75% (and only 23% of those who have not completed high school). Similarly, only about two consumers in three know that if a serving provides 25% of your needs for vitamin A it would require four servings to get 100% (Heimbach 1980).

Finally, a brief statement of the obvious before we get into actual data. "Concern" with food labeling, like most other things, is not an all-or-none phenomenon, but rather a continuum. We have found that the Likert-scale approach ("extremely concerned", "somewhat concerned", etc.) does not satisfactorily express this continuum, probably because it is too tied to social desirability. Rather, we have taken the approach of moving from general free-response questions to specific direct questions, with the assumption that relatively uncued expressions of concern or of label-information needs represent stronger feelings than do those elicited only by direct questioning.

The first question to examine is that of the extent and depth of general dissatisfaction with current food labeling.

The 1978 survey began by asking respondents:

"Aside from prices, please tell me about any particular problems, difficulties or concerns which you have with food these days."

It is assumed that mentions of dissatisfaction with labeling expressed in response to this exceedingly general question - which does not even mention food labels - reveal rather strong concern with existing labeling. Overall, about 8% of respondents had something to say about labeling, most often about ingredient listing. This climbs to 10% of those who have a food-related health problem in the household, and to about 20% of the highly-educated respondents who (as revealed in later questioning) pay most attention to labels).

The most frequently voiced complaints dealt with desired information not always being available; especially calorie content, listing of ingredients, and open-dating. There were also scattered mentions of displeasure with the complexity of label information and the use of chemical and technical terminology.

A little later in the interview the following question was asked:

"Are you satisfied with the kind or amount of information that's printed on food packages and cans, or can it be improved on in some way?"

This question, being considerably more direct than the previous one, naturally elicited responses from consumers with lesser degrees of concern or dissatisfaction with food labeling. One third (33%) of all respondents indicated less than complete satisfaction with current labeling. A follow-up question revealed that again the ingredient list (or occasionally the absence of an ingredient list) accounted for the majority of all complaints, although the desire to have information about calorie content on all products was also one of the most frequently offered single responses.

The largest source of dissatisfaction with the ingredient list is its failure to include all ingredients, i.e., the fact that it sometimes uses generic terms like "spices" or "artificial flavors." This problem was mentioned by nearly 20% of those who expressed dissatisfaction with any aspect of food labeling or about 5% of all consumers.

The next most frequent request with regard to ingredient listing is to provide quantitative information to tell how much of each ingredient - especially sugar - is in the product. This was mentioned by about 16% of those who were dissatisfied with current food labeling.

Finally, the third most frequent problem with the ingredient list - named by about 13% of those who complained about food labeling - is that the ingredient list is too complicated and uses too much technical or chemical terminology.

As an aside, it might be noted that this type of problem recurs throughout the research on labeling. Some people want more information, including information of an unavoidably technical nature, while others - or often the same people - want it all simplified. The two desires are obviously somewhat incompatible. The major goal of the label development program is to somehow provide necessary technical information in sufficient detail for potential users, and at the same time avoid overloading those who will be so repelled by so much information that they will abandon attempting to look at labels at all.

The only frequently mentioned desire other than those related to the ingredient list is to have information about calorie content available on all foods, expressed by 15% of those who were dissatisfied with current labeling.

These are not strikingly large numbers, but it must be noted that they represent only consumers with rather high levels of dissatisfaction. When those questions are asked more directly the proportion advocating changes rises substantially. Demand for quantitative ingredient listing, for example, climbs to 74% under direct questioning.

The reasons for these desires for more detailed information about ingredients are not far to seek. One of the most disturbing single findings from the research program is that over half (54%) of all consumers - and 70% of those who ever look at ingredient lists - report using the ingredients information for the specific purpose of avoiding consumption of a particular ingredient. In some cases the person only desires to limit consumption, while in others the avoidance is total; in either case, fear of adverse health consequences from consuming the ingredient is the dominant concern, although weight-loss diets, taste, nutritional concerns, etc., are also mentioned occasionally.

By a large margin the ingredient of most concern - and the most avoided - is sugar, which over 25% of all consumers report trying to avoid totally or to some degree.* This is followed by salt, preservatives, artificial colors and flavors, and fats, and there is at least one person attempting to avoid or reduce consumption of almost any substance which ever occurs in food.

At this point let us put the probability sample aside and turn to the 10,000 comments from consumers who chose to write to FDA or testify at the hearings. One mild surprise, considering the wholly unrepresentative nature of this sample, was the high correspondence which actually existed between the two approaches.

Naturally, all of the people who testified or commented are concerned to some extent about food labeling -- usually to a pretty great extent -- and consequently exhibit far higher frequencies of specific complaints, but their overall pattern of responses was surprisingly similar to that obtained in the consumer survey.

TABLE 1
PERCENT OF CONSUMERS REPORTS
VARIOUS NEEDS REGARDING LABELING

	Comments and Hearings ¹	1978 Labeling Survey ²
	%	%
Provide quantitative ingredient information	35	16
Require ingredient list on <u>all</u> products	35	8
Require calorie information on <u>all</u> products	34	15
List <u>all</u> ingredients in product	32	19
Give the amount of <u>sugar</u> in product	27	9
Give the amount of <u>salt</u> in product	25	5
Require complete nutrition information on all products	10	5
Give specific sources of fats/oils	9	2
Give the drained or fill weight of canned products	8	6
Use larger print on labels	6	7
Simplify the ingredient information	1	13

1. Source: DHEW, 1979

2. Source: Heimbach and Stokes, 1979. Percentages shown are based on respondents who are dissatisfied with current labeling.

* This, by the way, again points to the differences in levels of concern expressed in response to open vs. direct questioning: when consumers were asked directly if they were attempting to modify their consumption of any of a list of two dozen substances, 80% claimed that they are making an effort to reduce their intake of sugar.

Again, the ingredient list was the major source of problems -- it was discussed by 85% of commenters -- with the nutrition label second, being addressed by 52%. Indeed, the rank-order correlation between the top specific concerns from the two approaches was .49, with only two specific complaints accounting for most of the difference.

The first of these is the demand for ingredient listing to be extended to all products, which was the 2nd most frequent desire among commenters but the 6th among all consumers. The other is the need to simplify the ingredient list, which was virtually never mentioned by commenters but was the 4th most frequent concern of those in the consumer survey. In point of fact, this was the single complaint which was mentioned more frequently by "average consumers" in absolute as well as relative terms.

In many cases, it is not clear whether differences in desires for label changes between survey respondents and commenters are due primarily to the higher level of sophistication of the latter or to the fact that people with diet-related health problems are over-represented in this group. (Unfortunately, many letters did not explain the state of health of the writer nor his/her reason for writing; however, as noted in the report on the analyses of comments (DHEW 1979), "Those who wanted more ingredient information on food labels often cited a specific health problem as the basis for their request.") For example, commenters were both absolutely and relatively more likely to demand that the sodium or salt content of a food be stated on the package, and that the specific type of fat or oil be given (as opposed to the current "contains one or more of the following"). These data are useful primarily to those suffering from hypertension or cardiovascular diseases, but the survey did show a distinct correlation between education and income and desire for all type of additional label information.

When the data from the survey are compared with that from the analysis of comments in terms of the overall patterns, however, the impact becomes quite clear. Commenters show both absolutely and relatively more desire for information about disease-linked food characteristics: sodium and potassium content, cholesterol, saturated fat, and specific sources of fats. Survey respondents, on the other hand, exhibit greater concern with the overall information load: they want such things as calorie information and a list of ingredients to appear on more products, but they also stress that the label is already too complicated and too technical and should not be made more so.

There is much more detailed information about consumer desires available from the survey, the hearings, and the written comments, but this is not the place for discussion of all of the findings, and some of them will be addressed in following papers. Rather, let us look at the overall conclusions and their implications.

First, most consumers do not appear to be strongly motivated by a philosophy of "right to know." They are essentially pragmatic, and the information they want on food labels is that for which they have use. There is some recognition that too much information can be as bad as too little.

Secondly, they are not often motivated to use current label information or to request additional information by nutritional considerations. Indeed, the desire to obtain good nutrition from food appears to take a back seat to the fear of consuming substances which are perceived to pose a threat to one's health. Probably the key finding of this phase on the research program was that, where food labels are concerned, consumers are oriented almost wholly toward avoidance of risk rather than seeking of benefits. Use of the ingredient list is primarily directed toward avoiding possibly hazardous ingredients rather than toward obtaining desired ones: sugar, salt and preservatives dominate

FOOD LABEL INFORMATION: WHAT CONSUMERS
SAY THEY WANT AND WHAT THEY NEED

Edwin C. Hackleman, Food and Drug Administration

Abstract

This paper addresses the problem of isolating the important nutrient information desired by consumers on the food label. Since consumers have been criticized for their lack of nutrition knowledge, subject areas for additional learning are studied, and personal profiles based on learning interests are developed. Finally, comparisons between nutritional expert groups and household food buyers are made on the basis of desired nutrient information.

Introduction

Considerable attention in recent years has been devoted to the nature and types of information desired by consumers on food product labels. Actually, the concern lies within the broader realm of interest in the design of what Bettman (1975) dubs "information environments" in which consumers are being exposed to larger arrays of data describing all sorts of products. Wilkie (1975) and more recently McNeil and Wilkie (1979) and Westbrook and Fornell (1979) in their discussions related to formats for durable and energy product labels have shown how such information can affect appliance purchase decisions. One can conclude that most of the investigations surrounding the issue have addressed alternative formats for supplying consumers with information about products and brands to improve their ability to choose what to buy at the marketplace.

The literature abounds with studies concerned specifically with food product label information, including units of measurement concerns (Stokes and Haddock 1972), and the presentation of nutrition information for supermarket products (e.g., Bucklin and Asam 1973; Lenahan, *et. al.* 1973; Russo 1975, 1979; Goodwin and Etgar 1980). Bettman (1975) has proposed that a concentrated list of nutrient information, conveniently displayed in a prominent place adjacent to all brands of a product, would overcome a number of limitations to information processing imposed by separate labels of nutrient information. Such a scheme utilizing a matrix display in an effort to reduce memory demands and mental search costs was studied experimentally by Goodwin and Etgar (1980) yielding inconclusive results, apparently because of a lack of understanding of nutritional information by consumers to begin with.

Still, regardless of many consumers' lack of nutrition knowledge, there is considerable evidence that more detailed nutrition data will be referred to by some segment of shoppers and can result in purchases of more nutritious products (Stokes 1972). Research contracted for by the Food and Drug Administration has shown that nutrition-labeled products are noticed and used by enough supermarket customers to judge the labeling worthwhile (Adamsy 1972). Hence, as Barksdale and French (1974) and Day (1976) predict the future will see the label increasingly used for the dissemination of nutrition information, a strategy several large processors have already used to promote a competitive differential advantage.

Closely related to the problems of what to include on the food label is the literature concerned with information load and its effect on consumer brand choice (e.g., Russo 1974; Wilkie 1974; Summers 1974; Jacoby, *et. al.* 1974; Scammon 1977). These studies have addressed the quantity of product information display and its effect on a variety of selected response variables such as aided and unaided re-

call, "correctness" of choice, confusion of information, satisfaction with formats, and so on. Although results have been mixed, the tendency has been to suggest that too little information is worse than too much, thus paralleling the public policy of greater disclosure. Most of these studies, however, have operated under controlled experimental conditions and have not examined subject differences which would necessarily influence both the amount and type of information desired on the food label.

A major purpose of this paper, therefore, will be to further investigate food label information components desired by consumers, what types of label information to which they tend to pay most attention, and what food, health, and nutrition-related issues they desire to learn more about. An attempt will also be made here to utilize segmentation techniques to uncover relationships between desired label information, learning interests, and personal characteristics of chief food buyers and their households.

Methodology

Data Base and Background

The Food and Drug Administration in cooperation with the U.S. Department of Agriculture and the Federal Trade Commission launched a major effort in 1978 to evaluate consumer usage of and satisfaction with information on food labels. Three agencies conducted a series of five hearings across the nation, designed to solicit consumer requests for revisions to the label. Several media releases were issued covering the food hearings and stressing the need for consumer feedback to encourage comments by mail. Over 9000 letters were obtained, and 450 individuals testified at the hearings, thus indicating a high degree of interest in the subject of food labeling.

During the same year the FDA's' Consumer Studies Division, Bureau of Foods, began a comprehensive program of consumer research to examine both the quantity and content of information on the food label to allow recommendations for making the label more understandable and useful to consumers. The first phase of this program generated the data base for this study. A national survey, conducted by Response Analysis Corporation, yielding a probability sample of 1374 respondents. Personal interviews averaging just over 60 minutes each took place at home utilizing a questionnaire developed by the FDA and refined through a series of 18 developmental pretests, followed by a 50-case pilot study from two geographic locations. Upon completion of the interviews with the chief food buyers another questionnaire was dropped off for completion and mail-in at the respondents' convenience. A total of 864 questionnaires were recovered.

Late in 1979 another survey instrument was prepared to collect expert information from members of the American Institute for Nutrition (AIN), the Milk Industry Foundation (MIF), the Grocery Manufacturers of America (GMA), the American Meat Institute (AMI), the Food Marketing Institute (FMI), the American Bakers' Association (ABA), the National Food Processors Association (AFPA), and a segment of consumers on the FDA mailing list who, in the past, have exhibited heavy interest in such issues as nutrition labeling. A total of 820 questionnaires were recovered, including 536 AIN members, 177 industry representatives, and 107 from the consumer segment.

the reports of what ingredient-list information is most important. Of the twelve types of information on the nutrition label rated as "very useful" by at least 50% of survey respondents, only three - protein, vitamin C, and iron - are sought while the other nine are avoided. (This is particularly significant when it is noted that thirty-eight nutrients were rated, and these nine were the only "things to be avoided" in the group -- the other 29 are all substances seen as "good for you.") Given a choice of having information about things people often get too little of vs. information about things people often get too much of, survey respondents chose the latter by a 2-to-1 margin.*

These two points together point to the conclusion that the determination of the information to include on food labels should be based primarily upon consideration of the relationship of the information to health problems and concerns. That is, top priority must be placed upon information about substances which are known to impact upon diet-related diseases and diabetes. The next priority should be those substances to which people are most commonly or most severely allergic or intolerant, and to those - such as preservatives - which do not appear to be related to health, but which many people are concerned about regardless. Other information, while its inclusion might be laudable, must be considered dispensible in comparison.

Third, information about these substances must be presented in the correct level of detail. For example, fats must be broken into saturated vs. polyunsaturated because of the known health effects, while sugars are pretty much sugars: despite much publicity about fructose and other varieties of sugars, they are all about the same to diabetics - and to teeth.

Similarly, the amount of sugar, or sodium, or saturated fat is of concern and should be available, while that of preservatives or artificial colors is not and should not. Besides the fact that avoiders of these are not working under a set tolerance level in the way, say, hypertensives are with sodium, the amounts are tiny in any one package and thus require complicated and confusing quantitative statements.

Finally, it is evident that many consumers find current labeling incomprehensible. Only 10% have any idea of what "U.S. RDA" means (accepting any statement to the effect that it refers to the amount people should consume as correct), a rather basic necessity for understanding of much of the nutrition label; only about 75% know what "net weight" means; words like "riboflavin," "hydrogenated," and "polyunsaturated" are confusing to nearly everyone (Heimbach 1980). And this refers to the total population: the plight of the socially and educationally deprived is obviously far worse. Consumers themselves are aware of this when they demand that labeling be simplified, although they rarely realize quite how wide the gap is, probably because they have simply tuned out much of the terminology. Ignoring something is the sincerest form of rejection.

It is clear that severe constraints must be placed upon the amount of information placed upon labels, that this information must be directed toward real diet-related problems, that it must not be presented at a level of detail beyond what is really needed, and that every effort must be made to find alternative ways of expressing the information in terms that will be comprehended by at least a majority of the population.

*This discussion is not to say that consumers do not seek foods with good nutritional characteristics, only that they do not often use food labels to do it. My hypothesis is that consumers decide tentatively to purchase a food on other grounds, such as advertising, package appearance, taste, nutrition, friends' recommendations, etc., and use the label as a final hurdle; i.e., they check the label to make sure there is no good reason not to buy the product.

The next five papers will discuss in more detail the Government's efforts to fulfill these goals.

References

Health, Education and Welfare, Dept. of (1979), "Food Labeling: Report on the Analysis of Comments." Washington, D.C.: DHEW.

Hecht, A. (1979), "Consumers Speak on Cue, On 'Qube'." FDA Consumer, March.

Heimbach, J.T. (1980), "Report on the 1979 FDA Consumer Survey." (In preparation.)

_____ and R.C. Stokes (1979), "FDA 1978 Consumer Food Labeling Survey." Washington, D.C.: Department of Health, Education and Welfare.

The personal interviews and mail questionnaires addressed a number of concerns, including the following:

1. Consumer awareness and usage of current label information;
2. Problems encountered with usage of label information;
3. Experiences with food buying and usage;
4. Needs and desires for additional or revised information;
5. Knowledge and beliefs about food and nutrition;
6. Areas of interest for learning more about food-related issues;
7. Evaluations of media for learning more about food and nutrition;
8. Sources and types of untruthful information received about food;
9. Diet-related health problems in the U.S.

Nutrition Information

Just over 64% of the total sample claimed they paid attention to nutrient-label information. This group was subsequently asked to rate the information about 38 specific nutrients as being either "very useful," "of some use," or "of little or no use." A rating scale for each nutrient was derived by weighting these responses by 100, 50, and 0. In addition, the usefulness raw scores were subjected to principal components factor analysis in an effort to reveal underlying dimensions along which respondents may have perceived the nutrients. The average rating scores of the nutrients clustering in each relevant factor were then compared to personal characteristics of chief food buyers and their preferences for other aspects of the food label.

Learning Interests

Just under 63% of the total sample returned the mail questionnaire dropped off after the completion of the personal interview. One of the major sections of the questionnaire concerned a list of 17 food-related issues to which respondents indicated their degree of interest in learning more about.

Initially the items were ranked for a rough measure of overall interest and attention using a rating scale derived by weighting each response, similar to the procedures followed for nutrition information usefulness. Next, the items were subjected to principal-axis factor analysis to uncover any patterns in the data and to allow both sets of items to be rearranged to a smaller set of factors accounting for the inter-relationships. Factor scores derived from the analysis were regrouped into categories, and cross-tabulated with other personal characteristics of the respondents.

Results

Table 1 lists the simple frequency tabulations for each category of usefulness applied to the nutrients by the respondents. The nutrients are listed in the rank ordering derived from the rating scheme which closely parallels the fraction selecting the nutrient "very useful." Hence, a top-box analysis would produce virtually the same ordering of the nutrition information. The range of interest is quite wide for those items currently on the label. Calories were rated "very useful" by over 80% of the respondents, whereas riboflavin obtained 19%. Sugar and starch content, both not on the label, received substantial support and were above the median of those items not on the label. Respondents also desired, by over a 2-to-1 margin from a separate question, to have carbohydrates broken down into sugar, starch, and fiber.

Table 2 shows the varimax-rotated factor pattern derived from the first three factors extracted from the usefulness ratings given to the 38 nutrients. These three factors accounted for nearly 52% of the

TABLE 1

USEFULNESS OF NUTRIENTS BY THOSE PAYING ATTENTION TO THE NUTRITION LABEL

Nutrient Info.	Very Useful	Of Some Use	Little or No Use	Overall Rating
Calories	80.3	14.0	5.7	87
Protein	73.5	15.4	11.1	81
Vitamin C	65.3	24.9	9.7	77
Fat	68.1	18.3	13.6	78
*Sugar	64.2	21.8	14.0	75
Cholesterol	64.2	20.3	15.6	74
Iron	61.4	24.3	14.3	74
Sodium (salt)	62.1	21.7	16.1	73
Carbohydrate	60.7	22.4	16.8	72
*Starch	52.8	25.3	21.9	65
*Polyunsaturated Fat	52.1	24.7	23.2	64
Calcium	49.2	29.9	20.9	64
Saturated Fat	50.3	24.5	25.2	63
Vitamin A	48.3	28.9	22.9	63
*Vitamin D	47.6	28.8	23.7	62
Vitamin B-12	44.0	29.0	27.0	58
*Fiber	42.9	29.6	27.5	58
*Vitamin E	36.8	30.3	33.1	52
Vitamin B-6	35.5	28.1	36.6	50
*Potassium	33.7	26.7	39.5	47
*Iodine	31.1	31.0	38.1	47
Thiamine	27.3	25.1	47.5	40
*Vitamin K	25.2	20.8	54.1	36
Niacin	23.7	23.7	52.6	36
Riboflavin	19.0	21.8	59.2	30
*Phosphorus	17.9	20.4	61.8	28
*Choline/Lecithin	16.3	17.4	66.5	25
*Zinc	15.8	17.7	66.5	25
*Magnesium	14.0	17.3	68.6	23
*Copper	9.1	16.3	74.6	17
*Manganese	9.1	14.4	76.7	16
*Pantothenic Acid	8.8	10.1	81.1	14
*Folic Acid	6.9	9.4	83.8	12
*Molybdenum	6.3	7.4	86.2	10
*Chromium	5.8	9.1	85.1	10
*Biotin	5.1	10.2	84.8	10
*Selenium	5.8	7.0	87.3	9
*Inositol	4.3	10.0	85.7	9

N = 1374

* Currently not on nutrition label

TABLE 2

A THREE-FACTOR, VARIMAX ROTATED SOLUTION TO FOOD LABEL NUTRIENTS RATINGS

Nutrient	Factor I	Factor II	Factor III	Communality
Saturated Fat	.753	.074	.156	.597
Starch	.742	.137	.146	.591
*Polyunsaturated Fat	.703	.084	.187	.535
Total Fat	.685	.244	.050	.531
Sugar	.674	.155	.060	.482
Cholesterol	.665	.175	.072	.479
Total Carbohydrate	.658	.286	.096	.524
Calories	.554	.096	.090	.324
Sodium (salt)	.547	.298	.092	.398
Fiber	.533	.351	.034	.408
<hr/>				
Vitamin A	.111	.743	.157	.589
Vitamin B-12	.171	.721	.244	.609
Vitamin C	.286	.719	.005	.599
Vitamin D	.177	.719	.192	.585
Vitamin E	.153	.706	.256	.587
Iron	.208	.695	.075	.532
Thiamine	.088	.665	.324	.555
Calcium	.292	.660	.094	.530
Vitamin B-6	.112	.658	.298	.534
Phosphorus	.067	.555	.484	.547
Protein	.440	.502	-.055	.423
Vitamin K	.067	.461	.271	.290
Potassium	.355	.396	.392	.436
<hr/>				
Inositol	-.007	.056	.799	.638
Biotin	.023	.033	.798	.638
Chromium	-.009	.087	.768	.597
Selenium	.018	-.009	.755	.570
Molybdenum	.007	-.025	.689	.475
Folic Acid	.038	.007	.684	.469
Magnesium	.100	.336	.670	.572
Pantothenic Acid	.119	.125	.666	.473
Manganese	.083	.239	.661	.501
Copper	.037	.362	.645	.548
Choline/Lecithin	.112	.289	.622	.483
Riboflavin	.239	.332	.601	.529
Niacin	.222	.442	.505	.500
Iodine	.248	.431	.469	.467
Zinc	.143	.449	.469	.442
<hr/>				
Factor Labels	Macro-Nutrients	Important Micro-Nutrients	Unfamiliar Micro-Nutrients	

total variance in the system. All but three of the communalities exceed .4, indicating that nearly all the nutrients are closely related to at least one of the three principal components. Vitamin K and Calories possess the two lowest communalities, the latter perhaps because it is not really a nutrient. One would expect both of these variables to load higher on an additional factor.

Factor III appears to be a "who cares?" factor, since those items loading highly on it are many with which consumers are not familiar and tend to judge unuseful. The other two factors are more ambiguous, primarily because of the overlap between macronutrients often judged undesirable by consumers (e.g., saturated fat, sugar, calories, carbohydrates) and between micronutrients often judged desirable (e.g., iron, vitamin A, vitamin C, calcium). Two clear exceptions are fiber, a desired macronutrient, and sodium, an undesired micronutrient. Although protein is considered desirable, it is not clear from the double loading on factors I and II whether it is perceived as a macro- or a micronutrient.

Factor I contains "top of the label" nutrients: total fat, saturated and polyunsaturated fat, total carbohydrates, sugar, starch and fiber, calories and cholesterol. A conspicuous omission is protein which loads slightly higher on factor II. All the nutrients on factor I are frequently rated "bad for you," except for fiber and polyunsaturated fat, the latter of which receives mixed reviews. The presence of these two nutrients weighs against the argument of naming this factor, "undesirable nutrients." On the other hand, the presence of sodium and the absence of protein weighs somewhat against the label, "macronutrients."

Factor II envelopes all the lettered vitamins (A, B-6, B-12, C, D, E, and K), thiamin, protein, and minerals such as iron, calcium, phosphorus, and possibly potassium. An alternate name may also apply here, "desirable nutrient." The only argument negating this label is that sodium also loads somewhat highly here, despite a much higher loading on factor I.

Factor III is dominated by items many respondents have hardly heard of: inositol, biotin, selenium, molybdenum, folic acid, pantothenic acid, riboflavin, and so on. Also, several of these nutrients comprise elements or substances people are aware of but are not sure how useful they may be. Nutrients such as potassium, phosphorus, niacin, iodine and zinc would be examples of these.

Average usefulness scores were computed using the nutrients which loaded highest on a given factor. These scores, in turn, were broken down by the personal characteristics of households as shown in Figure 1. The slopes of the three lines permit comparisons between the demographically-defined segments. Highly educated food buyers tend to find all three types of nutrients more useful than the lower educated. High school graduates have equivalent interest in macronutrients and important micronutrients as those with some college or more. Those with college experience, however, tend to have a keener interest in unfamiliar micronutrients.

Perceived usefulness of macronutrients rises with age and does not decline until the food buyer reaches the senior citizen category. However, there is a steady decline in perceived usefulness of important micronutrients as age increases. This may, in part, be explained by an increase in interest in macronutrients and important micronutrients by food buyers with children. In addition, food buyers with household members taking vitamin or mineral supplements display greater perceived usefulness of all nutrients. Female food buyers find information on macronutrients and important micronutrients more useful than do male food buyers - indicating that obtaining nutritionally balanced diets for the family is regarded primarily as the woman's responsibility.

FIGURE 1

PERCEIVED USEFULNESS OF NUTRIENTS BY DEMOGRAPHIC SEGMENTS

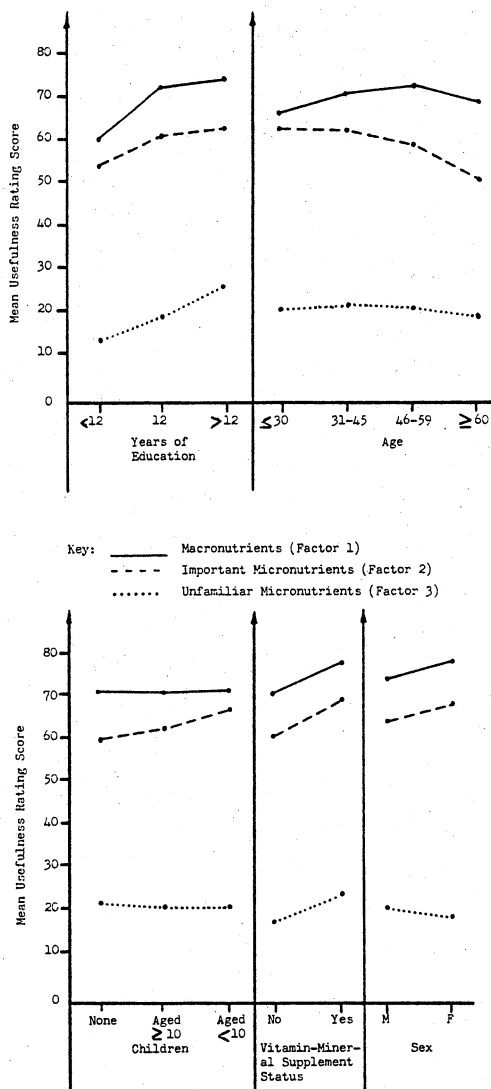


Table 3 shows the 17 consumer learning interest items, rank ordered by their overall rating derived from a five-point Likert scale. There appears to be substantial interest in general for all 17 items, as the modal response for all of them was "very interested." The heaviest interest, however, enter on relationships between food intake and health or diseases such as cancer, heart disease, blood pressure, and the "Delaney Clause" -- a law which prohibits the sale of any food product containing an additive causing cancer in animals or man.

A principal-axis factor analysis of the original data matrix uncovered a two-factor solution accounting for over 66% of the variance. Table 4 shows the varimax-rotated factor matrix and the communalities associated with the loadings on the first two factors. All but one of the communalities exceed .5, indicating that nearly all the variables are closely related to one of the two principal components. "Benefits and risks of reducing diets" may possess a higher loading on a factor not extracted here. Factor I is a food additive information factor, collecting together interest in the safety, regulations, types, rationale, and benefits/risks of food additives. Factor II is a health-disease

concern factor, highlighting a strong consumer interest in those issues connecting food with heart disease, cancer, high blood pressure, weight reduction and balanced nutrition.

TABLE 3
A RANK ORDERING OF INTEREST SCORES ON
VARIOUS FOOD-RELATED ISSUES

Learning Issue	Overall Rating
The relationship between the foods you eat and cancer	63
The relationship between salt intake and blood pressure	61
The relationship between the foods you eat and heart disease	61
The "Delaney Clause"	60
How to be sure the family diet is nutritionally balanced	59
Benefits and risks of food additives	57
How and why food additives that have been regarded as safe might be banned	56
How safe levels of food additives are determined	56
How and why being overweight contributes to health problems	55
How safety of food additives is determined	55
How to read nutrition labels	55
Why some food additives are needed in some foods	54
Where a list of unpermitted food additives can be obtained	54
Where a list of permitted food additives can be obtained	53
The truth about "health" or "organic" foods	53
How food additives are regulated	51
Benefits and risks of different reducing diets	49

It is suggested by these findings that two major groups can be isolated based on their learning interest orientations in the directions described by the two factors. To further describe them, factor scores were derived from the principal components to measure the propensity of each respondent to have an interest in these directions. The sample was split into thirds along each dimension and cross-tabulated against a host of descriptive variables to obtain a profile.

A cross-tabulation analyses yielding a number of significant differences is summarized in Table 5. Concentrating first on factor I, individuals who are more likely to be interested in obtaining information on food additives would also more likely be overweight, 46 to 59 years old, reside in a larger household, and have earned a college education. These individuals either are or live with a special dieter(s) and vitamin taker(s). They are less satisfied with the freshness of their food, tend to have greater knowledge of nutrition, are extensive users of the ingredients list, report greater incidence of stale or spoiled food, and are heavy users of open dates on perishable food labels. On the other hand, individuals who are disinterested in obtaining information on food additives would also more likely be overweight, 46 to 59 years old, reside in smaller households, and have at most a high school education.¹ These individuals are not or do not live with a special dieter(s) and vitamin taker(s). Although they claim to have little knowledge of nutrition and do not pay much attention to the list of ingredients, they are more satisfied with the freshness of their food, report fewer stale or spoiled food experiences, despite little reported usage of dates on food labels.

¹The implication here is that overweight, middle-aged respondents were sharply divided into their interests, one group having much interest and the other having little or none.

TABLE 4

A TWO-FACTOR, VARIMAX ROTATED SOLUTION TO LEARNING INTERESTS IN FOOD-RELATED ISSUES

Interest Variables	Factor I	Factor II	Communality
How safety of food additives is determined	.827	.268	.756
Where a list of permitted food additives can be obtained	.816	.292	.751
Where a list of unpermitted food additives can be obtained	.800	.287	.722
Why some food additives are needed in some foods	.716	.366	.647
How food additives are regulated	.801	.251	.705
Benefits and risks of food additives	.747	.417	.733
How safe levels of food additives are determined	.827	.298	.772
How to read nutrition labels	.631	.412	.568
How and why food additives that have been regarded safe might be banned	.666	.447	.644
The relationship between the foods you eat and heart disease	.437	.719	.708
The relationship between the foods you eat and cancer	.363	.723	.655
The relationship between salt intake and high blood pressure	.280	.809	.734
How and why being overweight contributes to health problems	.152	.819	.694
Benefits and risks of different reducing diets	.320	.592	.453
How to be sure the family diet is nutritionally balanced	.465	.619	.600
The "Delaney Clause"	.585	.456	.550
The truth about "health" foods	.553	.473	.530
Factor Labels	Food Additive Information	Health/Disease Concerns	

TABLE 5

RELATIONSHIPS BETWEEN LEARNING INTEREST FACTOR GROUPS AND THEIR PERSONAL CHARACTERISTICS

Descriptive Variable	Health/Disease Concerns - Factor II			Additive Information Interest - Factor I		
	Lo	Med	Hi	Lo	Med	Hi
Over-Underweight (Over → Under)	→			→		
Age (Younger → Older)	→			→		
Occupation (Blue Collar → White Collar)	→			→		
Household Size (One → 5 or more)	→			→		
Annual Income (Under \$5K → Over \$25K)	→			→		
Education (Under 8 yrs → College)	→			→		
Special Dieters in Household (None → One or more)	→			→		
Vitamin Takers in Household (None → One or more)	→			→		
Food Freshness Satisfaction (More → Less)	→			→		
Knowledge of Nutrition (Little → Extensive)	→			→		
Ingredients List Attention (Little → Extensive)	→			→		
Stale-Spoiled Food Experiences (None → Several)	→			→		
Usage of Dates on Labels (None → Frequently)	→			→		

Note: Relationships indicated by presence of arrow in table produced χ^2 significant at $P < .01$.

Interest in factor II, concerns for information related to health or disease, generates a somewhat different profile. A higher interest in these issues is shared by those who tend to be about the right weight or slightly under, white-collar workers, those living in smaller households, earning relatively high annual incomes (\$15K to over \$25K) and in possession of a college education. In addition, individuals in this group tend to be less satisfied with their food's freshness, and although they tend to use the ingredients list and dates on food labels, they claim to know little about nutrition. In contrast, consumers reporting a disinterest in health-disease information are more likely to be overweight, blue-collar workers, from larger households, earning relative low annual incomes (\$5K to \$15K) and possessing at most a high school education. These people are more satisfied with their food's freshness, although they tend not to use dates on food labels, pay little attention to the list of ingredients, and claim to have little knowledge of nutrition.

The overall ratings of the nutrient information supplied by the chief food buyers (see Table 1) were used to rank the items from 1 to 38 as shown in Table 6. Similar rating scores were obtained from the mail questionnaires which tapped industry experts, AIN members, and consumers heavily involved with the subject. Ranks from each of these groups are also shown in Table 6. Priority agreement between the judgmental groups seems high (Kendall's W = .95) overall, and pairwise agreement between the chief food buyers and each expert segment is also high. The lowest agreement (Spearman's R = .89) is obtained in a comparison with AIN members. Despite the apparent high overall agreement certain nutrients stand out from others as being judged quite differently. Vitamin C, sugar, and starch are given much higher priority by chief food buyers than by the expert groups. On the other hand, the experts tend to rate sodium (salt), calcium, iron, potassium, and riboflavin substantially more useful than household food buyers judge these nutrients.

TABLE 6

A COMPARISON OF RANKS GIVEN TO ALL NUTRIENT INFORMATION ITEMS BY CONSUMERS AND EXPERT GROUPS

Nutrient Info.	Chief Food Buyer (N=1374) Ranks	AIN Ranks	Industry Groups Ranks	Consumer Interest Group Ranks
Calories	1	1	1	1
Protein	2	4	2	4
Vitamin C	3	15	9	12
Fat	4	3	3	3
Sugar	5	14	13	11
Cholesterol	6	9	8	6
Iron	7	5	6	9
Sodium (salt)	8	2	4	2
Carbohydrates	9	7	5	5
Starch	10	20	22	20
Polyunsaturated Fat	11	8	17	8
Calcium	12	6	7	10
Saturated Fat	13	13	10	7
Vitamin A	14	12	12	15
Vitamin D	15	16	18	16
Vitamin B-12	16	22	20	17
Fiber	17	11	19	14
Vitamin E	18	27	23	23
Vitamin B-6	19	21	21	18
Potassium	20	10	11	13
Iodine	21	23	24	24
Thiamine	22	17	15	21
Vitamin K	23	30	26	25
Niacin	24	19	16	22
Riboflavin	25	18	14	19
Phosphorus	26	24	25	28
Choline/Lecithin	27	36	32	33
Zinc	28	26	30	27
Magnesium	29	28	27	26
Copper	30	31	31	32
Manganese	31	32	33	30
Pantothenic Acid	32	29	28	29
Folic Acid	33	25	29	31
Molybdenum	34	37	36	37
Chromium	35	35	38	36
Biotin	36	33	34	34
Selenium	37	34	35	35
Inositol	38	38	37	38

Spearman's R: Chief Food Buyer vs AIN: .89; Chief Food Buyer vs Industry Groups: .91; Chief Food Buyer vs Consumer Interest: .94; Kendall's W = .95

Tables 7, 8 and 9 contain the rank orderings of the usefulness rating scores bestowed by each group upon the macronutrients, important micronutrients and unfamiliar micronutrients derived from the factor analysis. Priority agreement within these three dimensions is again relatively high, but there appears to be substantially less agreement between the groups in the case of macronutrients (Kendall's W = .83) as shown in Table 7. In particular, AIN members show the greatest disagreement with chief food buyers (Spearman's R = .50). The rank for sugar shows the greatest deviation between chief food buyers and the experts, as households tend to give this nutrient higher priority. Turning to important micronutrients shown in Table 8, there is a tendency for minerals (iron, calcium, and potassium) to be ranked higher in priority by the experts than by chief food buyers, whereas the latter tend to rank vitamins (C, E, B-12) higher than the experts. One would have expected the households to agree the least with the experts on the unfamiliar micronutrients (Table 9), but agreement between the three expert groups and households was surprisingly high. The only nutrient which seems out of place in this list is choline/lecithin, which received a higher ranking by chief food buyers.

TABLE 7

A COMPARISON OF RANKS GIVEN TO MACRONUTRIENTS, BY CONSUMERS AND EXPERT GROUPS

Macronutrient	Chief Food Buyer (N=1374) Ranks	AIN Ranks	Industry Group Ranks	Consumer Interest Group Ranks
Calories	1	1	1	1
Fat	2	3	2	3
Sugar	3	9	7	8
Cholesterol	4	6	5	5
Sodium (salt)	5	2	3	2
Carbohydrates	6	4	4	4
Starch	7	10	10	10
Polyunsaturated Fat	8	5	8	7
Saturated Fat	9	8	6	6
Fiber	10	7	9	9

Spearman's R: Chief Food Buyer vs AIN: .50; Chief Food Buyer vs Industry Groups: .73; Chief Food Buyer vs Consumer Interest: .64; Kendall's W = .83

TABLE 8

A COMPARISON OF RANKS GIVEN TO IMPORTANT MICRONUTRIENTS BY CONSUMERS AND EXPERT GROUPS

Important Micronutrient	Chief Food Buyer (N=1374) Ranks	AIN Ranks	Industry Group Ranks	Consumer Interest Group Ranks
Protein	1	1	1	1
Vitamin C	2	6	4	4
Iron	3	2	2	2
Calcium	4	3	3	3
Vitamin A	5	5	6	6
Vitamin D	6	7	8	7
Vitamin B-12	7	10	9	8
Vitamin E	8	12	11	11
Vitamin B-6	9	9	10	9
Potassium	10	4	5	5
Thiamin	11	8	7	10
Vitamin K	12	13	13	12
Phosphorus	13	11	12	13

Spearman's R: Chief Food Buyer vs AIN: .74; Chief Food Buyer vs Industry Groups: .81; Chief Food Buyer vs Consumer Interest: .88; Kendall's W = .91

TABLE 9

A COMPARISON OF RANKS GIVEN TO UNFAMILIAR
MICRONUTRIENTS BY CONSUMERS AND EXPERT GROUPS

Unfamiliar Micronutrient	Chief Food Buyer (N=1374) Ranks	AIN Ranks	Industry Group Ranks	Consumer Interest Group Ranks
Iodine	1	3	3	3
Niacin	2	2	2	2
Riboflavin	3	1	1	1
Choline/Lecithin	4	13	9	10
Zinc	5	5	7	5
Magnesium	6	6	4	4
Copper	7	8	8	9
Manganese	8	9	10	7
Pantothenic Acid	9	7	5	6
Folacin	10	4	6	8
Molybdenum	11	14	13	14
Chromium	12	12	15	13
Biotin	13	10	11	11
Selenium	14	11	12	12
Inositol	15	15	14	15

Spearman's R

Kendall's W=.90

Conclusions

It would appear that the nutrition information on the food label can be revised to include a list of nutrients the consumer would find more useful in evaluating the contents of the package. There are several nutrients not included in most nutrition information lists which consumers would like to see (e.g., sugar, starch and fiber content, separate from total carbohydrates, and vitamin D). This may be one of the reasons why nearly 26% of the sample desired to see more label information, whereas slightly over 2% desired to see less. On the other hand, some of the "standard" information appearing on nutrition labels is judged to be of little or no use by the majority of consumers (e.g., riboflavin, niacin, thiamine). In addition, many producers include a number of micronutrients on the label which the consumer cares very little about (e.g., pantothenic acid, zinc, magnesium, phosphorus, chromium, copper, biotin, selenium, inositol, etc.). These findings may help explain why nearly 27% of the chief food buyers found the label information to be confusing and why over 34% expressed dissatisfaction with the current content of the food label information. What is desired, therefore, is a finer tuned, more relevant list of nutrients and not merely a more exhaustive list contributing mass to the information load. A typical criticism has been that consumers as a whole know very little about food and nutrition and are, therefore, incapable of properly evaluating nutritional information on food labels. Since over 73% of the respondents indicated they were interested in learning more about nutrition and over 77% were interested in learning more about food additives, their lack of knowledge appears overshadowed by the desire to learn more. The learning interests focus heavily on the effect of food intake on health and disease problems--the things which generate news headlines. However, a surprisingly strong support in learning more about food additives was shown here, although consumption side affects may be noticeable only in the very long run, and many are completely safe for use. Results indicate a powerful similarity between what nutrients consumers find most useful (and useless) on the label and what the experts judge they need to see (or not see). If the chief food buyers in households are to be criticized for their lack of nutritional knowledge, one cannot in the same breath criticize their ability to decipher which nutrients they should examine most carefully when

considering a purchase.

References

- Adamy, C.F. (1972), "The Retailer's Experience," Food and Cosmetic Law Journal, 27, 263-270.
- Bettman, James R. (1975), "Issues in Designing Consumer Information Environments," Journal of Consumer Research, 2, 167-77.
- Bucklin, L. and Asam, E. (1973), "Nutrition Labeling for Canned Goods: A Study of Consumer Response," Journal of Marketing, 37, 32-37.
- Day, G.S. (1976), "Assessing the Effects of Information Disclosure Requirements," Journal of Marketing, 40, 42-52.
- French, W.A. and Barksdale, H.C. (1974), "Food Labeling Regulations: Efforts Toward Full Disclosure," Journal of Marketing, 38, 14-20.
- Goodwin, S. and Etgar, M. (1980), "Alternative Organizational Formats for Nutritional Information: Processing and Policy Normative Perspectives," Marketing in the 80's: Changes and Challenges, 46, ed. R.P. Bagozzi, et. al., Chicago: American Marketing Association, 412-15.
- Heimbach, J.T. and Stokes, R.C. (1979), "FDA 1978 Consumer Food Labeling Survey," Washington, D.C.: Food and Drug Administration, Bureau of Foods.
- Jacoby, J., Speller, D.E., and Kohn, Carol A. (1974), "Brand Choice Behavior as a Function of Information Load," Journal of Marketing Research 11, 63-9.
- _____, "Brand Choice as a Function of Information Load: Replication and Extension," The Journal of Consumer Research, 1, 33-42.
- Lenahan, R.J., et. al., (1973), "Consumer Reaction to Nutritional Labels on Food Products," Journal of Consumer Affairs, 7, 1-14.
- McNeill, D.L., and Wilkie, W.L. (1979), "Public Policy and Consumer Information: Impact of New Energy Labels," Journal of Consumer Research, 6, 1-11.
- Russo, J.E. (1974), "More Information is Better: A Reevaluation of Jacoby, Speller, and Kohn," Journal of Consumer Research, 1, 68-72.
- Scammon, D. (1977), "Information Load and Consumers," Journal of Consumer Research, 4, 148-55.
- Stokes, R.C. (1972), "The Consumer Research Institute's Nutrient Labeling Research Program," Food and Drug Cosmetic Law Journal, 27, 249-262.
- _____, and Haddock, R. (1972), Interim Report for the First Two Phases of the CRI/FDA Nutritional Labeling Research. Washington, D.C.: Consumer Research Institute.
- Summers, J.O. (1974), "Less Information is Better?" Journal of Marketing Research, 11, 467-68.
- Westbrook, R.A. and Fornell, C. (1979), "Patterns of Information Source Usage Among Durable Goods Buyers," Journal of Marketing Research, 16, 303-12.
- Wilkie, W.L. (1973), "Analysis of Effects of Information Load," Journal of Marketing Research, 11, 462-6.
- _____, (1975), How Consumers Use Product Information: Assessment of Research in Relation to Public Policy Needs. Washington, D.C.: U.S. Government Printing Office.

FOOD LABEL INFORMATION: WHAT CONSUMERS
SAY THEY USE AND WHAT THEY ACTUALLY USE

Robert J. Vandenberg, Food and Drug Administration and The University of Georgia

Abstract

The purpose of this study was to assess the following question: Do consumers' self-reports of food label use correspond with what they actually use? Evidence provided here suggests they do. However, the extent of usage of specific information components of the food label (e.g., the ingredients list, the nutrition label, etc.) is dependent on individual differences.

Introduction

In 1978, a nationally representative sample of 1374 respondents was interviewed in a survey sponsored by the Division of Consumer Studies within the Food and Drug Administration. Although several other food related topics of consumer interest were addressed, the primary purpose of the survey was to assess consumer interest in, use of, and concerns with various aspects of food labels, including the ingredient list, nutrition information, package claims, open dating and other related components.

The results revealed that 76% of the respondents claim to pay some attention to the ingredient list of food products (Heimbach 1980). Additionally, over 70% (54% of all interviewed) of these consumers reported using ingredient list to avoid purchasing certain food products. In contrast to those reporting no attention paid to the ingredient list, attendors were more likely to be in good health, be very well nourished, be white, be between the ages of 31 to 59 years, be college educated and have above average incomes. The majority (94%) of the respondents also reported using open date information to tell them about the age or freshness of the product (Hackleman and Heimbach 1980). Furthermore, 80% of this group (74% of all respondents) stated they use date information in making purchase decisions. The users were again more likely to be in good health, be white, be between the ages of 31 to 45 years, be college educated, and possess above average incomes. Finally, 64% of the respondents stated they paid attention to the information on the nutrition label (Heimbach and Stokes 1979). Again, these consumers were characteristically more educated, between the ages of 31 to 59 years, white, and had higher incomes (Baumgardner, Heimbach and Stokes 1980).

Although these results are revealing, they do not address some fundamental questions concerning consumer food label use behavior. Basically, there was no means provided through which consumers' self-reports of food label use and actual usage of food labels could be independently assessed. In other words, is there a difference in what consumers say they use and what they actually use? One possible answer could be that respondents were answering the questions in a socially desirable manner (Cozby 1977). However, consumers who report attending to and using the food label are generally more educated and possess higher incomes. Indeed, of consumer demographics, education and income* account for the majority of the explained variance in reported use and satisfaction with food label information (Baumgardner, et. al. 1980). Aside from education, another factor could be the consumer's degree of concern with food label information. Some respondents may simply be more concerned with the quality of the food products they purchase and, as a result, they will be more likely to use the components of the food label.

*Education was the best single predictor with income being second best.

An extensive research project was designed, therefore, to further explore some of these issues. The basic design (to be detailed in the following section) was a two (high versus low education) by two (high versus low concern) factorial. A subsample (100) of the original nationwide sample of participants was accompanied on a shopping trip by an interviewer. The major purpose of the project was to pinpoint exactly what types of products were purchased, why the consumer made the purchase, and most importantly what information was used by them to make (or not make) a particular purchase.

Methodology

Sample

Respondents for this project were selected on the basis of three criteria. The first criterion was whether or not the respondent agreed during the original survey to participate in one or two follow-up interviews. Out of the original 1374 interviewees, 557 revealed a willingness to participate again. The second criterion was the respondent's education level (one of the independent variables). Respondents were equally divided into those relatively well educated and those less educated. Operationally, the well educated group possessed one or more years of college, and the less well educated had never attended college. The purpose of this classification was to include both respondents who are perhaps more articulate and more readily able to utilize package information and those who may have more difficulty using the information.

The final criterion (and second independent variable) was the consumer's concern regarding food label information. This was operationally defined on the basis of their responses to 11 questions from the original survey (Table 1).

TABLE 1

QUESTIONS USED TO SELECT THE HIGH
AND LOW CONCERN GROUPS

- A. Aside from prices, please tell me about any particular problems, difficulties or concerns you have with food these days?
- B. Will you please list or identify for me all of the different kinds of information you can recall that can be found on food packages and cans?
- C. Are you satisfied with the kind or amount of information that's printed on food packages and cans, or can it be improved on in some way?
- D. In what way would you like to see the information on food packages and cans improved?
- E. Which information, if any, printed on food packages and cans do you pay particular attention to or find helpful in any way? Please tell me the information you pay most attention to.
- F. Do you find any of the information on food packages and cans confusing or difficult to understand, or is it easy to understand?

- G. Have you ever noticed an ingredients list on food packages or cans?
- H. Do you ever use the ingredients information to avoid using a particular ingredient?
- I. Have you noticed any nutrition label information on food packages or cans?
- J. Do you ever pay any attention to the information shown on nutrition labels?
- K. Do you find any part of the nutrition labeling information confusing or difficult to understand, or do you feel that this information is easy to understand?

The respondent's answer to a particular question was scored, and all scores were summed to provide an overall Concern Score. Using this score, participants were then assigned to either a high concern category or low concern category. It was originally intended to acquire 100 participants with a distribution as shown in Table 2. However, due to faulty interviewing, equipment failure, etc. the final sample included 77 respondents distributed as shown on Table 2.

TABLE 2

DISTRIBUTION OF DESIRED AND FINAL
SUBSAMPLE OF PARTICIPANTS

	Well Educated	Less Well Educated	Total
High Concern	30 (30)	30 (15)	60 (45)
Low Concern	30 (13)	20 (19)	40 (32)
Total	50 (43)	50 (34)	100 (77)

Interviews

The overall interview procedure consisted of three components. The first component was an at-home interview during which the participant was asked a number of questions concerning meal planning, health problems, uses and concerns with food labels, planning for the shopping trip, and avoidance or preferences of certain food types. This portion of the study was designed to provide a profile of the consumer's shopping behavior, meal planning, etc. before leaving for the supermarket.

The second component of primary interest involved the interviewer actually accompanying the respondent on a regular shopping trip.

The interview was completely open-ended and the interviewer both tape recorded the ensuing interactions and took notes on relevant behaviors. Shoppers were asked to talk their way through the food purchases as they occurred in the store and explain in detail how they shop and use information. Shoppers were asked to describe and explain any and all sources of information used to make purchase decisions. This included information the shopper has learned from past experience, recent information from sources outside the store, (e.g., newspaper ads, information from shelf, tables or other locations within the store or information on the food label itself). The observer would prompt the shopper from time to time to clarify any confusing information.

The third component of the interview was again conducted at the respondent's home. First, the respondents were probed to determine whether they had modified their shopping be-

havior in any way, given the nature of the previous task (being accompanied on a shopping trip). Second, the respondent was asked (without seeing the product) to recall any label information on some of the products just purchased. The respondent was also asked to state any problems he/she might have had with the food label, anything that was of particular use and the interview terminated with a series of questions concerning food and meal planning. The taped interviews were transcribed into small booklets for ease of data coding so that type of product, information used for the product, whether or not it was purchased, etc. could be represented numerically.

Results

It should be noted at the outset that analysis of the data obtained in this study has barely begun, and what follows is a progress report including some preliminary summary data rather than a completed research report. Much more extensive and detailed analysis remains to be done with the data base.

Descriptive data

Table 3 depicts a simple frequency count of the top 19 products purchased by respondents. The 77 participants purchased a total of 1432 products with a mean of 18.6 products per shopper.

TABLE 3

THE TOP 19 PRODUCTS PURCHASED
ACROSS ALL RESPONDENTS

Product	Absolute Frequency	Relative Frequency (%) of Total Purchases
Fresh Vegetables	101	7.1
Canned Vegetables	86	6.0
Milk	56	3.9
Bread	51	3.6
Snacks (Pretzels, Chips, etc.)	51	3.6
Canned Goods (In General)	50	3.5
Beef/Hamburger	49	3.4
Fresh Fruit	48	3.4
Cheese/Cottage Cheese	43	3.0
Salt/Sugar	42	2.9
Margarine	41	2.9
Eggs	39	2.7
Flour/Meal/Rice	38	2.7
Fish	34	2.4
Unsweetened Break- fast Cereals	32	2.2
Fruit Juice	32	2.2
Oils/Shortenings	31	2.2
Spaghetti/Macaroni	31	2.2
Soup	31	2.2
	886	
	OR	
	62% of all purchased products (1432)	

As seen on Table 4, consumers used a total of 1132 pieces of information to purchase the products, yielding a mean of 14.7 pieces of information per shopper or about 0.8 pieces of information per product. This total does not include two types of information which were assumed to be virtually 100% used and therefore not coded: product identification (milk, cereal, etc.) and brand name.

The top category of information used (35%) was "Other", in-

cluding such classes as label claims, grades, product appearance, product description, etc. Of more importance though, was the weight placed by consumers on price in comparison to the ingredient lists and nutrition labels. Thirty-four percent of all the information was price relevant, with ingredients and nutrition information accounting for 12% and 4% respectively. Furthermore, open date information accounted for only 4% of the total information used by consumers. Finally, 11% of the information used by consumers was size and quantity of the product. This latter fact was probably related to price since consumers generally compare the size and quantity of products to get a better bargain. Again, of course, this is a preliminary analysis: not all information is available on all food products. Open dates, for example, are commonly not available on packaged foods.

TABLE 4
INFORMATION CLASSES

Type of Information	Absolute Frequency	Relative Frequency
Other Information (e.g., product appearance, description, etc.)	397	35%
Price Information	377	34%
Ingredients Information	140	12%
Size/Quantity Information	127	11%
Nutrition Information	46	4%
Open Date Information	45	4%
	1132	100%

Mean Pieces of Information per shopper = 14.7
Mean Pieces of Information per product = .79

Inferential Data

Analyses of variance were conducted on four components of the food label: (a) the ingredient list; (b) nutrition label; (c) the open date; and (d) the price. These components acted as the dependent measures in the analyses. Scores were computed such that one point was assigned per unit of information. For example, a shopper using the ingredient list 5 times for 5 different products during the trip was given a 5 for that particular dependent measure. The same process was used for the remaining three variables. The independent variables for the analyses were the two factors, education (high and low) and concern (high and low).

As seen in Table 5, only the results obtained from ingredient list use were significant to a large degree. The main effect for education was marginally significant [$F(1,73) = 3.26, p < .07$]. Consumers with a higher level of education used more ingredients information ($\bar{x} = 2.33$) than those shoppers with a lower education level ($\bar{x} = .96$). Greater significance was obtained when the concern level of the participants was analyzed [$F(1,73) = 7.33, p < .01$]. Consumers who reported a greater concern for more information used significantly more ingredients information ($\bar{x} = 2.67$) than those displaying less concern ($\bar{x} = .62$). The effects of education and concern level on ingredient list use was enhanced even more when the two factors interacted [$F(1,73) = 19.93, p < .001$]. This results because education had no real effect on ingredient list use when consumers were not concerned with it. However, participants with a high level of education and a

high concern for more information increased their use of the ingredients list substantially. Furthermore, even the less educated used the ingredient list more when they reported greater concern.

TABLE 5
RESULTS OF THE FOUR TWO-WAY
ANALYSES OF VARIANCE

Ingredients List Use					
Source	Sum of Squares	d.f.	Mean Square	F	P
Education	32.35	1	32.35	3.26	$p < .07$
Concern	72.75	1	72.75	7.33	$p < .01$
Education					
x Concern	197.90	1	197.90	19.93	$p < .001$
Error	724.86	73	9.93	---	---
\bar{x} High Education = 2.33			\bar{x} Low Education = .96		
\bar{x} High Concern = 2.67			\bar{x} Low Concern = .62		

Nutrition List Information

Source	Sum of Squares	d.f.	Mean Square	F	P
Education	.70	1	.70	.27	N.S.
Concern	8.08	1	8.08	3.11	$p < .10$
Education					
x Concern	0	1	0	0	N.S.
Error	189.67	73	2.60	---	---
\bar{x} High Education = .67			\bar{x} Low Education = .48		
\bar{x} High Concern = .92			\bar{x} Low Concern = .24		

Open Date

Source	Sum of Squares	d.f.	Mean Square	F	P
Education	3.83	1	3.83	2.92	$p < .10$
Concern	1.04	1	1.04	1.04	N.S.
Education					
x Concern	0	1	0	0	N.S.
Error	95.89	73	1.31	---	---
\bar{x} High Education = .83			\bar{x} Low Education = .37		
\bar{x} High Concern = .72			\bar{x} Low Concern = .48		

Price Information

Source	Sum of Squares	d.f.	Mean Square	F	P
Education	1.91	1	1.91	-.06	N.S.
Concern	9.34	1	9.34	.31	N.S.
Education					
x Concern	.04	1	.04	---	N.S.
Error	2232.59	73	30.58	---	---
\bar{x} High Education = 5.64			\bar{x} Low Education = 5.31		
\bar{x} High Concern = 7.00			\bar{x} Low Concern = 3.95		

Although ingredient list use was affected by consumer education and concern, nutrition list, open dating and price were not particularly influenced by these characteristics (See Table 5). Concern level had a marginal effect

on nutrition label use [$F(1,73) = 3.11, p < .10$]. Again, those who displayed a higher concern were more likely to use the nutrition list ($\bar{x} = .92$) than those with less concern ($\bar{x} = .24$). On the other hand, education had a marginal effect on open date use [$F(1,73) = 2.92, p < .10$]. More educated people were likely to use open dates ($\bar{x} = .83$) than those less educated ($\bar{x} = .37$). Education and concern levels had no effects upon participants' use of price information.

Discussion

These results tend to parallel those obtained in the past and suggest that self-reports of overall food label use correspond to a large degree with actual usage of the information. However, use and understanding of certain components (e.g., the ingredients list), were mediated by individual differences.

First, in a 1978 FDA Food Labeling Survey consumers were asked to rate on an 11-point scale how much they attended to various attributes of food products (Hackleman and Vandenberg 1980). These attributes included such aspects as price, nutrition label, product appearance, etc. Consumers ranked price first, product appearance and other similar attributes next highest, ingredient list next, and the nutrition label the least. This rank order corresponds almost perfectly with actual usage of these components on Table 4. In general, price and product appearance, etc. (i.e., the "Other" category) were used the most, the ingredient list next, and the nutrition label the least. Unfortunately, consumers were not asked on the initial survey to rate their attention to open date information. However, in pure descriptive terms, consumers' self-reports of food label use and their actual use does correspond.

The use of some components (especially the ingredient list) of the food label appear to be mediated by the education and concern level of the respondents. Education might have influenced whether or not consumers understand what the various ingredients are. However, when the consumers were not concerned with the food products, education had little effect. The consumer's concern could stem from several areas, such as the carcinogenic properties of various ingredients. These consumers might have had health problems or lived with those that do. As a result, they attend to the ingredients.

Why education and concern did not have an effect on nutrition label use and open date use is difficult to explain. Part of the reason is no doubt the fact that these label components do not appear on most products; consequently, in this study only 23 (30%) people used nutrition information and 27 (35%) open dates. And most of these people only used it once during the shopping trip. It is quite possible that further analysis will clarify the determinants of usage of these label components.

A rather simple answer is suggested for the nonsignificant results obtained when price was the dependent measure. Consumers, no matter their education, ethnic background, etc. are always seeking a good price.

In summary, consumers do appear to actually use the information on the food label they report to use. However, how consumers use various components of the overall label is mediated by certain individual differences. There is substantial evidence provided here, but future research needs to be conducted to uncover the nature of these differences and their effects on information use.

References

Baumgardner, M.H., Heimbach, J.T., and Stokes, R.C. (1980) FDA 1978 Consumer Food Labeling Survey: Demographic

Analysis of Survey Responses. Department of Health and Human Services, Food and Drug Administration, Bureau of Foods, Division of Consumer Studies: Washington, D.C.

Cozby, P.C. (1977), Methods in Behavioral Research. Mayfield Publishing Co.: Calif., 1977.

Heimbach, J.T. (1979), Preliminary Report on the FDA 1978 Survey of Consumer Food Label Awareness, Understanding, and Usage. Department of Health and Human Services, Food and Drug Administration, Bureau of Foods, Division of Consumer Studies: Washington, D.C.

Hackleman, E.C., and Heimbach, J.T. (1980), Food Freshness and Open Dating. Department of Health and Human Services, Food and Drug Administration, Bureau of Foods, Division of Consumer Studies: Washington, D.C.

Hackleman, E.C., and Vandenberg, R.J. (1980), The Relative Value of Different Components of Food Label Information. Department of Health and Human Services, Food and Drug Administration, Bureau of Foods, Division of Consumer Studies: Washington, D.C.

Heimbach, J.T., and Stokes, R.C. (1979). FDA 1978 Consumer Food Labeling Survey. Department of Health and Human Services, Food and Drug Administration, Bureau of Foods, Division of Consumer Studies: Washington, D.C.

COMMERCIAL MARKET TRACKING SYSTEMS: APPLICATIONS TO
POLICY FORMATION REGARDING NUTRITION LABELING

Michael L. Stewart, Food and Drug Administration

Abstract

The first part of this paper describes the construction of a data base integrating retail market share and food package information for a sample of packaged processed foods. The data base was then employed to estimate the extent of nutrition labeling in the retail food market as well as the amount of nutrition labeling carried out voluntarily and the degree to which food manufacturers were violating the food labeling regulation by not including a nutrition label where required.

Background

Through the early and mid-1960's, emphasis in food marketing and purchasing was primarily limited to product taste, appearance, price, and convenience. Nutrition did not constitute a dominant (or even important) theme beyond the basic rubric of attempting to select at least one item from each of several basic food groups to obtain "a well-rounded meal". Beginning in the latter part of the 1960's, however, nutrition began to acquire a more dominant position for both policy-makers and consumers. In 1967, the Senate Subcommittee on Employment, Manpower and Poverty uncovered widespread malnutrition and hunger in its travels through the Delta area of Mississippi. In 1967, the Citizen's Board of Inquiry into Hunger and Malnutrition held that as of that time ten million people across the nation were suffering from hunger and malnutrition. These conclusions were further supported by the findings of the federal government's 1968 Ten-State Nutrition Survey. Therefore, in 1968 Congress established the Senate Select Committee on Nutrition and Human Needs to act as watchdog and advocate for government food programs. In 1969, President Nixon called for a White House Conference on Food, Nutrition, and Health, and one panel was detailed to deal with food packaging and labeling (Austin 1977). This panel stipulated that the principal function of a food package is to protect the product and to provide an efficient means for its distribution, while the principal function of a food label is to identify and describe the product and to provide information to promote use of the product. The panel concluded that new packaging and labeling concepts for better nutrition and nutrition information would succeed to the extent that they supported these basic functions. Three of their recommendations were:

1. initiate a comprehensive review of laws and regulations to determine whether they insure and facilitate the delivery of sound nutrition to the consumer;
2. examine attitudes, economic barriers, ethnic preferences, geographic patterns, and other factors relevant to improving nutrition and health, including full utilization of the package and the label; and
3. appoint special committees to consider specific aspects of nutrition (e.g., ways to simplify labels while at the same time making them more descriptive nutritionally).

As a direct result of the White House Conference, the Food and Drug Administration (FDA) began to explore different methods of displaying nutrition label information. In 1972, the FDA presented a proposal for a nutrition labeling regulation which was modified and issued as a final regulation in 1973 (with an effective compliance date of July, 1975). The regulation emphasized voluntary rather than mandatory labeling in an attempt to speed acceptance and to avoid litigation which would delay implementation. The main elements of the regulation stated (Austin 1977, pg. 29):

1. the presence of a nutrition label is voluntary except when a nutrient is added to a processed food and/or a nutrition claim is made;
2. the nutritional information, if present, must be presented conspicuously on or to the immediate right of the principal display panel;
3. the information presented must include serving size, servings per container, calorie/protein/carbohydrate/fat content, and the percentages of the U.S. Recommended Daily Allowances (U.S. RDA) for eight specified nutrients;
4. a claim cannot be made that a food is a significant source of a nutrient unless each serving provides at least 10% of the U.S. RDA of that nutrient, nor can a product claim nutritional superiority in a nutrient unless it contains at least 20% of the U.S. RDA for that nutrient;
5. imitation foods are required to be labeled as imitation only when they are nutritionally inferior;
6. cholesterol and fatty acid labeling is permitted but must be accompanied by the other nutrition label information;
7. samples of fabricated or fortified foods to which nutrients have been added must contain at least 100% of their labeled value, while samples of food in which there are naturally occurring nutrients must contain at least 80% of their labeled value; and
8. fortified foods containing from 50% to 150% of the U.S. RDA of any nutrient are considered to be a dietary supplement rather than an ordinary food and are thus subject to additional regulations, while food with U.S. RDA content over 150% falls into the drug classification.

Food industry response to the labeling regulation ranged from confrontation to formulation of nutrition policies and organization of new consumer departments. Throughout the 1970's the pace of nutrition-oriented product development accelerated, with the appearance of "natural" foods and soy-based textured vegetable products. Various members of the food industry utilized the new consumer awareness of nutrition as a marketing strategy to expand their market shares. Other members of the food industry felt obliged to use nutrition

labeling as a defensive marketing strategy to prevent the loss of old customers who would otherwise switch to nutrition-labeled brands (Austin 1977).

Data-based estimates of the extent of nutrition labeling are limited. The Grocery Manufacturers of America (GMA 1975) conducted a survey of 110 member and 155 non-member companies prior to the 1975 effective date for the nutrition labeling regulation. Of the 79 companies that responded, most were food manufacturers, with a few retailers included. Eighty-five percent of those manufacturers responding indicated an intent to label some if not all of their products. Of the 15% responding negatively, 58% felt that their products were inappropriate for labeling (e.g., apple butter, fresh meat, margarine, cake flour, jellies, mustard, etc.). In nearly every product category, some manufacturers planned to institute nutrition labeling, and some did not. Larger companies tended to plan to institute nutrition labeling sooner than did smaller companies.

Austin (1977), for the Market Science Institute, integrated the GMA survey and other sources to conclude that \$17.5 billion (approximately 14%) of the \$127 billion representing total U.S. food consumption (including away-from-home consumption) was nutritionally labeled. Furthermore, approximately 30% of the \$57.6 billion remaining after exclusion of away-from-home consumption, condiments, alcoholic beverages, and fresh fruits and vegetables was nutritionally labeled.

Because of the scarcity of direct data estimating the total incidence of nutrition labeling and because of the importance of overseeing and perhaps modifying the nutrition labeling regulations, the Division of Consumer Studies, Bureau of Foods, FDA implemented the formation of the "Food Label and Package Surveillance (FLAPS) Data Base". Five objectives were formulated (Schucker 1978):

1. develop a sales-based estimate of the volume-importance of packaged foods currently sold under nutrition labeling;
2. extend the nutrition labeling conclusions to individual product classes;
3. estimate the percentage of packaged foods voluntarily labeled and assess the food industry initiatives in labeling;
4. determine the quantity and volume-importance of retail food sold in violation of nutrition labeling regulations;
5. make the data base open-ended to allow examinations of trends across years.

Data collection began in 1977.

Methodology

The FLAPS data base integrates two data sources: (1) the A.C. Nielsen Directory of Supermarket Products and (2) FDA in-house compilations of package and package-label information, off-label advertising, and package construction materials. The Nielsen Directory provided the sampling frame and market information. The in-house compilation supplied all non-market information (label information, package construction, package dimensions, etc.).

Nielsen Directory - Sample Specification

The sampling universe for the initial round of the FLAPS data base (FLAPS-1) consisted of the A.C. Nielsen

Directory of Supermarket Products for April-May, 1975-76. The Nielsen Directory, part of the Nielsen Early Intelligence System (NEIS), is a syndicated sales data base concerned with merchandise movement from central warehouses to a selected sample of U.S. food stores. The information in each directory represents bimonthly statistics for two consecutive years. Weights furnished by A.C. Nielsen enable the user to project the bimonthly sample to estimate total annual retail movement in the United States. The weights are based upon: (1) the ratio of number of stores in the sample to total Nielsen-type stores in the U.S.; (2) the percent of foods in a particular product class that move through Nielsen-type stores; and (3) the fraction of the total year during which the sample was collected. The information listed in the Nielsen Directory for any single item line within a brand includes can size, average retail price, and unit and dollar movement. The market share information is aggregated to the brand and product class levels. The 411 product classes included in the Nielsen Directory were aggregated by Nielsen at the request of the FDA into 51 major supermarket food groups (e.g., product classes "Spaghetti - Dry" and "Macaroni - Dry" became members of the product group "Pasta - Dry"). Sampling consisted of a multi-stage selection process.

Product Class Selection. Product classes within each product group were rank-ordered by decreasing retail dollar sales volume. Those product classes within a product group with largest dollar volume were selected into the data base until two-thirds of the retail dollar volume for each (and every) product group was accounted for. This procedure resulted in the selection of 158 of the 411 product classes. Forty additional product classes were randomly selected from the remaining 253, resulting in a total of 198 product classes, accounting for 85% of the total retail dollar volume of the 51 product groups.

Brand Selection. Six national brands were drawn from each of the 198 product classes: (1) those three brands in each product class ranked first through third (market leaders) in terms of retail dollar sales and (2) three additional brands selected randomly from the remaining brands (non-leaders) in each product class. (If six or fewer brands were available for any one product class, all brands were selected.)

Due to restrictive agreements between the A.C. Nielsen Company and retail organizations, sales information for individual private label brands was not included in the Nielsen Directory. Therefore, a modified sampling plan for private label brands was instituted, such that those 20 product classes with the largest private label retail dollar sales volume were selected. Twenty additional product classes were randomly selected from those remaining product classes that contained private label brands. Based upon a separate listing of retail firms with one or more private label brands, those three firms with the largest total annual retail food sales and three randomly selected firms were chosen in each product class. Field personnel were permitted to pick up any private label brand in a specific product class controlled by the sampled firm.

Item Selection. A "brand" was defined to include all sizes, flavors and form variations of a product under a particular label. Selection of a specific item within a brand (e.g., Brand X whole tomatoes in can size 303 rather than Brand X peeled tomatoes in can size 300) was accomplished by selecting the item with the largest sales volume within that brand. It was decided that those items would be most likely to reflect the manufacturers' current labeling policies and most recent label formats.

Special Samples - Meat and Convenience Foods. Processed meat and poultry and products containing meat and poultry, although regulated by the U.S. Department of Agriculture (USDA), were included in the FLAPS data base to determine the current extent of nutrition labeling for those products in anticipation of the issuance of labeling regulations by the USDA similar to those of the FDA. To segregate the products from those regulated by FDA, a separate product group was created. The sampling procedure was identical to that for FDA-regulated products.

Convenience samples (store-door delivery) of bread, fresh milk, and ice cream products were collected in three major markets: New York, Chicago, and Los Angeles metropolitan areas. All brands encountered in two leading chain stores and one leading independent store in each market were purchased. This sampling procedure was instituted due to lack of available retail sales information at the brand level for this category.

In-House Data Collection

Field personnel collected 1,713 individual brand items between April, 1977 and March, 1978. These items were sent to the Washington, DC office of the FDA for processing. Products without nutrition labels were sent as empty packages. Products with nutrition labels were sent intact (unopened) and were assayed for nutrient content upon arrival. All information on package labels (including all nutrition label information) was recorded. Access was obtained to the food products advertising file of the Bureau of Consumer Protection of the Federal Trade Commission. All available magazine and newspaper advertisements, radio scripts, and television commercials (in storyboard form) were analyzed for content and classified as to presence or absence of health and/or nutrition claims that would trigger mandatory nutrition labeling. When all information to be compiled in-house had been prepared, the resulting file was merged with the Nielsen marketing data and set up as a computer file data base.

FLAPS-1 Sample Coverage

The 51 product groups (411 product classes) in the 1975-76 Nielsen Directory represented an estimated \$45.9 billion in annual retail food sales. Excluding meat and meat containing products (regulated by the USDA), milk, milk products, snacks, and carbonated soft drinks (because of the significant amount of store-door delivery), remaining food products accounted for \$38.1 billion (360 product classes).

The FLAPS-1 data base (excluding meat and meat containing products, milk, milk products, snacks and carbonated soft drinks) totaled \$22.3 billion in projected annual retail sales (158 product classes, or 1,055 brands). Employing non-leader brands as unbiased estimators for all non-leaders in sampled product classes, the \$22.3 billion was projected to \$27.3 billion. The 862 market leader and non-leader brands, all of which had brand level sales data available, represented \$23.1 billion of the \$27.3 billion in retail market sales.

Results

The data in FLAPS-1 were examined using two dependent measures: (1) percent of brands (to estimate the extent of nutrition labeling among manufacturers) and (2) share of market dollars (to estimate the impact of labeling practices upon the retail market). Unless specifically noted, all analyses and figures are based upon packaged processed foods, excluding processed meat and

meat-containing products, milk, milk products, snacks and carbonated soft drinks.

General Analysis

Approximately one third (33.5%) of the 862 national brands in the FLAPS-1 data base displayed a nutrition label (see Table 1). These nutrition labeled brands

TABLE 1

PERCENT OF PROJECTED ANNUAL RETAIL MARKET DOLLARS (AND PERCENT OF BRANDS) FOR NATIONAL BRANDS*

Nutrition Label	Nutrition Label			Σ
	S (n)	Required	Not Required	
Present		27.6% (17.4%)	19.9% (16.1%)	47.5% (33.5%)
Not Present		0.6% (0.9%)	51.9% (65.4%)	52.5% (66.5%)
Σ		28.2% (18.3%)	71.8% (81.7%)	\$23.1 billion (n = 862)

*Packaged processed foods excluding processed meat, milk, milk products, snacks and carbonated soft drinks.

accounted for 47.5% of the \$23.1 billion dollars in projected annual retail sales for national brands sampled by FLAPS-1. Approximately the same proportion of brands displayed nutrition labels (16.1%) solely on a voluntary basis as fell under the mandatory nutrition labeling requirement (17.4%) because of either an on-label nutrition claim, an off-label nutrition claim, the addition of one or more nutrients to the product, or some combination of the above. However, those brands with required nutrition labels accounted for a greater relative share of total market dollars (27.6%) than did brands with voluntary nutrition labels (19.9%) and share of market for both was greater than the corresponding percent of brands falling into those two categories. Brands in violation of nutrition labeling regulations (required label not present) represented 0.9% of all brands and 0.6% of all retail market dollars. Over half of all brands (66.5%) were not required to and chose not to display a nutrition label (but accounted for only 52.5% of the retail market).

Market Leader vs. Non-Leader Brands

Labeled vs. unlabeled. Approximately equal numbers of market leaders (N=460) and non-leaders (N=402) were sampled in FLAPS-1 (see Table 2). Market leaders accounted for a greater relative share of market dollars (\$17.2 billion) than non-leaders (\$5.9 billion), as would be expected given the definition of leaders and non-leaders: leaders represented the three largest brands in each product class in terms of retail dollar sales. A greater proportion of leader brands (40.7% of the total number of leader brands) than non-leader brands (25.4% of the total number of non-leader brands) engaged in nutrition labeling. For both leaders and non-leaders, the share of total market dollars accounted for by nutrition labeling (51.6% and 35.7%, respectively) was greater than the corresponding proportion of all leader and non-leader brands with nutrition labeling (40.7% and 25.4%, respectively).

Mandatory vs. voluntary labeling. Market leader brands with nutrition labels were divided evenly between brands with required nutrition labels (20.0%) and those engaged in voluntary nutrition labeling (20.7%). Market leaders

with required nutrition labels accounted for a larger

TABLE 2
PERCENT OF PROJECTED RETAIL MARKET DOLLARS
(AND PERCENT OF BRANDS) FOR MARKET LEADERS,
NON-LEADERS, AND PRIVATE LABELS*

Market Leaders Only

N u t r i t i o n	Nutrition Label			
	\$ (n)	Required	Not Required	Σ
Present	29.3% (20.0%)	22.3% (20.7%)	51.6% (40.7%)	
Not Present	0.7% (0.9%)	47.7% (58.4%)	48.4% (59.3%)	
Σ	30.0% (20.9%)	70.0% (79.1%)	\$17.2 billion (n = 460)	

Market Non-Leaders Only

N u t r i t i o n	Nutrition Label			
	\$ (n)	Required	Not Required	Σ
Present	22.8% (14.4%)	12.8% (10.9%)	35.7% (25.4%)	
Not Present	0.4% (1.0%)	63.9% (73.6%)	64.3% (74.6%)	
Σ	23.3% (15.4%)	76.7% (84.6%)	\$5.9 billion (n = 402)	

Private Label Brands Only

N u t r i t i o n	Nutrition Label			
	\$ (n)	Required	Not Required	Σ
Present	17.1% (19.2%)	18.8% (24.4%)	35.9% (43.5%)	
Not Present	** (0.5%)	64.1% (56.0%)	64.1% (56.5%)	
Σ	17.1% (19.7%)	82.9% (80.3%)	\$4.2 billion (n = 193)	

*Packaged processed foods excluding meat, milk, milk products, snacks, and carbonated soft drinks.

** < .05%

proportion of total market leader dollars (29.3%) than did market leaders engaged in voluntary nutrition labeling (22.3%). In both categories, however, share of market was greater than the relative proportion of brands engaged in nutrition labeling.

More non-leader brands engaged in mandatory (14.4%) than in voluntary nutrition labeling (10.9%). Furthermore, the share of market dollars accounted for by mandatory labeling among non-leaders (22.8%) was more than a proportionate increase over the share accounted for by voluntary labeling among non-leaders (12.8%).

Labeling regulation violations. Approximately the same proportion of leader (0.9%) and non-leader (1.0%) brands did not display a required nutrition label. In both instances, share of market dollars (0.7% and 0.4%, respectively) was less than the percent of brands.

Private Label Brands

Nutrition labels were present on 43.5% of all private label brands (a greater overall proportion than for either market leaders or non-leaders). In contrast to market leaders and non-leaders, however, the proportion

of private label brands with required labels (19.2%) was less than the proportion with voluntary labels (24.4%). Because of the lack of private label brand-level sales information, share of market dollars for private label brands (see Table 2) was not comparable to share of market dollars for leaders and non-leaders. Total private label dollars within a product class had been divided equally among all private label brands within that product class for purposes of analysis and thus did not reflect individual brand share differences.

Fewer private label brands violated the nutrition labeling regulations (0.5%) than either market leader (0.9%) or non-leader brands (1.0%). At the same time, more private label brands engaged in voluntary labeling (24.4%) than either market leader (20.7%) or non-leader brands (10.9%).

Association Between Labeling and Share of Market

The fact that nutrition labeled brands (in aggregate) accounted for a greater relative proportion of the retail market dollars than did non-labeled brands (and mandatory more than voluntary) was indicative of a possible association between presence (and type) of labeling and brand share of market. A point-biserial correlation for all market leader and non-leader brands produced a modest but significant positive correlation between presence/absence of a nutrition label and brand share of market ($r = .14, p < .01, n = 862$). A point-biserial correlation between mandatory/voluntary labeling and brand share of market again produced a modest but significant correlation ($r = .11, p < .01, n = 862$). However, the mandatory/voluntary labeling by brand share correlation was not significant when limited to only those brands displaying a nutrition label ($r = .05, p > .10, n = 289$).

Mandatory Nutrition Labeling Triggers (National Plus Private Label Brands)

Labels present. Of the 1,055 national and private label brands in FLAPS-1, 196 were required by regulation to display a nutrition label. The most frequent single trigger for labeling (N=86) for those brands that did display a nutrition label (N=187) was an off-label nutrition claim (see Table 3). The two remaining triggers, added nutrients (N=35) and an on-label nutrition claim (N=30) occurred with about equal frequency. Nutrition labels on the remaining 36 brands were triggered by combinations: an on-label nutrition claim plus added nutrients (N=7) and an on-label nutrition claim plus an off-label nutrition claim plus added nutrients (N=29).

Labels absent. For the nine labeling violations, the most frequent causes were on-label nutrition claims (N=4) and on-label nutrition claims plus added nutrients (N=3). The last two violations were caused by added nutrients (N=1) and an on-label nutrition claim plus an off-label nutrition claim plus added nutrients (N=1).

USDA Regulated Products

A cross-tabulation was performed for only those 210 products in FLAPS-1 regulated by USDA (see Table 4). If the products had been subsumed under FDA regulations, 2.9% of the 210 brands would have been required to display package nutrition labels. Approximately one-third of those brands (1.0% of all 210 brands) did in fact display a nutrition label. The remaining two-thirds of the brands that would have been required to display a nutrition label (1.9% of the 210 brands) would have been in violation of the labeling regulation. Out of the 9.0% of the brands that did display nutrition

TABLE 3

DISTRIBUTION OF FREQUENCY OF OCCURRENCE FOR TRIGGERS OF REQUIRED NUTRITION LABELING*

	N	Trigger for Required Nutrition Labeling					Σ	
		(a) Label Claim	(b) Added Nutrient	(c) Off-Label Claim	(d) A+B	(e) A+B+C		Not Required
Nutrition Labeling Present	Leaders	12	13	35	6	26	95	187
	Non-Leaders	16	13	26	0	3	44	102
	Private Label	2	9	25	1	0	47	84
	Σ	30	35	86	7	29	186	373
Nutrition Labeling Not Present	Leaders	1	0	0	3	0	269	273
	Non-Leaders	2	1	0	0	1	296	300
	Private Label	1	0	0	0	0	108	109
	Σ	4	1	0	3	1	673	682
	Σ	34	36	86	10	30	859	1055

*Packaged processed foods excluding meat, milk, milk products, snacks, and carbonated soft drinks.

TABLE 4

PROCESSED MEAT AND MEAT-CONTAINING PRODUCTS ONLY

Nutrition Label	N	Nutrition Label		Σ
		Required	Not Required	
Present	2	17	19	
	(1.0%)	(8.1%)	(9.0%)	
Not Present	4	187	191	
	(1.9%)	(89.0%)	(91.0%)	
Σ	6	204	210	
	(2.9%)	(97.1%)	(100%)	

labels (either required or voluntary), 8.1% would have been considered engaged in voluntary labeling. Finally, 89.0% of the brands did not display nutrition labels but would not thereby have violated the regulation.

Conclusion

An estimated one-third of all national brand processed foods (excluding meat, milk, milk products, snacks, and carbonated soft drinks) packaged in the United States in 1976 displayed a package nutrition label. These nutrition-labeled brands accounted for slightly less than one-half of the estimated total retail dollar volume for national brand products. This figure appears consistent with Austin's (1977) estimate of 30% for national retail sales under nutrition labeling, given the differences in sample composition. (The FLAPS-1 estimate excluded all private label brands, meat and meat containing products, milk, milk products, snacks, and carbonated soft drinks in addition to the away-from-home dining, condiments, alcoholic beverages, and fresh fruits and vegetables excluded by Austin.)

Half of those brands with nutrition labels were required to display the information. Thus, one-half of nutrition labeled national brands (or one-sixth of all national

brands) either made on-label or off-label nutrition claims, added one (or more) nutrients to the product, or engaged in a combination of all three. A further one-sixth of all national brands engaged in nutrition labeling on a purely voluntary basis (i.e., although the package contained a nutrition label, no on- or off-label nutrition claims were made and no nutrients were added to the product). Brands engaged in voluntary nutrition labeling accounted for approximately one-fifth of the estimated retail dollar sales for national brands in 1976. Brands with mandatory nutrition labeling, while approximately equal in number to brands with voluntary labeling, accounted for half again (roughly three-tenths) of estimated national brand retail dollars. Thus, share of retail dollar sales was proportionately greater for nutrition-labeled than non-labeled brands and, within nutrition-labeled brands, for brands with mandatory rather than voluntary nutrition labeling. Significant albeit small positive correlations between (1) presence of nutrition labeling and share of market dollars and (2) type of nutrition labeling and share of market dollars supported the conclusion of a direct association.

A total of 0.9% of all national brands and 0.5% of all private label brands violated the nutrition labeling regulation by not displaying a required nutrition label. For national brands, the ratio of "share of retail dollars" to "percent of brands" in violation was less than 1.0. (Market share information at the brand level was not available for private label brands.) Thus, those manufacturers that violated the nutrition labeling regulation tended to be the smaller manufacturers. Label nutrition claims (with and without added nutrients) were the most frequent causes of the violations.

Whether nutrition labeling was instrumental in increasing brand share of market or whether large companies instituted nutrition labeling prior to smaller companies (or both) cannot be determined from the present data. However, the direct relationship between brand share and presence of labeling does appear consistent with the earlier GMA (1975) finding that large manufacturers intended to begin nutrition labeling before smaller companies. Later extensions of the FLAPS data base may provide insights as to the causal relationship between labeling and size of market share as trend data across time for both nutrition labeling and size of manufacturer become available.

References

Austin, James E. (1977), "Food Marketing: What Role for Nutrition?", Marketing Science Institute, Report No. 77-104.

Grocery Manufacturers of America (1975), "Results of the GMA Survey on Nutrition".

Schucker, Raymond E. (1978), "A Surveillance of Nutrition Labeling in the Retail Packaged Food Supply," Division of Consumer Studies, Bureau of Foods, Food and Drug Administration, Washington, DC.

White House Conference on Food, Nutrition and Health: Final Report, (1970), Washington: U.S. Government Printing Office.

IMPLEMENTATION OF A REVISED FOOD LABELING POLICY: EVALUATION AND TRACKING

Raymond E. Schucker, Food and Drug Administration

Abstract

The paper reviews the environment in which FDA is revising the food labeling regulations regarding nutrition information and ingredients. The problem of raising the salience of nutrition to the public is discussed from education and marketing perspectives. FDA plans are outlined for introductory consumer education on how to use the new label, including the need for evaluative research and market tracking of the rate of new label adoption by the food industry.

Background

The program of research and development in support of a revised food labeling policy will require approximately four years to bring to the point of national implementation. The first two years, 1978 and 1979, were concerned with identifying what changes should be made in food labeling information based on consumer expressions of interest from survey research, public hearings, representations of consumer need by organized consumer groups and consideration of public health as seen by food and nutrition professionals. Tentative positions on a number of food labeling issues were published in the Federal Register of December 21, 1979.

In the next two years three parallel efforts will complete the development phase of the revision of the food label. A series of experimental laboratory studies will be conducted to screen and identify the more promising label formats from among a large array of alternatives which will be created by a professional design firm. These studies are discussed in greater detail in another paper of this session. In a continuation of the tentative initiatives announced in the 1979 position paper FDA will publish, invite public comment, revise and issue specific regulations providing for certain expanded nutrition and ingredient information for which the Agency already has authority under the Food, Drug and Cosmetic Act. These include labeling the source of fat, requiring quantitative declaration of the amount of sugar, sodium and potassium, and establishing definitions for low and reduced sodium and cholesterol. A regulation defining levels for low and reduced calories became effective July 1980. For certain other desired food label changes FDA does not currently have legislative authority and will need to seek authority from Congress during this period, notably to require nutrition labeling on more foods and to extend specificity of ingredients declarations along a number of dimensions. The third effort during the remainder of the policy formulation and development stage will involve assessment of the economic impact of the proposed label changes on the food industry and the assessment of possible increased food costs against anticipated consumer benefits. Economic analysis of proposed regulatory actions is required where potential impact on the regulated industry is anticipated to exceed \$100 million.

Environment for Change

For a number of reasons, including lack of defined regulatory authority on the part of FDA, economic impact of the label revisions, or failure of the consumer laboratory research to sufficiently narrow the range of label format alternatives, it is not possible at this time to forecast in complete or final detail what the policy revisions will be or to mark a clear point in time when the process will move out of the policy formulation-

development stage into an implementation phase. However, for purposes of planning and discussion, it is convenient to establish the implementation phase as beginning in mid 1983 which is the uniform effective date set by FDA for industry compliance with final label change regulations growing out of the December, 1979 tentative positions.

Based on an FDA survey in the Fall of 1978 (Heimbach and Stokes 1979), overt consumer interest in improved food label information about nutrition or ingredients ranged between 8 - 15% among food shoppers, conservatively estimated through nondirective questioning. This is interpreted by some industry opponents of further change in food labeling as reflecting a low level of interest in label information and is raised as an objection to further regulation at this time. Whether one accepts FDA's more conservative estimate or other higher estimates of demand for information, such as have been published from time to time based on claimed use of or preference for food label information, consumer interest is not expected to increase in the near future and may even be eroding somewhat as other more pressing food concerns arise. In the general inflationary spiral of the late 1970's, accompanied by rising food prices, consumers began to become increasingly leery of Government regulatory action that might contribute to further increases in the cost of food. Most published consumer surveys continue to identify price as an overriding critical factor in food purchasing.

Of greater significance than the absolute level of consumer demand for nutrition and food ingredient information is that the nature of concern about food has generally turned away from emphasis on getting enough of the essential nutrients to avoidance of the consequences of overeating, particularly obesity. Also, uncertainty about the function, necessity and even safety of food preservatives and additives in general has led to an avoidance set among increasing numbers of consumers in their selection of food. Organized attempts are growing to press for the removal of questionable ingredients from the food supply where particularly vulnerable segments of the population are thought to be at risk. An example is the national network of parents groups that support the Feingold hypothesis linking certain color and flavor additives to hyperkinesis in young children.

The attention of the organized consumer movement in the last two years has also shifted somewhat away from the longer term issues of health and nutrition. With the exception of concern about the health effects of sugar consumerists are placing more emphasis on food safety, government control over environmental contaminants and removal of hazardous substances from foods. Nutrition labeling is not given quite the same priority as during the early and middle years of the decade, and there are no indications on the horizon that the situation will be reversed soon.

Although interest in revising the nutrition label has plateaued and counter pressures have appeared in some public sectors, interest in nutrition information has not declined among food retailers and in fact has been on the rise since publication in 1977 of suggested national dietary guidelines by the Senate Select Committee on Health and Human Needs under the chairmanship of Senator McGovern. Retailer activity accelerated in the latter 1970's as food chains began to develop and implement in-store information programs featuring nutrition. Retailer receptivity to promotional ideas and materials has

been great enough that at least one firm is now marketing an in-store nutrition information package to supermarket operators. Other chains have developed and implemented programs of their own.

Some of the retailer programs have been found to use nutrition information that is inaccurate or that makes unwarranted health claims and have been in conflict with existing FDA nutrition labeling regulations. Case law growing out of regulations issued under the Food, Drug and Cosmetic Act has led to the interpretation that food information presented in proximity to food products constitutes labeling. Thus, much if not all product-specific nutrition information that is presented in the supermarket is subject to the regulations. In the interests of minimizing consumer confusion while encouraging innovation, FDA has published a regulation permitting retailers as well as manufacturers to request temporary exemption from nutrition labeling requirements for purposes of testing alternative nutrition labeling formats at point of purchase. To date a number of retailers have sought exemptions for experimental programs, and the outlook is for food chains to continue to use in-store nutrition programs as competitive promotional and merchandising tools as long as retailers are convinced this is a service consumers want and will use.

Very little hard evidence is currently available on whether consumers actually modify their purchase behavior as a result of nutrition information disclosure at point of purchase. A recent collaborative effort (report in preparation) between Giant Food Inc. Washington, DC and the National Heart, Lung and Blood Institute tested a year long series of in-store pamphlets containing information on foods useful in an overall program of heart health maintenance. Supporting radio and newspaper advertising was also used. Sales in a matched panel of test and control stores were audited, and pre and post consumer interview surveys were conducted to track awareness and knowledge of the campaign, as well as specific elements of health information that had been presented. Some campaign effects were detected in the interview data, but no sales changes were observed that could be attributed to the nutrition information program.

Giant Food Inc. has also requested an exemption to conduct an additional two-year nutrition information program involving special shelf edge labeling for food brands that are reduced, low in, or free of sodium, cholesterol, fat and calories. Advertising will support the program in one market, while both shelf labeling and advertising will be withheld from a control market. Sales in a matched panel of stores will be tracked via electronic scanning of the universal product codes of nutrition labeled foods versus their unlabeled regular counterparts. Results of this and other experimental studies in supermarkets are not likely to be available in time to impact implementation of the label revisions in 1983.

Implementation

Although an effective date of mid 1983 has been set for industry compliance with the new or revised food labeling regulations, the individual regulations will actually issue at different points in time prior to the effective date. Some food manufacturers will begin to phase in revised labels as soon as necessary analytic information has been obtained to support newly required label declarations and when convenient with respect to label production and inventory cycles. Consumers will therefore experience a trickling in rather than an abrupt introduction of the new food label format.

Informing Versus Educating

The nutrition professional community and educators for a number of years have stressed the need for an overhaul of the basic four food groups approach to nutrition and for greater government support to develop curricula and teach nutrition at the primary and secondary levels. Viewed against the background of dissatisfaction with the level of nutrition knowledge and dietary practices of the U.S. population, FDA's program to revise the food label is seen in perspective as an effort to improve the information base which consumers need in order to make better nutritional food purchases. However, by no means does food labeling constitute education in the sense of teaching skills or modifying behavior. FDA has no legislative mandate or funding to develop curricula or otherwise to engage in basic long term education. The Agency currently envisions developing a short term, perhaps two-year program of introductory announcements and instruction on how to understand and use the revised food label. The program will be delivered through the print and broadcast media. In addition, instructional presentations such as slide shows will be distributed at the community level by FDA's field network of Consumer Affairs Officers.

Little guideline information is available on the resources and national level of effort that might be required to raise public awareness and change attitudes toward a subject so technically complex and burdened with difficult scientific vocabulary as nutrition. Even less is known about whether lasting behavioral changes can be achieved, let alone sustained, without massive support over a long period of time. The traditional education approach is generally criticized as having been less effective in raising the nutritional well being of the population than the direct interventions of food fortification or the various government supported feeding programs such as the school lunch programs.

An education framework may not be the most useful approach to determining what needs to be done to raise the nation's nutrition consciousness. The marketing approach, to take a different view, would focus on identifying then solving consumer nutrition-related needs and problems. This perspective on nutrition brings at least two issues into relief.

The first is that nutrition problems are generally perceived as not very immediate. Cardiovascular disease and hypertension may not produce a stroke or coronary for decades. Sugar in food is less popularly recognized as a promoter of calories than as a contributor to weight gain. Even overweight and obesity, although highly visible precursors of possible future health problems, may not be threatening today. In short, consumers don't have problems.

The second issue stems from the first and is the need to raise the salience of nutrition, i.e., to motivate consumers to want to behave as though tomorrow's health problems are really today's concerns. These two aspects of nutrition must be addressed in any national policy which seeks to achieve broadbased improvement in food and dietary behavior through consumer voluntary action. The magnitude of the requirement suggests that an educational approach alone will not be adequate to the task, nor does the superficial marketing analysis of the problem provide any ready suggestions as to how to solve it, or indicate the level of national effort and time frame that will be required to effect meaningful change.

Research Needs

In part because of the complexities of nutrition discussed previously, as well as time and resource

constraints, the experimental laboratory research to select an improved label format will be restricted mainly to criteria that demonstrate the consumer's ability to comprehend food label information and make appropriate product decisions regarding nutritional quality or ingredients composition. Consumer satisfaction with quality of decision making will also be used as a criterion measure in evaluating alternatives. On the other hand, the research conducted in the label development stage will not deal with consumer motivation to use information, or with intentions or other behavioral type criteria.

Unresolved Issues

Given the numerous elements of graphic design that might be used in formating nutrition and ingredients information, it is possible, even likely, the research will show that a number of alternative label information displays are equally effective in meeting the criteria of comprehensibility and utility in product decision making. Ultimately it may be appropriate to permit use of alternative formats to accomodate different package sizes and shapes. One school of thought proposes marketplace testing after the experimental laboratory type research in order to select a final label or two on the basis of demonstrated effects on food sales. However, to use a marketplace test for a run off of labels, in addition to being costly, would confound the issue of consumer ability to understand and apply information with the issue of willingness to use it. Moreover, in the typical 'noisy' market test employing a sales criterion measure, the odds do not favor the detection of consumer behavior change in response to revised nutrition label information alone in the absence of other efforts to begin to deal with the motivational side of the problem.

Use of multiple communication channels to deliver nutrition information has been proposed as a partial solution to the problem of raising the salience of nutrition, for example, by making a generic or even brand specific food information handbook available in the home for use in menu planning. Another suggestion has been to place similar information in a reference book in grocery stores, including nutrition information for fresh fruits, vegetables and meats, which are now exempt from nutrition regulations.

The current surge of retailer interest in featuring nutrition at point of purchase opens up possibilities for controlled store research using actual sales data to test variables that could enhance the impact of the revised food label as national implementation begins. An example might include use of the shelf edge to highlight selected new elements of nutrition information not previously appearing on food labels. Short term special floor displays in high traffic areas of the store might group an array of products having one or more of the same nutrition characteristics. Also, the food retailer's weekly newspaper feature ad might offer the opportunity to deliver nutrition information in conjunction with promotional pricing, which has high salience for the consumer.

Experimentation is also needed to establish the content of nutrition messages and weight of effort required in the broadcast media, use of which many believe is necessary if all sectors of the public are to be effectively reached and influenced by nutrition information. The grocery store in today's favorable retail climate is an appropriate focal point for designing and executing sensitive studies of behavior response to multi source communications, such as retailer and public service advertising, as well as locally produced radio and television shows covering nutrition related health subjects in community service time slots.

Program Evaluation

FDA's regulation permitting exemption from strict adherence to nutrition labeling for experimentation purposes includes the requirement of a formal evaluation component. On request the Division of Consumer Studies, Bureau of Foods, has been cooperating with food retailers in reviewing and offering comment on the design of such studies, which are expected to continue as in-store programs are developed to tie in with implementation of the revised labeling regulations. Evaluation studies have included the use of survey based consumer measures and store based sales criteria in pre versus post designs employing treatment groups only and also in more complex experimental designs using both treatment and control groups.

Through the mechanism of its annual food shopper survey the Agency will measure growth in consumer awareness and response to the introduction of the new label in relation to a benchmark point prior to the effective compliance date at retail. Evaluation criteria will include knowledge of newly added nutrition information, comprehension of nutrition vocabulary items targeted for revision, knowledge of new classifications for foods on key nutrients, such as calories, sodium and cholesterol, and consumer satisfaction with the quality of nutrition decisions made in food purchases.

The annual shopper survey will also be used in the evaluation of the two-year introductory education campaign now under development. Of particular concern is that the campaign, as well as the revised label, reach the poor, the undereducated and those who are not information seekers. The education campaign will include script and materials tailored to half-hour television and radio programs to be produced locally and featuring some aspect of the newly revised nutrition label as it relates to health, such as sodium and hypertension. These will be evaluated in terms of extent of acceptance by television stations for production, actual airing in local time slots, and size and composition of audience reached. Local consumer surveys will be conducted to determine relative impact of local versus national programming.

Market Tracking

For the near future, manufacturers will continue to have considerable latitude in deciding voluntarily whether to offer nutrition information at all on the food label. Until FDA obtains greater regulatory authority nutrition labeling will be required only if nutrients are added or advertising or nutrition label claims are made. Thus the rate at which the revised regulations impact the consumer will be tied in part to the extent of industry adoption and rate of changeover to the new format.

FDA will establish a benchmark level of nutrition labeling in the U.S. food supply under the current format via a food labeling and product surveillance survey which is conducted biannually based on a probability sample of packaged food products purchased from the retail shelf (Schucker 1978). Content analysis of the labels in terms of projected national sales of the sampled brands will permit tracking of the rate of changeover to the new food label by product class. The survey will also be used to assess quality of adherence to the specific labeling requirements of the regulations and to provide feedback to the regulatory compliance arm of the Bureau of Foods. Finally, label content will be tracked over time for indications of manufacturer use of nutrition and ingredient information for marketing, advertising or promotional purposes beyond simply meeting the requirements to list the information in the label section reserved for this purpose.

Summary

Consumer demand for nutrition information on the food label has plateaued recently, partly because of concern that government regulation will contribute further to inflation and rising food prices. Another factor has been the shift in concern from nutrition related food issues to health and safety questions about food additives and preservatives. Finally, concern about nutrition is turning away from nutrient deficiency to nutrient excess and health consequences of overconsumption.

In contrast, food retailers have shown growing interest in nutrition and are purchasing or developing their own in-store information programs. FDA has encouraged retailer experimentation with alternative ways to present nutrition information and by regulation requires an evaluation component to be included in such special programs. For these reasons the supermarket is an appropriate focus for experimentation during the implementation of the revised food label beginning in mid 1983. A basic need is for research to determine how to raise the salience of nutrition in the general public, to develop meaningful nutrition communications content, and to estimate the level of national effort that will be required, particularly in broadcast media.

FDA will use its annual food shopper survey to evaluate at the national level a consumer education campaign introducing the revised food label. Local consumer surveys and audience statistics will serve as a basis for evaluating television script packages designed for production by local television stations. The rate of food industry changeover to the new food label will be tracked via an ongoing biannual study based on probability sampling of packaged food products at retail.

References

- Heimbach, James T. and Stokes, Raymond C. (1979), "FDA 1978 Consumer Food Labeling Survey", Division of Consumer Studies, Bureau of Foods, Food and Drug Administration, Washington, DC.
- Schucker, Raymond E. (1978), "A Surveillance of Nutrition Labeling in the Retail Packaged Food Supply", Division of Consumer Studies, Bureau of Foods, Food and Drug Administration, Washington, DC.

FACTS AND FEARS: SOCIETAL PERCEPTION OF RISK¹

Paul Slovic, Decision Research, A Branch of Perceptronics
Baruch Fischhoff, Decision Research, A Branch of Perceptronics
Sarah Lichtenstein, Decision Research, A Branch of Perceptronics

Abstract

Subjective judgments, whether by experts or lay people, are a major component in any risk assessment. If such judgments are faulty, efforts at public protection are likely to be misdirected. The similarities and differences between lay and expert evaluations are examined in the context of a specific set of activities and technologies.

Introduction

People respond to the hazards they perceive. If their perceptions are faulty, efforts at public and environmental protection are likely to be misdirected. For some hazards, extensive statistical data are readily available; for example, the frequency and severity of motor vehicle accidents are well documented. The hazardous effects of other familiar activities, such as the consumption of alcohol and tobacco, are less readily discernible; their assessment requires complex epidemiological and experimental studies. However, even when statistical data are plentiful, the "hard" facts can only go so far towards developing policy. At some point human judgment is needed to interpret the findings and determine their relevance.

Still other hazards, such as those associated with recombinant DNA research or nuclear power, are so new that risk assessment must be based on complex theoretical analyses such as fault trees (see Figure 1), rather than on direct experience. Despite their sophistication, these analyses too, include a large component of judgment. Someone, relying on educated intuition, must determine the structure of the problem, the consequences to be considered, and the importance of the various branches of the fault tree.

Once the analyses have been performed, they must be communicated to the various people who actually manage hazards, including industrialists, environmentalists, regulators, legislators, and voters. If these people do not see, understand, or believe these risk statistics, then distrust, conflict and ineffective hazard management are likely.

In this paper, we shall explore some of the psychological elements of the risk assessment process that are critical to the management of hazards. Our basic premises are that both the public and the experts are necessary participants in that process, that assessment is inevitably subjective, and that understanding public perceptions is crucial to effective decision making.

Determinants of Perceived Risk

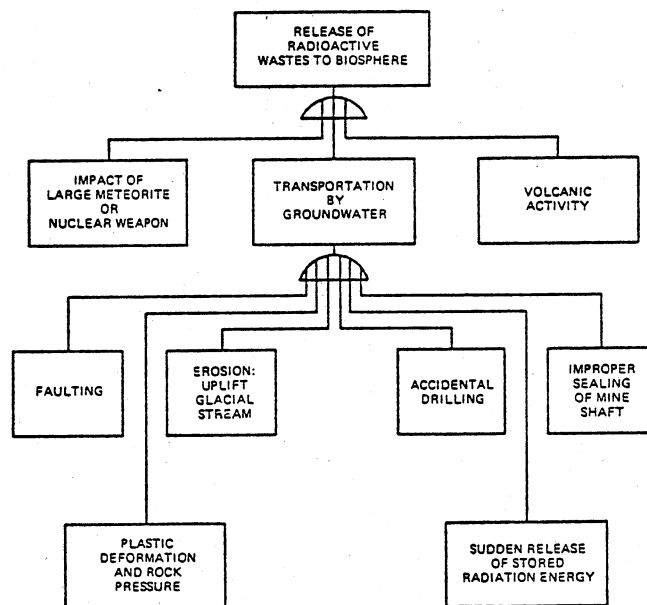
In order to aid the hazard management process, a theory of perceived risk must explain people's extreme aversion to

¹Much of the material presented in this paper is taken from our paper "Perceived Risk" in R.C. Schwing and W.A. Albers, Jr. (Eds.), *Societal Risk Assessment: How Safe is Safe Enough?* New York: Plenum Press, 1980.

²Eugene, Oregon.

FIGURE 1

A fault tree illustrating ways in which radioactive wastes might be released from a nuclear waste repository in bedded salt.³



some hazards, their indifference to others, and the discrepancies between these reactions and experts' recommendations. Why, for example, do some communities react vigorously against locating a liquid natural gas terminal in their vicinity despite the assurances of experts that it is safe? Why, on the other hand, do many communities situated on earthquake faults or below great dams show little concern for experts' warnings? Such behavior is doubtless related to the perceived probability of possible consequences from these hazards. The studies reported below broaden the discussion. They ask, when people judge the risk inherent in a technology, are they referring only to the (possibly misjudged) number of people it could kill or also to other, more qualitative features of the risk it entails?

Quantifying Perceived Risk

In one study, we asked four different groups of people to rate 30 activities (e.g., smoking, firefighting), substances (e.g., food coloring), and technologies (e.g., railroads, aviation) according to the present risk of death from each (Fischhoff 1978, Slovic 1980(a)). Three groups were from Eugene, Oregon; they included 30 college students, 40 members of the League of Women Voters (LOVW),

³Source: P.E. McGrath, "Radioactive Waste Management," Report EURFNR 1204, Karlsruhe, Germany, 1974.

and 25 business and professional members of the "Active Club." The fourth group was composed of 15 persons selected nationwide for their professional involvement in risk assessment. This "expert" group included a geographer, an environmental policy analyst, an economist, a lawyer, a biologist, a biochemist, and a government regulator of hazardous materials.

All these people were asked, for each of the 30 items, "to consider the risk of dying (across all U.S. society as a whole) as a consequence of this activity or technology." In order to make the evaluation task easier, each activity appeared on a 3" x 5" card. Respondents were told first to study the items individually, thinking of all the possible ways someone might die from each (e.g., fatalities from non-nuclear electricity were to include deaths resulting from the mining of coal and other energy production activities as well as electrocution; motor vehicle fatalities were to include collisions with bicycles and pedestrians). Next, they were to order the items from least to most risky and, finally to assign numerical risk values by giving a rating of 10 to the least risky item and making the other ratings accordingly. They were also given additional suggestions, clarifications and encouragement to do as accurate a job as possible.

Table 1 shows how the various groups ranked these 30 activities and technologies according to riskiness. There

TABLE 1
Ordering of Perceived Risk
for 30 Activities and Technologies^a

	Group 1 LOWV	Group 2 College Students	Group 3 Active Club Members	Group 4 Experts
Nuclear power	1	1	8	20
Motor vehicles	2	5	3	1
Handguns	3	2	1	4
Smoking	4	3	4	2
Motorcycles	5	6	2	6
Alcoholic beverages	6	7	5	3
General (private) aviation	7	15	11	12
Police work	8	8	7	17
Pesticides	9	4	15	8
Surgery	10	11	9	5
Fire fighting	11	10	6	18
Large construction	12	14	13	13
Hunting	13	18	10	23
Spray cans	14	13	23	26
Mountain climbing	15	22	12	29
Bicycles	16	24	14	15
Commercial aviation	17	16	18	16
Electric power	18	19	19	9
Swimming	19	30	17	10
Contraceptives	20	9	22	11
Skiing	21	25	16	30
X rays	22	17	24	7
High school & college football	23	26	21	27
Railroads	24	23	20	19
Food preservatives	25	12	28	14
Food coloring	26	20	30	21
Power mowers	27	28	25	28
Prescription antibiotics	28	21	26	24
Home appliances	29	27	27	22
Vaccinations	30	29	29	25

^aThe ordering is based on the geometric mean risk ratings within each group. Rank 1 represents the most risky activity or technology.

were many similarities between the three groups of laypeople. For example, each group believed that motorcycles, motor vehicles and handguns were highly risky, while vaccinations, home appliances, power mowers, and football posed relatively little risk. However, there were strong differences as well. Active Club members viewed pesticides and spray cans as relatively much safer than did the other groups. Nuclear power was rated as highest in risk by the LOWV and student groups, but only eighth by the Active Club. The students viewed contraceptives as riskier and mountain climbing as safer than did the other lay groups. Experts' judgments of risk differed markedly from the judgments of laypeople. The experts viewed electric

power, surgery, swimming and X-rays as more risky than did the other groups and they judged nuclear power, police work and mountain climbing to be much less risky.

What Determines Risk Perception?

What did people mean, in this study, when they said that a particular technology was quite risky? A series of additional studies was conducted to answer this question.

Perceived risk compared to frequency of death. When people judge risk, are they simply estimating frequency of death? To answer this question, we collected the best available technical estimates of the annual number of deaths for the activities included in our study. For some, such as commercial aviation and handguns, there is good statistical evidence based on counts of known victims. For others, such as nuclear or fossil-fuel power plants, available estimates are based on uncertain inferences about incompletely understood processes, such as the effect of low doses of radiation on latent cancers. For still others, such as food coloring, we could find no estimates of annual fatalities.

For the 25 cases for which we found technical fatality estimates, we compared these estimates with perceived risk. The experts' judgments of risk were so closely related to these statistical or calculated frequencies that it seems reasonable to conclude that they both knew what the technical estimates were and viewed the risk of an activity or technology as synonymous with them. The risk judgments of laypeople, however, were only moderately related to the annual death rates, raising the possibility that, for them, risk may not be synonymous with fatalities. In particular, the perceived risk of nuclear power was remarkably high compared to its estimated number of fatalities.

Lay fatality estimates. Before concluding that perceived risk does not mean annual fatalities, we investigated the possibility that laypeople based their risk judgments on subjective fatality estimates which were inaccurate. To test this hypothesis, we asked additional groups of students and LOWV members "to estimate how many people are likely to die in the U.S. in the next year (if the next year is an average year) as a consequence of these 30 activities and technologies."

These subjective fatality estimates are shown in columns 2 and 3 of Table 2. If laypeople really equate risk with annual fatalities, their own estimates of annual fatalities, no matter how inaccurate, should be very similar to their judgments of risk. There was, however, only a low to moderate agreement between these two sets of judgments ($r = .60$ for LOWV and $.26$ for students). Of particular importance was nuclear power, which had the lowest fatality estimate and the highest perceived risk for both LOWV members and students. Overall, laypeople's risk perceptions were no more closely related to their own fatality estimates than they were to the technical estimates. Thus we can reject the idea that laypeople wanted to equate risk with annual fatalities, but were inaccurate in doing so. Apparently, laypeople incorporate other considerations besides annual fatalities into their concept of risk.

Disaster potential. One striking result is the fact that the LOWV members and students assigned nuclear power the highest risk values and the lowest annual fatality estimates. One possible explanation is that LOWV members expected nuclear power to have a low death rate in an average year, but considered it to be a high risk technology because of its potential for disaster.

In order to understand the role played by expectations of disaster in determining laypeople's risk judgments, we asked these same respondents to indicate for each activity

TABLE 2

Fatality Estimates and Disaster Multipliers for 30 Activities and Technologies

Activity or Technology	Technical Fatality Estimates	Geometric Mean Fatality Estimates Average Year		Geometric Mean Multiplier Disastrous Year	
		LOWV	Students	LOWV	Students
1. Smoking	150,000	6,900	2,400	1.9	2.0
2. Alcoholic beverages	100,000	12,000	2,600	1.9	1.4
3. Motor vehicles	50,000	28,000	10,500	1.6	1.8
4. Handguns	17,000	3,000	1,900	1.6	2.0
5. Electric power	14,000	660	500	1.9	2.4
6. Motorcycles	3,000	1,600	1,600	1.8	1.6
7. Swimming	3,000	930	370	1.6	1.7
8. Surgery	2,800	2,500	900	1.5	1.6
9. X rays	2,300	90	40	2.7	1.6
10. Railroads	1,950	190	210	3.2	1.6
11. General (private) aviation	1,300	550	650	2.8	2.0
12. Large construction	1,000	400	370	2.1	1.4
13. Bicycles	1,000	910	420	1.8	1.4
14. Hunting	800	380	410	1.8	1.7
15. Home appliances	200	200	240	1.6	1.3
16. Fire fighting	195	220	390	2.3	2.2
17. Police work	160	460	390	2.1	1.9
18. Contraceptives	150	180	120	2.1	1.4
19. Commercial aviation	130	280	650	3.0	1.8
20. Nuclear power	100 ^a	20	27	107.1	87.6
21. Mountain climbing	30	50	70	1.9	1.4
22. Power mowers	24	40	33	1.6	1.3
23. High school & college football	23	39	40	1.9	1.4
24. Skiing	18	55	72	1.9	1.6
25. Vaccinations	10	65	52	2.1	1.6
26. Food coloring	-- ^b	38	33	3.5	1.4
27. Food preservatives	-- ^b	61	63	3.9	1.7
28. Pesticides	-- ^b	140	84	9.3	2.4
29. Prescription antibiotics	-- ^b	160	290	2.3	1.6
30. Spray cans	-- ^b	56	38	3.7	2.4

^a Technical estimates for nuclear power included normal operations and accidents from all phases of the fuel cycle. These estimates were found to range between 16 and 600 annual fatalities. The geometric mean of these estimates was used here.

^b Estimates were unavailable.

and technology "how many times more deaths would occur if next year were particularly disastrous rather than average." The geometric means of these multipliers are shown in columns 4 and 5 of Table 2. For most activities, people saw little potential for disaster. The striking exception is nuclear power, with a mean disaster multiplier in the neighborhood of 100.

For any individual, an estimate of the expected number of fatalities in a disastrous year could be obtained by applying the disaster multiplier to the estimated fatalities for an average year. When this was done for nuclear power, almost 40% of the respondents expected more than 10,000 fatalities if next year were a disastrous year. More than 25% expected 100,000 or more fatalities. An additional study (Slovic, in press), in which people were asked to describe their mental images of the consequences of a nuclear accident, showed an expectation that a serious accident would likely result in hundreds of thousands, even millions, of immediate deaths. These extreme estimates can be contrasted with the Reactor Safety Study's conclusion that the maximum credible nuclear accident, coincident with the most unfavorable combination of weather and population density, would cause only 3,300 prompt fatalities (U.S. Nuclear Regulatory Commission 1975). Furthermore, that study estimated the odds against an accident of this magnitude occurring next year to be about 3,000,000 : 1.

Disaster potential seems to explain much of the discrepancy between the perceived risk and the annual fatality estimates for nuclear power. Yet, because disaster plays only a small role in most of the other activities and technologies, it provides only a partial explanation of the perceived risk data.

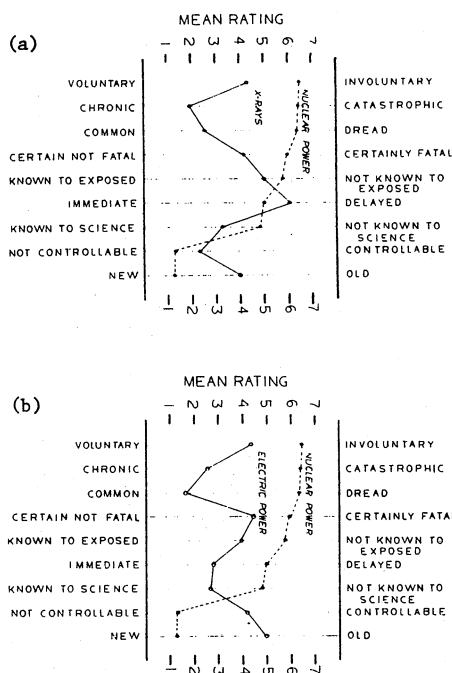
Qualitative characteristics. Are there other determinants of risk perceptions besides frequency estimates? We asked

experts, students, LOWV members and Active Club members to rate the 30 technologies and activities on nine qualitative characteristics that have been hypothesized to be important (Lowrance 1976).

The "risk profiles" made from mean ratings on these characteristics showed nuclear power to have the dubious distinction of scoring at or near the extreme on all of the characteristics associated with high risk. Its risks were seen as involuntary, delayed, unknown, uncontrollable, unfamiliar, potentially catastrophic, dreaded, and severe (certainly fatal). Figure 2 contrasts its unique risk profile with non-nuclear electric power and another radiation technology, X-rays, both of whose risks were judged to be much lower. Both electric power and X-rays were judged more voluntary, less catastrophic, less dreaded, and more familiar than nuclear power.

FIGURE 2

Rated characteristics of risk for nuclear power and related technologies.



Across all 30 items, ratings of dread and of the severity of consequences were closely related to lay judgments of risk. In fact, the risk judgments of the LOWV and student groups could be predicted almost perfectly from ratings of dread and severity, the subjective fatality estimates, and the disaster multipliers in Table 2. Experts' judgments of risk were not related to any of the nine qualitative risk characteristics.

Judged seriousness of death. In a further attempt to improve our understanding of perceived risk, we examined the hypothesis that some hazards are feared more than others because the deaths they produce are much worse than deaths from other activities. We thought, for example, that deaths from risks imposed involuntarily, from risks not under one's control, or from hazards that are particularly dreaded might be given greater weight in determining people's perceptions of risk.

However, when we asked students and LOWV members to judge the relative seriousness of a death from each of the 30 activities and technologies, the differences were slight. The most serious forms of death (from nuclear power and handguns) were judged only about 2 to 4 times worse than the least serious forms of death (from alcoholic beverages and smoking). Furthermore, across all 30 activities, judged seriousness of death was not closely related to perceived risk of death.

An Extended Study of Risk Perception

Our recent work extends these studies of risk perception to a broader set of hazards (90 instead of 30) and risk characteristics (18 instead of 9). Although data have thus far been collected only from college students, the results appear to provide further insights into the nature of risk perception. In addition, they suggest that some accepted views about the importance of the voluntary-involuntary distinction and the impact of catastrophic losses may need revision.

Design of the Study

For the extended study 90 hazards were selected to cover a very broad range of activities, substances, and technologies. To keep the rating task to a manageable size, some people judged only risks, others judged only benefits and others rated the hazards on five of the risk characteristics. Risks and benefits were rated on a 0-100 scale (from "not risky" to "extremely risky").

After rating the hazards with regard to risk, respondents were asked to rate the degree to which the present risk level would need to be adjusted to make the risk level acceptable to society. The instructions for this adjustment task read as follows:

The acceptable level of risk is not the ideal risk. Ideally, the risks should be zero. The acceptable level is a level that is "good enough," where "good enough" means you think that the advantages of increased safety are not worth the costs of reducing risk by restricting or otherwise altering the activity. For example, we can make drugs "safer" by restricting their potency; cars can be made safer, at a cost, by improving their construction or requiring regular safety inspection. We may, or may not, believe such restrictions are necessary.

If an activity's present level of risk is acceptable, no special action need be taken to increase its safety. If its riskiness is unacceptably high, serious action, such as legislation to restrict its practice, should be taken. On the other hand, there may be some activities or technologies that you believe are currently safer than the acceptable level of risk. For these activities, the risk of death could be higher than it is now before society would have to take serious action.

On the answer sheets, participants were provided with three columns labeled: (a) "Could be riskier: it would be acceptable if it were ___ times riskier;" (b) "it is presently acceptable;" and (c) "Too risky; to be acceptable: it would have to be ___ times safer."

The 18 risk characteristics included eight from the earlier study. The ninth characteristic from that study, controllability, was split into two separate characteristics representing control over the occurrence of a

mishap (preventability) and control over the consequences given that something did go wrong. The remaining characteristics were selected to represent additional concerns thought to be important by risk assessment researchers. As in the earlier study, all characteristics were rated on a bipolar 1-7 scale representing the extent to which the characteristic described the hazard. For example:

15. To what extent does pursuit of this activity, substance or technology have the potential to cause catastrophic death and destruction across the whole world?

very low catastrophic potential	1 2 3 4 5 6 7	very high catastrophic potential
---------------------------------------	---------------	--

Results

Risk characteristics. The mean ratings for the eighteen risk characteristics revealed a number of interesting findings. For example, the risks from most of these hazards were judged to be at least moderately well known to science (63 had mean ratings below 3 where 1 was labeled "known precisely"). Most risks were thought to be better known to science than to those who were exposed. The only risks for which those exposed were thought to be more knowledgeable than scientists were those from police work, marijuana, contraceptives (judged relatively unknown to both science and those exposed), boxing, skiing, hunting, and several other sporting activities.

Only 25 of the hazards were judged to be decreasing in riskiness; two of them (surgery and pregnancy/childbirth) were thought to be decreasing greatly. Risks from sixty-two hazards were judged to be increasing, thirteen of these markedly so. The risks from crime, warfare, nuclear weapons, terrorism, national defense, herbicides and nuclear power were judged to be increasing most. None of the hazards were judged to be easily reducible. The lowest of the 90 means on this characteristic was 3.2 (where 1 was labeled "easily reduced"); it was obtained for home appliances and roller coasters.

The ratings of the various risk characteristics tended to be rather highly intercorrelated, as shown in Table 3. For example, risks with catastrophic potential were also judged as quite dreaded ($r = .83$). Application of a statistical technique known as factor analysis showed that the pattern of intercorrelations could be represented by three underlying dimensions or factors. The nature of these factors can be seen in Table 3 in which the characteristics are ordered on the basis of the factor analysis. The first 12 characteristics represent the first factor; they correlate highly with one another and less highly with the remaining six characteristics. In other words, these data suggest that risks whose severity is believed not to be controllable tend also to be seen as dreaded, catastrophic, hard to prevent, fatal, inequitable, threatening to future generations, not easily reduced, increasing, involuntary, and threatening to the rater personally. The nature of these characteristics suggests that this factor be called "Dread." The second factor primarily reflects five characteristics that correlate relatively highly with one another and less highly with other characteristics. They are: observability, knowledge, immediacy of consequences, and familiarity (see Table 3). We have labeled this factor "Familiarity." The third factor is dominated by a single characteristic, the number of people exposed. This characteristic can be seen in Table 3 to be relatively independent of the other characteristics.

Just as each of the 90 hazards has a mean score on each of the 18 risk characteristics, it also has a score for each hazard on each factor. These scores give the location of

TABLE 3
Intercorrelations Among 18 Risk Characteristics
in the Extended Study

	Severity not controllable	Dread	Globally catastrophic	Little preventive control	Certain to be fatal	Risks & benefits inequitable	Catastrophic	Threatens future generations	Not easily reduced	Risks increasing	Involuntary	Affects me personally	Not observable	Unknown to those exposed	Effects immediate	New (unfamiliar)	Unknown to science	Many people exposed	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1. Severity not controllable																			
2. Dread	.82																		
3. Globally catastrophic	.78	.83																	
4. Little preventive control	.86	.72	.71																
5. Certain to be fatal	.80	.82	.73	.77															
6. Risks & benefits inequitable	.75	.76	.84	.63	.65														
7. Catastrophic	.77	.66	.77	.74	.67	.77													
8. Threatens future generations	.62	.76	.86	.52	.64	.81	.63												
9. Not easily reduced	.59	.67	.64	.67	.66	.59	.56	.59											
10. Risks increasing	.51	.63	.76	.48	.58	.76	.57	.75	.60										
11. Involuntary	.68	.58	.69	.52	.42	.77	.74	.62	.33	.44									
12. Affects me personally	.57	.64	.77	.43	.50	.71	.64	.78	.35	.67	.61								
13. Not observable	-.04	-.14	.04	-.19	-.28	.08	.00	.24	-.15	-.10	.33	.10							
14. Unknown to those exposed	.14	.05	.22	-.05	-.12	.28	.24	.35	-.20	.05	.63	.31	.79						
15. Effects immediate	.21	.15	.00	.34	.36	-.08	.11	-.29	.22	.01	-.25	-.14	-.87	-.77					
16. New (unfamiliar)	.32	.29	.32	.22	.20	.25	.18	.44	.13	.06	.36	.20	.63	.56	-.48				
17. Unknown to science	-.10	-.10	-.04	-.07	-.15	-.02	-.04	.02	-.23	.02	.13	.00	.43	.50	-.44	.32			
18. Many people exposed	-.04	.04	.23	-.14	-.11	.32	.23	.47	.03	.26	.34	.56	.37	.46	-.52	.07	-.01		

each hazard within the factor space. Figure 3 plots the hazards on Factors 1 and 2. Items at the high end of Factor 1 are all highly dreaded. Items at the negative end of Factor 1 are seen as posing risks to individuals and being injurious rather than fatal. The placement of items on the vertical dimension, Factor 2, intuitively fits the theme of familiarity and observability associated with the dimension label. Hazards lying at the extremes on Factor 2 (number exposed) are shown in Table 4.

This three-dimensional factor structure is of interest because it differs considerably from the two-dimensional structure obtained from ratings of 30 hazards on 9 characteristics (Fischhoff 1978). That structure, in which Factor 1 was labeled "severe" (i.e., certain to be fatal) and Factor 2 was labeled "high technology," had been found to be remarkably consistent across four different groups of lay and expert respondents (Slovic 1980(a)). The present results indicate that the particular set of hazards and the particular set of risk characteristics under study can have an important effect on the nature of the observed "dimensions of risk."

One point of commonality between the present analysis and the previous one is that nuclear power is an isolate in both. Although activities such as crime, nerve gas, warfare and terrorism are seen as similarly dreaded (Factor 1), none of these is judged as new or as unknown (Factor 2) as nuclear power.

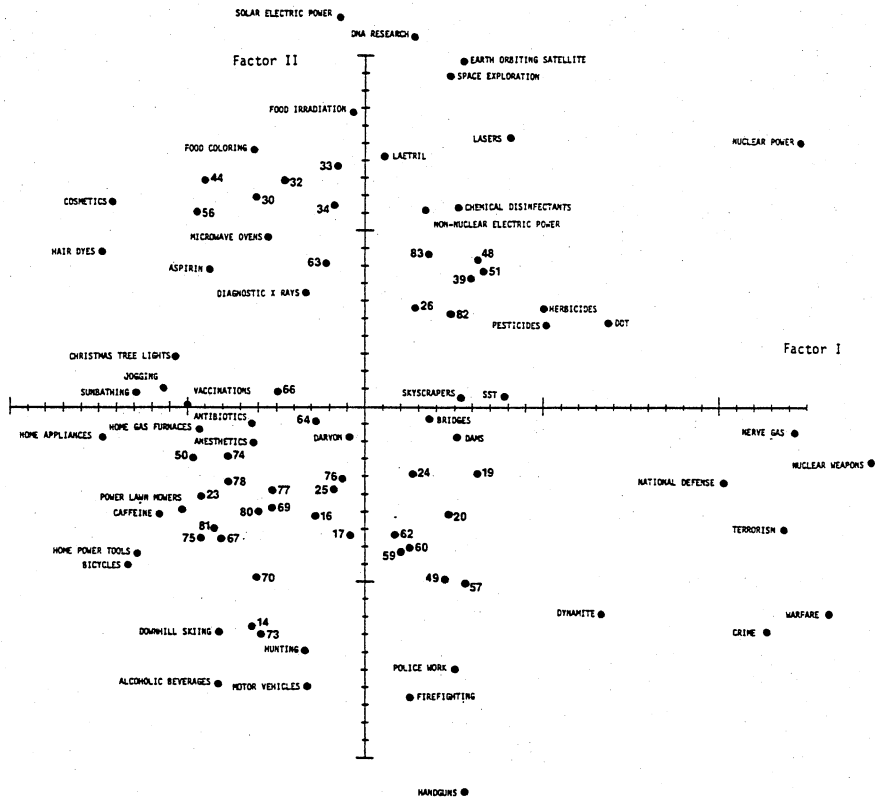
TABLE 4
Hazards With Most Extreme Scores on Factor 3^a

Highest Exposure	Factor Score
Alcoholic Beverages	2.9
Caffeine	2.0
Smoking	1.8
Food Preservatives	1.8
DDT	1.6
Herbicides	1.5
Motor Vehicles	1.5
Food Irradiation	1.4
Pesticides	1.3
Sunbathing	1.2
:	:
Nerve Gas	-1.2
SST	-1.0
Surfing	-1.4
Laetril	-1.5
Boxing	-1.5
Roller Coasters	-1.6
Scuba Diving	-1.8
Open-heart Surgery	-1.8
Lasers	-1.8
Space Exploration	-2.0
Solar Electricity	-2.0
Lowest Exposure	

^a Number of people exposed to the hazard.

FIGURE 3

Factors 1 and 2 of the three-dimensional structure derived from interrelationships among 18 risk characteristics in the extended study. Factor 3 (not shown) reflects the number of people exposed to the hazard.



Conclusions

Although research into the nature of perceived risk is still incomplete, we offer the following tentative conclusions:

1. Perceived risk is quantifiable and predictable.
2. Groups of laypeople sometimes differ systematically in their perceptions. Experts and lay persons also differ, particularly with regard to the probability and consequences of catastrophic accidents.
3. The degree of adjustment judged necessary to make risk levels "acceptable" is strongly determined by the perceived level of current risk; the greater the perceived risk, the greater the desired reduction. Perceived benefit plays a secondary role; all else being equal, somewhat less reduction in risk is deemed necessary to make highly beneficial activities "acceptable."
4. Many of the eighteen characteristics of risk hypothesized to be important to the public do correlate highly with perceived risk and desire for risk reduction. Certain clusters of characteristics are highly inter-related across hazards indicating that they can be combined into higher-order characteristics or "factors." Three factors, labeled Dread, Familiarity, and Exposure, seem able to account for most of the interrelations among the eighteen characteristics.

References

- Fischhoff, Baruch, Slovic, P., Lichtenstein, S., Read, S. and Combs, B. (1978), "How Safe is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits," *Policy Sciences*, 8, 127-152.
- Lorraine, W. (1976), *Of Acceptable Risk: Science and the Determination of Safety*, Los Altos, California: William Kaufmann Co.
- Slovic, Paul, Fischhoff, B. and Lichtenstein, S. (1980(a)), *Expressed Preferences*, Eugene, Oregon: Decision Research Report 80-1.
- _____, Fischhoff, Baruch and Lichtenstein, Sara (1980(b)), "Perceived Risk," *Societal Risk Assessment: How Safe is Safe Enough?*, R.C. Schwing and W.A. Albers, Jr., (eds.), New York: Plenum Press.
- _____, Lichtenstein, S. and Fischhoff, B. (in press), "Images of Disaster: Perception and Acceptance of Risks from Nuclear Power," *Energy Risk Management*, G. Goodman and W.D. Rowe (eds.), London: Academic Press.
- U.S. Nuclear Regulatory Commission (1975), *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*, WASH 1400 (NUREG-75/014), Washington, D.C.

CONSUMER BEHAVIOR AND ACCEPTABLE RISK:
SOME RESEARCH ISSUES*

M. Venkatesan, University of Oregon

Abstract

It is clear that the growing risk assessment literature is very relevant to the problems posed in the hazard management area for consumer products and services. However, there are some research issues that must be investigated now so that these researchers can proceed with a programmatic research emphasis. This paper has identified five such research issues.

Introduction

Product safety as an issue has reappeared vigorously with the recent disclosures and linking of "toxic shock syndrome" to the use of a certain brand of tampon. While the number of deaths reported is considerably smaller than a day's carnage on the highways, that brand of tampon is already off the market. The safety question relating to the use of saccharin is heating up. The recent legislation specifying minimum standards for the infant formula points up to another product area where lack of safety has produced tragic results, though not in vast numbers.

The recent Titan II mishap amply demonstrates that no technology is foolproof. Thus, technological risks are inherent in products developed for consumption by consumers. The right to safety has been made very explicit in the 1960s and the Delaney Amendments, the Product Safety Commission, and the like are the result of the prominence of the safety issue. Consumers are not unaware that perfect safety is impossible, and they are also cognizant of the risk-benefits relationships. Engel et al. (1978) pointed out that one of the important roles of consumer research is in establishing the level of safety that must be considered in decisions concerning the products.

As Rethans (1980) has pointed out, the assessment and management of hazards arising from technology have been of major concern to general public and to a number of governmental agencies. In order to help society cope with the technological risk, risk assessments have been undertaken in many areas. For example, the U.S. Nuclear Regulatory Commission made an assessment of the risk associated with nuclear power in their Reactor Safety Study. A number of other studies relate to hazards emanating from chemicals, other new technologies, etc. In the consumer arena, risk assessments for products are generally not made public. The only major visible effort is by the Product Safety Commission, which is concerned with technical facts of hazard assessment.

It appears from the newly emerging risk assessment literature that risk assessment has two components: one dealing with risk estimation and the other dealing with risk evaluation. As Rethans (1980) points out, the social facet of risk evaluation pertains to the importance attached to these hazards by society, and thus risk evaluation inevitably involves the judging of the "acceptability of the risk."

While some notions of perceived risk included the "physical risk" inherent in the product, no empirical work had been

undertaken to determine the "acceptable level of risk" with respect to products utilizing the "perceived risk" concept. The first study which has looked at the hazard management area for the products is by Rethans (1980). Since his paper has dealt with his study extensively, I will not detail his study here. Suffice to point out that his study makes a contribution to the important area of risk evaluation by consumers, its determinants and how to deal with product risks. The studies that are planned by Jacoby and his associates as part of their NSF-funded studies, are aimed at answering the question: "Do consumers actually consider health and safety information when making purchase, usage, and disposal decisions regarding technologically innovative products?" Thus, very little empirical evidence is currently available to make policy recommendations, nor is there enough evidence to make suggestions that will aid management decisions with respect to management of hazards relating to products. Therefore, this paper attempts to point out the relevance of this area to researchers in the consumer behavior area and to provide some guidelines regarding some of the issues that are emerging in this area.

Approaches to Acceptable Risk

It would appear that the greatest relevance is in the area of investigating the determinants of risk perception by consumers so as to aid in the evaluation of which hazards, and in what form, the consumers are willing to tolerate. Since there is no such thing as a completely risk-free product, consumers may be said to have an "acceptable level of risk." Some (Fischhoff, et al. 1980) contend that the term acceptable level of risk is a misnomer inasmuch as what we have at best is "that level of risk accepted according to a particular version of a particular kind of rationality" (p. 9). There are five approaches which appear to be currently utilized to determine the "acceptable level." They are: (1) formal analysis, (2) nonintervention, (3) procedural approaches, (4) professional standards, and (5) comparative approaches.

In the formal analysis approach, such familiar analyses as cost-benefit analysis, risk-benefit analysis, decision analysis and value-impact assessment are included. The nonintervention approach is characterized by the free market which sets prices on alternative hazard management option, and individuals choose according to price and personal definition of acceptance risk. In other words, this is an extreme position, as this approach opposes any intervention into dealings between buyers and sellers, which is conducted under the protection of a watchful and responsive legal system.

The common approach that is currently in vogue is the procedural approach. This approach relies on the political and economic forces molding the standards, by their interaction with one another and the varied feedback provided by their environment. The professional standards are provided by the technical experts most knowledgeable in a field. Finally, the comparative approaches look at the way risks have been balanced in the past or balanced at present, and they should guide the risk management decisions for the present and for future action.

As is evident, some of these approaches require some centralized decision-making body. Each approach has its pros and cons, and its limitations. None of these

* Portions of this paper are based on Fischhoff (1980) preliminary draft. Therefore, please do not quote from this paper.

approaches incorporate the subjective approach or the psychological elements of the risk assessment process that are critical in the evaluation and management of hazards relating to consumer products. Slovic et al. (1980(b)) paper examines the nature of perception of risks by individuals, which has direct relevance to researchers in consumer product safety area. They advocate a method of "expressed preferences" to get directly to the problem of perception of risk and in my judgment such a method is well suited to the study of acceptable risk for consumer products.

Slovic's (1980(b)) study attempted to ascertain what people meant when they said that a particular technology was risky. They found that the risk judgments of lay people related only moderately to the annual death rates and the risk perceptions were not closely related to their own fatality estimates. In other words, their study indicated that risk perception did not mean annual fatalities or frequency of deaths. It revealed that individual's perception of "disaster potential" played an important role in arriving at risk judgments. In addition, an array of qualitative considerations were associated with risk perception by individuals. The characteristics that are associated with risk perceptions in their study are: (1) voluntariness of risk, (2) immediacy of effect, (3) knowledge about risk, (4) to what extent are the risks known to science, (5) control over risk, (6) newness/familiar risk, (7) chronic or catastrophic nature of the risk, (8) common hazard or a dread hazard, and (9) severity of consequences.

Some Issues to Consider

Much of the individual decisions regarding choice of products relate to voluntariness of the risk. Where solutions are mandated, as in the example of air bags for automobiles, or standards of food additives, etc. such situations characterize the involuntariness of the risks on the part of consumers. While the Slovic (1980(b)) study did not find much difference in the evaluation of risk with respect to voluntariness, one would expect that in the context of consumer behavior, this consideration might play an important role. This is a major area which needs investigation by consumer researchers before proceeding to investigate other characteristics. The rationale is provided by Lorance (1976) who asserts that what is acceptable (by consumers) definitely depends on the degree to which we, as consumers, are free to opt for or decline the risk. Where individual's options are limited, such as the example where a governmental agency determines the acceptable level of risk and preempts the individual's voluntary decision-making, the involuntariness might greatly affect the perception of risk and its acceptable levels.

The second aspect which needs the attention of consumer researchers is the perception of "disaster potential." For example, in the recent incidences of toxic shock syndrome, the ailment rate is three in 100,000. However, preliminary surveys indicate that consumers may not have been concerned with average fatality per year (as in the Slovic study), but the potential catastrophic effect such products may have on the health and safety of consumers. Experts point out that not only for this product but for a variety of consumer products, testing before introduction of such products would not have been able to point out the possible hazard. That is, hazard information is unlikely to emerge before the products are marketed. This may also point up to the desirability of post-marketing surveillance in order to aid risk evaluation.

A related problem in the context of consumers is that they must rely on inferences based on what they remember hearing or observing about the hazards. Slovic (1979) points out that one heuristic that has special relevance for risk

perception is called "availability." Thus consumers can imagine or recall instances that occur frequently and numerous factors (such as recent incidences) affecting availability might distort their judgments.

Another factor that relates to consumers is the "overconfidence" factor. Many issues relating to products result in injuries rather than being fatal. In addition, their own experience as consumers without facing any hazards tend to strengthen their views that the hazards cannot be serious, particularly when an individual consumer's own hazard experience with respect to any of the risks associated with a product may be overwhelmingly benign. Thus, such overconfidence might affect their risk perception and the level of acceptable risk for a variety of products.

To aid consumers in risk perceptions, more and more warning labels are advocated for products with the hope that such information may be utilized by the consumer both for making purchase and use decisions. However, as Slovic et al. (1980(a)) point out, merely mentioning possible adverse consequences (no matter how rare) could enhance their perceived likelihood and make them appear more frightening. A recent study on a hypothetical Patient Package Insert by Keown (1980) lends some support to this notion. This is an area where considerable research is needed to help public policy makers and governmental agencies charged with investigating and recommending warning labels to products.

Another problem peculiar to consumer research may be the perception by a large number of consumers that risks associated with products are managed by competent professionals (both in the business organization and in the governmental agencies) that they need not be concerned about them. In addition, lay people seem to believe that our technological progress will solve the hazard problems. Such complacency on the part of consumers might well affect the efficiency of educational programs and the use of labels provided with the products in aiding consumer decision-making.

There are a number of other issues that should receive the attention of consumer researchers concerned with the perception of risk by consumers and how they handle such risks. Suffice to point out that the new discipline of "risk assessment" is very much relevant to researchers in the consumer behavior area, and we have a long way to go before we will be in a position to make recommendations for policy-makers.

References

- Engel, James F. et al. (1978), Consumer Behavior, Hinsdale, Ill.: The Dryden Press (3rd Edition).
- Fischhoff, Baruch, Slovic, Paul, Lichtenstein, Sarah (1978) "How Safe is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits," Policy Sciences, 9, 127-152.
- Fischhoff, Baruch, Slovic, Paul and Lichtenstein, Sarah, (1979), "Weighing the Risks. Which Risks are Acceptable?" Environment, 21, 4, 17-38.
- Fischhoff, Baruch, Lichtenstein, Sarah, Slovic, Paul, Derby, Stephen L., and Keeney, Ralph L. (1980), Approaches to Acceptable Risk (Preliminary Draft).
- Keown, Charles Jr. (1980), Attitudes of Laypeople, Physicians and Pharmacists Towards Prescription Drug Side Effects, Doctoral Dissertation, University of Oregon, College of Business Administration.

Lowrance, W. (1976), Of Acceptable Risk, Los Altos, California: William Kaufman.

Rethans, Arno (1979), An Investigation of Consumer Perceptions of Product Hazards, Doctoral Dissertation, University of Oregon, College of Business Administration.

Slovic, Paul, Fischhoff, Baruch, and Lichtenstein, Sarah (1979), "Rating the Risks," Environment, 21, 3, 14-39.

Slovic, Paul, Fischhoff, Baruch, and Lichtenstein, Sarah (1980(a)), "Informing People About Risk," Paper Presented at the Conference on Product Labeling and Health Risks, Banbury Center, Cold Spring Harbor, New York, May 21-23.

Slovic, Paul, Fischhoff, Baruch, and Lichtenstein, Sarah (1980(b)), "Perceived Risk," in R. C. Schwing and W. A. Albers, Jr., (Eds.) Societal Risk Assessment: How Safe is Safe Enough? New York: Plenum Press.

TOWARDS DETERMINANTS OF ACCEPTABLE RISK: THE CASE OF PRODUCT RISKS¹

Arno J. Rethans, The Pennsylvania State University
Gerald S. Albaum, University of Oregon

Abstract

Discussions of product safety should distinguish between risk estimation and risk evaluation. Whereas the former is an essentially scientific and objective activity, the latter involves normative and relativistic judgments on risk acceptability. The acceptability of consumer product risks is explored by examining the relevance of a risk acceptability model emerging from the marketing, product liability and risk assessment literature. The model is found to be deserving of attention by the Consumer Product Safety Commission.

Introduction

The assessment of risks associated with consumer products involves both risk estimation and risk evaluation. Risk estimation may be thought of as the identification of (product) hazards and the measurement of their probability and consequences (Slovic, Fischhoff and Lichtenstein 1977). Risk evaluation, on the other hand, is the complex process of anticipating the societal response to risk; this could be termed the "acceptability of risk" (Otway 1975; Rowe 1977).

This distinction between the two major dimensions of risk assessment becomes important in discussions of product safety. Indeed, recognition of the dimensions of measurement and evaluation allows us to clarify the term product safety. A product is deemed safe when its attendant risks are judged to be acceptable (Lowrance 1976). This conceptualization of safety differs sharply from the simplistic dictionary definitions which consider "safe" to mean "free of risk." A product can never be absolutely free of risk. Even if technically approachable, its accompanying price tag would be prohibitive.

More importantly, the distinction between the two facets of risk assessment provides a vehicle to characterize the activities of the Consumer Product Safety Commission (CPSC). The identification of products which pose risks to the American public constitutes risk estimation. This activity is essentially empirical in nature and is in part conducted through the maintenance of the National Electronic Injury Surveillance System (NEISS). Deciding whether the public might be or should be willing to bear the estimated risks constitutes risk evaluation.

The evaluation of risk, then, is more of a normative and political activity. It involves the judging of the acceptability of the product risks. Employment of the term "acceptability" emphasizes the fact that risk management decisions are relativistic and judgmental. Indeed, the term elicits some interesting and crucial questions: "Acceptable to whom?", "Acceptable in what terms", and "What are the guides to acceptability?"

The study reported here was designed to begin to provide some empirical information on the basis of which some of the above questions may be answered. Its underlying premise is that public attitudes and values must be reflected

in the Commission's decisions on what constitutes acceptable levels of risk associated with consumer products.

Conceptual Framework For the Study

The notion of acceptability is quite pervasive, even though it is not always given explicit emphasis. A review of the appropriate literature in the areas of marketing, hazard assessment and product liability revealed several guides to judge acceptability. In this section we briefly outline these guides and their origin.

Marketing²

The CPSC is charged with the responsibility of protecting the public against unreasonable risk of injury. Unfortunately, the Product Safety Act left it to the Commission to decide what "unreasonable risk" is and how to reduce it. The National Commission on Product Safety gave the Commission some guidelines by adopting a statement by Corwin D. Edwards:

"Risks of bodily harm to users are not unreasonable when consumers understand that risks exist, can appraise their probability and severity, know how to cope with them and voluntarily accept them to get benefits that could not be obtained in less risky ways. When there is risk of this character, consumers have reasonable opportunity to protect themselves; and public authorities should hesitate to substitute their value judgments about the desirability of the risk for those of the consumers who choose to incur it." (Final Report 1970)

This statement identifies a number of product risk characteristics which are believed to influence acceptability: knowledge of risk, ability to control the risk, and voluntariness of the risk.

Product Liability Litigation

Insight into the acceptability of risks may also be gained from an examination of product liability laws. A recent report by an Interagency Task Force of Product Liability (1978) noted that to determine whether a product is designed defectively (a plaintiff's argument), risks presented by the product must be weighed against the product's utility. Hence, a product which presents a substantial risk is not necessarily defective in the eyes of the law, since it may also have "great utility."

In balancing the risks and "utility" of the product, a list of factors has been proposed by Wade (1973) and has been adopted by courts in Oregon, Arizona, and Pennsylvania (Taskforce Final Report 1978). The factors to be considered are: usefulness and desirability of product, likelihood of injury, seriousness of injury, availability of substitutes,

²In the area of marketing there has been a substantial amount of empirical research on how consumers perceive the risks associated with choice situations (Ross 1975). This research conceptualizes perceived risk as a dual component, multifaceted phenomenon including the dimension of physical risk. The literature does not however address the notion of acceptability of the risk. One possible exception may be Bettman (1973) when he distinguishes between inherent and handled risk.

¹The authors would like to acknowledge the intellectual and financial support provided by Dr. Paul Slovic and the Decision Research, Inc., Eugene, OR, in the conduct of this study.

manufacturer's ability to eliminate unsafe character, user's ability to avoid danger and user's anticipated awareness of danger.

Risk Assessment Literature

The urgent need to help society cope with risks has produced a new discipline: Risk Assessment. When performing a risk assessment, the aim is to determine the magnitude and scope of a hazard and whether society should be exposed to it (Slovic and Fischhoff 1979). Researchers in this area have hypothesized and/or identified a number of variables which influence the acceptability of risks.

Chauncey Starr (1969) employing a method of revealed preferences found a number of variables to influence risk acceptability: voluntariness of risks, benefits accruing to society and the number of people involved in the risks. He concluded that the public is willing to accept "voluntary" risks roughly 1,000 times greater than "involuntary" risks. Furthermore, it appeared that the acceptability of risk is roughly proportional to the third power of the benefits. Finally, his data indicated that the social acceptance of risk was directly influenced by public awareness of the benefits of an activity, as determined by advertising and usefulness, and the number of people participating.

Starr's initial work stimulated further work, notably by researchers at the International Institute for Applied Systems in Vienna by research associates at Decision Research, A Branch of Perceptronics, Inc. and the Clark University Center for Technology, Environment and Development. Much of the work by these researchers centers around the question of how people perceive and respond to risks. The assumption underlying this research is that individual and societal acceptance of risk, in general, is predicated upon how risks are perceived, not necessarily by actual levels or risk (Otway and Pahner 1976; Slovic, Fischhoff and Lichtenstein 1979). Indeed, it has been found that people's perceptions of risks are only moderately related to actual levels of risk and qualitative risk factors are now being examined as possible explanatory variables for perceptions of risk and risk acceptability.

A review of the work by these researchers as well as the work by Lowrance (1976) and Rowe (1977) yields a number of risk characteristics which are thought to be responsible for judgments of risk acceptability: voluntariness of risk, degree of control over risk, knowledge of risk, potential for misuse, degree of risk to special groups, the benefit/risk ratio, ease of hazard reduction, extent of societal exposure, severity of consequences, and necessity of exposure.

The Emerging Model

The literature reviewed above suggests that perceptions of risk and acceptability of risks are determined by both the quantitative and qualitative characteristics of those risks. Specifically, acceptability perceptions are hypothesized to be negatively related to such "objectively" determinable characteristics as injury frequency and injury severity. Furthermore, a set of qualitative risk characteristics are thought to be perception influences. This set of characteristics and their hypothesized relationships with acceptability perceptions include: voluntariness (+), risk knowledge (+), risk control (+), product necessity (+), foreseeability (+), exposure (+), ease of risk reduction (-), user error (-), and risk to children (-).

Objectives of the Study

The study reported in this paper represents a first attempt to explore consumer perceptions of the acceptability of the hazards associated with consumer products. The general objective of exploration translates into two research ob-

jectives: (a) determine consumers subjective perceptions of product risks and (b) examine the relevance of the emerging model of risk acceptance to the judgments of the acceptability of product risks.

Methodology of the Study

To obtain the information necessary to accomplish the stated objectives, participants in the study were asked to complete a questionnaire. Participants were asked to rate a set of consumer products on a series of seven point scales, which included societal hazard and acceptability scales as well as scales representing the risk characteristics identified in the literature review described earlier. The scales are typologically modeled after the perceived risk studies in marketing (Ross 1975) and contentwise after studies by Green and Brown (1978).

A pretest of the questionnaire suggested that the task needed to be shortened in order to secure respondent cooperation as well as to avoid respondents' fatigue and its accompanying errors. Therefore, it was decided to run a split-run technique on the risk characteristic scales. That is, two versions of the research instrument (A and B) were administered. Each version contained the societal hazard and acceptability scales and six of the twelve risk characteristics.

The products which served as stimuli in the study were included on the basis of a number of criteria. First, the set of products was to reflect the frequency range found in the CPSC data tabulation. Second, the products selected must, at the same time, reflect the severity range of the NEISS data. Third, the product definition must be unequivocal. Many of the "definitions" in the NEISS tabulations do not constitute a succinct product, but rather a product category. Finally, products to be included in the questionnaire were to be selected so as to maximize the probability of product familiarity among respondents.

Application of these criteria led to the identification of the products shown in Table 1. The categorization of the severity of injuries into low, moderate, and high follows that of the CPSC and is based on the mean severity values of the products. A mean severity of 17 or less generally indicates low severity; the range 18 to 81 usually implies moderate severity; and a mean severity over 81 implies high severity. The severity values for the products of this study are shown in parentheses in Table 1.

TABLE 1
Product Stimuli for Acceptance Study

Severity of Injury	Number of Injuries		
	Less than 10,000 per year	10,000 - 60,000 per year	More than 60,000 per year
Low	Dehumidifiers (17) Power can openers (12) Sewing machines (17)	Hammers (14) Razors/shavers (14) Scissors (12)	Knives (14) Snow skis (17)
Moderate	Xmas lights (49) Hair dryers (43)	Cosmetics (32) Garage doors (31) Playground Equip. (37) Refrigerators (22) Television sets (32) Tricycles (27)	Bathtub/shower (62) Bicycles (40) Ladders (39) Skateboards (32) Stairs/steps (31)
High	Fireworks (158) Ammonia (152) Lighters (114) Portable saws (181) Sun lamps (125) Toy balloons (126)	Swimming pools (105)	Power lawn mowers (125)

Thirty five questionnaires of form A and forty questionnaires of form B were completed by members of three civic groups in a medium sized town on the West Coast. The questionnaires were hand-delivered to the groups and returned by mail in the case of one group and returned by a group representative in the case of the other two groups. In this way, seventy-five completed and usable questionnaires were collected. A donation in the amount of five dollars was made to the organizations' treasury of each completed

questionnaire.

The overall sample consisted of 52% males and 48% females. Their ages ranged from 20 to 52 years of age with a mean age of 30.5 and a modal age of 32. Sixty-five percent of the respondents were married and 24% of them were single. With regard to education, about 50% of the respondents completed a university degree. An additional twenty-five percent noted that they had completed some university education, and 12% completed high school and/or vocational school or less. Occupationally, the groups may be characterized as predominantly white collar rather than professional. Only seven percent were blue collar workers.

Findings of the Study

Risk Acceptability Ratings

Respondents were asked to judge whether or not risks associated with each of the twenty nine products were at a socially acceptable level. The mean scores of the products on the 7-point acceptability scale are shown in the first column of Table 2. This bipolar scale has been coded such that the higher the score, the greater the acceptability of the product risks. On an absolute level, the majority of the products are considered to bear risks which are perceived to be acceptable (i.e., greater than 4.0). The most acceptable risks are those associated with hammers, sewing machines, garage doors, refrigerators, and dehumidifiers. The lesser acceptable products are fireworks, skateboards, sunlamps, power lawn mowers and cosmetics.

TABLE 2

Product Risk Acceptability and Hazard Ratings

Product	Acceptability Mean Score	Rank	Rank on Societal Hazard
Hammers	6.29	1	21
Sewing Machines	6.08	2	25
Garage doors	5.99	3	27
Refrigerators	5.87	4	24
Dehumidifiers	5.83	5	29
Stairs/steps	5.75	6	12
Scissors	5.57	7	19
Bath/shower struct.	5.56	8	16
Tricycles	5.55	9	22
Bicycles	5.52	10	11
Snow skis	5.47	11	8
Knives	5.44	12	6
Power can openers	5.44	13	26
Toy balloons	5.41	14	28
Ladders	5.37	15	10
Television sets	5.35	16	20
Swimming pools	5.31	17	5
Razor/shavers	5.28	18	17
Playground Equip.	5.07	19	14
X-mas lights	4.99	20	13
Ammonia	4.98	21	9
Lighters	4.96	22	15
Hair dryers	4.77	23	18
Power saws	4.73	24	2
Cosmetics	4.68	25	23
Power mowers	4.52	26	4
Sunlamps	4.04	27	7
Skateboards	3.88	28	3
Fireworks	3.42	29	1

The third column of Table 2 shows the participants' ranking of the hazardousness of the products included in the study. Comparing the hazardousness rankings with the acceptability rankings suggests the presence of an inverse relationship between risk and acceptability. Indeed, when computing a rank correlation between these two variables, we find a coefficient of $-.62$. Reviewing the major rank discrepancies suggests that risk acceptability is not only influenced by perceived risk, but by other considerations as well. Swimming pools, knives, and snow skis from cases in point. Each of these products were considered to be quite hazardous from a societal perspective, yet were considered to be quite acceptable. In the case of swimming pools and snow skis, this result may be explained by the high ratings of these products along the voluntariness and foreseeability characteristic. The acceptability of the risks associated with knives is probably due to its high perceived necessity and the high degree of "knowability" of the risk.

Risk Characteristics

Participants were also asked to rate each of the products on twelve, seven point scales, each of which represented a characteristic of risk. The mean ratings on these characteristics for the more hazardous and the less hazardous products are shown in Tables 3 and 4.

TABLE 3: Risk Characteristics of More Hazardous Products

Qualitative Characteristic	Fire-works	Portable power saws	Skateboards	Power mowers	Swimming pools	Average
V1. Voluntariness	5.31	5.20	5.94	4.74	5.83	5.41
V2. Risk knowledge	4.21	4.83	4.74	4.80	5.00	4.72
V3. Risk control	3.97	4.97	4.20	5.11	5.20	4.69
V4. Severity of injury	5.37	5.63	4.89	5.54	5.83	5.45
V5. Personal exposure	2.63	3.03	1.69	3.00	2.34	2.54
V6. Ease of reduction	2.47	3.35	2.74	3.68	2.94	3.04
V7. Product necessity	2.13	4.53	1.88	4.00	3.83	3.27
V8. User error	5.83	4.92	6.13	5.23	4.78	5.38
V9. Risk to children	6.23	4.40	6.35	5.08	5.92	5.60
V10. Injury frequency	2.68	2.90	3.34	2.92	3.24	3.02
V11. Foreseeability	5.73	5.38	5.58	5.13	5.10	5.38
V12. Degree of exposure	3.98	3.83	3.90	4.98	4.50	4.23

TABLE 4: Risk Characteristics of Less Hazardous Products

Qualitative Characteristic	Sewing machines	Power can openers	Garage doors	Toy balloons	Dehumidifiers	Average
V1. Voluntariness	6.00	5.43	5.11	5.63	5.00	5.43
V2. Risk knowledge	4.66	4.29	4.31	4.14	2.97	4.07
V3. Risk control	6.31	5.49	5.57	5.91	4.61	5.58
V4. Severity of injury	2.97	3.09	3.66	2.97	2.97	3.13
V5. Personal exposure	1.37	1.51	1.66	1.14	1.26	1.39
V6. Ease of reduction	3.63	4.06	3.57	3.47	3.91	3.73
V7. Product necessity	5.85	2.08	4.65	2.40	3.50	3.69
V8. User error	3.38	3.40	3.18	3.45	3.18	3.31
V9. Risk to children	3.98	3.68	4.03	5.31	3.58	4.11
V10. Injury frequency	1.22	1.35	1.35	1.40	1.16	1.29
V11. Foreseeability	3.38	2.75	3.25	2.50	1.83	2.74
V12. Degree of exposure	5.25	4.38	5.80	4.68	2.83	4.59

An examination of individual product profiles provides insight into consumers' risk and acceptability perceptions. For example, the ratings for fireworks suggest that consumers feel that risks associated with this product are perceived as rather voluntary, somewhat known to the user, and highly foreseeable. At the same time, there is a high perceived likelihood of user error accompanied by major injuries if accidents do occur. In addition, children are perceived to be rather vulnerable to the product risks. Similar profiles emerge for the other more hazardous products such as skateboards, portable power saws, and power lawn mowers.

In contrast, when examining dehumidifiers, the least hazardous product, a rather different profile is found. Although consumers admit to relatively low levels of knowledge and consider the risk somewhat unforeseeable, the perceived minor injuries associated with the product and the low perceived probability of user error seem to make this product

relatively safe and acceptable.

Similar analyses may also be made at a group level. The data in the tables allow the reader to compare "group profiles" of more hazardous versus less hazardous products. The two groups differ substantially on the characteristics of risk control, injury severity, likelihood of user error, risk to children, frequency of injury and foreseeability.

Dependence Structure

The findings discussed in the preceding section suggest that a relationship exists between perceived and acceptable risk on the one hand, and the qualitative risk characteristics on the other. This section explores this relationship by means of a correlational analysis.

In Table 5, the product moment correlations between the mean risk characteristics scores and the mean risk and acceptability ratings are shown. The scores on the risk characteristics have been coded in a positive direction to facilitate interpretation of the data. The hazard and acceptability ratings correlated significantly ($p < .01$) with the characteristics of risk knowledge, risk control, injury severity, product necessity, user error, risk to children and injury frequency.

TABLE 5: Correlation Between Risk Characteristics And Hazard/Acceptability Ratings

	Personal Hazard	Societal Hazard	Acceptability
Voluntariness	.02	.06	-.08
Risk knowledge	.29	.30	.29
Risk control	-.27	-.38	.58
Injury severity	.65	.75	-.47
Personal exposure	.52	.50	.04
Ease of risk reduction	-.49	-.50	.03
Product necessity	-.16	-.18	.68
User error	.83	.86	-.66
Risk to children	.40	.52	-.25
Injury frequency	.70	.78	-.31
Foreseeability	.79	.83	-.36
Societal exposure	-.26	-.23	.44

The majority of the observed correlations conform to the direction of the influence hypothesized in the risk assessment literature described earlier. For example, the higher the likelihood of error, the less acceptable the product is considered. Similarly, the higher the controllability of the product risk the more acceptable the product was.

Yet, not all correlations between perceived and acceptable risk and risk characteristics behaved according to expectations. The characteristic of foreseeability shows a significant negative correlation with acceptability. One would expect, however, that the more foreseeable a hazard, the more acceptable the risk. It appears that foreseeability may have more than one dimension, and that the specific dimension that is operating is a function of the risk situation. For example, Ms. King, Chairperson of the CPSC, holds that the Commission's focus should be on risks which are not readily foreseeable by the public. The emphasis in this position seems to be on serious and unexpected hazards. Given this position, one would expect a positive correlation between foreseeability and acceptability. Respondents in this study, on the other hand, seem to regard foreseeable as "knowable", i.e., a sort of 'accidents are bound to happen and therefore the product is dangerous' attitude. In this latter perspective, the negative correlation seems reasonable.

A second characteristic that does not conform to expectation is the dimension of voluntariness. The more voluntarily assumed risks are presumed to be more acceptable. Although this relationship seems to be born out in the case of some products, e.g., snow skis, the correlation across all products is near zero. One explanation for this result is that the consumer goods studied here did not differ much along this dimension.

Finally, a comment is in order about the characteristic termed ease of risk reduction. The respondents were asked how easily hazards associated with a particular product could be reduced through product redesign. A priori one would expect a relationship that suggests the more easily reduced the risk, the less acceptable the risk. That hypothesized relationship was not observed.

Predictive Validity of Risk Characteristics

Earlier in the conceptual framework section a number of attempts to predict acceptable levels of risk were reviewed. In this section the approaches to these predictions are further generalized by an attempt to develop a formula which specifies acceptability of product risks as a function of the twelve risk characteristics under consideration in this study.

Table 6 shows the result of a stepwise regression analysis in which acceptability formed the dependent variable and the risk characteristics constituted the independent variables. Specifically, the table shows the entry sequence with the accompanying increases in r-squared as well as the resulting multiple correlation coefficients. Furthermore, in the column labeled "coefficient", are recorded the unstandardized regression coefficients for the variables in the final model.

TABLE 6

Summary Table of Stepwise Regression Risk Acceptability

Step	Variable	Coefficient	Multiple r	Increase in r^2	F to enter
1.	Product necessity	0.22	0.68	0.46	22.90
2.	User error	-0.28	0.83	0.23	19.84
3.	Voluntariness	0.33	0.89	0.10	12.60
4.	Risk knowledge	0.32	0.91	0.03	3.63
5.	Foreseeability	-0.26	0.92	0.01	2.04
	Intercept	3.29			

($F = 23.65$; $df = 5,23$; $Prof > F = 0.0001$)

Although caution should always be exercised when interpreting the results of stepwise regressions, it should be pointed out that the variables entered in the model do have theoretical justifications in the risk assessment literature. That is to say, although the order in which variables enter a stepwise regression procedure is not always reflective of the relative importance of the variables, it is felt that the evidence presented seems to be converging on the variables included. The possible exception here is the voluntariness characteristic. Although no significant correlations were observed between this characteristic and hazard and acceptability ratings, it plays an important role in the regression model.

Conclusions and Recommendations

In this paper we have introduced the distinction between risk estimation and risk evaluation in the determination of product safety. This distinction alerted us to the fact that product risk management decisions include objective and scientific components as well as more normative, relativistic and judgmental components. This latter realization leads us to develop the premise that public attitudes and values should be reflected in the Commission's decisions

as to what constitutes acceptable levels of risk associated with consumer products. Consumer research can play a role in the assessment of consumers' response to the product risks.

The results of the empirical study reported in this paper suggests the relevance of the risk acceptability model which emerged from a review of the marketing, product liability and risk assessment literature. The acceptability of product risks was found to be inversely related to perceived hazardousness. An analysis for the discrepant rankings, however, suggests that consumers may incorporate into their acceptability judgments dimensions other than perceived physical risk.

Several researchers in the area of risk assessment have noted the desirability of examining selected risk characteristics as possible explanatory variables for risk acceptability perceptions. The appropriateness of the characteristics for product risks assessment was examined by a stepwise regression analysis which indicated that about 84% of the variance in the acceptability judgments can be explained by the characteristics of product necessity, user error, voluntariness, and to a lesser extent product knowledge and foreseeability. From a decisional point of view it certainly must be concluded that these qualitative characteristics deserve specific consideration by the CPSC in their product hazard management decisions. User error in particular appears to be an important variable in both risk and acceptability perceptions. This finding may suggest that respondents would be receptive to educational campaigns regarding products rating high on this characteristic. Alternatively, the Commission may review the cost effectiveness of redesigning some of the products most prone to user error.

We further suggest that product "risk profiles" illustrated in Tables 3 and 4 be developed for products with high Hazard Index ratings. Studies of each of the risk characteristics for these products can provide considerable insight into individual and societal behavior toward risk acceptance.

References

- Bettman, James R. (1973), "Perceived Risk and Its Components: A Model and Empirical Test," Journal of Marketing Research, 10 (May), 184-90.
- Fischhoff, Baruch, Slovic, Paul and Lichtenstein, Sarah (1979), "Which Risks are Acceptable," Environment, 21 (May), 17-38.
- Green, C. H. and Brown, R.A. (1978), "Counting Lives," Journal of Occupation Accidents, 4 (May), 18-24.
- Lowrance, William W. (1976), Of Acceptable Risk, Los Altos, CA: William Kaufman.
- National Commission on Product Safety (1970), Final Report, Washington, DC: U.S. Government Printing Office.
- Otway, Harry J. (1973), "Risk Assessment and Social Choices," in Proceedings of the IIASA Planning Conference on Energy Systems, Laxenburg, Austria: International Institute for Applied Systems.
- _____ and Pahner, Philip (1976), "Risk Assessment," Futures, 8 (April), 122-34.
- Ross, Ivan (1974), "Perceived Risk and Consumer Behavior: A Critical Review," in Advances in Consumer Research, Vol. 2, M. J. Schlinger, ed., Urbana, IL: Association for Consumer Research.
- Slovic, Paul, Fischhoff, Baruch and Lichtenstein, Sarah (1977), "Risk Assessment: Basic Issues," in Managing Technological Hazard: Research Needs and Opportunities, R. W. Kates, ed., Boulder: Institute of Behavioral Science.
- _____, _____, and _____ (1979), "Rating the Risks," Environment, 21 (April), 14-39.
- _____ and _____ (1979), "How Safe is Safe Enough? Determinants of Perceived and Acceptable Risk," in The Management of Nuclear Waste, L. Gould and C. A. Walker, ed., New Haven: Yale University Press.
- Starr, Chauncey (1969), "Social Benefit versus Technological Risk," Science, 165 (September), 1232-8.
- U.S. Department of Commerce (1978), Interagency Task Force on Product Liability: Final Report, Springfield, VA: National Technical Information Service.
- U.S. Consumer Product Safety Commission (1979), Annual Report, Part I and II, Washington, DC: U.S. Government Printing Office.
- Wade, John W. (1973), "On the Nature of Strict Liability for Products," Mississippi Law Journal, 44, 825-37.

SOME PERSPECTIVES ON RISK ACCEPTANCE¹

Jacob Jacoby, Purdue University

Abstract

The theme of this session -- namely, that any hazard management system developed to protect consumers from the risks associated with products must take into consideration subjective (psychological) as well as technical factors -- was illustrated by considering the hazardous consequences that potentially could result from implementation of the U.S. Department of Agriculture's proposed rule on providing uncured frankfurters, bologna, and bacon to consumers. Several positive and negative aspects of the empirical research described at this session were then discussed. The paper concluded by presenting a framework for conceptualizing hazard management that is emerging from currently ongoing research into consumer awareness, comprehension and usage of health and safety information.

As many of us who teach consumer behavior tell our students, the modern era of consumer protection can be traced to President John F. Kennedy's Declaration of Consumer Rights in 1962. This Declaration described four rights: the right to be informed, the right to choose, the right to safety, and the right to be heard (or the right of redress). Partly because of its intriguing nature, to date, the lion's share of consumer research directed toward public policy issues has focused on the right to be informed (including the necessity for information to be free of deceptive or misleading connotations and the necessity for the information to be sufficiently complete so that the consumer is able to make an informed decision). It is time that a commensurate research effort be mounted in regard to product safety. As Feldman (1980, p. 73) points out: "Compared with other consumer protection issues, product safety is paramount because unlike other issues such as deceptive advertising or the failure to make good on warranties, product safety directly involves injury and loss of life."

Our objectives in this commentary are two-fold. The first is to consider selected aspects of the empirical papers presented at this session. The second is to indicate directions for future work.

Comments on the Papers Presented

The fundamental tenet which underlies these papers is that any hazard management system developed to protect consumers from the risks associated with products must take into consideration consumer psychological (that is, internal, subjective) factors as well as technical factors. As eloquently noted by Slovic, Fischhoff and Lichtenstein; "People respond to the hazards they perceive. If their perceptions are faulty, efforts at public and environmental protection [which do not consider these perceptions] are likely to be misdirected." Further: "Our basic premises are that both the public and the experts are necessary participants in [the risk assessment] process, that assessment is inevitably subjective, and that understanding public policy perceptions is crucial to effective decision making." We fully concur with this set of premises. Indeed, this point needs to be underscored; it

cannot be emphasized enough! It especially needs to be brought to the attention of policymakers and regulators who sometimes end up generating misguided and incorrect public policies through their failure to understand and take into consideration the relevant consumer psychological factors which may be operating.

On the Risk of Ignoring Consumer Psychological Factors

As a concrete illustration of how the disregard of consumer psychological factors might result in seriously increasing rather than decreasing the safety hazards associated with consumer products, consider the U. S. Department of Agriculture's efforts in regard to providing consumers with uncured (i.e., nitrite and nitrate free) meat products. Stimulated by research (which has since been discredited) that seemed to indicate that the ingestion of nitrites and nitrates was associated with unreasonably high levels of cancer, and also stimulated by consumer activists, the U.S.D.A. developed a proposed rule on the subject. The proposed rule, published in the Federal Register on August 21, 1979 and scheduled to become effective on September 20, 1979, provided that meat and meat products that were not cured with nitrates, nitrites, or other preservatives might be sold under their traditional names, so long as the word "uncured" appeared on the label as part of the product name and the label contained the following cautionary language: "No nitrate, or nitrite added; not preserved, keep refrigerated below 40° Fahrenheit at all times." (See Figure 1) The regulation also required that the uncured products be similar

FIGURE 1

USDA's Proposed Precautionary Language as it Would be Applied to Frankfurters

UNCURED FRANKFURTER
No nitrate or nitrite added
Not preserved - Keep Refrigerated Below
40°F at all times.

in size, shape, flavor, color, consistency and general appearance to the product commonly prepared with nitrite or nitrate. In other words, the consumer would have no other cues to the fact that the frankfurter (or bacon, or balogna) she was purchasing was somehow different than frankfurters (bacon, balogna) she had purchased before -- except for the cautionary labeling prescribed by the rule.

On September 20, 1979, the National Pork Producers Council, together with three members of the U. S. House of Representatives, filed a law suit challenging the proposed rule. These plaintiffs were subsequently joined by the National Independent Meat Packers Association. Their argument was heard in Federal Court in Des Moines, Iowa, in November of 1979. In presenting their case, these plaintiffs brought forth two expert witnesses.

The first witness, an internationally acknowledged authority on the subject of food toxins, testified that eight ounces of botulism spore were sufficient to wipe out the

¹Preparation of this paper was supported, in part, by a grant from the National Science Foundation (PRA 7920585); see Jacoby and Jaccard, in progress.

population of the entire planet. He went on to state that under certain conditions (such as might happen if a high school student took an uncured bologna sandwich to school, left it in his locker on a hot day, chose not to eat it for lunch but took it to his after school game instead, and then ate it at 5:00 p.m.) deadly botulism spore could develop in as little as 8 hours.

Participating as the second expert witness, the present author testified regarding likely consumer attention to, comprehension of, and use of the cautionary labelling information. In brief, and based upon evidence obtained across more than a score of consumer information usage investigations conducted to date, the testimony pointed out that: (1) Many consumers (perhaps as many as 20%) do not read any information on the labels for frequently purchased consumer products² beyond the brand name. (2) The 80% who do read package information rarely read all the available information but tend to focus on a very small percentage of this information (e.g., price). (3) Some of those who happen to read the proposed cautionary labelling would not understand its significance. From informal research conducted by this author (e.g., asking at least 500 people in his classes and in audiences to whom he has spoken whether they understood what the word "cured" meant when it was applied to a food product) and others, it would appear that fewer than 5% of the adult population comprehends that the term "cured" refers to the inclusion of preservatives. The overwhelming majority of people seem to have no idea as to what "cured" stands for in this context and a much smaller proportion (in the order of 5% to 15%) seem to believe that cured means "smoked." Further, even though most individuals would be expected to understand the literal meaning of "Not preserved -- keep refrigerated below 40° Fahrenheit at all times," this language provides no understanding of the dire consequences that might ensue from not keeping the product refrigerated below 40° at all times. (4) Finally, even if the consumer did comprehend the full significance of this statement, once the meat had been taken from the refrigerator (and, for example, been placed in a sandwich), subsequent conditions might be out of her control and it would still be possible for the meat to be handled unsafely.

After hearing the case and considering the record, on February 12, 1980 the District Court in Iowa entered a final order enjoining the government from enforcing or applying the challenged regulations. This decision was appealed by the U.S.D.A. and reversed just a few days ago, on September 23, 1980. In noting the reasons for its reversal, the Circuit Court cited the fact that warning labels have been generally shown to be effective (how come, then, people still smoke cigarettes? etc.), and that many of the comments submitted by consumers in response to the proposed rule "demonstrated consumer awareness of the potential dangerous consequences of marketing and consuming meat products that do not contain nitrites or nitrates." The Court apparently did not recognize the danger of extrapolating from such a small and unrepresentative sample as the group of consumers who chose to send in written comments on the proposed rule. The Court further argued that "the cautionary labelling will adequately inform the consumer of how to maintain such products in a wholesome condition until consumed." Preposterous! At the present time, the National Pork Producers are considering appealing this reversal to the next highest level of the judiciary which, in this case, is the Supreme Court.

²Remember: the uncured product had to be similar in size, shape, flavor, color, and general appearance to the cured product. The only distinguishing characteristic would be the cautionary language.

To my way of thinking, the proposed rule and subsequent reversal by the Circuit Court of the District Court's decision provides a frightening example of how public policy can be misguided when only technical factors are considered and subjective, consumer psychological factors are ignored. Having served as an expert witness in more than a dozen matters, I can honestly say that I have never felt so strongly (from my perspective as a consumer researcher) that the courts and policy makers acted in complete disregard for what is known regarding human behavior -- and that this might result in tragedy of catastrophic proportion for American consumers. True, if all consumers did indeed read and understand that uncured meat had to be refrigerated at all times and acted in accord with this belief, the risk would be reduced substantially.³ However, it is highly unlikely that all consumers will read and understand the cautionary language and behave as expected. By not taking these subjective considerations into account, the government is, in my opinion, opening up a Pandora's Box and substantially increasing the health and safety hazards to which the American populace will be subjected.

As the authors of these manuscripts have noted, the consumer element in hazard management is of critical concern and cannot be neglected. If the consumer fails to recognize that a risk exists, then his behavior can not be shaped accordingly. In other words, these authors are barking up a very important tree and their efforts need to be encouraged, applauded, and hopefully supplemented by considerable research from others. It is also commendable that their efforts, particularly in the case of Slovic, Fischhoff, and Lichtenstein, are clearly long term and programmatic.

On the Empirical Work Presented

Having acknowledged the significance of this work and the genuine debt that consumer researchers owe to these individuals, we still find several shortcomings associated with the papers presented. These tend to be shortcomings which come with the territory when researchers embark on journeys over relatively unexplored terrain. Fortunately, none of these problems are insurmountable.

One entry on the negative side of the ledger is in regard to the relatively small and unrepresentative samples employed. It is conceivable that this might have been a major contributing factor in Slovic et al.'s failure to replicate the findings of dimensional structure that had been obtained in the earlier investigation by Fischhoff. A second problem would be the controversial and somewhat questionable application of such high powered analytical techniques as factor analysis (Slovic et al. 1981) and stepwise multiple regression (Rethans and Albaum 1981) to such small samples, particularly given such large numbers of variables.

However, our major concern with these studies is not the issue of sample size or the questionable application of high powered statistics to samples of such size, but the operationalizations of the core concept of risk acceptance. The measures of risk acceptance used by Slovic et al. and Rethans and Albaum are provided in Figures 2 and 3, respectively. Consideration of these measures suggests that, while the researchers may have had a very clear idea in mind of what risk acceptance was, and while their operationalizations may be technically accurate reflections of these conceptualizations, it is highly questionable whether many of their subjects adequately

³Risk would still be associated with the stages of product preparation and handling prior to its reaching the consumer.

FIGURE 2
Slovic, Fischhoff and Lichtenstein's
Operationalization of Risk Acceptance

After rating the hazards with regard to risk, respondents were asked to rate the degree to which the present risk level would need to be adjusted to make the risk level acceptable to society. The instructions for this adjustment task read as follows:

The acceptable level of risk is not the ideal risk. Ideally, the risks should be zero. The acceptable level is a level that is "good enough," where "good enough" means you think that the advantages of increased safety are not worth the costs of reducing risk by restricting or otherwise altering the activity. For example, we can make drugs "safer" by restricting their potency; cars can be made safer, at a cost, by improving their construction or requiring regular safety inspection. We may, or may not, believe such restrictions are necessary.

If an activity's present level of risk is acceptable, no special action need be taken to increase its safety. If its riskiness is unacceptably high, serious action, such as legislation to restrict its practice, should be taken. On the other hand, there may be some activities or technologies that you believe are currently safer than the acceptable level of risk. For these activities, the risk of death could be higher than it is now before society would have to take serious action.

On the answer sheets, participants were provided with three columns labeled: (a) "Could be riskier: it would be acceptable if it were ___ times riskier;" (b) "It is presently acceptable;" and (c) "Too risky; to be acceptable: it would have to be ___ times safer."

FIGURE 3
Rethan's Operationalization of Risk Acceptance
Acceptability of Risk

Now that you have evaluated the risks associated with some consumer products, we would like you to judge the acceptability of these risks. Ideally, risk should be zero. However, products with zero risk do not exist. Rather, we decide whether or not risks associated with a product are at a socially (from a societal viewpoint) acceptable level. An acceptable level is a level that is "good enough," where "good enough" means that you think that the advantages of increased safety are not worth the costs to society of reducing the risk by altering the product or restricting its use.

Please indicate your views on the acceptability of the risks by circling the number on the scale that best describes your opinion.

	From a societal viewpoint, the present level of risk of this product is:
	Very Unacceptable Very Acceptable
1. Household ammonia	1 - 2 - 3 - 4 - 5 - 6 - 7
2. Sewing machines	1 - 2 - 3 - 4 - 5 - 6 - 7
3. Snow skis	1 - 2 - 3 - 4 - 5 - 6 - 7
4. Etc.	1 - 2 - 3 - 4 - 5 - 6 - 7

understood these operationalizations. Each of these operationalizations is long, complicated, contains several difficult concepts and are the kinds of concepts with which most lay people can be expected to be unfamiliar. Interpretation of these responses necessarily rests on the assumption that the respondent comprehends both the concepts and vocabulary being used. Yet how certain can we be that the respondents satisfactorily comprehend such sentences as: "On the other hand, there may be some activities or technologies that you believe are currently safer than the acceptable level of risk."? And just how confident can we be that the typical respondent has enough understanding of the issues to be able to assess whether "the advantages of increased safety are not worth the cost to society of reducing the risks by altering the product or restricting its use."? To interpret a response to these questions as being valid assumes that the respondent understands just what the "advantages of increased safety" are, knows what "the costs to society" are, has some idea of what these costs and benefits are "worth," has an understanding of what kinds of "alterations" of the product are possible, or what kinds of "restrictions on its use" may be imposed, etc. The questions further assume that the respondents can keep all these relatively unfamiliar and complicated considerations simultaneously in mind while answering the question. They further assume at certain parts that the respondent's thinking will be unidirectional, that is, if he thinks the product shouldn't be altered, he also thinks its use shouldn't be restricted. To my way of thinking, without any attempt to confirm and validate these numerous key assumptions, none of them are tenable.

In other words, I doubt whether the majority of subjects fully and accurately comprehended the questions they were responding to. Substantial evidence in the literature on questionnaire construction exists to indicate that even simple questions are miscomprehended by large proportions of the populace. Indeed, in a set of studies described at this Conference yesterday (cf. Jacoby, Nelson and Hoyer 1981), evidence was provided to indicate that 70 to 85% of subjects trying to understand a remedial advertising statement which was only one sentence in length either incorrectly understood that statement or extracted a confused meaning from it. How much greater, then, can we expect the levels of miscomprehension to be for a passage containing ten sentences referring to complicated and unfamiliar concepts? Until such time as evidence is supplied to indicate that the respondents accurately understood the operationalizations, we must entertain the possibility that the data reflect GIGO.

Future Directions

Clearly, work needs to be directed toward developing operationalizations which are something more than a collection of words which the investigators assume are capable of adequately assessing the concept of interest. It first needs to be established that these words are properly understood by the respondents before data obtained using these measures can be interpreted as being meaningful and programmatic work using these measures is undertaken.

Beyond these empirical considerations, I believe that the research thrust outlined by these authors would benefit considerably from a closer integration with the two decades old body of literature on the issue of perceived risk. This literature began with Bauer's (1960) seminal conceptualization and has since been summarized by Cox (1967) and Ross (1975), among others. At least one hundred studies have been conducted using this concept and a large body of findings have emerged. Among these findings is the fact that at least six different types of perceived risk (including perceived safety risk) have been identi-

fied and cross validated, and that perceived safety risk correlates with overall perceived risk .67 but contributes the least amount of explained variance in overall perceived risk scores (cf. Jacoby and Kaplan 1972; Kaplan, Szymbillo and Jacoby 1974). Consideration of this body of literature further reveals that the first three conclusions offered by Slovic et al. -- namely, perceived risk is quantifiable and predictable; lay people sometimes differ systematically in their perceptions of risk; and the greater the perceived risk, the greater the desired risk production -- already have a well-established history in the consumer literature. One would hope that a merging and integration of this earlier body of research with the current approach would enable us to go beyond what is already known.

Finally, I believe that the work that is conducted needs to be placed into a broader, conceptually richer framework. One emerging approach⁴ is to consider the subject of hazard management as involving two major components: technical and social issues (see Figure 4). The technical side of the ledger consists of the identification of hazards and the measurement of these hazards in terms of their probability of occurrence and the severity of their consequences. The social side of the ledger includes risk perception (that is the probabilities and consequences as perceived by the respondents) and risk acceptance.

FIGURE 4
A Perspective on Hazard Management

<u>Hazard Management</u>	
<u>Technical Issues</u>	<u>Social Issues</u>
Identification	Risk Perception
Measurement	Perceived consequences
Probability	Subjective probability
Consequence	Risk acceptability

Our own work on consumer awareness, comprehension, and usage of health and safety information (Jacoby and Jaccard, in progress) has involved developing an elaborated perspective based upon the earlier framework. This framework has continued to evolve and the current version is offered here for critique and use by others (see Figure 5). The perspective retains the notion of technical and social (or objective and subjective) components as being fundamental.

On the technical side, the framework begins with the actual identification of hazardous conditions and hazardous consequences. Hazardous conditions are assumed to be of two main types -- those associated with the product and those which occur with the interaction between the product and the environment (where the environment is meant to include the consumer as well). Table 1 illustrates these distinctions for insecticides. Note that various hazardous conditions are believed to be associated with the product and the manner in which it is being provided for consumer use. For example, the product may (or may not) contain ingredients which are toxic and pose a safety problem for humans. The product may be packaged in a container which does not have a childproof cap and/or it may contain inadequate labelling and usage instructions. In addition, many insecticides tend to be safe in terms of their contents and packaging but will become hazardous when certain environmental conditions are present. For example, some insecticides are safe except when used in the presence of exposed flames. Other insecticides are safe except when the environment in which they are used contains either

pregnant women, asthmatics, pollen sensitive people, infants, pets, people seriously ill, or people on prescribed medications. Yet other insecticides (particularly of the spray variety) only become dangerous to humans when used in the presence of exposed food stuffs.

TABLE 1
Illustrating the distinction between hazardous conditions and hazardous consequences for the product category "Insecticides"

I. Hazardous Conditions

A. Product

1. Does not contain childproof container
2. Contains problem ingredients
 - Kepona
 - Arsenates
 - Sodium Fluoride
 - Diazinon
 - Dichlorvos
 - Captan
 - 2, 4-D
3. inadequate labeling and instructions

B. Product X environment

1. exposed flames
2. exposed food stuffs
3. pregnant women
4. asthmatics and pollen sensitive people
5. infants and children
6. pets
7. people on prescribed medications
8. people seriously ill

II. Hazardous Consequences

- poisoning
- lasting nerve damage
- miscarriages and birth defects
- affect production of liver enzymes
- cancer

The actual presence of a hazardous condition (e.g. a bald tire on a car) does not necessarily lead to a hazardous consequence (e.g., an accident). The hazardous consequences which evidence suggests are associated with insecticide usage include poisoning, lasting nerve damage, miscarriages, birth defects, a deleterious impact on the production of liver enzymes, and cancer.

While the first major aspect of technical issues involves the identification of hazardous conditions and consequences, the second aspect involves assessment of: the probability of occurrence of each hazardous condition; the probability of each hazardous condition resulting in a hazardous consequence; and the severity of the consequences should they occur.

The third element on the technical issue side involves the identification of appropriate coping measures. These may be divided into those which are preventative in nature and those which are corrective. Preventative coping measures are those which enable the individual either to reduce or completely eliminate the existence of the hazardous condition or the hazardous consequence. (An example would be the use of safety goggles when using electrical tools). Corrective coping measures are those which should be employed to correct the existence of a hazardous condition or the occurrence of a hazardous consequence, e.g., adding air to an underinflated tire (preventative) or turning the wheel of a skidding automobile in the direction of the skid while pumping the brakes rapidly (corrective).

⁴We thank Rethans for bringing this to our attention.

FIGURE 5
Toward an elaborated perspective on hazard management

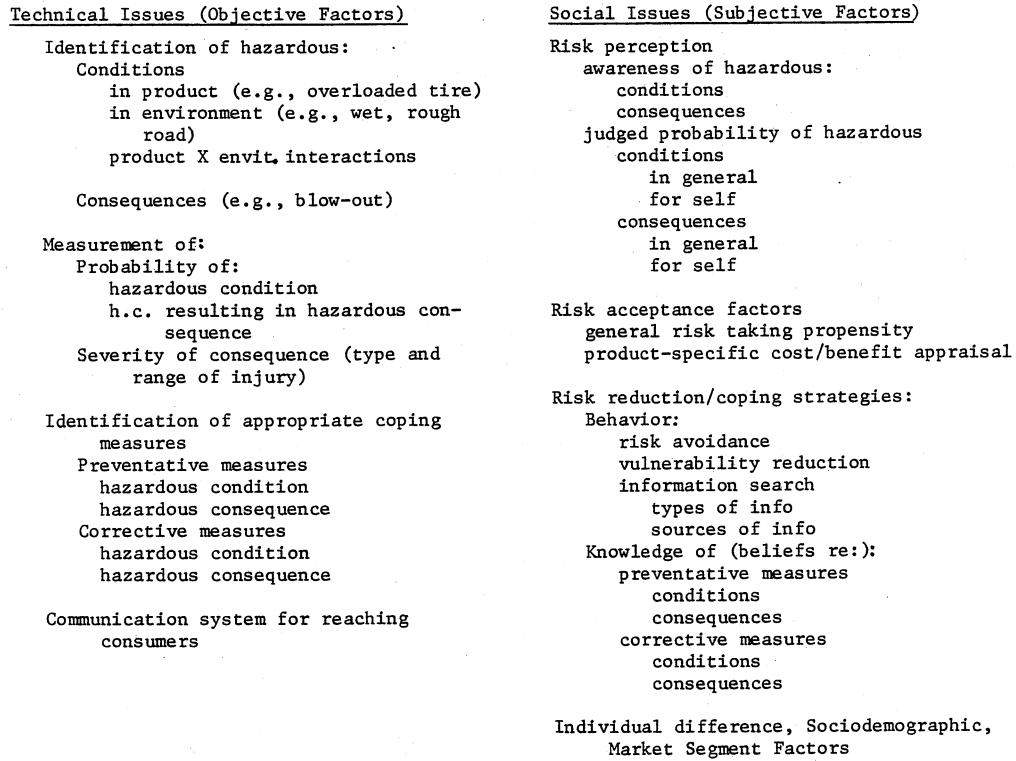


FIGURE 6
Toward integrating objective and subjective components.

		Hazardous Conditions									
		1	2	3	-	-	-	-	N		
Hazardous Consequences	1	a	b						c	d	e
	2										
	3										
	M										
	e										

- a = objective probability that the hazardous condition identified will lead to a hazardous consequence
- b = subjective (perceived) probability that the hazardous condition identified will lead to a hazardous consequence
- c = objective severity of that consequence
- d = subjective (perceived) severity of that consequence
- e = acceptance of the hazardous condition/consequence

The fourth element in the technical system would involve a consideration of the systems and procedures (media mix, execution strategies, etc.) that could/should be used for communicating the information regarding hazardous conditions, consequences, and coping measures to the consumer.

On the other side of the ledger are the subjective factors. These begin with risk perception and the consumer's awareness of the nature of the hazardous conditions and consequences associated with a particular product's use. In concert with Slovic et al. and Rethans, we firmly believe that if a consumer is unaware of either hazardous conditions or consequences, his behavior will not be affected by such conditions or consequences even though they may be serious and exist at high probability levels. In addition to awareness, one should consider the individual's probability judgments regarding these hazardous conditions and consequences as that individual believes they apply to others in general, and to himself in particular. As a preliminary attempt, Figure 6 outlines a framework for integrating some of these basic subjective factors with their objective counterparts.

Beyond the notion of risk perception, one has to take into consideration the factors affecting risk acceptance. While the consumer's perception of risk hazards and consequences may be accurate, he may still be willing (and even agree) to use a very risky product. In this regard, the authors of the papers presented at this session have identified at least a dozen factors seemingly associated with risk acceptance. To this list we would like to add the notion of "general risk-taking propensity." It appears to us that certain individuals are more inclined to be risk takers than others, and that this factor needs to somehow be taken into consideration in attempts to develop practical hazard management systems.

The subjective side of the ledger also needs to consider the risk elimination, risk reduction, and risk coping strategies that consumers can engage in. These include risk avoidance behavior (e.g., shunning the questionable product entirely), risk reduction behavior (i.e., reducing one's vulnerability to risk while maintaining at least some level of product usage), and continuing full use of the product while acquiring information regarding how to cope with hazards if and when they surface.

In this latter regard, it would be particularly useful for researchers who wish to influence public policy to identify the types of information consumers do use, and the sources to which they turn to in order to obtain this important information.

Beyond the behaviors that consumers actually engage in, one needs to assess the extent and accuracy of the beliefs consumers possess in mind regarding preventative measures and corrective measures. Again, to the extent that consumers have nonexistent or erroneous beliefs regarding corrective measures, different means for affecting their behavior will have to be employed.

Finally, any attempt to develop policies which will affect consumers at large needs to consider general individual difference and market segment factors. Certain risks tend to be much more likely, given certain combinations of market and individual difference factors. For example, it seems likely that risks are more probable with the elderly consumer, the disadvantaged consumer, the undereducated consumer, the very young consumer, etc.

Conclusion

In conclusion, we are enthused with the efforts engaged in by these researchers and would like to encourage both their efforts and others to embark on similar efforts.

However, we recommend considerable caution in accepting their results until we can be more confident that the measures of their key concept (namely, risk acceptance) are meaningful and valid.

References

- Bauer, R. A. (1960), "Consumer Behavior as Risk Taking." In R. S. Hancock (Ed.) Dynamic Marketing for a Changing World. Chicago: American Marketing Association.
- Cox, D. F. (Ed.) (1967), Risk Taking and Information Handling in Consumer Behavior. Cambridge, Mass.: Harvard University Press.
- Feldman, L. P. (1980), Consumer Protection: Problems and Prospects (2nd ed.). St. Paul, Minnesota, West Publishing Co.
- Jacoby, J. and Jaccard, J. J. "Assessing the Effects of Science Based Information on Consumer Technological Choices." Grant from the Division of Policy Research and Analysis, National Science Foundation. In progress.
- _____, & Kaplan, L. B. (1972), "The Components of Perceived Risk." In M. Venkatesan (Ed.), Proceedings, Third Annual Conference of the Association for Consumer Research. University of Chicago: Pp. 382-393.
- _____, Nelson, M. C., and Hoyer, W. D. (1981), "Correcting Corrective Advertising." In K. B. Monroe (Ed.) Advances in Consumer Research, Volume 8.
- Kaplan, L. B., Szybillo, G. J. and Jacoby, J. (1974), "Components of Perceived Risk in Product Purchase: A Cross-Validation." Journal of Applied Psychology, 59(3), 287-291.
- Rethans, A. J. and Albaum, G. S. (1981), "Toward Determinants of Acceptable Risk: The Case of Product Risks." In K. B. Monroe (Ed.), Advances in Consumer Research, Volume 8.
- Ross, I. (1975), "Perceived Risk and Consumer Behavior: A Critical Review." In M. J. Schlinger (Ed.), Advances in Consumer Research, 2, 1-20.
- Slovic, P., Fischhoff, B. and Lichtenstein, S. (1981), "Facts and Fears: Societal Perception of Risk." In K. B. Monroe (Ed.), Advances in Consumer Research, Volume 8.

INDIVIDUAL PREFERENCES AND PUBLIC POLICY

Robert T. Deacon, University of California, Santa Barbara

Abstract

Attempts to construct behavioral models of the public sector have resulted in the characterization of government as an exchange process, in which tax payments are the quid pro quo for public services received. The most widely discussed political-economic model follows from a characterization of citizen-voters as well informed, economically rational agents, embedded in a competitive political environment. Challenges to this theory question not only its positive implications, but also its normative content.

Introduction

The range of issues currently being raised in the field of state and local government finance is both diverse and volatile. The popular movement to limit taxes and public expenditures is probably the most visible of these, though court imposed changes in public school finance, and state and federal funding for local government programs are of continuing interest. Had a phenomenon such as tax limitation arisen a mere twenty years ago, it seems safe to speculate that the major topics of concern to the economics profession would have been with the distributional effects of tax cuts and the adequacy of alternative revenue sources to make up shortfalls. In contrast, the current literature is filled with studies of voting behavior on tax limit referenda, attempts to discover whether such phenomena are indicative of "political disequilibrium," and if so, what features a new equilibrium is likely to embody. This dramatic shift in emphasis reflects a rapid evolution in economic analysis of the public sector, one which has resulted from an attempt to conceptually bridge revenue and expenditure sides of the public budget with a behavioral paradigm (Buchanan 1975). In this emerging framework, government is characterized as an exchange process, in which taxes are the quid pro quo for services received. Individuals are endowed with essentially the same motivations, preference structures, and informational requirements faced in the private sector. But in the role of public sector consumer, the individual's opportunities and constraints are critically shaped by the nature of political institutions.

Viewing the citizen-taxpayer as consumer raises a variety of questions. What is the nature of his public sector demands, and how are they revealed? How accurately are these demands articulated, and how do public supply institutions respond to them? Can a concept of equilibrium, similar to that used to study private markets, be employed to characterize public sector outcomes? This general class of questions is addressed in the following pages. The next section briefly discusses the institution of voting, and its role as a device for revealing preferences. Following this, the concept of equilibrium in majority rule systems is explored, and competing views of the nature of the political process are examined. Concluding comments are presented in the final section. At each step, existing empirical evidence, as it relates to key conceptual questions, is briefly summarized.¹

¹More extensive discussions of several of the topics treated here appear in Deacon (1977).

Voting and Revealed Preferences

The notion that individuals act in self-interest when choosing among ballot alternatives has been implicit in public choice theory from its beginning (Bowen 1943, Downs 1957, pp. 27-45). Extension of the utility maximization hypothesis from the realm of the market to the voting booth has apparently been so natural that it is axiomatic in the theoretical literature. The view that individuals compare ballot alternatives and vote for those that offer preferred outcomes suggests the use of voting choices as information on revealed preferences for public goods.

To illustrate this approach, consider the following simplified situation (adapted from Rubinfeld 1977) involving a referendum to change the level at which a single public service is provided. Such choices are fairly common in public school districts in the U.S., where ballots offer alternative property tax levies. The preferences of individual i are represented by a conventionally shaped utility function in which the public service in question (q), and consumption of a composite private good (x_i) enter as arguments,

$$U_i = U_i(q, x_i) . \tag{1}$$

Normalizing so that the price of the private good is unity the individual's budget constraint is expressed

$$I_i = x_i + t_i E \tag{2}$$

where I_i is income, E is total public service expenditure by the jurisdiction in question, and t_i is the individual's tax share, the amount by which his tax will rise if public spending increases by one dollar. To complete the problem, spending is related to services received by a (convex) cost function for the public service

$$E = E(q) . \tag{3}$$

The consumer's preferred public sector outcome is characterized by maximizing (1) subject to (2) and (3). The solution to this problem is the consumer's desired public spending level,

$$E_i^* = D_i(t_i, I_i) . \tag{4}$$

The functional label D_i is used to indicate that the relationship is a conventional demand curve, where t_i is interpreted as a "tax price." Consequently, E_i^* is expected to be decreasing in t_i and increasing in I_i (if the public good is normal).

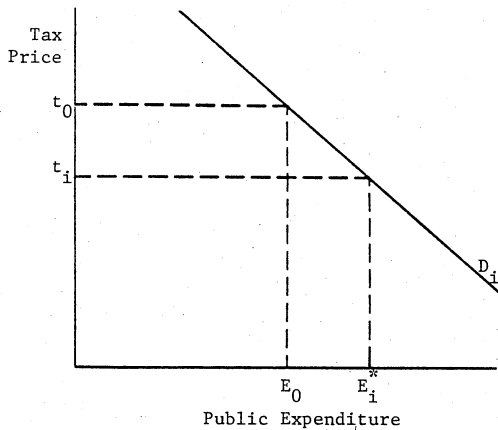
The demand curve is drawn in Figure 1. Suppose that the existing level of public expenditure is E_0 , and that the issue before the voter proposes a marginal increase in spending. Since voter i prefers a level of public outlays larger than E_0 , a yes vote on the proposed increase is expected. The general rule for voting on marginal increases such as this can be written:

$$\begin{aligned} \text{vote "yes" if } E_i^* > E_0 , \\ \text{vote "no" if } E_i^* < E_0 , \\ \text{abstain if } E_i^* = E_0 . \end{aligned}$$

Intuitively, if one had prior information on the shape of demand functions and on individual tax prices, then voting responses could be predicted. For example, among the class of voters who have demand curves similar to D_i

in Figure 1, those who face tax prices above t_0 would be expected to vote against the measure, while those with tax prices below t_0 would support it. Consequently, "the demand function (for public education) ought to be the most effective discriminant function for distinguishing 'yes' from 'no' votes that households cast" (Peterson 1975, p. 104).

FIGURE 1
Public Expenditure Demand



A number of public education demand functions have been estimated from school referendum data, though "probit" or "logit" techniques are usually employed in estimation rather than discriminant analysis. In some cases, individual attributes and voting choices have been available from survey data (Peterson 1975, Rubinfeld 1977). More often, however, only aggregate voting returns (e.g., by precinct) are reported, a condition that necessitates interpreting demand responses as pertaining to an average or representative voter in each group (Barkume 1978). In general, these studies have obtained significant price and income effects that are of expected sign. Moreover, since it is often reasonable to postulate that demands will depend upon other observable attributes such as age, education, family status, etc., a variety of additional hypotheses have been examined. For example, Rubinfeld (1977) found school employees much more likely to vote for tax increases than otherwise similar voters. Others have examined the effects of numbers of school age children in a family upon voting responses.²

If voting data could only be analyzed within the confines of the example developed earlier, the application of this approach would be severely limited. With noneducational referenda, diverse arrays of services may be involved in a single ballot, and payment may be exacted by changing incomes or market prices, rather than simply altering the level of a single tax. One technique that has been developed to analyze such complex alternatives employs the concept of a mixed indirect utility function (Deacon and Shapiro 1975). Within this framework, ballot choices are characterized in terms of utility comparisons, rather than

²In general, voting studies of the type described here, can only measure the relative magnitude of price and income effects. In a series of papers, Eli Noam has developed and applied a technique that permits estimation of the absolute effects of price, income, and other variables upon demands, and applied the method to the determination of optimal budgets. See Noam (1979) and references cited therein.

public expenditure demand functions, and utilities under alternative ballot outcomes are expressed as functions of income, prices, tax rates, and public goods levels. Application of this technique to referenda involving such issues as scenic resource conservation, rapid transit provision, and public hospital financing has yielded results that are generally in close accord with the underlying theory.

Prior to casting a ballot, the voter must decide whether or not to vote at all, i.e., whether the benefits of showing up at the polls exceed the costs (Downs 1957). Potential benefits depend upon the likelihood that a single vote will decide the outcome (as, for example, measured by the closeness of the election) and upon the expected difference in outcomes under the two alternatives on the ballot. Costs will vary with the imputed value of the individual's time, and the resources required to obtain and evaluate information regarding the alternatives. This economic model of the "turnout" decision has been successfully employed in a number of empirical studies (Barzel and Silberberg 1973, Chapman and Palda 1980). Moreover, it indicates that those who vote may not be a random sample of the populace. This conjecture was borne out in an analysis of the Headlee tax limitation amendment in Michigan by Gramlich *et al.* (1980), who found that nonvoters tended to be relatively uninformed or uncertain regarding the likely impacts of the measure, and generally inclined to be against it. The notion that voting costs may significantly influence electoral outcomes, by "selecting" the kinds of preferences revealed in elections, remains to be fully integrated into existing models of the collective choice process.

The analysis of voting behavior is of central importance to any model of the public economy, and must precede judgments concerning the ability of democratic systems to respond to citizens' preferences. To the extent that voting behavior is consistent with economic rationality, it becomes possible to predict outcomes and to formulate refutable hypotheses regarding the electoral process. In turn, it is the possibility of prediction that allows one to construct positive political models, and to make normative statements concerning actual political systems.

Majority Rule and the Median Voter

To address questions regarding the determinants of political outcomes and the nature of political equilibrium, an explicit model of the political process under majority rule is necessary. The first economist to suggest such a theory was Harold Hotelling (1929). The paradigm he offered was extended in a remarkable article by Howard Bowen (1943), and further refined by Downs (1957).

To motivate the development of this theory, consider the following problem. The range of possible public sector outcomes (policies) is represented as a set (a_1, a_2, \dots, a_n) , and elements of this set will be placed before the voters in a pairwise fashion. Each citizen-voter has preferences for the various alternatives, and in principle, these preferences can be represented in terms of individual utilities. Assuming that the preferences of voters are known, the problem is to predict which alternative will be chosen under simple majority rule. In general, the existence and characteristics of a solution to this problem depend upon the satisfaction of certain conditions regarding the shape of preferences and the way in which alternatives are offered.

It has been known, at least since the 18th century, that majority voting on pairs of alternatives need not exhibit transitivity. Cycles can develop in which a_1 defeats a_2 , a_2 defeats a_3 , and a_3 defeats a_1 , etc., even if the preferences of individual voters are fully transitive.

Without transitivity, equilibrium would not be possible unless some sort of stopping rule were invoked; but use of such a stopping rule makes the final outcome "path dependent," or sensitive to the order in which alternatives are offered. If, however, individual preferences satisfy a certain regularity property then the cycling problem disappears. For simplicity, suppose that indifference among pairs of alternatives is ruled out, and that the number of voters is odd. Then majority rule will be transitive if the list of alternatives (a_1, a_2, \dots, a_n) can be arranged in a sequence (a_1, a_2, \dots, a_n) such that the following condition is satisfied for each voter: If $U(a_i) > U(a_{i+j})$, then $U(a_{i+j}) > U(a_{i+j+k})$, for all $i, j, k > 0$. This statement of the so-called "single peaked" condition guarantees transitivity, though considerably weaker conditions will also suffice (Sen 1968). Intuitively, this condition is satisfied if each voter's utility, when plotted against the sequence alternatives, forms a single peaked curve, i.e., each voter has one uniquely preferred alternative, and utility declines monotonically as one moves away from this alternative in either direction.

If preferences are single peaked then the median of individually preferred outcomes is dominant in that it will be preferred by a majority of voters over any other alternative that may be paired against it. To see why this is true, let a_m be the median preferred alternative and suppose it is paired against some other alternative (a_{m+h}) that is higher in the sequence. Since a_m is the median preferred point, a simple majority of voters have preferred alternatives that are at or below a_m in the sequence; since their preferences are single peaked, each voter in this majority prefers a_m to a_{m+h} . Of course, the same argument applies to alternatives below a_m .³

For the median voter theorem to have any empirical content, the condition of single peaked preferences is clearly pivotal. In practice, there are a number of empirically important contexts in which the condition seems quite reasonable, and most theoretical and empirical applications have been restricted to such situations.⁴ The simple referendum example described in the preceding section is a case in point. The individual whose public expenditure demand is shown in Figure 1 prefers a public outlay of E_i^* . If actual spending is increased or decreased from this level, utility declines monotonically, as shown by the curve labelled U_i in Figure 2. Of course, the same argument applies to other individuals as well; correspondingly, Figure 2 also illustrates public spending preferences for two other voters, h and j . In this example, i is the median voter, and E_i^* is his preferred expenditure level. Notice that any proposal to increase (decrease) spending from E_i^* would be opposed by a majority of the electorate, h and i (i and j). (The other points labelled in Figure 2 are explained below.)

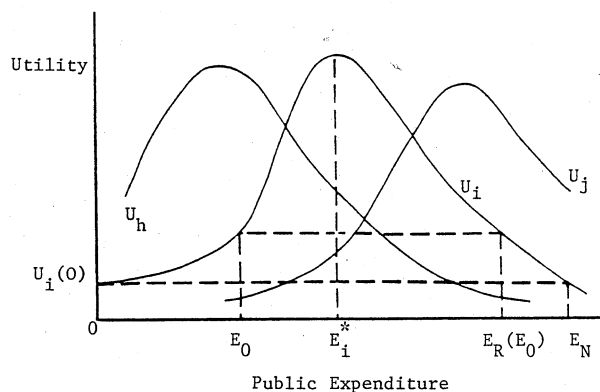
Majority Rule Equilibrium: Competitive or Noncompetitive?

If the median voter's optimum appears on the ballot it will defeat any alternative motion. But what guarantees that it will ever be placed before the voters? If motions,

³Hotelling (1929) speculated that political candidates and their platforms could be arrayed along a scale from liberal to conservative, and that this political spectrum formed a natural sequence over which individual preferences are likely to be single peaked. From this it follows that vote maximizing candidates in a two party system will be led to adopt similar platforms toward the middle (median) of the distribution of political opinion.

⁴See, however, Barzel and Deacon (1975), and Sonstelie (1979), for analysis of voting outcomes in situations where preferences are not single peaked.

FIGURE 2
Preferences and Majority Rule Outcomes



or political alternatives, are supplied competitively by political entrepreneurs who seek to maximize their chances of being elected, then the median outcome seems very likely. Nonmedian strategies simply invite defeat in this context. In noncompetitive environments, however, other outcomes are clearly possible. Since the competitive case has received primary attention in the literature, it is discussed first.

The median voter characterization of public sector outcomes has been used extensively in the analysis of state and local government spending patterns. From Figures 1 and 2, the competitive majority rule outcome (E_i^*) represents one point on the median citizen's demand curve. By relating public service expenditures in a sample of jurisdictions to the median voter's tax price, income, and other attributes expected to influence demands, the possibility of estimating an entire demand function emerges. Early attempts to apply this notion were limited to testing qualitative implications of the theory, e.g., that increases in the size of the non-residential component of the tax base reduce tax prices to all voters, and hence lead to larger expenditures (Barr and Davis 1966). Although the empirical models were somewhat loosely specified, the results obtained were sufficiently promising to encourage further research. The empirical methodology was refined considerably in a group of studies designed specifically to estimate parameters of demands for public services (Borcherding and Deacon 1972, Bergstrom and Goodman 1973). In each case, the median citizen's demand was assumed to be a log-linear (linear in logarithms) function of median income and median tax price in a jurisdiction. Specification of tax price usually presents the most difficult empirical task in such studies. An individual's tax price for a public service depends directly upon the cost of producing public output, and upon the share of total taxes paid by the citizen in question (see equations (2) and (3)). Empirically, variations in public service costs are usually related to prices of inputs (Borcherding and Deacon 1972, Deacon 1978). If services are financed by a property tax, then median tax shares can be estimated from information on median housing values and total assessed valuation in each jurisdiction (assuming that the median citizen owns and occupies a house of median value). Application of this methodology to state and local government expenditures has yielded estimates of price and income elasticities that are typically significant and of expected sign. Moreover, a degree of consensus has emerged regarding the actual magnitudes of these effects. Demands are almost invariably found to be price inelastic, with elasticities usually ranging from .2 to .6. Income elasticities are typically positive, but

less than unity, e.g., between .4 and 1.0.⁵

Identifying the relevant demand attributes for the median citizen in a jurisdiction is particularly difficult. If, for example, public service demands vary with age, the level of education, or the number of children in a household, the assumption that median demands can be uniquely related to median income and tax price alone, may well be inappropriate. Bergstrom and Goodman (1973) have developed a very useful theorem in this regard. Within the context of log-linear demands and certain regularity properties for income distributions, it provides conditions that are sufficient to specify the form of the median demand in a jurisdiction.⁶

The attempt to model demands in a public choice framework has shed light on a variety of related issues. For example, results obtained by Peterson (1975), Bergstrom and Goodman (1973) and others, are consistent with the proposition that renters perceive themselves as bearing only a fraction of the tax liability on rental property they occupy. Others have noted relationships between spending and the composition of the tax base that indicate the degree to which voters expect taxes levied on non-residential property to be borne by outsiders (Ladd 1975). These results are particularly important for public finance theory because they imply that questions regarding tax incidence and other distributive aspects of public policy cannot really be divorced from resource allocation issues, once a collective choice process is introduced. In this context, policies designed simply to redistribute tax burdens in the population, e.g., "circuit breaker" tax relief, or equalization formulas for school finance, may alter public service levels in ways that are, perhaps, predictable (Inman 1978).

In summary, a large body of theoretical and empirical research has been conditioned upon the median voter hypothesis. The fact that a measure of agreement in empirical results has emerged from diverse studies lends some support to the underlying paradigm.⁷ However, the model has by no means escaped criticism (Romer and Rosenthal 1980b). To many, the level of abstraction it employs is stratospheric; moreover, the basic premise that communities actually attain Hotelling-Bowen-Downs equilibrium has largely eluded direct tests. One source of concern is the informational requirements that the model imposes upon voters. Several authors have raised the possibility that complexity in taxing institutions may cause voters to incorrectly perceive tax liabilities. If such a misperception takes the form of "fiscal illusion," a systematic underestimation of tax bills, the effect would be to increase expenditure beyond the median voter's "fully informed" optimum. To date, however, attempts to shed empirical light upon this proposition have been rather inconclusive (Wagner 1976, Clotfelter 1978).⁸

⁵There are, however, notable exceptions. Public parks and recreation services often exhibit income elasticities greater than 1.0; estimated income elasticities for public welfare expenditures are sometimes negative.

⁶For further discussion of this theorem, and an empirical examination of the assumptions underlying it, see Inman (1978).

⁷Moreover, Peterson (1975) and Rubinfeld (1977) summarize evidence to indicate that demand estimates from median voter expenditure studies are in close agreement with results obtained from voting data.

⁸As Bergstrom and Goodman (1973) point out, a powerful law of large numbers applies to medians as well as to means. As a consequence, purely random errors in perception of tax liabilities or service levels tend to cancel out.

A methodologically more telling criticism concerns the fact that the median voter model minimizes the role of interest groups, and assigns politicians and bureaucrats the passive role of seeking out and supplying the demands of the voting middle class. This is essentially an assumption of competition in the political process, and it has been challenged by a number of writers who take a much more monopolistic view of the public economy. Niskanen (1971) was the first to develop a complete theory along these lines, and more recent attempts to formulate models of noncompetitive equilibrium in the public sector have adopted his basic approach.

According to Niskanen, government bureaus enjoy considerable monopoly power as sole suppliers of public services. Consequently, they can control the range of alternatives offered to the citizenry, or their elected representatives. Comparing market relationships in public versus private sectors, "the primary difference . . . is that the [government] bureau offers a total output in exchange for a budget, whereas a market organization offers units of output at a price" (Niskanen 1971, p. 25). In effect, bureaus have the power to confront their constituents with "all or nothing choices." In the context of a majority rule system, this amounts to a limited control over the agenda, or range of alternatives facing voters.

In Figure 2, individual i would receive utility of $U_i(0)$ if no public output were provided. If faced with an all or none offer, both i and j (a simple majority) would prefer expenditure levels up to E_N to the prospect of receiving no public service at all. By restricting the menu of available alternatives to a single pair, one of which involves no service, the monopoly bureau can effectively select any expenditure level between 0 and E_N . By way of characterizing the bureau's motives, Niskanen offers salary, perquisites of office, prestige and power as plausible maximands, but argues that in general these will be positively related to the size of the total budget. With budget maximization as an implicit goal, the Niskanen equilibrium expenditure level is E_N in Figure 2.⁹

Niskanen's budget maximization hypothesis has been recently extended by Romer and Rosenthal (1980a). In their model, the budget setting power of bureaucrats is somewhat attenuated, but significant nonetheless. The distinctive features of their approach hinge upon the existence of a "reversion level" to which public expenditures will return if the government's budget proposal is not approved. The reversion level is determined outside of their model, presumably at the constitutional stage. Returning to Figure 2, suppose the reversion level of spending is E_0 ; in effect, this is the bureau's minimum budget offer. Correspondingly, $E_R(E_0)$ is the maximum expenditure level that could gain a majority over E_0 , and thus represents equilibrium in the Romer-Rosenthal characterization. Introduction of a reversion level into the budget maximization framework yields a novel implication; in the absence of uncertainty, equilibrium expenditure is inversely related to the reversion level. "By facing the voters with a 'take-it-or-leave-it' choice, the setter [bureaucrat] exercises a threat over the voters. The worse the status quo [reversion level], the greater this threat, and consequently, the greater the gain to the setter [bureaucrat] from being able to propose the alternative" (Romer and Rosenthal 1978, pp. 35-36). Of course, this relationship only holds for reversion levels that are below the median optimum, E_i^* ; beyond that point, the equilibrium budget is

⁹It seems clear that Niskanen intended his model to apply primarily to federal governments. In any case, the degree of oversupply certainly depends upon the elasticity of demand and, hence, upon the availability of substitute services. At the local government level, the possibility of migration among jurisdictions presents substitution possibilities that are exceedingly costly at the federal level.

trivially equal to the reversion level.

Designing tests to distinguish between competitive and non-competitive political models is difficult at best. As is the case in the private sector, qualitative implications from competitive models are often fully consistent with non-competitive behavior as well. For example, in either Niskanen or Romer-Rosenthal versions of budget maximization, equilibrium public service levels are inversely related to service costs and directly related to incomes. The introduction of a reversion level in the Romer-Rosenthal framework, and its hypothesized effect upon budgets does, however, present an opportunity to separate the two models, since reversion levels should have little or no effect upon budgets in a competitive environment. Apparently the first attempt to examine this question empirically was carried out by Holcombe (1980). In Michigan, school boards have authority to levy property taxes up to a certain rate (which varies across districts) without approval of the electorate. Tax rates in excess of this statutory limit must be approved by majority vote. This system bears a very close resemblance to the Romer-Rosenthal non-competitive model with a reversion level. Holcombe (1980) presents anecdotal evidence to indicate that some districts attempted to exercise monopoly power by making non-incremental offers to the voters.¹⁰ On average, however, he found no evidence that budgets exceeded competitive levels. Intuitively, Holcombe (1980) found that excess levies were generally approved by substantial majorities (65.3 percent voting in favor, on average). Thus, school boards could evidently have obtained much larger school budgets (with smaller majorities) than were observed. More formally, Holcombe (1980) made certain simplifying assumptions regarding the general shape of individual preferences for alternative budgets, and assumed that the distribution of preferred budgets was normal within each district. Using these assumptions, together with information on election returns and the actual alternatives offered in school levy referenda, he was able to estimate the median preferred expenditure level in each school district. Comparing his estimates to actual budgets in over 200 districts, he found that observed budgets deviated from median preferred levels by a mere 2.4 percent on average. Ironically, actual levies were below his estimated median points, but the difference was not statistically significant.

The institutions of school finance referenda in Oregon are quite similar to those in Michigan. Romer and Rosenthal (1980) conducted a test of their model using data on educational expenditures and reversion levels in Oregon school districts. A negative relationship between the two variables was expected. After controlling for factors such as income, tax price, and state subventions to local districts they found that reversion levels were significantly related to school outlays, but the direction of the effect was not monotonic. Except for school districts with extremely low reversion levels, actual budgets were positively related to reversion levels, contrary to expectations.

Some additional indirect evidence on the budget maximization hypothesis has been obtained from comparisons of spending levels in jurisdictions where expenditures are set by elected representatives rather than by direct democracy. As Romer and Rosenthal point out, "the possibilities of both monopoly power acquired by elected officials and coalition of minority politics suggest that expenditures will be greater in representative democracies [than in direct democracies]" (1979b, p. 161). Both McEachern (1978) and Holcombe (1980) have examined this hypothesis

¹⁰ In one district, the reversion level was a tax rate of 6.0 mills, and the school board offered an alternative of 16.75 mills. The measure passed by a scant 1.1% (Holcombe 1980, pp. 267-268).

empirically. In the former study, debt levels in states that require a simple majority referendum to increase debt limits were compared to debt in states with no such requirements. After controlling for certain variables that might independently affect debt levels, no significant difference in per capita debt in the two groups of states was found. In a similar fashion, Holcombe (1980) compared per student educational expenditures in states that require spending referenda, to states where expenditure levels are chosen by elected officials; again, no significant differences were noted.

Finally, both Citrin (1979) and Courant, *et al.* (1980) conducted surveys of voting age citizens in states considering tax limitation referenda (California and Michigan). When asked about their preferences for increasing or decreasing government expenditures and taxes by various amounts, it was found that the median respondent desired no change; this was true not only for total spending, but for virtually all of the individual services surveyed. However, the observation that responses were consistent with competitive public sector equilibrium is rather surprising in this case, since both measures were adopted. This apparent anomaly may simply indicate the potential pitfalls in survey research, or imply that closer examination of the specific referenda in question is warranted.

Concluding Comments

The attempt to distinguish competitive from non-competitive outcomes in the public sector is clearly important. This is true for essentially the same reasons that support the study of market power in the private economy; competitive markets attain a degree of efficiency not present in monopolistic regimes. When considering the question of efficiency in the public sector, at least two concepts are important. The first, sometimes called "process efficiency," concerns the degree to which inputs are combined in a cost minimizing fashion. This topic has not been addressed in the present paper, though numerous studies have concluded that, due to the nature of the rewards system, public enterprises are not efficient in this sense. The second concept concerns whether, given the structure of costs, the correct level of public output is produced; it is this notion of efficiency that is of primary interest here. In a general sense, the level of output is "correct," if the marginal benefits of additional output just offset marginal costs. Under certain, not terribly unrealistic conditions regarding the distribution of demands and tax shares in the populace, the competitive Bowen equilibrium is efficient, or at least approximately so, in this sense (Bowen 1943).

From the empirical analysis presented to date, no strong indication of non-competitive budget maximization has emerged. However, given the limited number of attempts to test this hypothesis, final conclusions are clearly premature. In any case, the current degree of polarization in views of the public sector, with some analysts apparently believing that all outcomes are competitive and others that none are, seems inappropriate and counterproductive. It seems reasonable to speculate that the evidence on this question will eventually show some elements of monopoly power in public organizations, at least under certain conditions, as is the case in private markets. If so, the relevant questions will concern the degree of monopoly power observed and the conditions under which it is likely to occur (e.g., at which levels of government, for what kinds of services, under which types of institutional structure). With a bit of imagination, it is even possible to envision development of a public sector analog to the traditional study of industrial organization in private markets.

References

- Barkume, Anthony J. (1978), "Tax Prices and Voting Behavior: The Case of Local Educational Financing," Economic Inquiry, 16.
- Barr, J. L. and Davis, O. A. (1966), "An Elementary Political and Economic Theory of Local Governments," Southern Economic Journal, 33, 149-65.
- Barzel, Y. and Silberberg, Eugene (1973), "Is the Act of Voting Rational?" Public Choice, 19, 51-58.
- Barzel, Y. and Deacon, Robert (1975), "Voting Behavior, Efficiency, and Equity," Public Choice, 21, 1-11.
- Bergstrom, Theodore C. and Goodman, R. P. (1973), "Private Demands for Public Goods," American Economic Review, 63, 280-96.
- Borchherding, Thomas E. and Deacon, Robert T. (1972), "The Demand for the Services of Non-Federal Governments," American Economic Review, 62, 891-901.
- Bowen, H. R. (1943), "The Interpretation of Voting in the Allocation of Economic Resources," Quarterly Journal of Economics, 58, 27-48.
- Buchanan, J. M. and Tullock, G. (1962), The Calculus of Consent, Ann Arbor.
- Buchanan, J. M. (1975), "Public Finance and Public Choice," National Tax Journal, 28, 383-94.
- Chapman, Randall G. and Palda, Kristian S. (1980), "Voting Participation in a Public Consumption Perspective," Advances in Consumer Research, Vol. VIII.
- Citrin, Jack (1979), "Do People Want Something for Nothing: Public Opinion on Taxes and Government Spending," National Tax Journal, 32, 113-130.
- Clotfelter, Charles T. (1977), "Public Spending for Higher Education: An Empirical Test of Two Hypotheses," Public Finance, 8.
- Courant, Paul N., Gramlich, Edward M., and Rubinfeld, Daniel L. (1980), "Why Voters Support Tax Limitation Amendments: The Michigan Case," National Tax Journal, 33, 1-20.
- Deacon, Robert T. (1977), "Private Choice and Collective Outcomes: Evidence from Public Sector Demand Analysis," National Tax Journal, 30, 271-387.
- Deacon, Robert T. and Shapiro, P. (1975), "Private Preference for Collective Goods Revealed Through Voting on Referenda," American Economic Review, 65, 943-55.
- Deacon, Robert T. (1978), "A Demand Model for the Local Public Sector," Review of Economics and Statistics, 50, 185-93.
- Downs, A. (1957), An Economic Theory of Democracy, New York.
- Gramlich, Edward M., Rubinfeld, Daniel L., and Swift, Deborah (1980), "Voter and Non-Voter Preferences for Public Spending and Spending Limits," Institute of Public Policy Studies, University of Michigan.
- Holcombe, Randall G. (1980), "An Empirical Test of the Median Voter Model," Economic Inquiry, 18, 260-74; and preliminary mimeo with same title, Texas A. and M. Univ., 1976.
- Hotelling, Harold (1929), "Stability in Competition," Economic Journal, 29, 41-57.
- Inman, Robert P. (1978a), "Optimal Fiscal Reform of Metropolitan Schools: Some Simulation Results," American Economic Review, 68, 107-22.
- Inman, Robert P. (1978b), "Testing Political Economy's 'As If' Proposition: Is the Median Income Voter Really Decisive?," Public Choice, 33, 45-66.
- Ladd, Helen F. (1975), "Local Education Expenditures, Fiscal Capacity, and the Composition of the Property Tax Base," National Tax Journal, 28, 145-58.
- McEachern, William A. (1978), "Collective Decision Rules and Local Debt Choice: A Test of the Median Voter Hypothesis," National Tax Journal, 31, 129-36.
- Niskanen, William (1971), Bureaucracy and Representative Government, Chicago: Aldine-Aherton.
- Noam, Eli M. (1979), "Tax Limitations and Fiscal Federalism," National Tax Journal, 32, 181-88.
- Peterson, George M. (1975), "Voter Demand for Public School Expenditures," in Public Needs and Private Behavior in Metropolitan Areas, John E. Jackson (ed.), Cambridge, 99-115.
- Romer, T. and Rosenthal, H. (1978), "Political Resource Allocation, Controlled Agendas, and the Status Quo," Public Choice, 33, 27-45.
- Romer, T. and Rosenthal, H. (1979a), "Setters, Reversions, and Median Voters: Evidence from School Expenditure Referenda," Carnegie-Mellon University, Pittsburgh.
- Romer, T. and Rosenthal, H. (1979b), "The Elusive Median Voter," Journal of Public Economics, 12, 1-27.
- Rubinfeld, Daniel L. (1977), "Voting in a Local School Election: A Micro Analysis," Review of Economics and Statistics, 59, 30-42.
- Sen, A. K. (1968), Collective Choice and Social Welfare, San Francisco: Holden-Day.
- Sonstelie, Jon (1979), "Public School Quality and Private School Enrollments," National Tax Journal, 33, 343-54.
- Wagner, Richard E. (1976), "Revenue Structure, Fiscal Illusion, and Budgetary Choice," Public Choice, 31.

SUBSTITUTION BETWEEN PUBLIC AND PRIVATE GOODS: AN OVERVIEW
OF THE MARKET MEETING CONSUMER PREFERENCES¹

Charles L. Vehorn, U.S. General Accounting Office and Georgetown University

Abstract

It has often been suggested that a consumer may substitute private goods for public goods that are not provided at the consumer's desired level, but until recently research had not been undertaken on the degree of substitution. This paper presents an overview of the public-private substitution question, dealing with how researchers conceptualize the issues and how they attempt to observe consumers registering preferences.

Introduction

Many local officials make budgetary decisions based on the implicit assumption that the government has a monopoly on the provision of standard public services. However, this assumption that the government alone possesses the means to meet the preferences of citizen-consumers is often inaccurate. Instead, a number of private substitutes for public goods exist, although they may not be perfect substitutes. Thus, local decisions based on the monopoly assumption can lead to an inefficient allocation of public resources.

One fundamental difference between public and private provision is the way in which goods and services are rationed. When goods and services are provided in the private sector consumers and producers signal each other through the market process. If consumers, for example, desire more of a particular good, they bid up the price letting suppliers know it is advantageous to increase production. Consumers register changes in their preferences through this market mechanism and resources are allocated more efficiently to meet those new preferences.

In the public sector, on the other hand, middlemen drive a wedge between consumers and producers. These middlemen, whether elected or appointed public officials, have the authority to control the production process through their budgetary decisions. Thus the amount provided may vary from the amount that would have been supplied under market rationing. Because of non-market rationing and the collective nature of public goods many citizens are left dissatisfied, especially in communities where voters have heterogeneous preferences. This dissatisfaction creates incentives for a private market to spring up and fill the void. The mere existence of private substitutes for public goods indicates the market is working to meet consumer preferences.

The primary purpose of this paper is to present an overview of the public-private substitution process. First, a Lancasterian perspective is employed to provide an underlying conceptual framework for discussing the potential range of substitution (1966). Then a brief review of collective goods and their demand is given to introduce a more detailed treatment of recent attempts to observe consumers registering their preferences between public and private goods in the markets for police protection, fire protection, and education. The paper concludes by pointing out how further research in this area could provide information to aid public policy makers,

¹

The author wishes to thank Gary C. Anders for his helpful comments and suggestions. Views expressed in this paper are solely the author's and not necessarily the GAO's.

who must decide how to allocate the public's scarce resources efficiently.

The Range of Substitution between Goods

In a narrow sense of the word, substitution is simply the process of purchasing one commodity in place of another. This definition, however, fails to account for all those commodities that are not perfectly substitutable. Many commodities are closely related, yet in the eyes of the consumer not perfect substitutes, or for that matter perfect complements. It is perhaps useful to view pairs of goods as lying on a continuum according to how closely they are related. At one extreme are perfect substitutes, at the other perfect complements. In the middle are independent commodity pairs. In between either extreme and the middle is a wide range of closely related goods with some degree of substitutability or complementarity. Consumers do not make an either-or choice with these closely related goods, but select a commodity bundle with a desired mix, depending on prices, income, and tastes.

A key and sometimes neglected issue in discussing substitutability is the definition of a commodity. The traditional approach is to define a commodity as simply the physical item or service purchased -- a hamburger, a shirt, a glass of beer. If consumers desire commodities defined in this general way, it becomes difficult to explain why some consumers will drink only one particular brand of beer. If a beer is a beer, then there should be perfect substitution between brands. However, each brand of beer has its own particular taste, so consumers who are loyal to only one brand of beer are saying no other taste can substitute for the taste of my brand of beer. Thus the traditional approach to consumption overlooks the relationship between the purchase of goods and the purchase of attributes that are inherent in the goods.

A new approach to consumption, formally developed by Lancaster, specifically recognized that goods purchased in the market usually yield more than one well defined service. So it is reasonable to argue that consumers desire the satisfaction of a commodity's services or attributes when they make a market purchase. When consumers purchase housing they desire attributes like space, in the number and size of rooms, and location, with easy access to schools, transportation routes, and recreational facilities. Another example of a good with multiple attributes is the running shoe. When runners purchase shoes they are buying a series of attributes, some of which have only become available in the last few years through technological innovation. The attributes include flexibility, sole wear, rearfoot impact, forefoot impact, weight, rearfoot control, and traction.

Thus the new approach to consumer theory made attributes the object of choice by simply placing desired attributes instead of commodities into the utility function, and assuming an additive linear relationship between the two,

$$z_i = \sum_j b_{ij}q_j \quad (1)$$

where z_i = the attributes desired by consumers
($i = 1, \dots, m$)

q_j = the quantities of public and private goods ($j = 1, \dots, n$)
 b_{ij} = the amount of attribute i possessed by a unit of commodity j .

Public Goods and Consumer Demand

It can be shown that the new approach includes the traditional approach by writing equation (1) in matrix form

$$Z = BQ \quad (2)$$

where Z and Q are $(m \times 1)$ and $(n \times 1)$ column vectors respectively, and B is the $(m \times n)$ consumption technology matrix. Now if $m = n$ and the B matrix is diagonal with $b_{ii} > 0$ for all i and $b_{ij} = 0$ for $i \neq j$, then the new approach collapses to the traditional approach with a one-to-one relationship between z_i and q_j .

One important insight of the new approach to consumer theory is that changes in relative prices may result in a switch from one bundle of commodities to another because the former becomes an inefficient way to attain the desired attributes. Lancaster called this the efficiency substitution effect, which is different from, but reinforces the personal substitution effect in the traditional theory of consumption.

The efficiency substitution effect may be operative in substitution between public and private goods. Because of the nature of collective goods, which will be discussed more thoroughly in the next section, individual voters may desire more or less of the public service, i.e., under the new approach more or less of the public good's attributes. This dissatisfaction with public provision leaves the consumer with basically four options. The consumer can form a lower-level government, migrate from the locality, seek help from voluntary organizations as extra-governmental providers of collective goods, or resort to private market alternatives. The first three options imply high transaction costs, but the last option has relatively low transaction costs associated with the search for an acceptable substitute.

Even though private market alternatives exist, they are often far from perfect substitutes for public provision. In fact, as the following examples supplied by Weisbrod indicate, these goods are generally different in form from public goods (1975). If the provision of clean air is inadequate one can substitute air filters and purifiers in the home or work place. Consumers or firms can also substitute sprinkler systems for public fire protection; and locks, alarms, guns, watchdogs, or guards for public police protection.

Public police protection is an interesting example of a good with several attributes. Consumers desire crime prevention, protection activity, and a feeling of security. If the public sector does not provide the desired amount of these attributes, citizen-consumers can purchase various goods in the private market to increase protection and provide a better sense of security. These private market alternatives may be poor substitutes for policemen when it comes to apprehending a criminal, but a watchdog or burglar alarm enhances the sense of security. The private-good substitutes may be costly, but they give the consumer a degree of personal control, which is lacking in most public goods. The purchaser directly controls the watchdog or alarm for his own personal well-being. Collective goods by their non-excludable nature cannot provide this sense of individual ownership and control. In general, there are few private goods that are perfect substitutes for public goods, but the range of substitutability is wide and many private goods possess the same attributes found in public goods.

Public goods are distinct conceptually from private goods because public goods have a non-excludable and a non-exhaustible nature. A private country club, for instance, can exclude consumers from enjoying its recreational facilities, but a public park is shared by all citizens who desire to use it. The recent influx of Cuban refugees into Florida certainly caused a temporary strain on many private goods like food and clothing as suppliers' inventory became exhausted, but additional consumers cannot exhaust the supply of a "pure" public good, like national defense.

Most local public goods, however, have both private and public good characteristics. The use of a road is shared with other drivers, and due to congestion, a certain rivalness in consumption occurs. A public fire department does not exclude homeowners from protection, but it cannot simultaneously extinguish fires in separate locations because its resources are constrained. Thus studies of the demand for local public goods must incorporate this aspect of congestion into the analysis. The way Borcharding & Deacon (1972), Bergstrom & Goodman (1973), and others did this was to use a median voter demand model specifying the public good in terms of a crowding function.

Consider the total physical quantity of any local public good, q^* , which is subject to a certain degree of crowding due to the size of the population, N . Specify the crowding function as

$$q = q^*N^{-\alpha} \quad 0 \leq \alpha \leq 1. \quad (3)$$

If $\alpha = 0$, q^* is a Samuelsonian pure public good where every individual consumes the total quantity available. As α approaches 1 individuals are not able to enjoy the full amount, but must share q^* with others. When $\alpha = 1$, the individual receives the fraction $1/N$ of the total provision. Estimates of the crowding parameter for most local public goods have been closer to one than zero, indicating a low degree of "publicness."²

These estimates of crowding are consistent with another type of substitution, which is mentioned here but not pursued at length. Increasingly, the private sector is directly providing local public services. Some local governments have found it appropriate to contract out to the private sector for services like fire protection or garbage collection. The usual argument, most recently confirmed by Bennett and Johnson, is that private delivery of public services lowers costs, since the private sector operates under incentives that encourage cost effective provision (1979). Although the growth in "privatization" of public goods is an important occurrence, the substitution here is not in consumption but in the production process, which is beyond the scope of this paper.

In addition to crowding, the median voter's demand for a local public good is dependent upon income, y , the tax price or tax share, p_t , the prices of related goods (substitutes and complements), p_k , and tastes, T ,

$$q = f(p_t N^{-\alpha}, y, p_k, T). \quad (4)$$

Although many local public good demand studies recognized

²

However, there is some reason to believe that these estimates of "publicness" might be biased (King 1977).

the conceptual importance of private-good substitutes, empirical estimation of the degree of substitutability was not usually undertaken due to data unavailability. Research to shore up this empirical gap has just begun. The next section reviews recent studies that attempted to quantify the degree of substitutability.

Estimates of Public Good-Private Good Substitutability

Police Protection

Ann Bartel analyzed the demand for private protection by firms (1975). The main purpose of her study was to answer the following three questions. One, how is firm demand for protection related to business losses from crime and the probability of crime? Two, are public and private expenditures substitutes or complements? Three, does a firm choose self-protection as a substitute for market insurance, or will it spend more on protection if it has insurance?

She developed an extensive theoretical model that took into account expected losses and risk aversion, but was unable to test all of the model's implications due to data limitations. However, she was able to answer the three main questions through the use of logit analysis. A survey conducted by the SBA provided data on whether or not private firms hired guards. These data became the dependent variable in the logit analysis.

She found that firms do respond rationally to economic forces in making their protection decisions. Firms placed relatively more weight on the probability of loss than the size of anticipated losses. Also firms were willing to reduce their demand for private protection when public police protection was increased, so the two types of protection are substitutable. It appears that firms have a target level of desired protection, which is a mix between public and private protection services. Insurance, however, was not a substitute for private protection because firms could purchase insurance at a lower price if they hired private guards or protective services.

Lancaster's new approach to consumer theory is helpful in understanding Bartel's results. The attribute that firms desire is a certain level of protection at the lowest cost. So firms combine public services and private services until they obtain their desired level of protection, while at the same time judiciously choosing the mix of private protection services to minimize costs.

Clotfelter employed a conceptual approach similar to Lancaster's, except he viewed the attribute as the final output and commodities as intermediate goods (1977). This approach allowed Clotfelter to formulate the private-public choice issue in terms of a production function

$$z = f(q*N^{1-\alpha}, q') \quad (5)$$

where z = the ultimate output or attribute
 $q*N^{1-\alpha}$ = the total amount of the public good adjusted for crowding
 q' = the total amount of the private good.

Within the production framework, output can be maximized by equating the marginal rate of technical substitution between public and private inputs to the ratio of their marginal costs. This type of model indicated that the question of substitution depended on the relative costs of inputs, the state of technology, the public good characteristics of the public input, and the public institutions that decide how to raise revenue and distribute benefits.

Clotfelter's next step was to develop an empirical test that would measure the degree of substitutability between a public and private good. He chose protection against crime as the final output, z , with employment in public and private protection as the inputs. Using a CES production function, he estimated the elasticity of substitution under two assumptions. First, he assumed that the wage rates for public and private policemen were exogenous and found, through ordinary least squares on cross-sectional state data, that the elasticity of substitution was positive, .58, implying a low degree of substitutability. Then he assumed that the wage rates were endogenous and used the instrumental variable estimation technique. Under this assumption the elasticity of substitution was much larger, 2.47, suggesting considerable latitude for substitution when input prices change. An elasticity of substitution above one implies the demand for private substitutes will increase proportionately more than public wage increases, thus decreasing the relative share of public spending on a given activity. Clotfelter's estimate, 2.47, was significantly greater than zero, but the standard error was too large to confirm that the estimate was also greater than one. He concluded by suggesting that other forms of private protection besides protective employment merit consideration if we are to understand how the consumer achieves a desired level of protection.

Consumer use of various means of private security measures like locks, alarms, handguns, etc. was examined in a more recent article by Clotfelter (1978). Here he analyzed various effects from increased consumption of private protection devices. One obvious effect from increases in private protection activity is the deterrence of crime. Criminals will have to search longer and their expected returns will be smaller because of private protection activity. This gross deterrence effect, however, may be partially offset by what is termed as the isolation effect. Increases in private protection activity may erode the citizen's social responsibility for routine surveillance and reporting of suspicious activity. In addition to the gross deterrence and isolation effects, private protection activity by some could simply divert crimes to other unprotected households. This is termed the displacement effect.

Lack of appropriate data restricted Clotfelter's empirical analysis of the various effects of private protection on crime levels. Also he was unable to obtain any data regarding market interaction between private protection activity and public protection.

Fire Protection

Vehorn extended the median voter approach (Equation 4) to the demand for public goods by examining market interaction between publicly-provided and privately-provided goods for the case of fire protection (1979). The long standing difficulty, as mentioned previously, has been lack of data. In the case of fire protection, however, it was possible to obtain private good price data for homeowners insurance and fireproofing materials in different SMSAs, allowing for a cross-sectional analysis. Since public fire protection is a shared good, each consumer must make his own personal choice about whether publicly provided fire protection is his optimal amount of fire protection. Hirsch hypothesized that a substitute for public protection, to a certain degree, is the private market good insurance (1970). Vehorn conducted an empirical test of Hirsch's conjecture by estimating cross-price elasticities between public and private services in the fire protection market using a simultaneous equations framework. A simultaneous equation model was appropriate because the price of insurance

along with the firemen's wages are endogeneous factors.

In order to test for the existence of market interaction, the signs and statistical significance of the cross-price elasticity estimates had to be considered. If these elasticity estimates were positive, then it could be concluded that increases in private-good prices lead to increases in the demand for public fire protection, ceteris paribus. Vehorn found that the cross-price elasticity estimates for both insurance and fireproofing materials were positive and in most cases significant. The range for insurance was from .44 to .81 depending on how the tax share was measured and whether the observations were for SMSAs or central cities. The range for fireproofing materials was from 1.48 to 1.82.

The evidence that Vehorn presented showed price changes in private markets do indeed effect the voter's demand for public goods. His general conclusion was that in areas where the prices of private market alternatives were high the demand for public fire protection was also high.

Education

Erekson also used a median voter model to examine the interrelationship between public and private education (1978). In 1976-77, 16.8 percent of all elementary and secondary schools were non-public, so private education does provide a viable market alternative to consumers. The question remains, what is the effect of private education on the provision of public education?

One expected effect in public schools is an increase in per pupil expenditures as non-public schools attract more students. Offsetting this effect, however, is the taxpayer's decreased support for public education as more students switch to non-public education. Erekson's empirical results support the former effect, but he could not confirm the existence of the latter effect for his sample of 671 school districts in the state of New York. He also estimated a cross-price elasticity between the demand for public education and non-public tuition rates. This estimate was positive and significant in most cases indicating that private education was indeed a substitute for public education. The magnitude of the estimate, however, was small suggesting that public and private education are far from perfect substitutes. Perhaps the vast difference in the quality dimension restricts the degree of substitutability. Erekson concluded by noting that any attempt to model the demand for public education must take into account the market for private education.

Summary and Conclusions

The purpose of this paper was to present an overview of consumer substitution between public and private goods. Initially, one might think that such substitution is impossible due to the unique nature of collective goods. However, if one views the consumer as a utility maximizer who chooses between attributes rather than physical commodities, it becomes apparent that a wide range of substitution possibilities exist.

Literature in the area of public-private substitution is sparse, but growing. In general, these studies have been able to observe a moderate degree of substitution occurring. Thus citizen-consumers do have other alternatives to public provision, and some do make use of these alternatives whenever they feel that the public provision does not meet their own personal needs. However, more detailed research on specific private markets is needed to further our understanding of exactly how these markets spring up to meet consumer preferences.

Local public officials who recognize the existence of private market alternatives, instead of assuming that the public sector is a monopoly provider, can use these private markets to their advantage. Encouraging the growth of private markets to help meet the needs of citizens with diverse preferences can reduce the pressure to increase public services without expanding the budget. Budgetary pressures, in part, have been the impetus in the move towards "privatization" of public goods, and tight budgets could be the stimulus for an effort to expand the use of private market alternatives. Public sector resources could be allocated more efficiently if research provided local officials with estimates of consumer responses to such an effort.

References

- Bartel, Ann P. (1975), "An Analysis of Firm Demand for Protection Against Crime," Journal of Legal Studies, 4, 443-478.
- Bennett, James T. and Johnson, Manuel H. (1979), "Public Versus Private Provision of Collective Goods and Services: Garbage Collection Revisited," Public Choice, 34, 55-63.
- Bergstrom, Theodore C. and Goodman, Robert P. (1973), "Private Demand for Public Goods," American Economic Review, 63, 280-296.
- Borchering, Thomas E. and Deacon, Robert T. (1972), "The Demand for Services of Non-Federal Governments," American Economic Review, 62, 891-901.
- Clotfelter, Charles T. (1978), "Private Security and the Public Safety," Journal of Urban Economics, 5, 388-402.
- Clotfelter, Charles T. (1977), "Public Services, Private Substitutes, and the Demand for Protection Against Crime," American Economic Review, 67, 867-877.
- Erekson, O. Homer (1978), "An Estimate of the Impact of Private Education on the Provision of Public Education," Paper presented at the Southern Economic Association Meetings.
- Hirsch, Werner Z. (1970), The Economics of State and Local Government, New York: McGraw-Hill.
- King, A. Thomas (1977), "Why Public Goods Look Like Private Goods," Mimeo.
- Lancaster, Kelvin J. (1966), "A New Approach to Consumer Theory," Journal of Political Economy, 74, 132-157.
- Vehorn, Charles L. (1979), "Market Interaction Between Public and Private Goods: The Demand for Fire Protection," National Tax Journal, 32, 29-39.
- Weisbrod, Burton A. (1975), "Towards a Theory of the Voluntary Non-Profit Sector in a Three Sector Economy," in E. S. Phelps (ed.), Economics and Altruism, New York: Russell Sage Foundation.

CONSUMER INFORMATION ABOUT PUBLIC GOODS
AND THE WORKABILITY OF THE WELFARE STATE¹

Martin Pfaff, University of Augsburg and INIFES²
Ernst Kistler, INIFES

Abstract

The modern welfare state is faced by fiscal crises. Commonly scientists consider an inflation of wants the principal cause thereof. Consumer information on the public transfer system as one part of a most urgently needed consumer policy for the public sector can on the one hand cause more demand, it can help, on the other hand, to create a more rational and effective transfer system. This paper deals with the consequences of an extension of information on public goods and services. It concludes that increased understanding of and more information about public transfers is needed on normative as well as practical grounds.

The Fiscal Crisis of the Welfare State

In capitalist as well as in socialist countries discussions have arisen about the efficiency of social policy as it is practiced. Inefficiency and ineffectiveness, suboptimality in allocational, distributional, and aggregative matters have been diagnosed. During the period of world wide prosperity and growth these effects could be tolerated more readily; however, during the last few years there are definite symptoms indicating limits in the willingness of financing such inefficiencies. The reduction in economic growth thus has led to suggestions such as privatization. Not only the New Left speaks of such boundaries in financial means and of the workability of the public sector. Today social policy, in its dependence on tax revenues is faced with the sheer impossible task reconciling actual or perceived wants with available economic and social resources. Irrespective of the actual causes thereof we have to realize the dangers brought about by excessive burdens on the public transfer system. The suggested possible alternatives need further investigation. The supply of public goods, however, is not only limited in terms of economic feasibility. There are furthermore doubts, whether the goods and services provided really meet the true needs and wants of the consumer. Empirical studies show (e.g. Katz et al. 1975) that frequently people do not or cannot realize their claims on certain real or monetary grants. Other urgent problems are the growing tendencies towards bureaucratic failure and the lack of critical participative and complaining behavior by public sector consumers.

Some social scientists (Scherhorn 1975, Young 1978) suggest an introduction of consumer policy instruments into the public sector and increased education, information, and participation of consumers may solve the aforementioned problems to a large extent (INIFES 1979). This short paper cannot deal with the limitations and advances of a consumer-oriented perspective of public transfer policy. It is confined to the very specific aspect of consequences of increased consumer information for the public budget

¹ Research supported by the Federal Ministry of Research and Technology (VER006).

² International Institute for Empirical Social Economics (INIFES), Haldenweg 23, 8901 Leifershofen, Federal Republic of Germany.

The Problems

Two antithetic theses can be stated when we discuss the effects of consumer information:

- 1) A theory of inflation in aspiration and demand, or
 - 2) A theory of more effectiveness of public grants.
- Unquestioned, however, by proponents of both points of view and irrespective of secondary effects, we expect that more and better informed consumers imply an increase in consumer sovereignty.

An inflation in demand, leading to an overburdening of the public budgets, must be expected when we proceed from the principles of traditional economic psychology and assume strictly egoistic behavior. Consumers or representatives of interest groups in the wake of getting more information and insights into the transfer system would tend to ask for more public goods and transfer payments. On the one hand they will attempt to receive transfers hitherto de facto or formally confined to other groups; a growing awareness of feeling handicapped in comparison to other groups leads to attempt to acquire privileges for themselves as well.

On the other hand more information concerning public goods might bring about increased utilization of transfers for which they had been eligible before, for which, however, they had not exercised their rights - leading to excessive burdens on public budgets.

In contradistinction to the aforementioned arguments one can suppose that increased information - and corresponding to this more feed-back, e.g. in the form of complaining behavior - may bring about greater effectiveness in public transfers and to reductions in public spending. Not only would increased information sharpen the awareness of capacity limits of the budget; citizens may become especially aware of the character of the transfer systems as a zero-sum-game. Consumers might reduce their demands inspite of the high level of self-financing.

A consumer-oriented transfer policy must be oriented towards the preferences of consumers which, in turn, requires that consumers be well informed and able to participate in decision processes. Consumer information - in that sense - could help to improve the efficiency and effectiveness of transfer policy. The advantages of more consumer sovereignty in a "voucher-system" (Beyme 1977, p. 115) and the improvements in consumer-technology (Skarpelis-Sperk, 1978) may help overcome the selective preclusion effects brought about by lack of information. Both schools of thought mentioned before would concur on this point.

The following discussion deals with some indications regarding the validity of these hypotheses in terms of practical experience. Some conclusions for consumer policy in the public sector are drawn.

Dangers of Inflation in Demands

The recent discussion of the distributive effects of public transfers indicates that some of the assumptions of fiscal theory on the nature of public goods have to be reconsidered (Pfaff 1978). Certain groups are often excluded from the receipt of certain transfers and often rivalness in consumption does, in fact, exist; these effects are group-specific.

The common models of political decision-making, especially those of new political economy, proceed on the assumption of a truly economic man. The economic psychologists, too, believe, that people vote especially for those transfers that benefit primarily themselves. Farmers vote for farm-subsidies and turn down projects to improve urban infrastructure; the poor want more public income maintenance payments and less subsidized theatres - because they rarely visit theatres; all voters vote against higher expenditures for underdeveloped countries.

Better understanding of and information about the transfer budget, one's own and others' benefits and burdens could increase the common feeling of being treated unjustly and could further selfish motivations leading to increased demands on the system. In a parliamentary democracy the consequence would be an inflation of wants; the budgetary limitations may well lead to a neglect of urgent needs, especially of underprivileged groups which would be suppressed in the process of democratic competition. This argument seems self-evident. But in spite of this some aspects suggest that further consideration is required.

Table 1 shows the results of a 1978' opinion poll on public transfers. This table shows weighted and aggregated indices of public performance concerning monetary transfers. The specific differing interests of different groups become evident; however, we also discern, at the same time, that to a large extent people accept transfers benefiting other persons or groups. Other surveys as well indicate a high level of general endorsement on public transfer and social policy (Deimer, Kistler 1980); quite obviously, there is no complete absence of something like compassion or acceptance of the needs of the underprivileged, something like an insight into the necessity of the need of the public aid to other persons and groups - a sense of identity with others.

TABLE 1

Selfishness of Groups Regarding Public Transfer System: People's Opinion on Reducing Public Expenditures (in %)

Transfer Group of Population	Public-Grants-in-Aid to Old-Age Insurance				Public-Grants-in-Aid to Unemployment Insurance					
	a	b	c	d	e	a	b	c	d	e
Population et al.	73	17	4	2	1.55	49	32	11	6	1.07
Persons Receiving Old-Age Insurance Payments	82	11	2	2	1.69	50	32	10	6	1.16
Persons Receiving Public Assistance	88	6	4	-	1.78	74	22	1	-	1.59
By Educational Level: Elementary School	78	14	3	3	1.61	51	30	11	7	1.07
Elementary School and Vocational Training	75	17	3	3	1.58	50	32	10	6	1.10
High School	71	19	5	2	1.56	49	32	13	5	1.07
University	64	25	4	3	1.43	48	33	10	6	1.07

Explanatory Notes:
 a - reductions in no case
 b - reductions might be considered
 c - reductions agreeable
 d - reductions in any case
 e - weighted index
 Source: INIFES 1980.

Possible Ways of Improving the Effectiveness of Transfer Policy

These arguments indicate that more information especially about the situation of other people and groups could lead to a reconsideration and reevaluation of one's own relative position compared to that of others with regard to the public budget constraints. There is a chance to induce more understanding of the limitations in financing public transfers and the connections between taxes and transfers. Understanding the necessity to assume burdens and to aid others could curb an excessive inflation in demand.

"Consumers better understanding of the interdependencies of needs, their causes, their changing, and their satisfaction" as a "prerequisite of critical cooperation" (Czerwonka, Schöppe, Weckbach 1976, p. 279) could lead to a more effective and rational supply of and demand for public transfers.

This assumption which is the central argument in favor of more participation of citizens/consumers in urban planning, administrative decision-making, etc. is based on diverse empirical evidence.

With regard to our specific question we shall cite two results that suggest demand-reducing or at least satisfaction-increasing effects of increased information for consumers.

A 1980 survey of young persons in the city of Augsburg showed a highly significant (.01!) relation between the satisfaction concerning transfer policy and the satisfaction with informational policy about city planning. Table 2 shows that of the persons satisfied with the informational practices only 27.4 per cent are dissatisfied with the policy measures for the young. However, 47.9 per cent of the ones dissatisfied with urban information policy are also dissatisfied with the supply of public goods.

TABLE 2

Connection between Satisfaction with Public Information Policy, Public Transfer Performance, Political Interest and Democratic Apathy - The Case of Policy Measures for the Young in the City of Augsburg (in %)

Satisfaction with Informational Activities	Satisfaction with Transfer to Own Group		Political Apathy (Belief One Can Influence Democratic Processes)		Political Interest		E
	Satisfied	Dissatisfied	Yes	No	High	Low	
Satisfied	36.9	27.4	21.2	68.2	41.7	53.6	13.9
Dissatisfied	16.2	47.9	29.3	57.1	43.7	54.1	68.2
E	19.8	41.2	26.5	58.6	45.1	51.9	-

Source: INIFES 1980.

Considering that there is only a limited connection between the satisfaction with public information and political interest, and that political apathy and lack of belief that democratic engagement can be of any influence, are higher with dissatisfied people (28.3 per cent) than with satisfied ones (21.2 per cent), we can conclude that consumer information is an important prerequisite for systems performance.

Similar considerations may apply to other groups and other transfers. In some investigations, for example, we find the assumption that better information may reduce moral-hazard behavior within the social insurance. Kruse (1979, p. 134) reports in an inquiry that 64 per cent of the persons insured in the Public Health Insurance believe that more and better information may reduce necessary and irresponsible demand for medical services.

The conclusion that informed consumers would help create a more rational transfer policy and reduce selfish aspirations surely is obvious, but difficult to prove. General practice certainly suggests that consumer information is considered important - also in the public sector. How else would we explain the flood of fancy information brochures emanating from diverse agencies.

Such information campaigns take place in all areas from energy-saving to more economy in demand for medical care. The belief in the rationality of consumer behavior goes

hand in hand with the hope that better information via more extensive participation may result in a transfer system that is better geared to consumer needs.

This aspect of higher effectiveness in public supply and individual demand is also directed towards the third of the assumptions discussed above based on the theory and practice of planning.

The idea of consumers as informational or even economic and social resources of decision-making and production of human services (see for example Gartner 1980) is based on the normative premise that consumers know best what's good for them (Pfaff 1977).

The Need for Consumer Sovereignty in the Public Sector

Transfer and social policy procedures in Western countries differ from those in socialist countries; in the latter transfers are allotted while in Western countries extensive steps on the part of the recipient in terms of applications, provision of documents and contacts with offices are necessary before receiving transfer payments. This presumes a high degree of competence and information on the part of the consumers/citizens. Especially because of these necessities some public transfers, such as public aid and assistance programs to the poor or preventive medical examination are not fully used, especially by people who really need them most.

Consumer sovereignty is dependent on a high level of knowledge - especially in the public sector.

Only informed consumers are able to realize their rights to make the best of public transfers and to articulate and perhaps to assert their needs.

Apart from the whole range of effects of more information and on the public budgets it is certain that such procedures will help consumers in their roles as potential transfer recipients.

Conclusion

This discussion could lastly not prove - especially in quantified and monetary figures - the consequences of improved information activities on the budget, last not least because secondary effects (for example employment effects in the service sector) remain outside the consideration.

However, when we view the public sector not only as a system with satisfying and pacifying tasks, but more as a "democratic social provider" ("Demokratischer Sozialstaat", Art. 20, Constitution of the Federal Republic of Germany), as a viable democratic society, there must be criteria of evaluation referring especially to the third of the hypotheses discussed above.

Under normative aspects the ongoing practice of guaranteeing rights to public transfers without providing the potential recipients according to their limitations with the necessary aid to exercise those rights is grossly deficient. Such ways of public supply, relying upon the nonrealisation of existing rights are not at all satisfactory. Furthermore, it is wrong to expect that if public agencies publish more booklets about their activities that this will necessarily lead to better information.

Discussions about the workability of the welfare state cannot convince unless they also take into consideration the perversions of social policy by the lack of information on the part of consumers. Transfer policy requires an elaborated information system. Even if an increase in demand for public payments were to be triggered especially within existing programs, we cannot accept the argument that more extensive information provides dangers to the workability

of the welfare state.

The budget might be burdened further, however, efficiency and effectiveness of transfer policy would increase - especially with regard to distributional objectives. The workability of the welfare state increases when hidden poverty is reduced and non-realised rights are exercised equally by all.

Consumer policy in the public sector might help!

References

- Beyme, Klaus, v. (1977), Sozialismus oder Wohlfahrtsstaat?, München: Piper.
- Czerwonka, Christine, Schöppe, Günter and Weckbach, Stefan (1976), Der aktive Konsument: Kommunikation und Kooperation, Göttingen: Schwarz & Co.
- Deimer, Klaus and Kistler, Ernst (1980), Verbraucherperspektiven in der Sozialpolitik, Paper presented at the International Symposium 1980 in Augsburg (will be published in the proceedings).
- Gartner, Alan (1980), Private and Cooperative Alternatives to the Public Transfer Policy, Paper presented at the International Symposium 1980 in Augsburg (will be published in the proceedings).
- INIFES (1979), Der Einfluß verbraucherpolitischer Instrumente bei staatlichen Anbietern, Unpublished Interim Report, Leitershofen.
- INIFES (1980), Zusammenstellung von subjektiven Indikatoren zum Transfersystem, Unpublished Manuscript, Leitershofen.
- Katz, Daniel et al. (1975), Bureaucratic Encounters, Ann Arbor: University of Michigan Press.
- Krüse, Udo (1979), "Meinungen und Einstellungen der Bevölkerung zur GKV", Sozialer Fortschritt, 6, 131-134.
- Pfaff, Martin (1977), "Who Decides What is Good for Whom?", Journal of Consumer Policy, 2, 138-142.
- Pfaff, Martin, Ed. (1978), Problembereiche der Verteilungs- und Sozialpolitik, Berlin, München: Duncker & Humblot.
- Scherhorn, Gerhard (1975), Verbraucherinteresse und Verbraucherpolitik, Göttingen: Schwartz & Co.
- Skarpelis-Sperk, Sigrid (1978), Soziale Rationierung öffentlicher Leistungen, Frankfurt, New York: Campus.
- Young, Dennis R. (1978), "Public Sector Consumer Problems: A Reply to the Responses", Journal of Consumer Policy, 3, 265-277.

VOTING PARTICIPATION IN A PUBLIC CONSUMPTION PERSPECTIVE¹

Randall G. Chapman, University of Chicago
Kristian S. Palda, Queen's University

Abstract

Data from ten Canadian provincial elections and census-derived socioeconomic variables are used to estimate a model of voting participation. The model is developed, and the empirical results are interpreted, within the context of rational voter and household investment-consumption theories.

Introduction

Voting decisions, an integral part of the household investment-consumption process, can be studied from a consumer behavior perspective. Theories of rational voting and consumer behavior are invoked in this paper to develop a voting participation model. This model is a component of a larger integrative electoral market model currently under development (Chapman and Palda 1980). The empirical results presented here relate the number of votes cast in a constituency to a range of political and socioeconomic variables using data from ten Canadian provincial elections and from the 1970 census. A goal of this paper is to test the extent to which generalizations can be made about the forces affecting voting participation. The results obtained show substantial consistency across time and provinces and are of particular interest with respect to income and education effects on voting turnout.

Theories of Voting Turnout

Participation in the political process is generally viewed as indicating the vigor of an individual's commitment to the democratic process. The most obvious manifestation of such participation is the act of voting. It is not surprising, therefore, to find that studies of electoral turnout stand at the center of attention of students of politics. Not unexpectedly, there are competing theories of voting turnout behavior.

The dominant voting paradigm among political scientists and non-economists is based on a social-psychological perspective which sees social integration as leading to a psychological attachment to the political system and thence to participation (Irvine 1976, Popkin et. al. 1975). The key concept is socialization: if an individual is socialized into the prevailing norms of society or its subgroups, he is likely to vote due to feelings of obligation. In empirical studies, education, income, age, race, and membership in certain organizations have been employed as indicators of the degree of political socialization (Kim, Petrosik, and Enokson 1975).

Economists and some political scientists prefer a theory of voting behavior which rests on the assumption, proposed by Downs, that voters are "rational": they evaluate available alternatives in light of their benefits and costs and implicitly employ a decision calculus based on maximizing expected utility (Downs 1957). In deciding whether or not to vote, the Downsian rational individuals cast their votes only when the expected utility of so doing is positive. This decision calculus can be

represented compactly by the following behavioral equation (Frolich et. al. 1978):

$$U = p_1 |D| + p_2 L - C \quad (1)$$

where:

- U = the net value (expected utility) an individual associates with voting
- p_1 = the probability of an individual's vote affecting the outcome of the election
- D = the expected differential between the winning and the second-place party
- p_2 = the probability that an individual's vote will make a difference in the survival of the democratic political system
- L = the long-run value of participation
- C = the cost of voting.

The Downsian view of the voting participation decision implies that the act has combined consumption and investment features (Stigler 1975). Not only would lower costs lead to a higher turnout (as might be expected on both consumption and investment grounds) but, in addition, we would expect a larger turnout the greater the prospective investment return to the voter.

The accent on the investment perspective is more than a pedantic quibble because the socio-psychological and economic predictive turnout models use intersecting sets of variables to operationalize their underlying constructs. The socio-psychological school holds that income, education, and age level represent suitable proxies for social integration. From the Downsian perspective, these same variables represent measures of opportunity and information processing costs of voting. The differentiating elements between the two theories are most pronounced on the side of the long-run investment costs and benefits yielded by voting, such as taxation to the haves and government transfers to the have-nots.

Why should the student of consumer behavior contemplate these two conflicting theories of voting participation with other than benign indifference? The answer lies in the enormous economic consequences of voting in our post-industrial society. The share of government expenditures in the GNP ranges from 25 percent in Japan to 33 percent in the United States to more than 50 percent in the UK, among the OECD countries, and the taxes collected to finance such expenditures represent the single largest bite out of the income of the consumer-voter-taxpayer (Palda 1980). To the extent that the consumer can influence the size of the tax bite or his social benefits by his vote, his decision to vote (and for whom) can be viewed as a discretionary "buyer-type" activity. In this sense, the sphere of voting is of legitimate concern to the marketing and consumer behavior student.

To these writers, at least, the Downsian approach appears to lead more naturally to a marketing perspective of the voting participation decision. Such a perspective would stipulate that voting represents the selection of a purchasing agent (elective representative) for a period of one or more years. This agent is endowed with powers to decide not only on the mix of public goods to be provided but also on the amount of the voter's income to be appropriated. The voting decision can thus be seen as an investment in a purchasing agent. It exhibits standard investment aspects such as a multiperiod time horizon,

¹Funding for this study was provided by a grant from the Social Science and Humanities Research Council of Canada.

revenues (public goods) and costs (taxes) to be compared on a net present value basis, scrap value of the agent's incumbency, a rate of discounting which reflects personal preferences, and risks associated with the prediction of the capital equipment's (the agent's) expected performance. (The riskiness could be proxied with survey-obtained and regression-estimated standard deviations of linear multi-attribute model coefficients. The D term in equation (1) seems most amenable to being modeled in terms of ideal point distance models and estimation procedures (Ordeshook (1976).)

To bring in consumption aspects of voting behavior, an appropriate integrating framework is that of the household production function (Wittman 1975). In the Beckerian (1965) perspective, households purchase intermediate goods in the market and combine these with their labor inputs and domestic capital equipment to produce final outputs which yield utility. The "capital equipment" within the voting context (elected purchasing agents) is not a property of the household "factory" but rather a part of the "industrial" infrastructure which, unlike in the orthodox industrial case, the household members can influence by their vote. By virtue of their personal labor skills, perhaps acquired by education, certain households will be more proficient at processing electoral information and so will be more likely to vote. Practice may lead to consumption preference and so to direct utility derived from voting. Other household characteristics, such as the presence of small children, may raise electoral consumption costs by allowing less time to acquire political information and to go to the polls.

To sum up, with some additional effort the existing theory of rational voting can be brought closer to the traditional preoccupation of consumer behavior scholars. The central economic importance of voting behavior is a sufficient justification for their attention to the voting participation phenomenon. The next section outlines a model employed to account for electoral participation in Canadian provincial elections and sketches its relationship to a Downsian-household production function approach.

A Model of Voting Participation

A comprehensive theory of voting must necessarily apply to both the decision to opt for a particular candidate (an elective purchasing agent) and the decision to vote, this paper's main concern. The Downsian rational voter theory has been substantially clarified, tested, and elaborated upon over the last 20 years. Its integration with household production theory is in its infancy. We cannot do entire justice to a complete Downs-Becker specification because we do not have access to individual level survey data which are needed to assess the party differential variable D and the p_2L term in equation (1). The model of turnout which we propose and test in this paper does, nevertheless, draw its inspiration from the rational voter theory, supplemented by recourse to the economics of information.

Our model of voting participation may be described in the following general terms:

$$\text{CAST} = f(\text{REGVOT}, \text{CLOSE}, \text{INC}, \text{ED}, \text{TOTEXP}, \text{UNEMP}, \text{FEMPAR}) \quad (2)$$

where:

- CAST = the number of votes cast (in 000s)
- REGVOT = the number of registered voters (in 000s)
- CLOSE = the difference between the vote share of the winner and the runner-up (ex post)
- INC = average per capita income (in \$000s)
- ED = percent of population over 15 years of age with at least complete high school education

- TOTEXP = the total reported campaign expenditures of all major candidates (in \$000s)
- UNEMP = the percent unemployed workers in the labor force
- FEMPAR = proportion of employed females in the total labor force.

All of the variables in equation (2) refer to a constituency in a given Canadian provincial election. The constituency socioeconomic variables -- INC, ED, UNEMP, and FEMPAR -- were obtained from special Statistics Canada runs in which the 1970 census data were geocoded along the boundaries of the provincial constituencies.

REGVOT serves as a constituency-size scaling variable. It should be noted that in Canada, unlike in the United States, voter registration is done by enumerators and does not, therefore, require voluntary effort on the part of the individual qualified citizen-voter.

CLOSE, an ex-post variable, measures the probability (p_1 in equation (1)) of an individual's vote affecting the outcome of the election (Barzel and Silberberg 1973). Theory predicts a negative sign on its coefficient: the narrower the margin, the greater the impact of each vote, the higher the incentive to turnout.

INC, the average per capita income (as of the 1970 census), has at least a three-fold interpretation. Simple economic theory suggests that there is both a substitution effect, in terms of opportunity costs facing higher income voters, as well as an income effect, related to an explanation of voting as a consumption good. These effects cut in opposite directions and offer no unambiguous qualitative conclusion as to the expected sign of this variable (Tollison, Crain, and Pautler 1975, Crain and Deaton 1977). In the absence of information on property and income subject to provincial taxation (i.e., on variables which would throw light on the investment aspect of voting), INC might reflect the taxation stake of the voter and be expected to exhibit a positive coefficient. The composite of all of these forces will determine the sign of INC. If it turns out to be negative, opportunity cost aspects predominate.

The economic theory of information would predict, ceteris paribus, a positive association between turnout and the level of education possessed by voters-consumers: the higher the education level, the lower the costs incurred in acquiring and processing political information (Goldman and Johansson 1978). As for ED, so for TOTEXP: campaign expenditures would be expected to have a positive influence on turnout as they represent the lowering of costs of information acquisition (Abrams and Settle 1978).

An economics of information perspective would also predict that the unemployed, likely to be on average less educated, could encounter higher costs in acquiring and processing electoral information. (The socio-psychological approach to electoral participation would suggest that the unemployed are disaffected with the political process, and thus would have a tendency to turn out in lesser numbers.) They would have, however, lower income opportunity costs of voting and possibly a greater stake in voting for left-leaning parties to the extent that provincial governments, rather than the federal government, can be held accountable for employment opportunities. Thus, theory is ambiguous with regard to the sign of UNEMP.

Finally, FEMPAR also has different possible interpretations as to expected sign. On one hand, high levels of female participation would be associated with higher incomes and possibly a greater "tightness" in household time allocation thus increasing opportunity costs. On the other hand, on-the-job discussions of politics and world events might lead to lower information costs.

Empirical Results

The model in equation (2) was estimated using data from each of ten Canadian provincial elections. A linear-additive functional form was chosen on the grounds of parsimony. To avoid constituency size affecting absolute campaign spending, the expenditure variable was formulated in dollars spent per registered voter (i.e., as TOTEXP/REGVOT). The results of estimating the model are reported in Table 1. It should be noted that our a priori theory led us to use one-tailed tests of significance for REGVOT, TOTEXP/REGVOT, ED, and CLOSE (where we expected three positive and one negative signs, respectively) and two-tailed tests of significance for UNEMP, INC, and FEMPAR (where theory was ambiguous with respect to signs).

Two comments are in order regarding the linear-additive functional form of the voting participation model. The model was also estimated in logarithmic form to assess the possible presence of significant interactions among the independent variables. The results obtained were consistent with those from the simpler linear-additive functional form, so the more parsimonious additive version was chosen. A second functional form issue concerns the dependent variable. The usual dependent variable in a voting participation model is TURNOUT (=CAST/REGVOT). A linear-additive model with TURNOUT as the dependent variable was estimated for comparison purposes. Its results were also consistent with our results using CAST as the dependent variable. Again, parsimony led us to prefer the simpler version.

In interpreting the empirical results, patterns of generalizability were looked for. An important feature of the research reported in this paper is the large amount of data assembled to estimate the voting participation model. Most politometric research is based on a single electoral event only. The ability to generalize from such results may be open to question. With ten elections as a data base, the results of this study lend themselves to

attempts to document generalizable and consistent patterns of voting behavior.

REGVOT, a scaling variable which standardizes for the size of the constituency, has naturally the largest (positive) impact. The high explanatory power of the model derives from simple correlations between CAST and REGVOT which are 0.90 or more in magnitude.

In three of the ten elections, per registered voter campaign expenditures have a significantly positive impact on turnout. A significantly negative effect is found in one election; no obvious explanation is available for this anomalous result. It may be noted that these four significant results occur in the largest and most urbanized Canadian provinces (Ontario and Quebec). A public policy question relates to the possible impact of campaign expenditure ceilings on voting behavior. Legislated ceilings on individual candidate expenditures were in effect in all provinces except British Columbia. The avowed purpose of spending limits is to restrict the "buying" of seats on the assumption that campaign expenditures influence voting outcomes. This assumption has been repeatedly empirically verified (Chapman and Palda 1980). Spending limitations might, therefore, have the unintended and socially undesirable side effect of constraining the flow of information from candidates to voters, thus effectively increasing the information acquisition costs to the voters and possibly leading to reduced voter participation. Our empirical results here do not support, on the whole, the hypothesis that overall constituency spending affects voting participation.

The closeness of the electoral contest has a significant impact on participation in six of the ten elections. (The negative sign of CLOSE means that as the difference between winner and runner-up narrows, turnout increases.) This result is in line with previous empirical tests of the Downsian theory and supports the view that a stronger probability of influencing the election outcome

TABLE 1
RESULTS OF ESTIMATING THE VOTING PARTICIPATION MODEL

Variables	Canadian Provincial Election									
	B. C. 1972	B. C. 1975	Manitob 1973	Manitob 1977	Ontario 1975	Ontario 1977	Quebec 1973	Quebec 1976	Saskatc 1975	Saskatc 1978
REGVOT	0.689*	0.728*	0.873*	0.808*	0.702*	0.664*	0.838*	0.878*	0.771*	0.721*
TOTEXP/REGVOT	-0.617	-0.204	-0.039	0.187	0.736*	1.145*	1.473*	-1.379*	0.138	0.071
CLOSE	0.024	1.809	-0.841*	-1.419*	-2.192*	-4.029*	-2.402*	0.330	-0.525	-1.530*
UNEMP	-0.659*	-0.725*	-0.023	-0.034	-0.637*	-0.656*	-0.254*	-0.235*	-0.085*	0.054
INC	-1.435*	-2.473*	-0.507*	-0.537	-0.260	-0.609	-1.507*	-1.928*	-0.136	-0.129
ED	11.016*	16.008*	6.412*	7.377*	-0.881	0.691	19.354*	20.156*	2.599*	2.549*
FEMPAR	-0.082	-0.185*	-0.025	-0.040*	-0.160*	-0.066	-0.074*	-0.072*	-0.037*	-0.011
Constant	10.260*	15.129*	0.298	0.984	8.370*	6.701*	1.698	5.807*	1.329	1.054
R ²	0.981	0.978	0.974	0.969	0.947	0.932	0.927	0.975	0.881	0.951
n	48	48	57	57	125	125	110	110	61	61

* Significantly different from zero at the 0.05 level: two-tailed test for UNEMP, INC, and FEMPAR and one-tailed test for REGVOT, TOTEXP/REGVOT, CLOSE, and ED.

moves the voter to cast his ballot.

Perhaps the most interesting results are derived with regard to the income and education variables, taken together. All of the ten INC coefficients are negative, of which five are significant. Nine of the ED coefficients have a positive sign; the one negative coefficient is not statistically significant. As far as we can tell, this is the first time (certainly with aggregate-level data) that the separate effects of income and education on voting participation are assessed -- and this in face of the high positive correlations between them (from 0.60 to 0.94 in our sample). The strong consistency of our results militates for the following interpretation. The substitution or opportunity cost effect predominates with regard to income: as income increases, holding education constant, voters tend to participate less since they value their time more highly. (This result comes through even though the average income measure used is not totally "pure" for it would have been preferable to use income derived from employment, unavailable to us, and to have had access to an investment-type taxation-stake variable (Denver and Hands 1974).) Higher education, on the other hand, leads to increased voter turnout as it reduces the costs of acquiring and processing political information.

Nine out of the ten coefficients of UNEMP are negative, seven of these significantly so. Six of these significant results are obtained in the industrially-oriented provinces of British Columbia, Ontario and Quebec. Since education and income determinants are held constant at their sample means, the only Downsian interpretation plausible is that the unemployed do not perceive provincial governments as capable of job creation. (Note, however, that the socialization perspective would consider the structurally unemployed as participating less in the life of the polity.)

Six of the ten FEMPAR negative coefficients are statistically significant. While there is a pattern of positive correlation between FEMPAR and INC and between FEMPAR and ED (in the 0.35 to 0.65 range), the added influence of feminine participation in the labour force is suggestive of non-trivial household time allocation costs.

In summary, the model's empirical results bear out rather convincingly the economic view of rational man faced with the decision to vote in light of his costs, benefits and the uncertainty he faces and taking account of the household context. The generally consistent pattern of the results over the ten elections has been arrived at despite the aggregate, constituency-level nature of the variables used and some strong collinearity among the socioeconomic determinants.

Concluding Remarks

The model of voting participation reported upon here constitutes the initial component of a larger recursive electoral market model which will attempt to account for party candidate success (with special emphasis on the role of campaign expenditures), candidate expenditures and the contribution generation mechanism (Chapman and Palda 1980). The elaboration and testing of this integrated model with data from the ten Canadian provincial elections is being undertaken within the framework of the rational voter theory. The pattern of consistently plausible results obtained with regard to voting participation augurs well for the forthcoming study.

References

- Abrams, Burton A. and Settle, Russell T. (1978), "The Economic Theory of Regulation and Public Financing of Presidential Elections," Journal of Political Economy, 86, 245-257.
- Barzel, Yoram and Silberberg, Eugene (1973), "Is the Act of Voting Rational?" Public Choice, 22, 51-58.
- Becker, Gary S. (1965), "A Theory of the Allocation of Time," The Economic Journal, 75, 493-517.
- Chapman, Randall G. and Palda, Kristian S. (1980), "Voting Patterns and Campaign Expenditures in Canadian Provincial Elections," Forthcoming in Proceedings of the 52nd Annual Meetings of the Canadian Political Science Association.
- Crain, W. Mark and Deaton, Thomas H. (1977), "A Note on Political Participation as Consumption Behavior," Public Choice, 26, 131-135.
- Denver, D. T. and Hands, H. T. G. (1974), "Marginality and Turnout in British General Elections," British Journal of Political Science, 4, 17-35.
- Downs, Anthony (1957), An Economic Theory of Democracy, New York: Harper and Row.
- Frolich, Norman, et. al. (1978), "A Test of Downsian Voter Rationality: 1964 Presidential Voting," American Political Science Review, 72, 178-197.
- Goldman, Arieh and Johansson, J. K. (1978), "Determinants of Search for Lower Prices: An Empirical Assessment of the Economics of Information Theory," Journal of Consumer Research, 5, 176-186.
- Irvine, William A. (1976), "Testing Explanations of Voting Turnout in Canada," in Ian Budge, et. al. (eds.), Party Identification and Beyond, New York: John Wiley and Sons.
- Kim, Jae-On, Petrocik, J. R., and Enokson, S. N. (1975), "Voter Turnout Among the American States: Systemic and Individual Comparisons," American Political Science Review, 69, 107-131.
- Ordeshook, Peter (1976), "The Spatial Theory of Elections: A Review and a Critique," in Ian Budge, et. al. (eds.), Party Identification and Beyond, New York: John Wiley and Sons.
- Palda, Kristian S. (1980), "The Public Sector and Marketing: Should the Twain Meet?" Forthcoming in Proceedings of the Second Annual Special AMA Conference on Marketing Theory.
- Popkin, Samuel L., et. al. (1976), "Comment: What Have You Done For Me Lately? Toward an Investment Theory of Voting," American Political Science Review, 70, 779-805.
- Stigler, George J. (1975), "Comment," Journal of Law and Economics, 18, 743-744.
- Tollison, Robert, Crain, Mark, and Pautler, Paul (1975), "Information and Voting," Public Choice, 24, 43-49.
- Wittman, Donald (1975), "Comment," Journal of Law and Economics, 18, 735-741.

MEASURING SUBJECTIVE VALUATION AND
DEMAND FOR GOVERNMENT SERVICES

Eli M. Noam, Columbia University

Abstract

The paper describes a method by which demand equations for public goods and services can be estimated. A model is described which is based on the revealed preferences in voting behavior by different groups, taking in particular account of abstentions and their frequency. The results permit the derivation of the absolute valuation for governmental services by different income groups.

Introduction

The problems of estimating demand functions in private markets, vexing as they are, still pale in comparison with those encountered in the public sector. Here we have no markets, no customers, no sales, no products, and usually no data. We have to deal with "public goods" that are fuzzy conceptually, indeterminate distributionally, and measurable indirectly. At the same time it is important for decision makers in the public sector to have some sense of the demand for the many alternative public services because of the huge governmental outlays. Some sense of the demand for competing public services would be most helpful. This is not necessarily a matter of trial and error; mistakes in determining demand functions for public goods are not self-correcting. With private goods, sooner or later a wrong demand function will come home to roost, provided that some decisions are based on them. With public goods, however, a government can go on for a long time supplying hospital beds and canals that no one wants in particular except for some small vocal groups.

Because the knowledge of public demand functions could be of considerable help in the allocation of the huge governmental budgets, a number of attempts to determine them have been made by researchers in the past. A first method of analysis is an opinion survey or controlled experiment (Bohm 1972). The difficulty with this approach, besides its cost, is that no real concept of cost exists in peoples' minds when it comes to large public expenditures. Sometimes, too, a survey is subject to manipulation by the subjects, since there is little to stop them from vastly exaggerating their politically oriented likes and dislikes. In order to avoid this problem, some authors have developed an entire theory of the type of questions that removes the incentive for manipulation (Green and Laffont 1976), or proposed methods to assign cost shares to the individual (Tideman and Tullock 1976). These methods are remarkably ingenious but probably too complex for use in practice.

A third approach is to take a look at the real world and to assume that its reality is an expression of public desires (Borcherding and Deacon 1972, Bergstrom and Goodman 1973). Thus one can compare expenditures in different jurisdictions and analyses and their relation to community characteristics such as income, education and tax rates. If affluent towns spend twice as much as poor ones on law enforcement, we have information about the relative preferences with respect to income. The question remains, however, whether the underlying assumption that political decision making accurately expresses public demand is reliable. Small bands of bureaucrats, insiders, and notables can obtain decisions according to their own preferences rather than of those of the public at large. In a society with free election contests and easy mobility

such possible divergence will probably be bridged over time. The literature on log-rolling and election-posturing (Buchanan and Tullock 1972) explain the public choice mechanism by which public preferences become policy decisions, while the so-called Tibout hypothesis (1956) of locational mobility (voting with one's feet) provides a public finance mechanism.

A fourth approach looks at decisions on the individual level, namely by voting. Concretely, this means in the case of public good issues the use of results from referenda. The advantages here are, first the realism of data: they reveal preference in concrete expenditure matters, and typically after the voter has been subject to a fair amount of information about both sides of the issue at hand. Neither is there usually a reason to suspect that voters would disguise their true preferences for some strategic purpose. Recognizing these advantages, several studies of referenda on public finance issues were undertaken. They fall into two broad categories, of which the first is an essentially empirical use of correlation studies of aggregate approval rates (Wilson and Banfield 1965, Frey and Kohn 1970, Birdsall 1965), and the second more theoretical, such as Deacon and Shapiro's study (1975) with a sophisticated model as an underpinning, but with application to only two referendum issues.

In these studies an important problem remains unresolved: voting results indicate the direction of public preference, but they do not necessarily reflect its intensity.¹ An issue may be mildly preferred or passionately desired by the same percentage of people; this will not be obvious from the voting results.

This paper describes a method to overcome such problems by deriving a model by which measures of preference for public goods can be obtained. There, preference measures can be analyzed for factors that influence them, and demand functions can then be found.

The Model

The key element of the model is its recognition of a previously completely ignored source of information from voting data, namely the rate of abstention. It is obvious that the active voting in favor or in opposition to a proposal reflects some preferences; but so does non-voting. Implicit in non-voting by a large group of people who vote on other occasions is an indication that the issue is of limited importance to them. This is particularly true with voters who are frequently called to the polls and where, therefore, the consumption aspect of voting--the exercise of the franchise--becomes secondary. In such a situation, voters are more selective in their voting; they go to the polls when the issue is subjectively important to them, and do not when they perceive the issue to be of only minor significance. It is important to contrast this analysis of differential behavior of the same group of voters over elections with those studies that compare voting behavior of different voting jurisdictions within the same election.

It may be argued that another reason for abstention exists, namely that the outcome of a vote is so predictable that an

¹Deacon and Shapiro (1975) use "Utility" to overcome this problem, but only as an intermediate theoretical construct.

individual's vote hardly matters anymore, thus reducing participation (Frohlich et al. 1978). Empirically, however, this claim turns out to be untrue, at least for the jurisdiction which was analyzed. In the course of this study an analysis of the participation rates in hundreds of referenda showed practically no correlation (.081) with the closeness of the vote. Maybe the reason is that in local voting normally no sophisticated polls are taken that predict the outcomes of the voting with a reasonable degree of accuracy; or perhaps because it requires some stake in the outcome to mount a drive in order to "get the vote out", so that such attempt will succeed only in important votes.

Thus we make two assumptions about voting behavior, denoting the perceived benefit with B and defining a voting "threshold" S such that a voter acts according to the following rules:

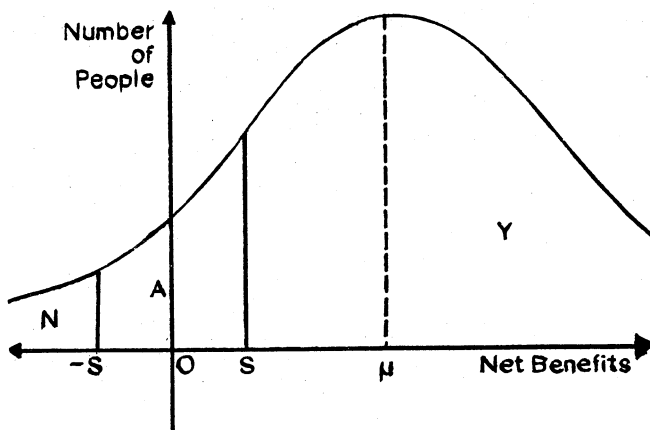
- (A) Vote Yes if $B > S$
- (B) Vote No if $B \leq -S$
- (C) Abstain from voting in $-S < B \leq S$

Let us postulate next that the voter is a member of a subgroup of the total population, of which we assume that

- (D) Their perceived benefits are normally distributed around some mean μ and with some variance σ^2
- (E) They follow decisions rules (A) - (C)

These assumptions are represented by Figure 1. The abscissa represents the magnitude of benefits, both positive and negative, around point 0. The units are as yet undefined. The vertical axis represents the frequency distribution of these benefits in the subgroup. Points to the right of 0 reflect positive benefits, and points to the right of S represent "yes" voting. Similarly, at points left of -S "no" voting occurs, while people whose potential benefits lie between -S and S abstain. The abstainers' proportion of total voting population is the area A, and Y and N represent yes and no voting proportions.

FIGURE 1



The objective for analyses is to determine the mean benefit for the group. Formally, this amounts to determining

²Normal distribution is a common assumption in much of the public choice of literature on spatial electoral analysis. See, e.g., Hinich and Ordeshook (1970) and Hinich, Ledyard and Ordeshook (1972).

the unknown mean μ of a normal distribution whose variance σ is also unknown, but where we know the areas Y, N, and A to exist, respectively, right of S, left of -S, and between S and -S. This can be expressed as the probabilities.

$$P\{B \leq -S\} = N \quad (1)$$

$$P\{B \leq S\} = N + A \quad (2)$$

These equations can be standardized into

$$P\{Z \leq \frac{-S-\mu}{\sigma}\} = N \quad (3)$$

$$P\{Z \leq \frac{S-\mu}{\sigma}\} = N + A \quad (4)$$

Let Z_X be the cumulative distribution function of the standardized normal distribution, and therefore

$$\frac{-S-\mu}{\sigma} = Z_N \quad (5)$$

$$\frac{S-\mu}{\sigma} = Z_{N+A} \quad (6)$$

Thus, the variance σ can be expressed, after substitutions, as

$$\sigma = \frac{-ZS}{Z_N - Z_{N+A}} \quad (7)$$

and the mean μ , using (5) and (7), is found to be

$$\begin{aligned} \mu &= -S - Z_N \sigma \\ &= -S - Z_N \left(\frac{-ZS}{Z_N - Z_{N+A}} \right) \\ &= S \left(\frac{Z_N + Z_{N+A}}{Z_N - Z_{N+A}} \right) \end{aligned} \quad (8)$$

This means that μ can be readily calculated as a multiple of S once the proportions of N and A are known and once their cumulative Z values are obtained from tables of normal distribution.

This procedure can be repeated for any number of subgroups i, resulting in a series of mean benefits μ_i which can be analyzed for their susceptibility to factors of demography, economics, etc. For most of such estimations the actual magnitude of the voting threshold S is immaterial, as long as we assume that it is constant for all subgroups. But this assumption can be relaxed, and S can instead be estimated by reversing the above procedure. Let there be observed voting behavior for certain issues, where the mean benefit μ of the normally distributed random variable B is known from some outside information as well as the proportions Y, N, and A. Then, from equations (5) and (6) we have

$$\sigma = \frac{-2\mu}{Z_N + Z_{N+A}} \quad (9)$$

which substituted into (5) results in

$$S = Z_N \frac{2\mu}{Z_N + Z_{N+A}} - \mu = \mu \left(\frac{Z_N - Z_{N+A}}{Z_N + Z_{N+A}} \right) \quad (10)$$

Thus S_i can be determined if N_i , A_i , and μ_i are known for subgroup i. Of course, for most issues μ_i will be unknown. To deal with this situation, it is assumed that

- (E) An individual's voting threshold S is constant in the short and medium term.

This assumption implies that a rational voter remains rational and that his decision calculus as to whether it is worthwhile to vote is stable, at least over a few years. The reasonableness of this assumption will receive support in the empirical part of the paper. Hence, if S can be determined in several issues, and provided that the

results are reasonably similar to each other, the average S will be used for other issues as well.

The variables that are investigated for their influence on the attitude toward government services are income and other demographic factors such as education, number of children, age, occupational category, etc. Let this be expressed by the equation³

$$\mu = \Omega \left(\frac{Y}{\bar{Y}}\right)^\delta C^\lambda X_i^{\eta_i} \quad (11)$$

where $\frac{Y}{\bar{Y}}$ = relative income as a measure for socio-economic status

C = tax cost to the individual

X_i = other demographic factors

and δ, λ, η_i = elasticities.

A special problem in the analysis of public goods is that under almost all tax systems their cost to the individual varies with his income. To the individual the tax price depends both on the total cost of the public good, V, and on his relative share of the tax burden, T/R, where T is his tax payment and R is total revenue. The tax price is therefore

$$C = \frac{T}{R} \quad (12)$$

Normally, tax contributions are an increasing function of income, either through taxation of income, property, or consumption such that

$$T = \epsilon Y^\psi \quad (13)$$

Substituting into equation (12) we then have

$$C = \frac{\epsilon Y^\psi}{R} \quad (12')$$

Substituting this in turn into the earlier equation (11)

$$\mu = \Omega \left(\frac{Y}{\bar{Y}}\right)^\delta \left(\frac{\epsilon Y^\psi}{R}\right)^\lambda X_i^{\eta_i} = \Omega \left(\frac{\epsilon Y}{R}\right)^\lambda \bar{Y}^{\psi\lambda} \left(\frac{Y}{\bar{Y}}\right)^{\delta+\psi\lambda} X_i^{\eta_i} \quad (11')$$

λ and ψ can be estimated exogenously by

$$t = \epsilon Y^\psi \quad (13)$$

and an expression for the cost-elasticity λ

$$\mu = \rho C^\lambda \quad (14)$$

which is subject to independent estimation. If ψ and λ are obtained exogenously, equation (11') simplifies into

$$\mu = \alpha \left(\frac{Y}{\bar{Y}}\right)^\beta X_i^{\eta_i}$$

where the parameters α and β are shorthand for

$$\alpha = \Omega \left(\frac{\epsilon \bar{Y}}{R}\right)^\lambda \bar{Y}^{\psi\lambda} \quad \text{and} \quad \beta = \delta + \psi\lambda \quad (15)$$

β is the "raw" or observed income elasticity. It has been corrected in the way described for changes in the tax price in order to yield the "true" income elasticity, which takes into account that as income rises the tax price also increases. $\delta = \beta - \psi\lambda$, i.e., the "real" income elasticity δ is equal to the observed "raw" income elasticity minus the product of price elasticity and the elasticity of the tax system.

³A linear function does not seem likely in light of some of the high incomes observed.

The data analysis utilized voting results from Switzerland, a country with an unusual degree of direct participation by the electorate. Among Swiss jurisdictions, Basel-Stadt is the smallest urban canton(state);⁴ it was chosen in order to reduce the effect of geographic location on voting behavior. Basel is a highly developed middle-sized city with a long civic tradition and an international location, and the preference of electorate should be illustrative.

The main method of estimation involves a cross-section analysis across relatively homogenous polling districts, where the relative income and the demographic characteristics of the district are the independent variables and the intensity of preference that is found through the model is the dependent variable.

Voting results for referenda by polling place are available in the official gazette.⁵ Demographic information is found in the Swiss national census; specifically analyzed, in addition to income, were education, number of dependent children, age, and occupation (self-employed vs. employees). The source of income data is a market research survey.⁶ Cost data for the public goods were found, quite laboriously, in the resolutions by the cantonal government to the electorate (Ratschlaege) which precede voting.⁷ The active electorate--i.e., the electorate excluding habitual non-voters against which percentages of Y, N, and A are calculated--is defined as the highest number of participants that voted in a referendum concerning cantonal matters.⁸ In order to avoid a systematic bias, one must ascertain that the average voting participation of the active electorate is not different for different income groups. An analysis of this question determined that the defined L did not display bias; average participations by the active electorate in the nearly two hundred issues that are investigated in this paper are nearly identical: for the lowest income district, 55.6%; for the median income district, 56.5%; and for the highest income district, 56.2%. This active electorate's participation rate must be distinguished from average general voting participation rate, in which low income groups display the usual lower rates.

To estimate the threshold value S_i , the procedure described earlier in equation (10) is used and issues are chosen for which the mean benefits can be estimated independently. The referendum issues that are used are votes on public utility charges and taxation,⁹ for which the financial consequences are known or are officially forecast in the above mentioned Ratschlaege or resolutions; we

⁴The use of Basel was inspired by Frey and Kohn (1970).

⁵Kantonsblatt, Basel-Stadt, on days following the referenda.

⁶PROGNOS, Konsumpotential and Filialnetz, Untersuchung über die Investitionsplanung des ACV, includes the distribution of household income of 1965 in districts. Tables made available by L. Kohn.

⁷Collected at the Cantonal Chancery and the Cantonal Archives.

⁸Respectively in each decade, maximum participation rates are for referenda held on 12/11/49; 12/5/54; 2/28/65; 10/20/74.

⁹Referenda held 12/11/49; 1/19/50; 11/24/68; 7/12/75; 6/12/77.

use these as μ_i in order to determine S_i . Table 1 shows an example for such calculation. The resultant S_i are all relatively similar to each other, with values in a band 23-34 Sfr. Very similar results were also found for the other voting precincts. Thus we find support for our assumption of a constant S . The average value (28 Sfr. or approximately \$20) is therefore assumed as the voting threshold. As mentioned before, a different magnitude of S would not change the elasticities.

TABLE 1
APPLICATION OF MODEL TO UPPER INCOME
PRECINCT (BRUDERHOLZ)

Issue	$\%Y$ (Y/L)	$\%N$ (N/L)	$\%(N+A)$	$\frac{Z_N+Z_{N+A}}{Z_N-Z_{N+A}}$	μ	S	σ
Tax Reduction	52.3	35.0	47.7	1.369	40	29.2	183.38
In-creased Progressivity in Tax Rates	17.1	61.1	82.9	-1.835	-50	27.2	88.95
Gas Utility Rates	32.2	11.6	67.8	.444	15	33.8	36.01
Water Utility Rate Change I	47.6	25.8	42.4	1.826	50	27.4	129.56
Water Utility Rate Change II	44.0	10.9	56.0	.782	25	31.9	43.18

Also calculated from statutory tables of income tax is the progressivity of the tax rate¹⁰ with respect to income as $T = \epsilon Y^{\psi} = .0000036 Y^{1.98}$. This progressivity is assumed to apply in general to the cantonal tax system, there being no sales or consumption tax and only a moderate property tax.

To find the cost elasticities λ , time series for benefits are estimated in the following way. From the multitude of referenda, those voting issues were selected that were over the years repetitive or similar to each other so as to make comparison possible over time. For example, there are several referenda about the public support to the municipal theater. Similarly, public expenditures for roads, for administrative buildings, and for schools are a frequent subject of voting. We chose therefore seven categories of public goods and services for which a number of similar referenda are available and run time series of the form (14) over the median income precinct, with different cost figures as variable. The results are listed in Table 2. The cost elasticities are found to have the negative sign which one would expect. The magnitude of the coefficients is small and fairly similar for each category. It is interesting to note that preferences are less elastic for upper class issues such as law enforcement and culture, and more elastic for lower class categories such as social programs. For the latter the cost-sensitivity is larger.

The results for the cross-section analysis of equation (12) are listed in Table 3. Nearly all of the coefficients of "true" income elasticity are statistically significant, and

¹⁰Kanton Basel-Stadt, Steuertafeln, 1977.

TABLE 2
COST ELASTICITIES OF PUBLIC GOODS

Transportation	-.1012 (-2.7753)
Social Programs	-.1395 (-1.3191)
Sport	-.1754 (-2.7185)
Redistribution	-.2314 (-.9263)
Police and Law Enforcement	-.0183 (-1.8219)
Theater	.0549 (-2.1391)

t-statistics in parentheses.

are reported. The other variables, however, do not show any consistent significance. Thus the most important variables are reported, while the others were omitted and the regression re-calculated. The results follow.

TABLE 3
ELASTICITIES OF PREFERENCES FOR PUBLIC
GOODS AND SERVICES

	Income Elasticity*	Other Significant Elasticities*	R ²
Unemployment Compensation	-4.1283 (-1.9101)	Children .0319 (1.8268)	.9872
		Occup. .0270 (1.593)	
Hospital	-.0613 (-.4214)	Age .0933 (3.1921)	.6525
Welfare Subsidies	-.6529 (-.7165)	Occup. -.2114 (-2.327)	.7922
Highways	.6421 (.6142)	Age -.1231 (-2.312)	.6129
Mass Transit	-.0742 (-2.2121)	Age .2120 (2.057)	.5703
		Children .0278 (1.9128)	
Prosecutor	2.9509 (3.6136)		.8002
Prison	2.0471 (5.2218)	Occup. .0124 (1.8294)	.9162
Old Age Home	.0317 (1.7212)	Age .2915 (2.5941)	.4672
Sports Facilities	-.1512 (-1.8632)	Children .2033 (2.191)	.5990
Museum	1.5231 (2.1327)	Educ. .0829 (4.1642)	.7219
Educational Expenditures	1.5826 (1.7313)	Educ. .2172 (1.5272)	.7219
		Children .6171 (1.9597)	
Foreign Aid	.7172 (4.3191)		.8241

t-statistics in parentheses.

Turning first to the income elasticities, we can observe that most are of a good size and either statistically significant or nearly so. Positive income elasticities are found for law enforcement (prosecution, prison), for education and culture, foreign aid, and highways. Negative elasticities exist for mass transit and redistributive social programs such as unemployment compensation, welfare subsidies and public hospitals. None of the other demographic factors shows consistent statistical importance; age is significant for road and mass transportation (younger people prefer highways, hospitals and old age homes), however the coefficients are fairly small. Self-employment occupation contributes to the negative attitudes towards welfare and unemployment compensation. The number of children affects the preference for mass transit, unemployment compensation, sports facilities, and education (in favor of all). Finally, education affects the preferences for support of an art museum as well as of education itself.

Because all the coefficients are fairly small, we find that relative income explains the predominant part of the preferences for public goods and services.

It is also interesting to estimate and compare preferences in absolute terms. One way to interpret the meaning of such preference values is to think of them as the maximum payment the individual would be willing to make to have the public good. Another interpretation is to think of the individual as being indifferent between having the public good, or having the stated sum of money. A larger sum would be preferable to the public good. Thus the value can also be seen as a schedule of "bribes" to have a voter change his mind if he thinks that his vote makes the difference. The income of the highest and of the lowest 20 percent of the population are substituted into the equations, and other variables are kept at median values. The results are shown in Table 4. As can be seen, the

TABLE 4
VALUATION OF PUBLIC GOODS

	Highest 20% of Income	Lowest 20% of Income
Unemployment Compensation	.76	12.20
Hospital	12.02	37.90
Welfare Subsidies	-13.52	31.32
Highways	11.02	4.40
Mass Transit	5.88	9.04
Prosecutor	25.84	3.56
Prison	24.06	5.66
Old Age Home	2.91	7.81
Sport Facility	- 3.22	9.62
Museum	12.61	- 1.29
Educational Expenditure	64.38	2.02
Foreign Aid	10.32	- 1.34

high-income individuals have a strong absolute valuation for education and law enforcement expenditures; their preferences for roads, and the art museum is positive but moderate in size. For social programs their preference is normally negative, but not strongly so. It does not equal the strong positive demand by the low income group for these expenditures. For low income people, not surprisingly, the most important issues involve services from which they benefit: public hospitals, old age homes, unemployment compensation and welfare support. Nearly all other public goods are demanded positively--which is not irrational considering their low cost to poor people--but not very intensely.

Summary

This paper describes a method to derive measures of the demand for public goods. It uses the results of referenda and incorporates variations in non-voting into a model and calculates both the elasticities of the preference intensities with respect to income and other demographic variables, and finds absolute measures of preference. The model itself is a useful extension of the analysis of voting outcomes to demand function for public goods. The empirical results are specific to a Swiss jurisdiction; similar studies should be undertaken for other countries, and this is a next step for research.

References

- Bergstrom, Theodore C. and Goodman, Richard E. (1973), "Private Demand for Public Goods," American Economic Review, 63, 780-796.
- Birdsall, William C. (1965), "A Study of the Demand for Public Goods," in Richard Musgrave (ed.), Essays in Fiscal Federalism, Washington, D.C.
- Bohm, Peter (1972), "Estimating Demand for Public Goods: An Experiment," European Economic Review, 3, 111-130.
- Borcherding, Thomas E. and Deacon, Robert T. (1972), "The Demand for the Services of Non-Federal Governments," American Economic Review, 62, 891-901.
- Buchanan, James N. and Tullock, Gordon (1972), The Calculus of Consent, Ann Arbor: University of Michigan Press.
- Deacon, Robert and Shapiro, Perry (1975), "Private Preferences for Collective Goods Revealed Through Voting on Referenda," American Economic Review, 65, 943-955.
- Frey, René and Kohn, Leopold (1970), "An Economic Interpretation of Voting Behavior on Public Finance Issues," Kyklos, 23, 792-804.
- Frohlich, Norman, et al. (1978), "A Test of Downsian Voter Rationality: 1964 Presidential Voting," American Political Science Review, 72, 178-197.
- Green, Jerry and Laffont, Jean J. (1976), "On the Revelation of Preferences for Public Goods," Journal of Public Economics, 8, 79-93.
- Hinich, Melvin J. and Ordeshook, Peter (1970), "Plurality Maximization vs. Vote Maximization: A Spatial Analysis with Variable Participation," American Political Science Review, 64, 3, 772-791.
- Hinich, Melvin J., Ledyard, John O., and Ordeshook, Peter (1972), "Nonvoting and the Existence of Equilibrium Under Majority Rule," Journal of Economic Theory, 4, 144-153.
- Tibout, Charles (1956), "Economic Theory of Fiscal Decentralization," Journal of Political Economy, 64, 416.
- Tideman, T. Nicholas and Tullock, Gordon (1976), "A New and Superior Process for Making Social Choices," Journal of Political Economy, 84, 1145-1159.
- Wilson, James Q. and Banfield, Edward C. (1965), "Voting Behavior on Municipal Public Expenditure: A Study in Rationality and Self-Interest," in Julius Margolis (ed.), The Public Economy of Urban Communities, Washington, D.C.

FISCAL ILLUSION AND CONSUMER SOVEREIGNTY: AN EXPLORATORY STUDY¹

Werner W. Pommerehne, University of Zurich

Abstract

Data on the Australian federal government's performance as perceived by the voters and government revenue and expenditure variables are used to estimate a model of interaction between the voters' and the government's behavior. Special emphasis is given to the role of fiscal illusion on the voters' side in evaluating government performance, and to its manipulation by the government toward securing reelection. The model is developed and the empirical results are interpreted under the perspective of consumer sovereignty.

Introduction

Consumer sovereignty refers to a process of choice in which the choice is free (most major characteristic), informed (so that the act of choice is meaningful), personal (made by and not for people), and responsible (choosers must be aware of the consequences of their choice). It is often said that these preconditions for consumer sovereignty are not fulfilled in the marketplace. Consumer advocates argue that buyers are faced with high information costs, offensive and often misleading advertising, and shoddy and unsafe goods. They have sharply attacked these obstacles to consumer choice, also criticizing the whole market economy for, among other things, the monopolistic practices applied by many producers and suppliers.

However, as was pointed out by Roland McKean (1973) in his address to the Southern Economic Association, the government sector has for the most part been ignored in these analyses and attacks, though both goods and services which are provided by market and by the government are relevant to the consumer's well-being. As he further says (1973, p. 285), "I conjecture that consumers lose more from monopoly power and non-profit incentives in the provision of most government services than from monopoly power in activities of similar magnitudes in the private economy." Under the very best of circumstances, which are rarely attainable, government sector output may approach the market in the extent to which the suppliers adapt to the consumers' desires. In reality, however, the supply side push may dominate the forces of the demand side pull in terms of the structure and pattern of public sector output, this for several reasons (see e.g. Diamond 1970, Breton 1974, Wagner 1978, Niskanen 1979): Though in a democracy there are at least two parties competing at the electoral level for the support of the voters, following an election there is only one majority for the duration of the legislative period and thus the winning party has a legal monopoly. Moreover, politicians and parties cannot be held legally liable for their promises, all the more so as the particular commodities promised are usually not very tangible entities but only vague notions. In addition, there is seldom any other recourse possible than changing one's vote at the next election. There is rarely any alternative to the government-provided goods and services; private options are often legally prohibited. Last but not least, as long as the

¹The author would like to thank Friedrich Schneider for his most helpful comments, and Sandra Stuber for editing the paper for English.

public services are financed through taxes (instead of through user charges), there is no free choice. Richard Wagner (1978, p. 95) concludes from this: "Whereas the natural state of the private sector seems to be competition, the natural state of the public sector in a democratic society appears to be monopolistic." Thus under a perspective of consumer sovereignty one would expect that consumer research, when extended to the government sphere, would have a large (and growing) area for analysis in the near future.

Though a monopolistic government has discretionary power which it can use for achieving its own ideological goals, it cannot do whatever it wishes. In the same way as a monopolist on the market is faced with the constraint of a decreasing demand curve, a monopolistic government is subject to constraints, of which the most important one in a democracy is the need to be reelected. One way for it to improve its reelection chances is to exploit fiscal illusions held by the voters. Fiscal illusion as used here is taken to mean the systematic misperception by voters of the public revenue burden they bear and the benefits they receive through government spending. It can be seen as a kind of rational ignorance in consumer choice, going back to the different natures of the various public revenue extraction items and the public services, and thus to their different information costs. It also relates to the only minor extent to which the consumers can influence the output of the public sector as compared to that of the market sector, and the weak incentives for informing oneself on the public policy under discussion and its financing that result therefrom.

In the following fiscal illusion will be analyzed in the framework of a model of interaction between the voters and the government² with Australia being taken as the concrete example, mainly because its political framework at the federal level is highly suitable for our analysis.³

Fiscal Illusion and Voters' Behavior

It is assumed that, in evaluating the government's performance, voters maximize their own utility: i.e., the extent of their support for the governing party - as measured by current popularity or by voting for the government at election time - depends on how satisfied they are with the government. Because of the high costs and low benefits involved, the consumer/voter in general has little or even no incentive for becoming fully informed about the government's performance. This is true with regard to informing oneself not only about one's present and future burden of govern-

²The general framework used here was developed by Frey and Lau (1968). Its usefulness for empirical research has been demonstrated by a series of applications, see Pommerehne, Schneider and Lafay (1980) for a recent survey.

³The federal government controls over 70% of all public revenues and expenditures, and also has the power to quickly make discretionary changes on the revenue side and, to a lesser extent, in public expenditures. For more details see Schneider and Pommerehne (1980a).

ment receipts and the benefits received through public spending programs, but also about the past, which may be discounted by the voter.

Besides this, one can expect that the information on the various revenue items which is absorbed on a day-to-day basis may be biased in a systematic way because of the different costs involved in becoming informed. These information costs are dependent on the revenue extraction system's degree of fragmentation. A three hundred year old tradition in public finance suggests that the more complex the mix of public revenue sources is, the more the individual will tend to underestimate the true burden of the cost of government. A series of small public receipts may be felt less than a single large tax, particularly if it is extended over time since small losses are more easily forgotten than large ones. Moreover, because of the varying degrees of visibility, of timing of extraction, and of hidden shifts in the revenue burden, it can be concluded that voters/taxpayers will perceive certain public revenue items (such as indirect taxes and public debt incurred) less than others (such as e.g. a pay-as-you-earn income tax). All this opens up opportunities for the government to install extraction institutions that will ceteris paribus decrease the perceived cost of government and thus favorably influence the voter's evaluation of the government's performance (Pommerehne and Schneider 1978).

Voters may also be subject to systematic misperception of the benefits of public spending. These also involve positive information costs, which - due to the nature of the public goods and services - are often much higher than in the case of market goods. Many public goods have "credence" qualities (Darby and Karni 1973) which are characteristic of goods and services that are utilized in combination with other goods composed of uncertain properties. Because the quality of many public goods is difficult to judge even after provision, their evaluation may require a greater degree of costly expertise. However, it seems to be very difficult to derive a hypothesis as to what kinds of individually perceived public services will be systematically more or less strongly felt. This problem arises because of the very great difficulty in discriminating between actual and perceived benefits. Whereas the cost of government activities to a taxpayer via taxes and other forms of fiscal extraction is, to a large extent, identifiable - so that if the perceived and actual costs can be reasonably determined, the difference can be attributed to fiscal illusion - this does not hold for the spending side. For this reason the following analysis deals mostly with taxes and other forms of government finance, and does not try to distinguish between preferences and favorably biased perception when government spending is considered.

It is also assumed that voters rarely draw a close connection between public services and their actual financing sources. This assumption may be valid as the principle of non-affectation is used at the federal level. The link between federal revenue and expenditure is further obscured by the government's ability to incur debt.

As a measure of how Australian voters/taxpayers perceive and evaluate the government's overall performance, a monthly Gallup Poll was used showing the percentage of the citizenry "willing to vote for the Australian government at a federal election" at that point. This data series for current government popularity (POP) gives us the dependent variable. Turning to the independent variables, first the revenue system's fragmentation is measured by $\frac{1}{n} \sum_{i=1}^n (REV_i)^2$, where REV_i equals the i -th revenue amount for all federal revenue items. This measure of complexity moves closer to 1 the more the government relies on only one source of income; it is $1/n$ if it uses all n revenue

sources equally (each accounting for $1/n$ of the total). In order to capture the influence of the different items on the public revenue side, the various fiscal extraction devices are grouped into two classes of direct taxes (a pay-as-you-earn income tax and assessed income tax, and a tax on dividends and interest); one class of indirect taxes (sales tax, company taxes, license fees); one of customs and various duties; and one of government debt incurred. All these variables are measured as shares of total government revenue. The various spending items are also grouped into classes: transfers (social security and welfare payments to individuals, payments to disabled people, and retraining programs); health programs, education, general and scientific research, and cultural and recreation measures; investments in transportation and communication, water and electricity supply, other public utilities, and investment grants to the states; public administration, law and order, public safety, and legislative services; and defense expenditure. These variables are all measured as shares of total expenditure. Finally, the ratio of total federal revenue to GNP is introduced to take into account perception of the increase in the overall federal revenue burden that took place during the 1970s.

Government's overall performance, as perceived by the voters, relates however not only to the provision of public goods and their financing, but also to its ability to steer the economy. This aspect, i.e. that the federal government is held responsible for the state of the economy, is captured by using a monthly Gallup survey taken since January 1970 which asks "Are you satisfied (or not) with the current economic performance of the government?" as an additional independent variable.⁴

For government popularity a negative sign and a significant coefficient are expected for the ratio of total federal revenue to GNP, and for the measure of the complexity of the whole revenue system (i.e., the lower $\sum_{i=1}^n (REV_i)^2$ is, the more the government relies on all revenue items alike and thus the higher is ceteris paribus the current popularity standing). Negative signs are also expected for those individual revenue shares which are strongly perceived. Similarly on the spending side, a positive sign is expected for those individual expenditure items which are favorably perceived by a large part of the voters. A positive sign is also expected for the perceived government economic performance (i.e. the better the perceived ability of the government to steer the economy, the more voters tend ceteris paribus to support the governing party). All data are monthly and a three-month lag is taken for all independent variables as it is assumed that the citizen needs some time to notice a change in taxing and spending (using weighted lags up to one year, no major significant changes occurred in the following estimation results). To test how good the voters' memory actually is, it is assumed for simplicity's sake that it can be measured by a lagged endogenous variable and that the same size of discount rate holds for all independent variables (i.e. a Koyck transformation is used). The results for the GLS estimation of the government popularity function for the period 1971.1 through 1977.9 is given in equation 1.⁵

⁴Perceived government economic performance is again influenced by popular indicators of the current and past states of the economy, such as the rates of unemployment and of inflation, which can be easily culled from day-to-day sources; for a detailed analysis of this relationship see Schneider and Pommerehne (1980b).

⁵More exactly, both the equation for explaining perceived economic performance (fn. 4) and that for government's popularity are simultaneously estimated by GLS technique. This also explains why the variable "Government economic performance" in equation 1 is not lagged.

POP _t	=	0.49	Intercept	(1)
(Current government popularity)				
	+	0.47** (8.99)	POP _{t-1}	
	+	0.38** (5.47)	Government economic performance _t	
	-	1.21** (-4.70)	Total revenues as a share of GNP _{t-3}	
	-	0.51** (-3.89)	Complexity of the revenue system _{t-3}	
	-	0.76** (-5.41)	Taxes on earned income _{t-3}	
	-	0.33* (-2.50)	Taxes on dividends and interest _{t-3}	
	+	0.18 (1.12)	Indirect taxes _{t-3}	
	+	0.80 (1.64)	Government debt incurred _{t-3}	
	+	0.42** (4.19)	Transfer payments _{t-3}	
	+	0.12* (2.19)	Expenditure on health, education, recreation _{t-3}	
	+	0.01 (0.89)	Expenditure on transportation, water supply and electricity _{t-3}	
	-	0.42 (-1.36)	Expenditure on public administration, law and order, public safety _{t-3}	

d.f. = 62, $\bar{R}^2 = 0.97$, $h = 1.27$. The figures in parentheses below the parameter estimates indicate the *f*-values (one asterisk indicates statistical significance at the 95%, two asterisks at the 99% confidence level, using a two-tailed test); d.f. shows the degree of freedom; \bar{R}^2 is the corrected coefficient of determination; and *h* indicates the test statistics for autocorrelation.

The results show first that citizens do indeed seem to highly discount changes in the revenue and spending structures: The coefficient of the lagged endogenous variable indicates that over 95% of what happens is forgotten within the space of one year. Secondly, the government's economic performance as perceived by the voters does influence the government's overall performance (popularity) in a positive and significant way.

As expected, total revenue burden and the complexity of government's revenue system have both a negative and strongly significant influence on current government popularity. The marginal impact of these variables is also quantitatively important: An increase in, for example, total revenues as a share of GNP by 1 percentage point *ceteris paribus* reduces current popularity by 1.21 percentage points; in contrast, the marginal decrease in government economic performance of 1 percentage point is only about one-third of this (0.38 percentage point). Going to the individual revenue items, one only finds a highly significant influence on government's current popularity in the case of the two direct taxes. This result is fully in line with the

classic argument of public finance literature that direct taxes are most strongly felt, whereas indirect taxes and most other revenue items (especially government debt) are much harder to detect. Moreover, a stronger dampening impact on government support of direct (income) as compared to more indirect (sales) taxes has also been found recently for other countries, including the United States (Eismeier 1979). Looking at the spending side, one finds a significant positive marginal effect on government popularity only in the cases of transfer payments (mostly to private households) and of expenditure on health and education programs. The impact of the latter, however, is only one-third of the first. The relatively high marginal influence of transfer payments is not implausible as it may be argued that they strongly reduce uncertainty concerning future direct benefits and thus may be valued highly by voters. Even if not all voters/taxpayers end up being public transfer beneficiaries, the great majority of them may still expect to. Comparing the influences of both sides of the government budget, the total marginal influence of the revenue items is almost five times as large (in absolute terms) as that of the spending items. Again, a similar result showing a much stronger influence on voters' support for the government for the revenue side as compared to the spending side has also been found recently for the United States (Niskanen 1979).

In order to do an additional check of these empirical results, an *ex ante* forecast for the period 1977.10 through 1978.12 was done based on GLS estimates for 1971.1 through 1977.9. The prediction is clearly superior to naive forecasts (Theil's inequality coefficient is 0.05; the average root mean squared error, 0.62%; and the average mean error of deviation 0.63%), indicating that the most important variables determining evaluation of the government's overall performance have been taken into account.

To summarize: The empirical findings show that voters/taxpayers seem to be unaware of the full cost of government activities when the revenue system is fragmented and especially when those types of taxes and other means of revenue extraction are used which are harder to detect than others. There is also evidence that some spending items, such as additional transfer payments, seem to be more popular and more strongly felt than increases in other expenditures. These effects on both sides of the budget and the fact that the voters highly discount past government activities provides opportunities for the government to use its fiscal instruments in a systematic way to retain the popularity it needs to allow it to secure reelection.

Government's Behavior: The Strategic Use of Fiscal Illusion

Let us assume the existence of a government that wants to maximize its utility, here understood mainly as the pursuit of its ideological goals. Even if it is assumed that the government cannot be voted out of office in the middle of a legislative period, it is still subject to various constraints in trying to achieve its ideological goals, of which the most important is the reelection constraint. Thus the government faces a dynamic maximization problem of when to undertake what fiscal policy action in order to maximize its utility.

It is assumed that the government is not able to resolve this problem, and that it will behave in a satisficing manner. It takes the results of popularity surveys as the best current available indicator of its reelection chances. If the current popularity level of the government is high, and/or if there is plenty of time left until the next election, the government will use its various fiscal instruments

to pursue its ideological goals. In comparison with a right-wing government, a left-wing government will generally increase public sector activity including new and/or expanded spending programs. A look at the Australian governments in the 1970s shows that Labor (left-wing) governments explicitly stated preferences for more spending on education and improvements in the welfare and health care systems, and decreased outlays for national defense. The Country-Liberal (right-wing) governments stated preferences for a much smaller level of growth in current and future government activities. There were also major differences in preferences as to how to finance public expenditures. Whereas a Country-Liberal government tends to favor tax financing, a Labor government relies more on incurring additional debt.

If the reelection chances are indicated to be poor, the government will concentrate on securing reelection rather than on pursuing its ideological goals. For this purpose, the government - regardless of who is in power - will try to create favorable fiscal illusions on the part of the voters by means of a systematic revenue and spending policy, counting in the voters' short memory to aid it in this. Before an election it will decrease the more direct and strongly felt taxes in particular (such as personal income tax), and increase harder to detect revenues (especially public debt) in order to finance additional popular spending or to cover the deficit caused by the lowering of strongly felt revenues.

When formulating the use of fiscal instruments, it is assumed that in a state of low popularity the government will react all the more strongly the greater is the ratio POP*/POP (the critical level of current popularity POP* having the value of 51%). The second important factor is reflected in the discretionary variable "Time since last election" (TSLE), the obverse of time left until the next election, which takes the values 1, 2, ..., 30, 1, 1, starting with the beginning of each legislative period.⁶ The government has also to take legal obligations and the behavior of the public administration into consideration in using its revenue and spending instruments. It seems realistic to assume that the public administration does not simply follow the wishes of the government but rather tries to maximize its own utility. It tends to resist major changes because this may threaten its own position and prefers to make only small and incremental changes. It would therefore seem to be useful to take the past spending and revenue structures into consideration as these are the starting points for changes therein. In addition to the legal and administrative constraints, the government has to take changes in the balance of payments into account, as well as the size of the budget deficit (a positive sign in the case of deficit, and a negative sign for budget surplus), whose maximum equals the maximum incurable additional debt, which is set through the legal framework.

It is now possible to formulate the following equation for the i 'th fiscal instrument ($INST_{i,t}$):

$$INST_{i,t} = \begin{cases} a_1 & INST_{i,t-12} \\ a_2 & \text{Change in balance of payments}_{t-6} \\ a_3 & \text{Budget deficit deviation from the long term trend}_{t-6} \\ a_4 & \text{Current popularity standing}_{t-6} \\ a_5 & \text{Time since last election}_t \end{cases} \quad (2)$$

Administrative and economic constraints

Political constraints

⁶This formulation implies that the government's reaction six months before an election will be stronger than when the election is still far off.

$$\text{Ideological preferences} \begin{cases} + a_6 & \text{Ideological preferences of Country-Liberal governments}_t \\ + a_7 & \text{Ideological preferences of Labor governments}_t \end{cases}$$

for $i = 1, 2, \dots, 26$ ⁷.

The 26 instruments refer to 10 federal revenue and 16 spending items, and are calculated as shares of total revenue and total expenditure respectively in order to capture the changes in the structure and eliminate the typical time trend effects. The explanatory variables are lagged by six months as it is assumed that the Australian government needs at least half a year to react to political and economic changes (when a longer or more complex lag structure is used, the estimation and the forecast results do not improve significantly).

The theoretically expected signs of the administrative and economic constraints are $a_1 > 0$ (for both the revenue and spending sides), $a_2 < 0$ and $a_3 > 0$ for revenue items, and $a_4 > 0$ and $a_5 < 0$ for expenditure items. In the case of a_4 and a_5 , a decrease is expected in the share of those revenue items which are strongly felt by the voters, and an increase in popular spending items. For the two ideological parameters a_6 and a_7 the signs already discussed are expected, for example that a Labor government will have a greater preference for increases in education than Country-Liberal governments ($a_6 \text{ education} < a_7 \text{ education}$). The results of the simultaneously estimated revenue instruments are given in Table 1.

Table 1 shows that the two types of government have significant differences in their ideological preferences regarding which instruments are to be used. When there is leeway for action, as for example after an election, the right-wing government (Country-Liberal) reduces in particular the assessed income tax, the income tax on dividends and interest, and, to the greatest extent, government debt. A left-wing (Labor) government acts in the opposite manner. It increases indirect revenue items (including debt), and, among the direct taxes, especially those which will be reduced by a right-wing government. It is interesting to observe that both governments deviate from their ideological preferences regarding the revenue instruments when trying to secure reelection. Goaded by a low popularity level and/or an upcoming election, both governments significantly reduce direct taxes, the earned income tax the most. They also reduce excise taxes, which are shown separately on purchase receipts. The harder to detect revenues are deliberately not reduced; in some cases, government debt for example, they are even increased ceteris paribus.

Another interesting finding is that compared to a left-wing government, right-wing governments show a greater preference for a simple and comprehensive revenue structure that relies on relatively few and quantitatively important revenues items. However, when trying to secure reelection, both types of government follow a policy of fragmentation of the total revenue, thus creating a more favorable perception of the government on the voters' part.

An ex ante forecast was also done for 1977.10 through 1978.12 to provide an additional check of the empirical results. With one exception (government debt) the predictions for the 10 revenue items are far superior to those achieved with naive forecasting methods. The best predictions are those for direct taxes (Theil's inequality measures being

⁷The method used here to measure ideological preferences regarding the fiscal instruments is rather simple: The intercept is broken up into two dummy variables in order to capture the different use of instruments by Country-Liberal and Labor governments with a shift parameter. It is assumed that the ceteris paribus conditions are fulfilled so that ideological preferences can be measured in this way.

TABLE 1
ESTIMATION OF AUSTRALIAN GOVERNMENT POLICY FUNCTION^a

Instruments on the revenue side (as a share of total revenue)	Legal, admin. and economic constr.			Reelection constr.		Ideological constr.		Test statistics	
	Lagged endogenous instrument (t-12)	Change of balance of payments (t-6)	Budget deficit deviated from long term trend (t-6)	Current popularity standing (t-6)	Time since last election (t)	Country/Liberal (t)	Labor (t)	\bar{R}^2	h
Tax on earned income	1.08** (8.85)	-0.14** (-2.96)	0.12** (2.78)	-0.61** (-3.54)	-0.78** (-6.03)	0.89 (1.54)	-1.64* (-2.06)	0.89	1.07
Assessed income tax	1.03** (6.59)	-0.09** (-2.84)	0.07** (2.98)	-0.18** (-2.89)	-0.31** (-3.03)	-0.84* (-2.32)	1.96* (2.41)	0.84	1.19
Tax on income from dividends & interest	1.02** (7.54)	-0.10** (-2.99)	0.12** (3.03)	-0.11* (-2.54)	-0.17** (-3.84)	-0.59* (-2.11)	2.34* (2.12)	0.81	1.33
Company income tax	1.03** (7.98)	-0.06* (-2.71)	0.10** (2.84)	-0.03 (-1.03)	-0.09* (-2.11)	1.59 (1.68)	3.47* (2.58)	0.81	1.27
Sales tax	0.95** (5.41)	0.24 (1.58)	-0.19* (-2.08)	0.25 (1.84)	0.26 (1.91)	1.86 (1.30)	3.59* (2.64)	0.76	1.30
Excise taxes	0.96** (5.65)	-0.03 (-1.31)	0.02 (1.89)	-0.12* (-2.36)	-0.20* (-2.46)	1.36 (1.84)	3.47** (2.89)	0.74	1.36
Customs and duties	0.91** (5.49)	-0.08* (-2.71)	-0.03* (-2.14)	0.04 (0.86)	0.08 (1.03)	-1.34* (-2.11)	2.91* (2.64)	0.75	1.41
Estate, gift and stamp duties	1.02** (7.27)	0.22 (1.21)	-0.12* (-2.02)	0.17 (1.90)	0.32* (2.06)	1.03 (1.86)	1.84 (1.71)	0.71	1.47
Fees and fines	0.96** (6.59)	0.13 (1.39)	-0.18* (-2.21)	0.09 (1.84)	0.20 (1.64)	0.94 (1.64)	2.57 (1.89)	0.68	1.51
Government debt	1.04** (8.46)	-0.09* (-2.56)	-0.10* (2.61)	0.50 (1.58)	0.69* (2.03)	-1.57* (-2.59)	2.89* (2.57)	0.66	1.49
Complexity of the revenue system	0.41** (3.10)	---	---	-0.27** (-2.99)	-0.34** (-3.69)	0.89** (3.63)	0.27** (2.81)	0.58	1.47
Complexity of the revenue system	0.44** (3.54)	-0.21 (-1.89)	-0.34* (-2.54)	-0.24** (-2.77)	-0.30** (-3.19)	0.94** (3.88)	0.36** (2.89)	0.58	1.44

^a As it is assumed that the various instruments are not independently used by the Australian government it is very likely that the disturbances from them are correlated. If so, the technique of multivariate regression gives more efficient estimates, and is used here. For notes see eq.1.

ca. 0.20; the average root mean squared errors ca. 1.75%; and the average mean errors of deviation ca. 1.20%), i.e. those instruments which are discretionarily changed before an election (and there was a general election in March 1978).

For reasons of space the empirical results for the spending side are not presented here but only briefly discussed. Looking at the ideological differences between both types of government, Country-Liberal governments favor additional expenditure for capital formation, defense, and law and order. A Labor government prefers to decrease the last two and favors instead additional transfer programs (including foreign aid), education, and health care. Again, if a government is afraid that it will not be reelected, all transfer payments to individuals and households (but not foreign aid) are significantly further increased in addition to the most favorably perceived expenditure on education and health. This strategy of increasing especially transfer payments before an election seems to be quite common, and can be observed to operate similarly for instance for the United States (Tufté 1978, chapter 2). The ex ante forecasts which have been made to test the model's predictive ability in an election year lead to superior results for 11 of 16 spending items (evaluated by Theil's inequality coefficient which is smaller than 1). The instruments for se-

curing reelection again provide the best forecast results (with an average mean error of deviation of less than 1.5%).

If these results are compared to the findings for the revenue side, it can be seen that the use of the spending instruments is of less importance in securing reelection. This is not implausible and may simply reflect the often stated greater rigidity of the expenditure side as compared to the revenue side.⁸

To summarize: The general hypotheses in this paper are that (1) voters will be rationally underinformed and thus subject to systematic misperceptions of fiscal variables, and that (2) the government in a representative democracy will try to exploit such fiscal illusions in order to strengthen its position, especially when it is trying to secure reelection. The general framework of the analysis is that of a monopoly, though one its need for periodic popular approval. Empiri-

⁸ Additional evidence for this argument is gained when a regression is carried out with the lagged endogenous variables of the revenue and spending items (in billions of Australian dollars), thus capturing only the legal and administrative influences. The mean of the explained variance is then 82.5% for the revenue items, but 93.1% for the expenditure by categories.

cal evidence for systematic misperception of the cost of government is offered by the results for the government's popularity function: The findings show that voters/taxpayers seem to be unaware of the full cost of government activities when a complex revenue system and indirect and more hidden revenue items are used. There is also some evidence that a few spending items are favorably perceived by the majority of voters, though it remains open how much of this is due to preferences and how much due to misperception. This plus the fact that voters discount past government activities all helps to provide opportunities for the government to act as if it had a monopoly position. As the empirical results show, significant differences between the two parties' use of fiscal instruments for ideological purposes appear after an election was taken. However, when trying to secure reelection each government deviates from the pursuit of its ideological goals if necessary and uses the fiscal instruments in a predictable fashion in order to exploit fiscal illusions held by the voters.

Conclusions and Suggestions

Several conclusions and suggestions arise from this analysis, all relating to the provision and diffusion of more and better information on the government sector to politicians and consumers. However, it seems useful to distinguish between information dealing with government sector output, i.e. its product and cost, and with the institutions and processes of real world public decision-making.

To start with the first: Enough has been said in the above to conclude that the true cost of government to the consumer/taxpayer has a large grey area and may be systematically misperceived in many circumstances. Under the perspective of consumer sovereignty, which stresses the right to information leading to meaningful purchasing decisions, better knowledge of the true and of the perceived cost of government to the individual would seem desirable. Some work has begun in this direction, for example with the development of a consumer tax index for Canada (Walker 1976). Moreover, it can be expected that especially the analysis of the perceived tax burden and its determinants may profit from applied perceptual and learning psychology, and from consumer research and its techniques of direct consumer inquiry. The right to information also requires knowledge of the true and the perceived nature of government services, their quality, how long they are provided, and, for example, their spatial distribution. It would seem that this kind of information could be strongly improved using techniques of consumer and marketing research, e.g. multiattribute product analysis and hedonic price approaches (Palda 1980).

However, even if this kind of information is greatly bettered, the problem remains that there may be still only be weak incentives to exploit it. Non-neglectable costs arise to the individual voter in making collective decisions, and the additional cost of doing so on a well-informed basis may far outweigh the gains. As there are usually only a limited number of alternatives to be chosen among, a system of mandatory payments prevails, and the number of voters is large, it may even be irrational for the individual to become sufficiently informed to make an intelligent decision. Under this perspective it would be desirable to improve in particular information on the functioning of the government's institutions and on how the political process really works. This information relates to such questions as to how strong (or weak) the competition between parties in a representative democracy really is, and thus how small (or large) the leeway is that an elected government has to deviate from the voters' wishes. Moreover, it would also provide the basis for research on how the real-life political process and the functioning of government institutions can be influenced so

that the wishes of the voters are - quasi automatically - better taken into account by the government.

References

- Breton, Albert (1974), The Economic Theory of Representative Government, Chicago: Aldine.
- Darby, Michael R. and Karni, Edi (1973), "Free Competition and the Optimal Amount of Fraud," Journal of Law and Economics, 16, 67-88.
- Diamond, Harold S. (1970), "The Budget and Consumer Sovereignty," American Journal of Economics and Sociology, 29, 163-177.
- Eismeier, Theodore J. (1979), "Budgets and Ballots: The Political Consequences of Fiscal Choice," in: Douglas W. Rae and Theodore J. Eismeier (eds.), Public Policy and Public Choice, Beverly Hills and London: Sage.
- Frey, Bruno S. and Lau, Larry J. (1968), "Towards a Mathematical Model of Government Behavior," Zeitschrift für Nationalökonomie, 28, 355-380.
- McKean, Roland N. (1973), "Government and the Consumer," Southern Economic Journal, 39, 481-489.
- Niskanen, William A. (1979), "Economic and Fiscal Effects on the Popular Voter for the President," in: Douglas W. Rae and Theodore J. Eismeier (eds.), Public Policy and Public Choice, Beverly Hills and London: Sage.
- Palda, Kristian S. (1980), "The Public Sector and Marketing: Should the Twain Meet?" Forthcoming in Proceedings of the Second Annual Special AMA Conference on Marketing Theory.
- Pommerehne, Werner W. and Schneider, Friedrich (1978), "Fiscal Illusion, Political Institutions, and Local Public Spending," Kyklos, 31, 381-408.
- Pommerehne, Werner, Schneider, Friedrich, and Lafay, Jean-Dominique (1980), "Les interactions entre économie et politique: synthèse des analyses théoriques et empiriques," Forthcoming in Revue Economique.
- Schneider, Friedrich and Pommerehne, Werner (1980a), "Politico-Economic Interactions in Australia: Some Empirical Evidence," Economic Record, 56, 113-131.
- Schneider, Friedrich and Pommerehne, Werner (1980b), "Illusions in Fiscal Policy: A Case Study," Forthcoming in Statsvetenskaplig Tidskrift.
- Tufte, Edward R. (1978), Political Control of the Economy, Princeton: Princeton University Press.
- Wagner, Richard A. (1978), "Advertising and the Public Sector Economy: Some Preliminary Ruminations," in: David G. Tuerck (ed.), The Political Economy of Advertising, Washington, D.C.: American Enterprise Institute.
- Walker, Michael ed. (1976), How Much Tax Do You Really Pay? Introducing the Canadian Consumer Tax Index, Vancouver: Fraser Institute.

COMMENTS

Robert Ferber, University of Illinois

The three papers on which I have been asked to comment deal with a variety of different issues related to consumer behavior in the public sector. The first paper, by Deacon, deals with opposing views on whether the consumer, as a voter, exhibits economic rationality in making choices at the polls. The second paper, by Vehorn, raises the question of the extent to which consumers are likely to substitute private goods for public goods in their purchasing behavior. In contrast to these first two largely theoretical papers, the third paper, by Pfaff and Kistler, seeks to present empirical information on the desirability of providing consumers with better information about public programs. Since these papers are so distinct, it seems best to discuss each of them separately.

The focus of the paper by Deacon is to question whether voters are indeed as well informed and economically rational as is made out for them in voting models. Although I am not that well acquainted with this literature, the studies cited in this paper, especially those referring to the desire by voters for maintenance of the status quo, would seem to cast doubt on the hypothesis of economic rationality as usually defined. Certainly, the latter finding suggests that voters are influenced more by habit persistence than by any desire to bring marginal benefits into line with marginal costs.

One other key question discussed in this paper is whether majority rule leads to competitive or noncompetitive equilibrium. As the paper points out, the evidence on this point is mixed. My own hunch is that the evidence will continue to be mixed, since much is likely to depend on the particular situation. If a governmental body is in firm control, with an electorate that is either passive or relatively satisfied, noncompetitive equilibrium is much more likely. On the other hand, if things are in a state of ferment or the governmental body is unsure of having the support of the voter, competitive equilibrium would seem to be more probable.

From an empirical point of view, I am not convinced that the finding that voters prefer "no change" indicates competitive equilibrium. Survey experience suggests that such statements are as likely as not to reflect lethargy on the part of the respondent, or lack of information, as well as satisfaction with the status quo. It is not clear from Deacon's paper whether any attempt has been made in these prior studies to distinguish among these different possibilities.

It is also not too clear from this paper what exactly is the meaning of "majority rule equilibrium." In theory, this equilibrium is based on the median of the individually preferred outcomes. In practice, however, it is not clear how well defined this median may be, how it may be estimated and how stable it may be over time. Possibly the other studies mentioned in the paper have dealt with this problem. To me, however, this seems to impose all sorts of informational requirements on the voters that may be highly unrealistic.

The Vehorn paper deals with another aspect of market efficiency, namely, the substitutability between public and private goods. In fact, this paper might just as easily have been titled, "Complementarity and Substitutability between Public and Private Goods." In many instances, it is clear both from the paper and personal observation that

private goods do not seek to substitute for public goods, but to complement them. This is true, for example, in the case of protection services, purchase of insurance and various aspects of transportation.

It might also be noted that the finding of substitutability was based on analyses where governmental divisions were the unit of observation. A very different result could be obtained if the unit of observation were the individual family, something which would seem to be more meaningful if the focus is on understanding consumer behavior.

At the same time, there is no question that substitutability is definitely present and, as Vehorn points out, exists in the case of education. On that topic, however, I feel that one question posed in the paper is a rather odd one, namely, "What is the effect of private education on the provision of public education?" Certainly, the experience of the last few decades suggests that private education springs up in response to inadequate public education, so that perhaps the more meaningful question should be the effect of the provision, and quality, of public education on private education.

Perhaps the basic question relates to the type of services being demanded by the consumer. If these services are not being supplied by the public sector, the chances are very high that an industry will arise in the private sector to meet this need. At the same time, it should be recognized that particular services may be sought from the private sector even though they may also be available from the public sector, if the latter services are not adequate or sufficient in the mind of the consumer. It is for this reason, for example, that the life insurance industry has been flourishing despite the growing importance of the Social Security System.

Whether one refers to these services or goods as attributes or something else, one principle is still the same. (Incidentally, it is about time that people stopped referring to the attribute approach of Lancaster as "a new approach to consumption"; it is now nearly 15 years since this theory first appeared.) The fact remains that even though public services may be adequate for the majority of the people, they may still not be adequate for enough others, so that it pays for the private sector to get involved in providing similar services.

The conditions under which the provision of such private services will flourish has received some study, as noted by Vehorn, but it is also clear that much more study of this question is needed. Moreover, with the growing taxpayer revolt and the desire to minimize government involvement in the economy, such studies should hopefully also provide information on the extent to which various services should be provided by the public sector, and the extent to which supplementary services should be left to the private sector.

The paper by Pfaff and Kistler presents some survey data and comes to the conclusion that, "increased understanding of and more information about public transfers is needed on normative as well as practical grounds." These are very logical conclusions. The only problem is that it is not at all clear how they follow from the data presented in this paper.

In fact, two sets of data are presented. One set of data (Table 1) gives the results of an opinion poll taken of

some undefined population, with no indication of sample size or possible statistical significance, and with no explanation of the data. What it seems to provide is estimates of the percentages of different population groups having different attitudes toward the retention of grant-in-aid programs to old-age insurance, and to unemployment insurance, respectively, though the percentages do not total 100 percent. On the basis of these data, the conclusion is drawn that "more information especially about the situation of other people and groups could lead to a reconsideration and reevaluation of one's own relative position compared to that of others with regard to the public budget constraints . . . understanding the necessity to assume burdens and to aid others could curb an excessive inflation in demand."

Actually, what this table shows is that, for example, people receiving old-age insurance very much favor the continuation of such a program, and a substantial proportion also favor unemployment insurance. Similarly, people receiving public assistance (many of whom may also be receiving old-age insurance, but this is not clear) also heavily favor unemployment insurance.

If these data indicate anything, therefore, it is that people are very much influenced by self interest in their attitudes on public transfer programs. Moreover, they are likely to favor public transfer programs other than what they are receiving, possibly out of a feeling of sympathy and the need for mutual support to ensure maintenance of their own program. Since there is no indication in this table as to which respondents have more information or less information, the authors' earlier cited inference on the importance of information would seem to have no basis whatsoever.

The second conclusion, drawn from Table 2, seems equally doubtful. It is that "we can conclude that consumer information is an important prerequisite for systems performance." This conclusion is based on a table which shows a positive relationship between satisfaction with informational activities by young people in Augsburg and satisfaction with transfers to their own group. However, the same table shows almost equally strikingly that those who are satisfied with informational activities are much more likely than the dissatisfied to believe that one cannot influence democratic processes. Moreover, those who are satisfied with informational activities seem to have no more political interests than those who are dissatisfied.

Perhaps this table can be interpreted differently, for once more the data are not explained, and percentages do not add up to 100 percent. Also, a symbol, epsilon, is presented with no explanation of its meaning.

Under the circumstances, one can only agree with the later statement in this paper, "the conclusion that informed consumers would help create a more rational transfer policy and reduce selfish aspirations surely is obvious, but difficult to prove." Whether this point can be proven or not, at the least one would hope for a more meaningful framework for investigating this question and a reasonably careful empirical analysis.

THE IMPACT OF PROGRAM EVALUATION NEEDS ON RESEARCH METHODOLOGY

R. Bruce Hutton, University of Denver
Dennis L. McNeill, University of Denver

Abstract

This paper presents a perspective on impact evaluation research that stresses the diagnostic role of evaluation in addition to the more common end-results evaluation. This leads to several implications for research--most notably that the research should require multiple methods to accurately diagnose program effects. An example of a recent program and the attendant methodological decisions illustrate the implications of the application of this perspective to evaluation research.

Introduction

Recently an increased emphasis in program evaluation has emerged in the public sector. While several different forces have spurred this emphasis, the motivation comes principally from the growing dissatisfaction with seemingly ineffective programs (e.g. Howard and Antilla 1979) and the desire for a more systematic approach to policy decisions (e.g. Mazis and McNeill 1978).

However, the use of evaluation research by policy makers remains naive and is characterized by a justifiable hesitation as to the ability of research to evaluate complex social programs. Some of this hesitation comes from a lack of experience with research design, but there is a good portion of this reserve which is justified by the embryonic practice of program evaluation. There are inherent problems in the research field which revolve around the selection of a proper research design, development of valid dependent measures, use of appropriate analysis techniques, and drawing valid conclusions from the data. These problems are compounded by a lack of systematic feedback to the relevant disciplines regarding evaluation research that has been completed. However, as evaluation research becomes a larger part of the academic literature, it is clear that key methodological decisions have rendered many of the past evaluation efforts meaningless (see, for example, Phillips and Caider 1980).

This paper describes the design of a recent evaluation project (Hutton and McNeill 1980) which incorporated a multi-method approach to answering the questions about program impact. This approach has been suggested in the past (see Campbell and Fiske 1959, and Heeler and Ray 1972) but has received virtually no support through implementation in evaluation research. Reasons for this oversight range from cost constraints to a myopic view of the evaluation problem.

The Evaluation Problem

The literature on evaluation is replete with various terms describing evaluation typologies. Scriven (1967) describes evaluation research as formative (pretesting or developmental research) and summative (measuring program effects) evaluation. Additionally, Freeman (1976) indicates a form of evaluation called process evaluation which investigates the procedures that were used to implement a policy program. Essentially, the research questions which can be answered by evaluation are of three generic types:

• End-Result Evaluation

Did the program accomplish the goal for which it was originally designed?

• Diagnostic Evaluation

Why did the program results occur?

• Formative Evaluation

What components of the program should be included?

It should be noted that evaluation research often occurs after the program goals and components have been determined. Consequently, the role of formative evaluation is likely to be most aided by careful diagnosis of past efforts. As noted by Van Maanen (1979) much can be gained by findings with comparative groundings--those conclusions developed by contrasting programs with similar goals but different means of achieving these goals. Generically formative evaluation is different from the attempt to chronicle program effects through end-results or diagnostic research efforts.

End-result evaluation is the most common approach found in the design of evaluation studies, and, of the two tasks, is the easiest to operationalize. The end-result analysis requires a specification of the level of ultimate effect of the program (often behavior) and the design moves to that level once the program has had time to work. Clearly, goals of the program and timing of the results key the methodological decisions. Descriptive survey methods dominate the approach to this first task, although there is some use of quasi-experimental designs. However, these methods often suffer from an inherent inability to explain why results occurred (e.g., surveys) or from threats to internal validity in certain commonly used quasi-experimental designs. In this case the policy decision must then rely on post-hoc explanations of results which can never be fully defended.

Diagnostic evaluation is a different and more complex task. The goal of this research is to profile program impact in a way that understanding of the reasons for that impact can lead to adjustments in the current program and better program decisions in the future. The increased complexity of this type of research question requires the researcher to consider many more facets of program effectiveness that include the following:

- goals of the program;
- timing of program effects;
- timing of, and effects of individual program components;
- external factors which could influence program success or failure (e.g., conflicting programs already in place);
- individual difference variables in the population;

- the potential uses of the program results; and
- the time frame for future policy decisions.

The diagnostic evaluation process makes a more concentrated attempt to profile the environment in which the program must work and takes a more systematic approach to the evaluation than an end-result perspective. In addition, a diagnostic evaluation recognizes the limitation of a single methodology to answer all research questions.

The following sections will describe a case study of research design decisions which were intended to address both the diagnostic and end-result evaluation needs of a public policy program. In addition, results of the evaluation will be highlighted.

The Program to be Evaluated

The program evaluated was called the Low Cost/No Cost Energy Conservation Program (LC/NC) and was designed and implemented by the U.S. Department of Energy. The program was introduced in the six New England states in the fall of 1979 and consisted of the following components:

- A LC/NC booklet which described eleven categories of home energy conservation tips. These tips could be completed by the household for low or no cost and the eleven tips could save 25% of the home fuel costs in the winter.
- Paid advertising in radio, television, and newspapers. This activity consisted of professionally designed ads run in the New England region from November 8 to December 2, 1979.
- A showerflow control device given as an incentive with each LC/NC booklet. The device was a small plastic cone-shaped object which, when installed in a showerhead, restricts the flow of water (and use of hot water) from the shower. This was also one of the eleven LC/NC tips.
- Some public relations efforts to initially introduce the program to the region.

The focus of the program was the LC/NC booklet which was hand delivered with the showerflow device by the U.S. Postal Service to 4.5 million households in New England. All of the other program components were designed to inform the population about the program and to motivate the population--through the expected savings--to complete the LC/NC tips.

Several research questions were relevant to this program but two were prominent for the research methodology:

- Did the program stimulate LC/NC energy conservation behavior?
- Did the advertising contribute to the program success?

The first question is the end-result evaluation which essentially asks for a tally of energy conservation behaviors that resulted from the program. However, the timing of these behaviors is critical. The end of the year presents many competing demands on the consumer pocketbook and time, but due to the nature of future policy decisions results were necessary by January 1980.

The second major question which impacted on research revolved around the role of advertising in program success. Since paid advertising was a relatively new and controversial component of public programs, it was felt that the role of advertising in improving program effects should be documented. The difficulty now lay with matching methodology to evaluation needs.

Matching Evaluation Needs with Methodology

The End-Results Evaluation

Table 1 presents the basic research questions and the multi-method designs which were used to gain diagnostic answers from program results.

The end-results evaluation posed some difficult problems in design. First, the after-only design (a survey) alone would not be able to provide a systematic basis for judging what the conservation behaviors would have been without the LC/NC program (Campbell and Stanley 1963).

TABLE 1

Description of Research Designs

I.	What were the program effects?	An after-only quasi-experimental design with New England (n=1207) as the treatment condition and upstate New York (n=604) as a control were used to test the effects of the LC/NC program on conservation behaviors.
II.	What was the effect of paid television ads?	A panel study was used to gain an unobtrusive measure of advertising exposure. Of those completing the diary, 92 households were given the same questions as above to correlate television exposure to LC/NC behavior.
III.	Do the LC/NC ads have the potential to work?	The two LC/NC ads were tested in a day-after recall test to ascertain the ads' potential to inform New England and thus to contribute to program success. Males and females were approximately equally represented in the sample (n=305) testing recall of the commercials.

Thus, both high success or failure would have been easily assailed without a basis for defense. Consequently, it was decided that a comparison group design be employed.

There are many critical decisions to make in the selection of a matched control condition for this evaluation. In order to make a proper selection and to avoid the negative artifacts of this process, the goal of the evaluation must be clarified. This evaluation must show whether the LC/NC program stimulated energy conservation behavior. Thus, the matching criteria must be those that would increase the probability of response to energy conservation (e.g., severe winters, dependency on expensive fuel, etc.). In other words, the control must be at least equally ready to respond to this program due to the need for these specific conservation behaviors. It was decided that New York, particularly the upstate area which included Rochester and Buffalo and rural Genesee county, matched New England well on several a priori characteristics:

- . heating oil dependence;
- . severity of winter;
- . age and structure of homes;
- . availability of previous conservation efforts;
- and
- . demographic profile of the population.

However, because New York matched New England on the readiness to respond to conservation, the region also carried with it a considerable amount of conservation activities. These conservation activities were more prominent than New England, and, as the results were to show, caused New York to evidence more conservation behaviors prior to the program than the level found in New England. In a strict sense the groups were not matched on actual prior conservation behavior but were matched on the more critical dimensions which characterized the readiness of the region to respond to the LC/NC program. The unequal level of conservation activity cannot be ignored for it introduces some alternative hypotheses to the results that were measured after the program was in place. The alternate hypotheses affected both the analysis and the questionnaire design.

Three alternative hypotheses were most critical to this study: 1) a different group hypothesis, 2) a ceiling effect hypothesis, and 3) an alternative causal element hypothesis.

The different groups hypothesis states that people who are currently eligible for LC/NC behaviors in New York and New England are not alike. Since the current proportion of the population completing LC/NC behavior in New York is higher, the easy to persuade segments of the population may have already been reached. Comparing the responses since the program inception is probably an unfair test due to the increased difficulty of persuading later adopters in New York (Shoemaker 1979).

The ceiling effect hypothesis operates in the same direction but for a different reason. A ceiling effect would suggest there exists some asymptotic level of behavioral response to energy conservation beyond which only minor changes in behavior are possible. Thus, this potential alternative hypothesis also makes inappropriate a comparison of the responses since the LC/NC program if the initial levels of response are not comparable. The changes in New York would be smaller by definition because these responses are closer to the ceiling.

The alternative causal element states that whatever caused New York to be high is now causing the behavioral responses in New England to change.

All of the above have implications that lead the analyst to consider only the overall behavioral penetration

levels in the test between the treatment (New England) and control (New York). If overall behavioral change is greater in New England than New York for the LC/NC tips, the alternative hypothesis become unlikely explanations for the results. The different groups hypothesis is not operable since New England and the LC/NC program would be able to better penetrate the hard-to-persuade segments. The ceiling effect hypothesis would be a poor alternative since the responses would have to exceed the hypothetical ceiling for New England to produce significant behavioral change. Finally, the alternative causal element would not be a likely explanation since the LC/NC program would produce changes greater than any hypothesized other causal element.

In the case where comparison groups cannot be equated on behavior prior to the program, the analysis then must be constrained to a test of overall behavior--rather than behavior since the program. This does not overcome the matching problem but does eliminate the alternative hypothesis endemic to a pre-post comparison, and serves an additional benefit of providing a rather stringent test of program impact.

The lack of matching on prior levels of behavior has implications for the questionnaire development. The questionnaire must be developed in order to measure behavioral response to the LC/NC energy saving tips. However, because of the focus on behavior, the short duration between program introduction and measurement, and the lack of a perfect match on prior levels of LC/NC behaviors, the questionnaire content, as shown in Figure 1, must also address the following:

- . the ability of respondents to do the LC/NC tips,
- . the reading of the LC/NC booklet,
- . the timing of the completion of the tip,
- . intentions for completion of the tip, and
- . reasons for not intending to complete the LC/NC tips.

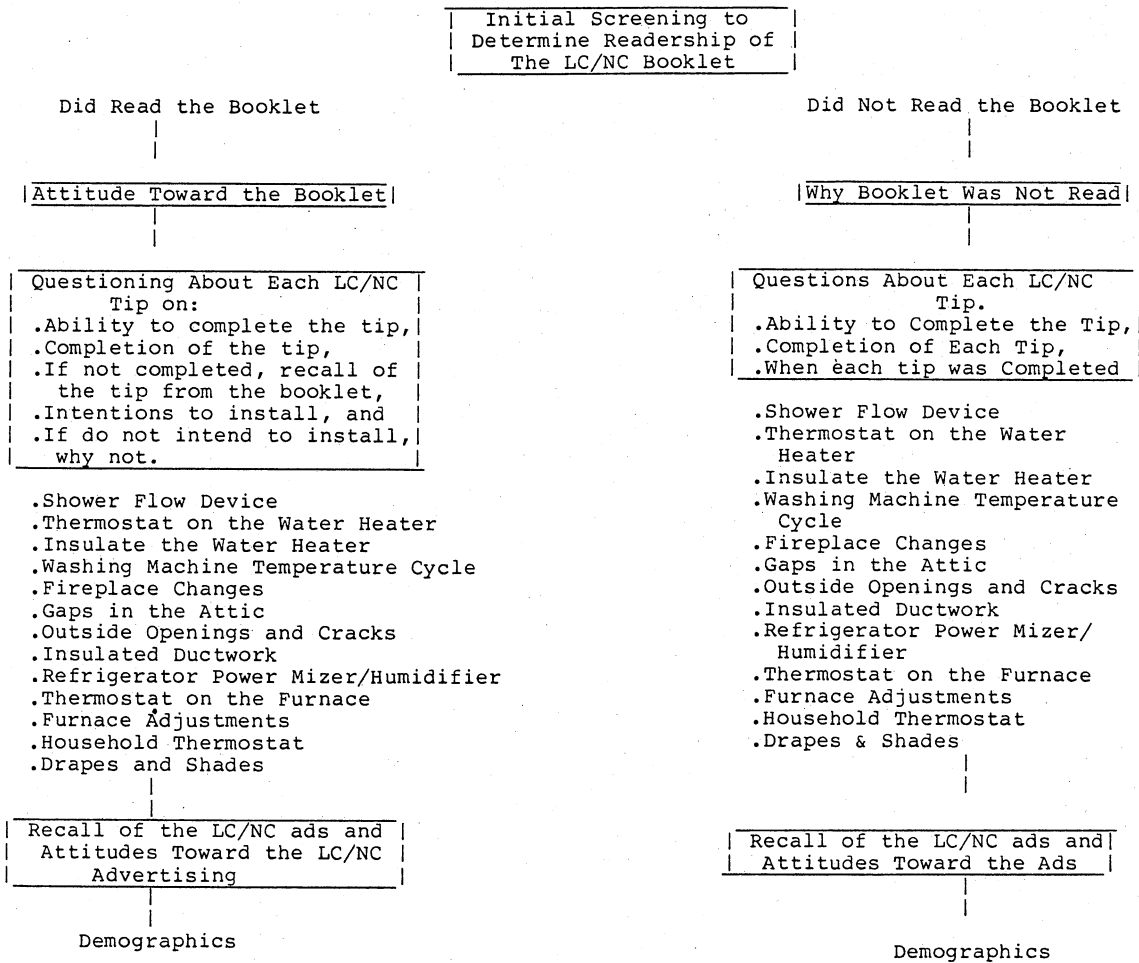
As in most studies particular attention must be given to the validity of dependent measures. In this case the concern is with the reliance upon self report behaviors, and the potential cue for "right" answers given the need to reference the LC/NC booklet. The validity of these responses was examined in two ways. The first is face validity following from the logical flow of the questions. For example, a yea-saying bias may push the respondent to say they have completed the LC/NC tip recently. If a subject said no, that the LC/NC tip had not been completed since the program, they were then asked if they saw the tip and intend to complete it later. A no answer to these two questions was followed by "why don't you intend to complete the tip?" It is here that respondents can indicate it has already been completed. In order to assess the validity of the timing of the completion of the tip, a percentage of the respondents should have followed the above sequence to indicate that the tip had been done before the program. On the face of the logic of these patterns of responses the researcher has some increased confidence in the validity of the timing of the behavior.

However, this does not address the presence of a bias to report completion of the tip when this is not the case. This was handled by a second phone call three weeks following the initial interviews to a sample of households to test the responses to key LC/NC tips. In a sample of 180 households in the treatment and control condition, 96% of the original responses were confirmed, and the remaining 4% were explained by an inability to query a knowledgeable member of the household regarding the LC/NC tips.

The final methodological consideration involves sampling in both the treatment and control condition. A probability sample of 1207 and 604 in New England and New York

FIGURE 1

The Components of the LC/NC Questionnaire



respectively was prepared from telephone exchanges in two strata—big cities, and non-metropolitan areas. The method included unlisted numbers and new listings in their correct proportion.

The Diagnostic Evaluation

The diagnostic evaluation focused on the contribution of paid television advertising and resulted in two separate studies being completed. The first answered the question of the relationship of exposure to LC/NC television ads on the impact of the LC/NC program. The second study addressed the issue of the potential of the LC/NC ads to aid program impact.

The Relationship of the LC/NC Ads to Program Impact. This study had considerable importance to this program and to others desiring to use paid advertising as a part of an integrated policy program. There were three specific goals to this aspect of the evaluation. The first was to describe the relationship between the exposure to the ads and program results. The secondary goals of this study were to examine the efficacy of the media-buying decisions and to serve as cross validation for the end-results study. The key issue for the first goal was to implement

a methodology whereby exposure to the LC/NC ads could be unobtrusively measured and related to the overall program impact. In order to accomplish this, a sample of households in the greater Boston metropolitan area were selected from an existing mail panel to keep a diary of their adult T.V. watching behavior for a period November 8 to December 2, 1979. This period corresponds to the period when paid television advertising was in operation. A sample of the diary format is found in Figure 2. An original sample of 250 households was selected and by December 15 154 (61.6%) households had mailed in the set of four completed diaries.

The task then was to code the diaries so that the television watching could be matched to the times when the LC/NC ads were shown. The relationship to the LC/NC program effects was accomplished by a telephone interview using the questionnaire from the end-results evaluation. This questionnaire was administered to 92 households from December 15 to December 24. The exposure to the LC/NC ads was then related to the impact of the LC/NC program effects.

FIGURE 2

(Sample Format)

TV Viewing Questionnaire for Saturday

	Male/Female Head of Household Watching TV?		Station No.	Program Name	Head of Household	
	No	Yes			Male	Female
7:00 AM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>
7:30 AM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>
8:00 AM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>
8:30 AM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>
9:00 AM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>
⋮	⋮	⋮	⋮	⋮	⋮	⋮
11:00 PM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>
11:30 PM	<input type="checkbox"/>	<input type="checkbox"/>	→ _____	_____	<input type="checkbox"/>	<input type="checkbox"/>

Several important considerations need to be addressed with this design. First, the amount of T.V. watching is likely to be correlated with LC/NC ad exposure. However, the key goal of this design is the assessment of increased "real world" exposure to the ads on LC/NC conservation tips. Thus, while the design cannot causally link increase repetition with program effects, it can: 1) describe the pattern of behaviors and ad exposure found in the sample population; and 2) through careful analysis constrain the ability of alternative hypotheses to explain results. In addition, the correlation between ad exposures and LC/NC behaviors is most likely to be explained by overall television watching behavior in cases where ad exposure and LC/NC behavior are linearly related.

As mentioned earlier there are two secondary goals for this study. The exposure to the ads can confirm the advertising buying decision used in this program. The advertising was purchased on the basis of cumulative gross rating points that indicated 81.5% of the population would be exposed to LC/NC ads at least 12 times. The pattern of exposure of this panel to the ads can confirm (or disconfirm) the efficacy of this media buying decision.

Further the results of the survey can serve as a comparison to the end-results evaluation methodology in terms of the penetration of the LC/NC behaviors.

The Memorability of the LC/NC Ads. In order to determine the role of paid advertising in the effects of the LC/NC program, it is necessary to rule out an alternative hypothesis that the ads cannot contribute due to poor communication effectiveness. This essentially asks "Do the LC/NC ads have the communication potential to contribute?"

This task was accomplished by testing the recall of the LC/NC ads with a random sample of commercial viewers the day after the first exposure to the ads. These tests were run on November 9 in Boston with a sample from each of the

two versions of the LC/NC ad. The sample (n=305) was screened to make sure they were in the commercial audience, and were then asked to recall the commercial and its messages. The commercial recall is then compared to industry norms for 30 second product commercials, and, in addition, the messages recalled were compared to the intent of the LC/NC program.

Thus, the diagnostic goals approach has resulted in a multi-method approach to the evaluation of the LC/NC program. This multi-method design has the advantages of removing much of the need for post-hoc guesses as to the reason for program effects and, as the next section indicates, the results provide a more complete profile of program impact.

Relating Research Output to the Research Questions

As stated earlier two research questions guided the research design. Presented below are several highlights of the results of the diagnostic evaluation of the LC/NC program.

What were the program effects?

- .. Readership of the LC/NC booklet was high (71.1% of the sample in New England).
- .. Four out of eleven LC/NC energy conservation tips evidenced significant (p<.05) improvement when comparing New England to New York.
- .. New England had completed more LC/NC energy saving tips than New York (New England 4.63 v. New York 3.96; t = 5.0, p<.001).

- .. The biggest predictor of the degree of response to the program was the amount of prior behavior. Those with the greatest response to the program had completed significantly fewer LC/NC tips prior to the program.

The results of the second research question yielded similar diagnostic findings:

- What was the relationship of paid television advertising to program effects?
 - .. The recall study indicated the LC/NC ads communicated as well as the normal 30 second commercial.
 - .. The 25% savings claim was recalled as a poor third selling point. This took on added meaning when it was found that those believing this claim did more LC/NC tips.
 - .. Advertising exposure was related to enhanced program effects when increasing exposure from very low levels (0 to 1 exposure per week) to 1 to 2 exposures per week. However, further increases in exposure were found to be a significant decrease in LC/NC behaviors. This raises the specter of an early wear out for this type of ad, and questions the use of a decision rule which is founded in a "more the better" philosophy for the use of advertising in conservation programs.
 - .. Only 9% of the males and 22% of the females reached the level of exposure that the cumulative gross rating point media-buy decision rule had indicated.

The above are highlights of the results of the LC/NC impact evaluation which attempted to empirically diagnose as well as evaluate the effects of the program. Several implications emanate from this example of perspective on evaluation.

Implications

Evaluation research is not a discipline but rather a conglomeration of applied scientists attempting to provide empirical data to policy makers regarding the effects of a public program. This lack of organization results in a very complex and important area receiving little attention and almost no systematic feedback to the impacted disciplines.

In addition, for several reasons the researcher cannot go to the policy maker for guidance. The policy maker does not have adequate training to judge research methodology and policy makers are, at the very least, hesitantly supportive of evaluation. The implication of the two previous points places a great burden on the internal standards of the researcher.

A number of factors make the decisions in research methodology difficult ones. Several of these factors are attendant to the policy programs themselves. The timing of the planning for research is often short due to the fact that evaluation is generally one of the later program decisions. This often precludes taking premeasures and places a constraint on the time for planning of the research design. In addition, decisions for future policy often have to be made before the research is completed so the research often fails to play its proper role in program planning. Also, the nature of policy programs often preclude the development of a proper control condition or make difficult the

identification of properly matched groups. Finally, the orientation of evaluators is heavily toward survey-type designs and to single-method evaluation studies. Both of the latter are severely limiting decisions.

There are clear advantages to attempting causal designs in establishing program effectiveness, ruling out alternative hypotheses, and removing the need for post-hoc theorizing in explaining results. Of equal significance is the need for multiple-methodologies to fully profile program effects. The most common perspective to evaluation is to provide end-results data. However, this ignores the fact that programs are part of an on-going activity which needs to understand program effects as well as document them. This highlights the inherent deficiency of any single study to measure impacts as well as diagnose the results for understanding. But, this often means that the individual studies undertaken may, when taken alone, evidence deficiencies. These deficiencies, however, must be taken in light of the complete profile of the evaluation effort.

This paper has presented the methodological decisions of a recent evaluation effort for the LC/NC program. In this evaluation multiple-methods were used to both strengthen the conclusions of program effects through convergence of the results of the various studies, and to provide diagnosis so that future programs could benefit from the analysis of the LC/NC tactical decisions.

References

- Campbell, Donald T. and Fiske, Donald W. (1959), "Convergent and Discriminant Validation by the Multi-Trait Multi-Method Matrix," Psychological Bulletin, 56, p. 81-105.
- _____ and Stanley, Julian C. (1963), Experimental and Quasi-Experimental Designs for Research, Chicago, Rand McNalley.
- Freeman, Howard E., "The Present Status of Evaluation Research," Evaluation Studies Review Annual, p. 17-51.
- Heeler, Roger M. and Ray, Michael L. (1972), "Measure Validation in Marketing," Journal of Marketing Research, 9, p. 361-370.
- Howard, Niles and Antilla, Susan (1979), editorial, Duns Review, September, p. 49-57.
- Hutton, R. Bruce and McNeill, Dennis L. (1980), "An Empirical Evaluation of the Low Cost/No Cost Energy Conservation Program: A Draft," U.S. Department of Energy, Washington, D.C.
- Mazis, Michael B. and McNeill, Dennis L. (1978), "The Use of Marketing Research in FTC Decision-Making," in S. C. Jain (ed.) Research Frontiers in Marketing: Dialogues and Directions, American Marketing Association, Chicago.
- Phillips, Lynn W. and Calder, Bobby J. (1979), "Evaluating Consumer Protection Programs: Part I. Weak but Commonly Used Research Designs," Journal of Consumer Affairs, Winter, 13, p. 157-185.
- Scriven, M. (1967), "The Methodology of Evaluation," in R. E. Stake (ed.) AERA Monograph Series on Curriculum Evaluation, 1, p. 39-83.
- Shoemaker, Floyd R. (1979), "Diffusion and the Marketing of New Products," presentation handout, University of Denver, Denver, Colorado, February 15, 1979.
- Van Maanen, John (1979), "The Process of Program Evaluation," The Grantsmanship Center News, January/February, p. 29-74.

RISK-BENEFIT ANALYSIS AND THE DETERMINATION OF ACCEPTABLE RISK¹

Rachel Dardis, University of Maryland
Julie Stremel, Arthur Young and Company

Abstract

Consumer safety has been regarded as a major consumer issue for many years. In recent years, this has resulted in considerable debate concerning "how much risk" is acceptable. The objectives of this paper were to develop a methodology for assessing risk and to apply the methodology to a selected consumption activity. The results indicate the role of risk-benefit analysis in the identification of "high-risk" consumption activities. This type of screening is important in view of resource limitations which preclude the development of risk reduction strategies for all consumption activities.

Introduction

Consumer safety has been regarded as a major consumer issue for many years. While there is general agreement that consumer safety can and should be increased, there is less agreement concerning the level of safety that should be achieved. Proponents of zero risk argue that no risk is acceptable though the attainment of zero risk in a complex technological society such as the United States has been challenged with respect to both feasibility and desirability. The rejection of zero risk leads, in turn, to the question of "how much risk" is acceptable. Unfortunately, the frequently used argument that any risk is acceptable as long as the expected benefits are greater than the expected costs neglects the costs of developing risk reduction strategies for a wide spectrum of consumption activities. The costs of obtaining such information may be far greater than the potential benefits.

The objectives of this paper were to develop a methodology for assessing risk and to apply the methodology to a selected consumption activity in order to demonstrate the feasibility and utility of risk-benefit analysis. The results should provide policy makers with a method for assessing risk. Risk assessment may then be used a) to obtain information concerning the level of risk to which consumer are exposed and b) to rank activities with respect to risk.

Need for Risk Assessment

Deaths and injuries involving consumer products pose serious social and economic problems in the U. S. It is estimated that approximately 20 million Americans are injured each year in the home as a result of accidents involving consumer products (National Commission on Product Safety 1970, National Safety Council 1978). The annual cost to the nation has been estimated at more than \$5.5 million.

The Consumer Product Safety Act which was passed in October 1972 was designed to protect the consumer from unreasonable risk of injury. A Consumer Product Safety Commission (CPSC) was established with jurisdiction over more than 10,000 consumer products. The Commission was granted authority to "develop, promulgate and enforce safety standards for consumer products and to inform the public concerning significant hazards associated with consumer products".

The number of consumer products which are subject to regulatory activities presents problems as the Commission acknowledges.

"It is not difficult to know when the need exists to ban an obviously dangerous or hazardous product.

But it is difficult to determine which of more than 10,000 consumer products are of the greatest risk to consumers all over the country and which of these products are the likeliest candidates for safety standards without undue cost to the consumer." (U.S. Consumer Product Safety Commission 1973)

In view of the impossibility of dealing simultaneously with all consumer products, it is necessary to establish a basis for ranking products with respect to unreasonable risk. One method for ranking products was the hazard index developed by the Consumer Product Safety Commission in 1973 and revised in 1975. The index was based on frequency and severity of injury. Accident frequency was obtained from NEISS data while severity values were based on type of injury and necessity for hospitalization. Severity values in the 1975 and 1976 hazard indices are given in Table 1. (U.S. Consumer Product Safety Commission 1976).

TABLE 1

SEVERITY VALUES USED IN CPSC HAZARD INDEX

Severity Category	Severity Value
1 = Mild injuries/small areas, dermatitis and sprains	10
2 = Punctures - fractures	12
3 = Contusion - scalds	17
4 = Internal organ injury	31
5 = Concussions - cell and nerve damage	81
6 = Amputations - crushing and anoxia	340
7 = All hospitalized category sixes	2,516
8 = All deaths	2,516

Categories 1 to 5 were increased by one severity value if the injury resulted in hospitalization. Similar values were used in the 1973 hazard index with the exception of category 8 where a value of 34,721 was assigned in place of 2,516.

A Frequency Severity Index (FSI) was obtained by multiplying the frequency of injuries in each severity category by the corresponding severity values. The Age Adjusted Frequency Severity Index (AFSI) differs from the FSI in that injuries to children under 14 are first multiplied by 2.5 in order to provide special consideration for this age group. The AFSI is the hazard index employed by the CPSC. In issuing the hazard index, the Commission noted that "a product does not have to be high on this list to receive Commission action but those high on the index are more likely to receive early attention".

¹University of Maryland Agricultural Experiment Station
Article No. A-2828.

The hazard index was criticized with respect to the particular severity values since there was no indication that the values were related to the economic consequences of a particular type of injury. As a result, the hazard index was replaced in 1978 by numbers of injuries involving various products. While the use of a simple numerical index avoids the bias associated with an incorrect severity scale, it provides less information to the policy maker. The new numerical index also has many of the same limitations as the discontinued hazard index.

First, neither index considers the number of products in use nor the degree of product usage by the consumer. Thus, products with the same number and types of injuries could be viewed as equally hazardous though hours of exposure might differ to a considerable extent. Both Lowrance (1976) and Rowe (1977) have discussed the need for including hours of exposure in any risk assessment.

Second, consumption benefits are ignored. However, the need for including product benefits in any evaluation of risk has been pointed out by Starr (1968, 1972), Clark (1969), and Dickerson (1968). Starr used consumer expenditures for an activity as a measure of benefits and concluded that the public was willing to accept a relatively high level of "voluntary" risk. Clark examined the conditions under which a product might be considered reasonably safe. His hazard index included injury severity, number of products used per year, costs of avoidable injury, costs of safety improvements and importance of product use. The first four terms in the hazard index represented the ratio of the costs of injuries avoided by a safety improvement to the costs of making the safety improvement. The fifth term - importance of product use - reflected the degree of product substitution since "a hazard is more grimly accepted as "reasonable" if there is little alternative to the use". Dickerson pointed out that "the mere fact that a product is dangerous is not enough; it must be unreasonably so". He also noted that it is not possible to achieve complete safety and that the need for and utility of the product should be considered in product regulation.

It is important, therefore, to consider both the costs and benefits from a particular consumption activity in evaluating risk. Risk-benefit analysis is a technique which may be employed to assess the costs and benefits of a given activity, which involves risk. Risk assessment includes both the probabilities of various outcomes and the consequences of such outcomes expressed in dollar terms. Benefit assessments measures the benefits to the individual and society from the given activity.

Interest in risk-benefit methodology as a decision-tool for evaluating risks with respect to public health and safety was evidenced in two conferences which were held in 1971 and 1975. The first conference was organized by the Committee on Public Engineering Policy, National Academy of Engineering who noted that there "had been neither extensive public discussion or formal attention devoted to the subject or the emerging technique of benefit-risk analysis" (National Academy of Engineering 1972). Topics discussed at the conference included data requirements for decision-making, the process of risk-benefit analysis and the problems of implementing good analyses.

The second conference was part of a National Science Foundation funded study at UCLA entitled "A General Evaluation Approach to Risk-Benefit for Large Technological Systems and Its Application to Nuclear Power" (Okrent 1975). The major emphasis of the conference was risk assessment and two large technology areas were considered - nuclear reactors and the shipment and storage of liquified natural gas. Risk assessment utilized probabilities of various outcomes and the consequence of such outcomes to estimate

the expected loss from a particular activity. Consideration was also given to the determination of acceptable levels of risk. Besuner and Gibson focused on risk per hour of exposure and concluded that the risk associated with a product or activity was acceptable if it fell at the lower end of the scale. Wiggins noted, however, that several factors might be responsible for discrepancies in risk per hour of exposure including benefit differences which would warrant a higher risk for some activities. In the case of man-made hazard, he emphasized that benefits must be weighed against the risks. This was also the opinion of Oberbacher who concluded that it was important to assess both the risks and benefits from a given technology.

Risk assessment was also the focus of a recent symposium which was held at the National Meeting of the American Chemical Society (American Chemical Society 1977). The introductory paper by Lowrance included a discussion of acceptable risk and procedures for determining acceptable risk. Lowrance defined safety as a judgement of the acceptability of risk but noted that "it is not often clear who should decide the acceptability of what risks, for whom and on what terms and why". Lowrance also stressed the need for "adequate socio-political decision-making tools" commenting that while the concept of "safety" is poorly understood, the concepts of "efficiency, cost, benefit, and other decisional attributes may be even less well understood".

The advantages of risk-benefit analysis with respect to risk evaluation are as follows:

- a) It serves as a basis for comparing products since it includes both the costs and benefits of a consumption activity. In contrast, hazard analysis ignores the need for and the utility of the product to the consumer.
- b) Risks and benefits pertain to the same period of time so that it is not necessary to estimate hours of exposure. Benefit analysis is based on consumer expenditures per unit of time (hour, day, week, year). Risk analysis is also based on the same unit of time. Thus, the number of hours of exposure per unit of time is immaterial. While risks and benefits pertain to product use per unit of time is not necessary to estimate hours of exposure since it affects both risks and benefits equally. In contrast, hazard analysis is confined to risk so that hours of exposure are important. Lack of information concerning consumer product usage may limit the employment of hazard analysis in risk evaluation.

Risk Assessment Model

The following assumptions were used for the risk assessment model proposed in this paper.

1. Accident probabilities and economic consequences of accidents vary by accident severity, e.g., fatal injury, major injury, minor injury.
2. Each product in a consumer product category (e.g., bicycles, ovens) has an equal probability of being involved in an accident of a given severity and will incur the same accident costs. However, this probability may vary between product categories.
3. The value attached to a consumption activity is equal to its ownership costs.
4. Risks and benefits from ownership pertain to the time of consumption, i.e., only risks and benefits are incurred during product use.

Relative risk is then given by $w'x/v$ where x is a column vector of increases in various types of injuries due to the activity, w is column vector of the costs of injuries and v is the value attached to the activity by the consumer. The term $w'x$ may be considered the expected loss from a particular consumption activity. Ownership costs reflect the willingness to pay for a consumption activity and include both acquisition and maintenance costs. In contrast to ownership costs which are incurred by all consumers, accident costs are borne by only a small group of consumers. Expected loss is a measure of this uncertain prospect.

In the derivation of risk-benefit ratios, the term $w'x/v$ may be re-written as follows:

$$w'x/v = (P_f C_f + P_n C_n) / V \quad (1)$$

where P_f = increase in probability of fatal injury
 P_n = increase in probability of non-fatal injury
 C_f = Cost of fatal injury
 C_n = cost of non-fatal injury, and
 V = ownership cost

The above equation may be re-written as

$$\frac{(N_f/Q)(C_f) + (N_n/Q)(C_n)}{V} \quad (2)$$

$$\frac{N_f C_f + N_n C_n}{VQ} \quad (3)$$

where N_f = number of fatal injuries
 N_n = number of non-fatal injuries
 Q = number of products in use, and
the other terms are as defined earlier.

The numerator in equation (3) is the total costs of accidents involving consumer products while the denominator is the total ownership costs associated with such products.

Application of Model

The risk-benefit model was applied to seven categories of adult clothing. The seven items selected were based on their consideration as possible candidates for flammability standards. (Product Safety and Liability Reporter 1974, National Cotton Council of American 1974, Simon et al. 1975).

The seven categories examined were: a) women's robes and housecoats, b) women's nightgowns, c) women's pajamas, d) men's pajamas, e) women's dresses, f) men's shirts, and g) men's pants.

Ownership costs and accident costs were derived for purchases in 1976 and all cost estimates were expressed in constant 1976 dollars.

Accident Costs

The direct costs of accidents were the medical costs while the indirect costs were the earnings foregone due to accidents (Dardis 1978). The present value of foregone earn-

²Market prices are used to represent the value attached to the last unit purchased of a good or service. Such prices reflect "value in exchange" as opposed to "value in use". However, the two are inter-related since the consumer's willingness to pay is based on the utility from the product or the perceived value in use.

ings was used in the case of death. Two discount rates - 5% and 10% were used to obtain present value estimates.

This measure neglects the pain and suffering component associated with accidents. While some agencies such as DOT have used court awards for pain and suffering to reflect such costs, this measure has generally been regarded as unreliable due to the extreme variability of court awards and the fact that only a small proportion of accident cases are decided in court (Dardis 1978). The omission of pain and suffering costs need not invalidate the ranking of products with respect to acceptable risk if it is assumed that this cost component is proportional to the sum of direct and indirect costs. Under this assumption, all risk-benefit ratios will be increased by a constant factor leaving product rankings unchanged.

Ownership Costs

It was assumed that the various clothing categories were non-durable goods with an expected life of one year. Ownership costs were based on total consumer expenditures for the particular clothing category in view of lack of data concerning maintenance costs for clothing. However, maintenance costs for most clothing items are likely to be small in view of the importance of easy care clothing in 1976. In addition, specific clothing items comprise only a small portion of total clothing consumption so that their contribution to total maintenance costs is likely to be small.

Extension of product life does not have a significant impact on risk-benefit ratios. For example, if the expected life is assumed to equal two years, then a first approximation of the present value of costs from purchases in 1976 is given by

$$0.5C_{1976} + (0.5C_{1977}) / (1+i) \quad (4)$$

where C = costs of accidents due to the activity and
 i = discount rate.

If the number and severity of burn accidents is similar in 1976 and 1977, then the present value of costs from purchases in 1976 is given by

$$0.5C_{1976} + (0.5C_{1976}) / (1+i), \text{ or } 0.95C_{1976} \quad (5)$$

for a 10% discount rate. The assumption of an expected life of one year when the average product life is greater than one year will thus serve to underestimate accident costs. However, such underestimation is likely to be small in the case of non-durable goods.

Provision for Imports

Consumer expenditure data were based on U.S. value of shipments. However, accident costs pertained to total U. S. consumption which included imported garments. Accident costs pertaining to U. S. shipments were based on the relative importance of U. S. shipments. According to a recent study by the Council on Wage and Price Controls (1978) domestic production accounted for approximately 90% of domestic consumption in 1976. Thus, 90% of U. S. accident costs were attributed to U. S. production.

Results

Consumer expenditures on U. S. merchandise in 1976 were based on U. S. value of shipments and retail mark-ups for the different clothing categories (U.S. Department of Commerce 1977, National Cotton Council 1974). Expenditures ranged from \$713.46 million for men's pajamas to \$3,735.22 million for men's pants. Based on the assumption that garments have an expected life of one year and zero maintenance costs, total consumer expenditures in 1976 are equal to total ownership costs in 1976. Ownership costs are given in Table 2.

TABLE 2

RETAIL VALUE OF U. S. CLOTHING SHIPMENTS IN 1976^a
(\$ Million)

Category	Value of Shipments
Women's Robes and Housecoats	577.08
Women's Nightgowns	710.53
Women's Pajamas	209.29
Men's Pajamas	173.46
Women's Dresses	3,198.13
Men's Shirts	2,366.70
Men's Pants	3,735.22

^aWholesale value of shipments increased by retail markups (National Cotton Council 1974)

Accident Costs, 1976

The number of injuries for each category was based on a total burn injury estimate of 60,000 for all ages in 1976 (U. S. Department of Health, Education and Welfare 1976). Data from the National Center for Health Statistics was then applied to this estimate to obtain the number of adult burn injuries, ages 13 and above (U. S. Department of Health, Education and Welfare 1977). The number of adult clothing burn injuries was based on a low estimate from the New York State Department of Public Health (1977) and a high estimate from the National Burn Information Exchange (U.S. Department of Health, Education and Welfare 1971). The resulting low and high adult clothing burn injury estimates range from 5,213 to 17,798. It should be noted that the high estimate from the National Burn Information Exchange (NBIE) reflects the fact that most cases from this data source are characterized by severe burns and hence are more likely to involve clothing. The New York State study includes all types of burn injuries and probably presents a more realistic picture of clothing involvement.

Adult clothing death estimates were based on data from the National Center for Health Statistics (U. S. Department of Health, Education and Welfare 1977) and the National Fire Protection Association (1969). The number of deaths ranged from a low of 459 to a high of 801.

The distribution of injuries and deaths for each adult clothing category was based on data from flammable fabrics investigations conducted by the Consumer Product Safety Commission (1975). The results are given in Table 3. The number of injuries is considerably higher than the number of deaths in all instances. The men's shirt category has the highest number of burn injuries while men's pajamas have the lowest number of injuries. The women's robes and housecoat category has the highest number of deaths while women's pajamas have the lowest number of deaths.

TABLE 3

NUMBER OF CLOTHING BURN INJURIES AND DEATHS
ATTRIBUTED TO U. S. SHIPMENTS IN 1976^a

Category	Injuries		Deaths	
	Low	High	Low	High
Women's Robes and Housecoats	460	1570	92	159
Women's Nightgowns	346	1179	43	76
Women's Pajamas	258	882	16	28
Men's Pajamas	104	355	21	37
Women's Dresses	255	870	27	49
Men's Shirts	683	2332	79	138
Men's Pants	339	1158	35	60

^aBased on the assumption that domestic production equals 90% domestic consumption (Council on Wage and Price Stability 1978, p.58).

The direct costs of burn injuries and deaths were based on length of hospital stay, which varied according to burn severity. Length of hospital stay ranged from 32-124 days which included an allowance for readmission and retreatment while a weighted average of 17 days was used for deaths (Dardis 1978). Hospital costs for flammable fabrics victims were estimated at \$333.00 per day and served as a basis for estimating total medical costs. Data from Rice (1966) and the Social Security Administration (Gibson and Mueller 1979) indicated that hospital costs accounted for 59% of total medical expenses for injuries.

The indirect costs of injuries were derived from the probability of working or housekeeping, number of days lost from working or housekeeping, and mean earnings. In the case of death, the present value of future lifetime earnings was obtained using 5% and 10% discount rates. Total direct and indirect costs of burn injuries and deaths are given in Table 4 for a 10% discount rate. Injury costs based on low injury estimates range from \$4.35 million for men's pajamas to \$27.94 million for men's shirts. Injury costs for high injury estimates range from \$14.84 million to \$95.41 million for the same two clothing categories.

TABLE 4

COSTS OF CLOTHING BURN INJURIES AND DEATHS
ATTRIBUTED TO U.S. SHIPMENTS IN 1976
(\$ Million)

Category	Injuries		Deaths ^a	
	Low	High	Low	High
Women's Robes and Housecoats	19.66	67.14	2.04	3.47
Women's Nightgowns	13.98	47.74	1.36	2.45
Women's Pajamas	9.55	32.60	0.70	1.12
Men's Pajamas	4.35	14.84	0.75	1.40
Women's Dresses	10.93	37.33	0.61	1.26
Men's Shirts	27.94	95.41	4.02	7.03
Men's Pants	13.58	46.39	2.17	3.73

^a10% Discount rate.

The costs of burn deaths are considerably lower reflecting the small number of burn deaths relative to burn injuries. As a result, the use of a 5% discount rate in place of a 10% discount rate had little impact on the total costs of burn injuries and deaths.

Risk-Benefit Ratios: 1976

The ratio of total accident costs and total ownership costs is given in Table 5 for each of the seven categories. In the case of low cost estimates, the ratios range from 0.004 for men's shirts and pants to 0.049 for women's pajamas. This means that every dollar spent on men's shirts or men's pants has an expected loss of less than half a cent. In contrast, an expected loss of approximately 5 cents is associated with every dollar spent on women's pajamas. If one wishes to consider the expected loss as a type of risk tax, then the risk tax for sleepwear is the same order of magnitude as a sales tax while the risk tax for daywear is considerably lower. However, an essential difference between the risk tax and the sales tax should be emphasized. All consumers pay a sales tax while no consumer pays the specific risk tax. Some consumers (those involved in accidents) pay considerably more than the risk tax while other consumers pay nothing. For society as a whole, the risk tax is as meaningful as a sales tax since it indicates the total costs of accidents associated with a particular consumption activity.

TABLE 5

RISK-BENEFIT RATIOS FOR VARIOUS CLOTHING CATEGORIES

Category	Risk-Benefit Ratio	
	Low ^a	High ^a
Women's Robes and Housecoats	0.038	0.122
Women's Nightgowns	0.021	0.071
Women's Pajamas	0.049	0.161
Men's Pajamas	0.029	0.094
Women's Dresses	0.013	0.043
Men's Shirts	0.004	0.012
Men's Pants	0.004	0.013

^aPertains to low and high estimates of the number of burn injuries and deaths.

In the case of high cost estimates, the risk is higher as might be expected. The risk tax for sleepwear ranges from 7% to 16% while the risk tax for daywear ranges from 1% to 4%. However, risk taxes based on high cost estimates are probably overstated since they are based on data from the National Burn Information Exchange which are biased in favor of clothing burns.

It is also of interest to compare the data in Table 3 and 4. Absolute costs of accidents in the case of daywear are highest for men's shirts followed by men's pants and women's dresses. However, when risk-benefit ratios are examined, women's dresses emerge as the highest risk category followed by men's shirts and pants. Similar reversals occur with respect to sleepwear with women's pajamas replacing women's robes and housecoats as the highest risk category.

Discussion

The results of this study indicate that considerable variations exist with respect to risk benefit ratios for the various adult clothing categories. Risk-benefit ratios for sleepwear are far higher than risk-benefit ratios for daywear. If low cost estimates are accepted, intervention in the case of men's shirts or pants is unnecessary. In contrast, the data for women's sleepwear provide support for intervention. This intervention may be a flammability standard similar to that used for children's sleepwear or it may be an educational program. Both types of intervention could be undertaken by the Consumer Product Safety Commission. An alternative strategy might be the voluntary marketing of more flame retardant garments by the apparel industry. Such marketing would have to be accompanied by an educational program since consumers, in many instances, are unaware of the hazards from flammable fabrics or the need for protection.

Other results of the study are also of interest. First, the proposed method for assessing risk appears feasible. The assignment of economic consequences to various types of injuries permits comparison between products with more than one type of injury. Ownership cost data may be obtained from consumer expenditure data, product life estimates, and maintenance cost data.

Second, risk-benefit analysis provides insight concerning the level of risk to which consumers are exposed. The issue of prevailing risk is of particular importance since a society that is not risk-free must decide what level of risk requires intervention. If there is no basis for determining when intervention is necessary then risk reduction activities may be ineffective, i.e., we may concentrate on low risk areas and ignore high risk areas.

Finally, it should be noted that risk-benefit analysis is only part of the risk reduction process. This type of analysis serves to indicate prevailing levels of risk and those areas where intervention may be desirable. Once the necessity for intervention is perceived, the selection of an appropriate strategy requires additional analysis. The main advantage of risk-benefit analysis pertains to its identification of risk levels and, in turn, those areas where intervention is necessary. This type of screening is important in view of resource limitation which preclude the development of risk reduction strategies for all consumption activities. Thus, in the case of adult clothing, alternative strategies should first be investigated for women's sleepwear. This investigation would, in turn, indicate if intervention were justified and the optimal intervention strategy.

References

- American Chemical Society (1977), "Symposium on Risk Assessment and Hazard Control," National ACS Meeting, New Orleans, Louisiana.
- Clark, C. C., (1969), "When is a Product Reasonably Safe?" Appliance Engineer, 3, 17-19.
- "CPSC Working Draft of Standard for Flammability of Fabric for Specified Apparel Items," (1974), Product Safety and Liability Reporter.
- Council on Wage and Price Stability (1978), Textiles/Apparel, Washington, D. C.
- Dardis, R., et.al. (1978), Cost-Benefit Analysis of Consumer Product Safety Programs, NSF-RANN, Final Report.
- Dickerson, R. Reed (1968), Product Safety in Household Goods, Indianapolis: Bobbs-Merrill Company.
- Gibson, R. M. and M. S. Mueller (1979), "National Health Expenditures, Fiscal Year 1976," Social Security Bulletin, 49, 3-22.
- Lowrance, W. W. (1976), Of Acceptable Risk, California: W. Kaufman, Inc.
- National Academy of Engineering (1972), Perspectives on Benefit-Risk Decision-Making, Washington, D. C.
- National Cotton Council of American (1974), Possible Economic Impact of Flammability Standard on Selected Items of Apparel.
- National Safety Council (1978), Accident Facts.
- New York State Department of Public Health (1977), An Epidemiologic Study of Burn Injuries and Strategies for Prevention.
- National Fire Protection Association, Fire Record Department Study (1969), "The Single Fatality Fire," Fire Journal, 63, 34-35.
- Okrent, D. (1975), "Risk-Benefit Methodology and Application: Some Papers Presented at the Engineering Foundation Workshop," California.
- Rice, D. P. (1966), Estimating the Cost of Illness, U. S. Department of Health, Education and Welfare, Public Health Service, Health Economics Series, Number 6.
- Rowe, W. D. (1977), An Anatomy of Risk, New York: John Wiley and Sons.
- Simon, S. R., et.al (1975), Study of the Feasibility of Applying Flammability Standards to Adult Nightwear and Other Apparel, Research Report, Batelle Columbus Laboratories.

Starr, C. (1968), "Social Benefit versus Technological Risk", Science, 165, 1232-1238.

Starr, C. (1972), "Benefit-Cost Studies in Sociotechnical Systems," in Perspective on Benefit-Risk Decision-Making, National Academy of Engineering, Washington, D. C.

U. S. Consumer Product Safety Commission (1973), "Consumer Product Hazard Index".

U. S. Consumer Product Safety Commission (1976), "Consumer Product Hazard Index".

U. S. Consumer Product Safety Commission (1975), Flammable Fabrics, 7th Annual Report.

U. S. Department of Commerce, Bureau of Census (1977), Current Industrial Reports -- Apparel, Series MA-23A(76)1.

U. S. Department of Health, Education and Welfare (1976), Burn Care, A Report to the Senate.

U. S. Department of Health, Education and Welfare (1971), Flammable Fabrics, 3rd Annual Report.

U. S. Department of Health, Education and Welfare, Public Health Service, National Center for Health Statistics, U. S. Hospital Discharge Survey, unpublished burn injury data.

SCIENTIFIC PROGRESS AND RESEARCH TRADITIONS IN CONSUMER RESEARCH

Liisa Uusitalo, International Institute for Environment and Society, Science Center Berlin
Jyrki Uusitalo, University of Helsinki

Abstract

This paper takes up some programmatic points related to the question of scientific dynamics in the field of consumer research. The article first takes a few steps in the direction of general conceptual analysis concerning the notion and problematics of progress in the field of social science, and then proceeds to discuss specific themes of present dominant research traditions in consumer research. Finally, it is argued that scientific progress in this area presupposes the development of new theories by comparing and weighting the relative merits and demerits of various traditions. Progress is not attainable merely by applying traditional models of consumption to new areas of phenomena. Alternative traditions and concepts should be proliferated on the field; this demand is also supported by the emergence of new societal problems pertinent to the field.

Scientific Progress and the Study of Social Science

Presently, the analysis of cognitive growth and progress discernible in various fields of scientific enterprise has become one of the key themes in the general philosophy of science. The "blockbuster" opening up this discussion was, as is generally known, the study by the American historian of natural science, Thomas S. Kuhn (1970). His study deals with so-called scientific revolutions, that is to say, global changes in the "paradigms" of natural science. These are changes in the scientific community's conceptual and methodological commitments that underlie scientific practice and that concern the subject matter and the methodology.

Kuhn's work itself brought about (or triggered off) some sort of "revolution" in the arena of science studies. Until then, in the wake of logical empiricism, the phenomenon of scientific progress itself had by no means been denied. However, it had been taken to consist mainly in a stepwise cumulative factgathering, and in the extension of previously well-confirmed theories, to cover new empirically observable phenomena. (For a good summary of this standard view, see e.g. Nagel 1961). Kuhn's dynamized vision of science also drew attention to the significance of the scientific community by stressing that the paradigm facilitates the professional communication and judgments. Kuhn's idea was also supported by a newly arisen interest in Kantian philosophies of science that stressed the dependence of scientific observation on some conceptual structure and emphasized the "theory-ladenness" of empirical facts and findings. (Such a philosophy is advocated by e.g. Popper 1965, ch. 3.)

All these arguments called the standard view forcefully into question. The new approaches also differed from one another in some significant points. For instance, Kuhn was stressing the significance of cumulative "normal-scientific" stages of work between the conceptual revolutions, whereas Popper and his disciples laid emphasis on a continuous active "proliferation" of several potentially falsifiable theories and the scrutinization of them against one another. (See Lakatos and Musgrave (eds.) 1970.) At any rate, nowadays at least, in the domain of natural sciences, a generally accepted view is that the philosophic work on foundational questions takes place in "dynamized" frameworks of some kind or another.

(For recent documentations of discussion, see e.g. Suppe 1978 and Niiniluoto & Tuomela (eds.) 1979).

One of these foundational issues is the problem of scientific progress. The new dynamic views of scientific activity have effectively helped scientists realize that progress in science is not necessarily guaranteed just by sticking to theories or theoretical structures that in the past have yielded the greatest number of successful predictions of phenomena. The new contributions in the philosophy of science have rather drawn general attention to the aspect of doing science which one might call, following Wartofsky (1976), "rational hopefulness" - that is, one's being able to choose on good grounds between rival research traditions and picking out those ones that most effectively helped one to tackle with the most important problems which the domain is facing. This precisely has been the background motive for some recent dynamic models of scientific activity that sharply dissociate themselves from the time-old empiricist tradition (Lakatos 1970, Laudan 1977).

Since the beginning of the above-described wave of discussion, attempts have also been made to thematize the field of social or "actional" sciences in a corresponding way. (For the case of sociology, see e.g. Friedrichs 1970 and Gouldner 1971; for economics, see Latsis 1976; for psychology, see Herrman 1976.) However, it seems that the overt plausibility of most of these studies (but not all of them) has often regrettably turned out to be illusory. Instead of first dealing with the important issue about the possible differences and contrasts between natural and social sciences, some authors have rather tended to pick out a priori some dynamic model of science and impose it on the discipline under study. Consequently, some of these studies have been inadequately treating the crucial question - that is, what kind of scientific knowledge is involved when the domain of human action is being studied.

For instance, Kuhn's dynamic model of science has frequently been used to describe paradigm change in social sciences mainly in terms of external factors (e.g., societal goal-settings and interests) affecting it. It is interesting to note, however, that Kuhn himself obviously did not intend his model of science to be understood in this sense only. Rather, he wanted to draw attention to the fact that science essentially is an activity of solving specific kinds of cognitive problems and always take place in a more or less well-established context of beliefs and assumptions about the disciplinary subject matter shared by the community of scholars in the field.

However, the notion of problem solving as a criterion of paradigm choice should not be understood in a straightforwardly "sociopragmatic" sense. For example, in the case of social science, it does not suggest the simple idea that science has just to solve externally determined "problems", but rather that cognitive activities in science also contribute to the structuring of the problems themselves.

In Kuhn's opinion, as is well known, the periods of stable "normal science" activity consist of attempts at making freshly observed phenomena intelligible in terms of the generally adopted set of beliefs about the subject matter of inquiry. Now, this idea of Kuhn's is obviously not directly applicable in constructing a "dynamized" model of

social sciences. In these sciences, we are not dealing with a stable domain or reality which determines the research in the same way as "nature" does for the natural sciences; the social reality is always essentially constituted by human actions as well. (On this criterion demarcating between natural and social science, see Krüger 1979.) And, this is why it is hardly conceivable that one single "normal scientific" tradition could be reigning in any particular social-scientific discipline.

Social sciences would also seem to dispose over normative criteria of theory choice that are themselves continuously changing. The overall idea here, then, is that in social science, especially when a newly emerging and largely interdisciplinary field like consumer research is at stake, the choice of the most promising - i.e. progressive - research traditions is not a matter of arbitrary decision but must itself be grounded on rational appraisal concerning the long-term problem-solving efficacy of various theoretical alternatives as well as their capability of initiating inquiry in new fields.

Research Traditions in Consumer Research

Consumer research, as a newly emerging discipline, consciously thematizes itself in terms of a specific, even restricted sub-area of social behavior - that is, consumer behavior. Here, we meet a rather clear-cut case where a previously unknown area or subject matter has been repeatedly shaped up more or less by analogy with the practices of some already well-established social science, be it economics, psychology or sociology. For the theory formation and the choice of consumer theories taking place within any of the thus constituted traditions, different and often extensive lists of criteria have been suggested. (see e.g., Zaltman et. al. 1973, Bagozzi 1976a). An empiricist philosophy of science has often been dominating these texts. Also, some ideas from more recent, so-called realist philosophy of scientific explanation have been utilized in some cases, but the perspective has been mainly on a rather general programmatic level (e.g. Bagozzi 1976b).

When it comes to the consumer theory appraisal internal to some tradition, the term 'theory' in this context, is taken to mean various systems of propositions or models that explain certain given phenomena, i.e., a given subject matter. (For instance, Zaltman et. al. 1973 compare the consumer models presented by Howard & Sheth, Nicosia, and Engel & Kollat & Blackwell, all of these describing the decision processes of individual consumer choices.) In all, scientific criteria have been used predominantly in evaluating theories within one consumer research tradition, rather than in evaluating different research traditions or paradigmatic belief systems.

Actually, one might venture to suggest that in this discussion, the famous Kuhnian view of one dominating "paradigm" at a time has been taken for granted, even though rather implicitly, and consequently, consumer research has been treated as being constituted by a single paradigm. However, the Kuhnian idea of the sequential development of a given discipline between the stages of stabilized "normal science" has also been explicitly disputed even by authors active in this area, maintaining that consumer research as a field, in fact, consists of several simultaneous and competing "paradigms" or research traditions. (For example, Bagozzi 1976b presents this view by analyzing competing paradigms in "macro marketing". All macro paradigms, in turn, seem to compete with the prevailing micro approaches which unfortunately have not been discussed in detail.)

In fact, the view of several competing traditions in explaining different kinds of consumption phenomena appears

also in the critique of psychological reductionism in present-day consumer research as well as in several calls for a different research approach that would allow an analysis of social phenomena which the psychologically oriented models never do bring into focus (e.g. Glock & Nicosia 1964, Levy & Zaltman 1974, Zaltman & Sternthal 1975, Angelmar & Pinson 1975, Arndt 1976, Mayer & Nicosia 1977, Bagozzi 1978, Jacoby 1978).

In a cross-scientific discipline like consumer research, it is, in fact, obvious that more than one research tradition will be present simultaneously rather than sequentially. As the situation now is, it seems, however, that these research traditions develop rather isolated from one another, and there seem to be very few integrating links between them. Therefore, it would seem to be more motivated to talk about a multi-disciplinary rather than an interdisciplinary discipline.

While some of these research traditions have been and still are very dominating, some others, in their turn, have for some reason or another been receiving less emphasis than they would obviously earn if certain criteria would be adopted which evaluate scientific progress based either on the potential problem-solving efficacy or on the capability of initiating theories for new fields of scientific inquiry.

When one takes a look back in time at the proceedings of all the ACR conferences or a look at some of the principal scientific journals where consumer research and its findings have been documented over the last ten years or so, one easily gets an impression of the overwhelmingly strong position of psychological research tradition in consumer research. In addition, most of the social-psychologically oriented research, say, research on social group influence or personal influence and its impact on consumption, is classifiable under this heading.

The subject matter of this tradition is constituted by individual product choices or decision processes leading to a choice. Among choices, the brand choice has been given an astonishingly strong emphasis in empirical research. The rapid development of the psychologically oriented research tradition - in a close connection to marketing research - is associated with the expansion of consumption in the 1960's and 1970's and the corresponding needs of competitive marketing. The formal development in cognitive psychology, from which the most significant models were taken, and the development of experimental research methods and multivariate analysis techniques, of course, have also contributed to the cumulative efforts and relatively high consensus on the constitution of the subject matter in this research tradition.

The approach being basically more or less predictively - and often even manipulatively - oriented in its workings, the realism of its background assumptions about the nature of human behavior has not been much of a problem to the advocates of this school.

The stronger a tradition becomes institutionally, the more difficult it often seems to be to bring up alternative research traditions that do not rely on the methodological individualism or readily received quasi-natural-scientific or positivistic view of science, cherished by the psychological paradigm and its underlying concepts of human action.

Actually, here, even the basic scientific terminology has been "psychologically biased". For example, if one presents the idea that consumption should be studied more than it has been hitherto in association with social factors or in itself as a social phenomenon, one easily meets the objection that such things already have been taken into account. Now, social, cultural and institutional factors, are, however, in these psychological or "social influence"

models frequently reduced to an individual level and often presented in a more or less ad hoc fashion as so-called situational micro-level factors. Social factors associated with any macro-level social theory do not appear, with the possible exception of some eclectic studies concerning the significance of certain separate variables (like social class).

In fact, it is somewhat ironic that the defenders of present-day dominant psychological or social-psychological approaches in consumer research, who in their work implicitly accept this as the only legitimate research, do not seem to take into account that this tradition was originally started as (and still is) a competing tradition to the "pure" neo-classical economic consumption theory which was then considered to make too strong assumptions concerning consumer rationality and to lay too little emphasis on the changes in consumer preferences. The idea of the formation and change of preferences - however, mainly applied to brand preferences only - became, for pragmatic reasons, one of the main subjects of interest in the marketing-oriented consumer research during the 1970's.

There seems to be only a few connections between the psychological "behavioral" consumer research and the neo-classical economic consumption theory, which both have developed in separate scientific surroundings - the former among marketing-oriented researchers and the latter among economists. However, some models within the two traditions seem to lend support to one another very clearly. For instance, Lancaster's model of the maximization of utility received from different product characteristics is in accordance with the cognitive attitude or brand preference models. Basically, all "psychological" brand preference models are based on the same neo-classical idea of individual utility maximization as is the economic theory of consumption functions. (Therefore, there must also be more or less institutional, external reasons which could explain the divergence between economists on one hand and behavioral consumer researchers on the other.)

A Challenge for Alternative Research Traditions

A positivistically modelled view of science seems to be common to both the neo-classical economic theory of consumption and the psychologically oriented research tradition. For example, the use of empirical findings and empirical research dealing with tendencies of human behavior on a societal scale as an aid to developing concepts or testing their construct validity is strikingly unusual in these approaches. And, empirical research is more often being considered just a case of hypothesis testing even in situations where there do not yet exist any theoretical concepts or propositions grounding the formation of hypotheses.

Now, a first step towards evaluating the merits and demerits of these discussed traditions is to take their ideas about the subject matter of consumer research under a critical scrutiny. The domain of economic activity seems to have been in fact treated in a quasi-naturalistic way here, and by means of a methodology modelled on natural science.

The quasi-naturalistic character of economic and behavioral models can be seen to consist primarily of two inter-related tendencies: (1) First, the tendency to stress that the basic task of economic science is to work out predictions of future observable phenomena and not bother about the possibly lacking realism of theoretical assumptions, i.e. the tendency to adopt an instrumentalistic philosophy of science. (2) Second, the tendency to make particular ontological and metaphysical assumptions, about the behavior of economic agents. In these

assumptions, the laws of rational economic activity are modelled on the idea of laws of nature in physical science. (For a critique of related ideas about the "plastic man", see Hollis 1977). Naturally, one here also should point out that the particular view of natural science adopted by the "quasi-naturalists" in social science is a through-and-through positivistic one, and that there are philosophies of natural science that sharply dissociate themselves from the above-mentioned basic tendencies (see e.g. Harré 1970).

Moreover, the dominance of individualistically and psychologically oriented models have in fact excluded many questions from the problem-solving field of the established traditions. Such questions obviously deserve consideration in light of recent work in social science generally - i.e., research in the general field of social studies that has contributed to a deeper understanding of societal processes which obviously have an influence on consumption activities (for example, studies on structural change in society, the emergence of so-called post-industrial life patterns).

Recently, in this connection, the qualitative aspects and consequences of consumption have emerged as a new subject matter about which we still possess relatively little scientifically grounded information. An analysis working to explain these features of consumption activity must include variables that deal with the institutional conditions and surroundings of consumption. Without resorting to such variables, for instance, a historical analysis of consumption does not seem to be possible; and moreover, these variables are apt to prove fruitful for an investigation of differences in existing consumption and life-style patterns.

Naturally, one must also add here that a mere aggregate-level analysis (in the spirit of the traditional pure economic approach) is not sufficient in analyzing the qualitative features and social consequences of consumption in modern societies; the study of consumption features typical of the different segments of society and its population is also necessary.

All of this would seem to signal the emergence of a fresh research tradition in consumer research, the historical-institutional tradition, which lays emphasis on the interrelations and interdependencies between consumption, economic structures of society, and the prevailing value and belief systems of human agents and groups. This approach is by no means totally new, as there have been several schools of thought since the last century that emphasize these interdependencies. (Most well known are perhaps the historic school in Germany, the Marxian school and the North American institutionalists, e.g. Veblen 1899. See also Seckler 1975.) Some other efforts to integrate the views of the economic and social sciences are exemplified in the works of Pareto 1935, Weber 1947, and Parsons & Smelser 1956. Thus far, however, these approaches have seldom inspired consumer research.

A cognitive choice in favor of any newly "budding" tradition, in contemporary circumstances, obviously cannot be grounded on only the quantitative appraisal of the number of significant empirical problems this tradition can solve or on phenomena it is able to explain by generating appropriate theories in the domain. The implementing of scientific progress in the presently discussed field involves, rather, the interesting task of finding a way of weighting the significance of empirical questions posed or, perhaps, even successfully answered by this approach, against the significance of problems posed or solved by the other traditions. Laudan (1977, p. 32) has put this general point in the following way: "If a philosophy of science, or a model of scientific progress, is going to be satisfactory, it must provide some guidelines not only for counting, but also of weighting, scientific problems on a scale of relative importance and cruciality."

One can easily see that the problem types of different traditions are not by far similar, which of course may make comparisons more difficult. However, even then it may be possible to measure and weight the efficiency of different traditions in the field by looking at the number of problems they solve and also the number and significance of problems they leave anomalous, that is, to be solved by some competing tradition. (On the role of problem-weighting in general, see Laudan 1977, 31-40.)

Ultimately, the weighting obviously also involves the comparison of the merits and demerits of the quasi-naturalizing and largely empiricist traditions on the one hand, and an interdisciplinary tradition enabling one to formulate several sorts of research tasks dealing with the social phenomenon of consumption, on the other hand. And, in dealing with a discipline that concerns itself with aspects of the human domain - i.e., consumption activities and their societal context - one might do well in opting for traditions that are most likely to do justice to the peculiar subject matter of social reality. The significant aspects of this reality can hardly be captured in terms of the utility-maximizing assumptions of established theories of economic rationality, even though the work done by these research traditions, when measured purely in terms of the quantitative magnitude of problems solved by them, may be impressive. After all, these research traditions are likely to leave many empirical problems about consumer behavior in its societal context unanswered.

The Challenge of Solving New Social Problems

The above presented criticism concerning the two most dominating theoretical research traditions in consumer research, economic consumption models and psychologically oriented consumer behavior models, is not to deny their obvious advantages in solving a number of research problems in the future as well. But, one cannot avoid seeing that they - bound to their specific type of models and methods - have also very much influenced the idea of what kind of problems have been considered to be "important" and "worthwhile" as a target of research in consumer science. Moreover, in both established traditions, a danger exists of treating existing theories as capable of solving all new problems too, and of explaining all new empirical social phenomena as well. That would mean sticking to the old, non-dynamic view of progress in science.

As an example of new social phenomena which call for new theory building, we can mention consumption-related environmental problems. It is true, for example, that economic demand functions or consumer attitude models can be applied in predicting consumers' reactions to certain parameter variables in affecting the environmentally relevant choices. The explaining and understanding of the causes of ecological damages, however, is possible neither in the pure framework of economic theory of consumption functions nor in any psychological information processing or group influence models. Neither do some eclectic middle range theories, borrowed from social science, prove very satisfactory as a basis of explanation.

What in fact are needed in this case are theories that lay emphasis on historic-institutional, cultural and economic development of societies or theories that describe the intermitting mechanism by which this development is reflected in micro-level consumer behavior. Of course, new concepts and theories do not emerge out of nothing. Therefore, they essentially have to integrate existing knowledge from different research traditions.

This integration work, at the same time, should be based on the motive of working out a methodology and philosophy of consumer research that dissociates itself from the research practice modelled on a purely positivistically

understood idea of natural science. If this point is not stressed clearly enough, the misunderstanding would easily emerge that the call for theoretical innovation in consumer research actually were a call for increasing what is already a high level of eclecticism in consumer behavior studies.

What is really desirable and necessary, in our view, is the breakthrough of a new metascientific way of thinking and the subsequent proliferation of new theoretical approaches and research methods that are at present not generally accepted. Often, new theories must also start from the basic level of theory building, that is, from the level of heuristic theorizing and the development of theoretical frameworks that subsequently can be expected to give rise to fruitful hypotheses.

References

- Angelmar, R. and Pinson, C. (1975), "The Meaning of Marketing," Philosophy of Science, Vol. 42, 2, 208-214.
- Arndt, J. (1976), "Reflections on Research in Consumer Behavior," B. Anderson (ed.), Advances in Consumer Research, Vol. III, Cincinnati: Association for Consumer Research, 213-221.
- Bagozzi, R.P. (1976a), "Science, Politics and the Social Construction of Marketing," K.L. Bernhard (ed.), Marketing: 1776-1976 and Beyond, Chicago: American Marketing Association, 586-592.
- _____. (1976b), "Macro-marketing - Distributive Processes from a Societal Perspective," C.C. Slater (ed.), Proceedings of the Macro-Marketing Seminar, University of Colorado, August 15-18, 1976.
- _____. (1978), "Marketing as Exchange: A Theory of Transactions in the Marketplace," American Behavioral Scientist, Vol. 21, 4, 535-556.
- Friedrichs, R.W. (1970), A Sociology of Sociology, New York: The Free Press.
- Glock, C.Y. and Nicosia, F.M. (1964), "Uses of Sociology in Studying 'Consumption' Behavior," Journal of Marketing, Vol. 28 (July), 51-54.
- Gouldner, Alvin (1971), The Coming Crisis of Western Sociology, London: Heinemann.
- Harre', Rom (1970), The Principles of Scientific Thinking, London: Routledge and Kegan Paul.
- Herrmann, Theo (1976), Die Psychologie und ihre Forschungsprogramme. Göttingen: Hogrefe.
- Hollis, Martin (1977), Models of Man, Cambridge: Cambridge University Press.
- Jacoby, J. (1978), "Consumer Research: How Valid and Useful are All Our Consumer Behavior Research Findings?" Journal of Marketing, Vol. 42, 2, 87-96.
- Krüger, Lorenz (1979), "Does a Science Need Knowledge of Its History," I. Niiniluoto & R. Tuomela (eds.), The Logic and Epistemology of Scientific Change, 51-61. (see its ref.)
- Kuhn, Thomas S. (1970), The Structure of Scientific Revolutions, 2nd enlarged edition, Chicago: University of Chicago Press.
- Lakatos, Imre (1970), "Falsification and the Methodology of Scientific Research Programmes," I. Lakatos and A. Musgrave (eds.), Criticism and the Growth of Knowledge, (see its ref.)

Lakatos, Imre and Musgrave, Alan (eds.) (1970), Criticism and the Growth of Knowledge, Cambridge: Cambridge University Press.

Latsis, Spiro (ed.) (1976), Method and Appraisal in Economics, Cambridge: Cambridge University Press.

Laudan, Larry (1977), Progress and Its Problems: Towards a Theory of Scientific Growth, London and Healey: Routledge and Kegan Paul.

Levy, S.J. and Zaltman, G. (1974), Marketing, Society and Conflict, Englewood Cliffs, N.J.: Prentice-Hall.

Mayer, R.N. and Nicosia, F.M. (1977), "The Sociology of Consumption: Searching for Useful Perspectives on Society," European Research, 5 (July), 143-152.

Nagel, Ernest (1961), The Structure of Science, London: Routledge and Kegan Paul.

Niiniluoto, Ilkka and Tuomela, Raimo (eds.) (1979), The Logic and Epistemology of Scientific Change. Acta Philosophica Fennica, Vol. XXX (1978), Amsterdam: North Holland Publishing Company.

Pareto, V. (1935), The Mind and Society, London: Harcourt, Brace and Company.

Parsons, T. and Smelser, N.J. (1966 (1956)), Economy and Society: A Study in the Integration of Economic and Social Theory, London: Routledge and Kegan Paul.

Popper, Karl (1965), Conjectures and Refutations, 2nd ed., New York: Basic Books.

Seckler, D. (1975), Thorstein Veblen and the Institutionalists, London: The Macmillan Company.

Suppe, Frederick (1974), "The Search for Philosophic Understanding of Scientific Theories," Frederick Suppe (ed), The Structure of Scientific Theories, Urbana-Chicago-London: The University of Illinois Press, 6-232.

Wartofsky, Marx W. (1976), "The Relation Between Philosophy of Science and History of Science," R.S. Cohen, P. Feyerabend and M.W. Wartofsky (eds.), Essays in Memory of Imre Lakatos: Boston Studies in the Philosophy of Science, Vol. XXXIX (1976), Dordrecht: D. Reidel Publishing Company, 717-737.

Weber, M. (1947), The Theory of Social and Economic Organization (Wirtschaft und Gesellschaft), New York: The Free Press.

Veblen, T. (1965 (1899)), The Theory of the Leisure Class, London: The Macmillan Company.

Zaltman, G., Pinson, C.R.A. and Angelmar, R. (1973), Metatheory and Consumer Research, New York: Holt, Rinehart and Winston.

_____ and Sternthal, B. (1975), Broadening the Concept of Consumer Behavior, Atlanta: Association for Consumer Research.

TOWARD A REPLICATION TRADITION IN CONSUMER BEHAVIOR:
CROSS-CULTURAL REPLICATION OF BENNETT AND MANDELL'S
STUDY OF THE LEARNING-INFORMATION SEEKING HYPOTHESIS

Johan Arndt, Norwegian School of Economics and Business Administration
Kjell Grønhaug, Norwegian School of Economics and Business Administration
Richard E. Homans, University of Missouri, St. Louis
R. Neil Maddox, University of Missouri, St. Louis
Frederick E. May, University of Missouri, St. Louis

Abstract

Two replications of Bennett and Mandell's classic study of the effects of prior product/brand experience on the information-seeking behavior of new car buyers are reported. Bennett and Mandell found significant support for the notions of classical learning theory but only non-significant evidence for a hypothesis drawn from the Howard-Sheth theory. Here none of the hypotheses were supported in the first replication conducted in the Mid-West. However, the results of the second replication in Norway conformed with the findings of Bennett and Mandell.

Introduction

The purpose of the study reported here is to conduct a cross-cultural replication of the often-quoted study of Bennett and Mandell (1969) on the relationship between learning and information-seeking.

As pointed out some years ago by Kollat and his colleagues (1970, p. 329), progress in consumer behavior research is hindered by the lack of a strong replication tradition. Many propositions have shaky empirical foundations as they are based on findings in a single study conducted at a special time, on a special location, in a special research setting, and by a special researcher. Nevertheless, despite their humble origin, such single-shot findings often take the form of "generalized knowledge" as they are quoted and requoted. Reservations as to the generality of the results tend to be leveled out.

Objectives of Replication

In general, a purpose of replication is to isolate universals from phenomena which are merely idiosyncratic to particular populations or methods of measurement.

More specifically, the role of replication may be discussed in terms of Campbell's distinction between internal and external validity of empirical findings (Campbell, 1957). Internal validity relates to the impact of the experimental manipulation on the set of independent variables. Hence, the validity may be threatened by lack of controls for extraneous variables producing spurious effects. Replication studies concerned with internal validity would emphasize rigorous design, particularly better controls and allowance for other explanatory variables.

The replication study to be reported here, however, focuses essentially on the external validity of the findings in the original investigation. Therefore the critical issue is the generalizability of the findings. It is true from the philosophy of Hume that induction or generalization can never really be justified logically. However, it is possible to approach the problem pragmatically by examining the sensitivity of the findings originally uncovered to variations in population, physical setting, and different operational definitions of

the variables.

All replication studies involve a dilemma. To simplify, the situation is as follows: If the replicator is too "conservative" and the design of the new study is too similar to the original one, the findings risk being only trivial reconfirmations of what is already known. Such research is unlikely to be interesting to the scientific community. If on the other hand the replicator yields to the temptation of extending the study too far, deviating too much from the original design, such "radicalism" may produce non-comparable data. In this case, a middle road was chosen.

Choice of Study for Replication

At least three important criteria should be used in the choice of study to be replicated. First, it is imperative that the study has addressed a substantial and central theory area within the field of interest. Second, there should be some minimum of consensus among researchers in the area that the particular study is a significant contribution. These two substantiality criteria are needed as safeguards to avoid the risk of having the journals flooded by trivial, quick-and-easy replications. A third (and practical) criterion is that the methodology should be reported in sufficient detail to allow the desired comparability of design.

The study chosen for replication was Bennett and Mandell's (1969) investigation of the effects of learning or prior product/brand experience on the information-seeking behavior of new car buyers.

Learning theories have lodged themselves firmly in marketing and consumer behavior as explanations of brand choice and information-seeking and handling (Alderson, 1952; Kotler, 1965; Ray, 1973). The models employed have ranged in complexity from Watson's (1913) simple Stimulus-Response formulation to the elaborate Howard-Sheth model (Howard & Sheth, 1969). In addition to being more complex, the Howard-Sheth model differs from the main body of learning theory in a major way which is of importance here.

Howard and Sheth argue that all experience with a class of objects, whether positively or negatively reinforced, are instructive. The consumer's information requirements and the intensity of his search efforts decline as experience in buying and using the generic product accumulates.

In classical learning theory, learning is portrayed as a function of the number of reinforced trials or the number of successive reinforced trials with a specific stimulus object (brand) (Hall, 1966, Hilgard and Bower, 1966).

The Bennett and Mandell (1969) study was chosen as it appears to be a classic in its own light. It was one of the

first major studies addressing the Howard-Sheth theory. As such it is routinely quoted in standard textbooks such as Hansen (1972, p. 333), Howard (1977, p. 141), Engel, Blackwell, and Kollat (1978, p. 239), and Berkman and Gilson (1978, p. 409-10) and comprehensive reviews of consumer information-seeking behavior, such as Newman (1977). A third point was that the methodology seemed to be basically sound.

Fourth, the reporting of the method was detailed enough for the purposes of the present study.

Bennett and Mandell's Study

In their study Bennett and Mandell formulated the following hypotheses:

- H₁: The buying experience itself is instructive, whether or not the choice is positively reinforced. As this experience increases, evidenced by the total purchases in the individual's history, the amount of effort expended on information search will decrease.
- H₂: As the total number of reinforced purchases of a brand increases, the amount of information-seeking before purchase of that brand will decrease.
- H₃: As the number of sequential reinforced purchases of a brand increases, the amount of information-seeking before purchase of that brand will decrease (Bennett and Mandell, 1969, p. 431).

The first reflects the Howard-Sheth model. The other two were drawn from the nations of classical learning theory.

Bennett and Mandell tested the hypotheses with self reports of information-seeking collected from 146 recent new car buyers chosen from new car registrations in the Spring of 1967 in Harrisburg, Pennsylvania. The respondents were requested to reconstruct their entire car ownership history, were given ample time, and were also encouraged to examine any legal or financial records that might assist their recall. The independent variable, "number of trials", was defined in three ways. The first (T₁), which was used to examine H₁, was simply the total number of cars the respondent had owned. For T₂, the total number of purchases of the brand currently owned were counted. T₃ was the length of the current sequence of purchases of the brand currently owned.

To illustrate, consider the history COOCCOCC where C is the brand of automobile currently owned and O is any other brand. The values for this individual would be T₁ = 7, T₂ = 4, and T₃ = 2.

The second set of critical data was related to the extensiveness of information-seeking. The measurement instrument was a structured questionnaire specifying ten classes of information sources, see Table 1. Responses were weighted on a judgmental basis (by a panel of experts) according to a scheme believed to reflect the effort involved in consulting various information sources (Table 1). The first hypothesis was not supported ($\chi^2 = 4.46$, $df=6$, $p<.62$).

Strong support was found for H₂ ($\chi^2 = 17.80$, $df=6$, $p<.007$) and H₃ ($\chi^2 = 22.12$, $df=6$, $p<.001$), the relationships being monotonic and in the predicted direction.

Replication

As previously mentioned, an objective of this replication was to test the robustness of the original findings to variations in the operational definitions and to changes in physical setting or culture.

The first replication study which was also conducted in the U.S. (St. Louis) introduced some variations in field design.

To test the spatial validity of the results, analysis was also performed on a second set of data collected in Norway. As far as we can judge, the design of the Norwegian study coincided with the Bennett and Mandell study in most respects, the only main difference being the one of location. It should be noted that at that time, Norway was lagging somewhat in national income per capita as compared with the U.S. Because of high fiscal taxes making cars very expensive, the incidence of car ownership was smaller, while the length of ownership for each car acquired was much longer. Hence Norwegian consumers were expected to have experienced less learning as car buyers than American counterparts.

St. Louis Study

The first replication study which used data collected from a cluster probability sample of 130 St. Louis households during the Spring of 1966, introduced some variations in the measurements made.

Bennett and Mandell employed a structured questionnaire, while the data in the St. Louis study were derived from responses to a series of open-ended questions. Bennett and Mandell sampled current new car registrations. In the St. Louis replication randomly selected households were asked about their most recent new car purchase.

Responses indicating that the buyer had consulted any of the information sources in Table 1 were recorded. Two different summated measures of information seeking were developed for each subject. In the first, the effort weights reported by Bennett and Mandell were applied. In the second score, no weights were employed; all sources were given an equal value. Bennett and Mandell recorded the total amount of information-seeking for all cars considered combined. The first replication, however, only obtained measures of information sources used for the car bought.

The measures of prior product/brand experience were identified to those employed by Bennett and Mandell.

Bergen Study

The source of the data in the second replication was personal interviews conducted among a sample of 96 recent new car buyers, selected by random sampling among all new car registrations in a given time period in the late Fall of 1969 in Bergen (the second largest city in Norway).

Like in the original study a structured questionnaire was used, and the buyers were asked about their information-seeking for all the brands considered. Also other parts of the procedure paralleled the Bennett and Mandell investigation. The only main difference was that auto shows (see Table 1) was not included as a possible source of information (owing to the absence of this source in Bergen at that time).

Therefore, while the three data sets are contemporaneous, there were certain differences in procedure. These differences are summarized in Table 2.

Table 1

Sources of Information and Importance Weights

Information Source	Weight ^a
<u>Consumer Reports</u>	18
Dealer visit	18
Expert opinion ^b	12
Friends' opinion	10
Reading brochures ^c	10
Discussion with spouse	9
Auto show	7
Advertisements	7
News articles	6
Discussion with children	3
Total	100

Source: Bennett and Mandell (1969).

^aWeights indicate the relative amount of effort required to use the particular source.

^bExpert could be mechanic or "purchase pal" with special knowledge of automobiles, either of a technical or economic nature.

^cIndicates the actual reading of brochure, usually picked up at "Dealer visit".

Table 2

Summary of Main Differences in Design in the Three Studies

Study	Design			
	Location and Time	Sample Frame	Method of Measurement of Extent of Information Seeking	Range of Information Seeking Included
Bennett and Mandell	Harrisburg, Pennsylvania Spring, 1967	New car registrations	Structured questionnaire	Information seeking for all brands (cars) in the evoked set
First replication	St. Louis Missouri, Spring 1966	Most recent new car purchase of households	Open-ended questionnaire	Limited to information seeking for the brand chosen
Second replication	Bergen, Norway Fall, 1969	New car registrations	Structured questionnaire	Information seeking for all brands in the evoked set

Theoretical Expectations

At this point it is appropriate to indicate the theoretical expectations for the replications and to comment on the meaning of confirmation and disconfirmation in this respect.

Since the procedure was practically the same (and since there was no reason to believe that the Harrisburg and Bergen car buyers would differ significantly in terms of relationships among the variables of interest), the Bergen replication was expected to produce findings supporting the original results.

For the St. Louis replication the expectations were less clear. In this case it is necessary to speculate about the impact of the variations in procedure introduced. The use of open-ended questions (instead of a structured questionnaire) in the St. Louis replication would seem to mean a downward bias in reported information seeking and more error due to inaccurate recall. Hence, the correlations obtained would be suppressed. Limiting the measure of extent of information-seeking to information gathering about the car actually bought (rather than about all cars in the evoked set) was believed to have similar effect. Therefore, the St. Louis study was expected to show weaker relationships for the two hypotheses (H_2 and H_3) supported by the original study.

In this case, confirmation of the results originally found would strengthen the belief that there is a relationship between learning and extent of information seeking. Spatial validation by the Bergen data would make it more likely that the Harrisburg findings did reflect an underlying behavioral law, and that the findings were not a chance result of some idiosyncrasies of a given setting. Disconfirmation of the findings by the St. Louis study would imply that the underlying "true relationships" were not strong enough to withstand major variations in procedure weakening the measures.

Findings

The correlations of the measures of product/brand experience with the information seeking indices for the St. Louis and Bergen samples are shown in Table 3.

As seen in Table 3, the St. Louis and Bergen data were analyzed by computing product moment correlation coefficients instead of the Chi Square statistic as in the original study. This did not affect the conclusions, but was done to simplify reading and easing comparison. As shown in Table 3, all the signs for the St. Louis data, with the exception of one, were as predicted. However none of the correlation coefficients were significantly different from zero. There was clearly no compelling evidence for any of the three hypotheses.

For the Bergen sample, on the other hand, the results showed a remarkable similarity to the Bennett and Mandell findings. While no support was obtained for H_1 , the data conformed with the theoretical expectations for H_2 ($p < .05$) and H_3 ($p < .05$).

Discussion

The main finding common to all three studies was the lack of support for the H_1 hypothesis based on the Howard-Sheth formulation that all buying experiences in the product category are instructive. A possible explanation is that the development of learning theory as applied here has mainly been concerned with low-involvement or low complexity issues, such as repetitive buying or frequently purchased convenience goods. It may be that for a big-ticket items such as cars (being presumably higher

in importance, complexity, and involvement) there is a non-monotonic relationship between learning and information-seeking. At low levels of learning there may be little search for information because consumers, though objectively "needing" information, lack ability to process and evaluate information.

Table 3

Correlations of Product/Brand Experience with Information Seeking

(Product moment correlation coefficients)

Product/Brand Experience	Information Seeking			
	St. Louis Sample (n = 130)		Bergen Sample (n = 96)	
	Weighted	Unweighted	Weighted	Unweighted
T ¹	-.02	+.04	+.12	+.12
T ²	-.05	-.03	-.24 ^a	-.20 ^a
T ³	-.07	-.07	-.24 ^a	-.20 ^a

^aSignificant at the .05 level or better

A second finding was that the weights assigned to the sources of information did not make any difference.

The main difference in results in the two replication studies was that the Bergen study as expected did support the patterns uncovered in the original study. However, also in line with our expectations, there was no relationship for H₂ and H₃ in the St. Louis replication.

Hence, it appears in conclusion that the Bennett and Mandell findings were not sensitive to change in physical setting and culture but very sensitive to changes in the operational definitions and way of measurement. The fact that the previous findings were reproduced in a different setting, increases our confidence in the generality of the original results.

The findings of this attempt to replicate the Bennett and Mandell study underscore the need for a replication tradition in consumer research. Further replications should bring in more variables and/or alternative hypotheses to establish under what conditions the original effects found exist. Hence, comprehensive systematic replication is necessary before it is possible to obtain a high degree of confidence in a lawlike generalization.

One solution is to carry out what Lykken (1968) calls "constructive replication", letting second and third researchers start from the conclusion of the original researcher using only the fact or law that the former claims to have established and then let the replicators formulate their own conceptualization and methods for sampling, measurement, and data analysis. To illustrate, instead of using a cross-sectional design, the Bennett and Mandell findings could be put to a more fundamental test by designing a longitudinal study. A panel study could follow a sample of persons over time through several car purchases.

References

Alderson, W. (1952), "Psychology for Marketing and Economics," Journal of Marketing, 17, 119-35.

Bennett, P. D. and Mandell, R. M. (1969), "Prepurchase Information Seeking Behavior of New Car Purchasers - The Learning Hypothesis," Journal of Marketing Research, 6, 430-3.

Berkman, H. W. and Christopher, C. G. (1978), Consumer Behavior: Concepts and Strategies, Encino, California: Dickenson Publishing Company

Campbell, D. T. (1957), "Factors Relevant to the Validity of Experiments in Social Settings," Psychological Bulletin, 54, 297-312.

Engel, J. F., Blackwell R. D., & Kollat, D. T. (1978), Consumer Behavior, Third Edition, Hinsdale, Ill.: The Dryden Press

Hall, J. F. (1966), The Psychology of Learning, Philadelphia: Lippencott

Hansen, F. (1972), Consumer Choice Behavior, New York: The Free Press.

Hilgard, E. R., and Bower, G. H. (1966), Theories of Learning, Third Edition, New York: Appleton-Century-Crofts.

Howard, J. A. (1977), Consumer Behavior: Application of Theory, New York: McGraw-Hill Book Co.

Howard, J.A., & Sheth, J.N. (1969), The Theory of Buyer Behavior, New York: John Wiley & Sons.

Kollat, D.T., Engel, J. F., & Blackwell, R. D. (1970), "Current Problems in Consumer Behavior Research," Journal of Marketing Research, 7, 327-32.

Kotler, P. (1965), "Behavioral Models for Analyzing Buyers," Journal of Marketing, 29, 37-45.

Lykken, D. T. (1968), "Statistical significance in Psychological Research," Psychological Bulletin, 70, 151-9.

Newman, J. W. (1977), "Consumer External Search: Amount and Determinants," in A. G. Woodside, J. N. Sheth, & P. D. Bennett (Eds.), Consumer and Industrial Buying Behavior, New York: Elsevier, 79-94.

Ray, M. L. (1973), "Psychological Theories and Interpretations of Learning," in S. Ward & T. S. Robertson (Eds.), Consumer Behavior: Theoretical Sources, Englewood Cliffs, N. J.: Prentice-Hall, 45-117.

Watson, J.B. (1913), "Psychology as the Behaviorist Views It," Psychological Review, 20, 158-77.

METHODOLOGY AND META-METHODOLOGY IN CONSUMER RESEARCH: A COMMENTARY

Peter D. Bennett, The Pennsylvania State University

When asked to serve as a discussant in a session such as this one, there is a tendency to search for a common thread that holds it all together. That thread is there, though it's weak. In these brief comments, I will try not to stretch that thread to its breaking point.

All three papers plea for creative insight into new or revised methodologies, or in one case an even more fundamental plea to open our attention to new or different approaches to consumer research--our philosophies of science. All argue, more or less persuasively, for consumer researchers to not only look hard at our methodological biases and proclivities, but to develop new methodologies to fit unique problems which are beyond the evaluation of individual choice behavior.

But there the similarities end, so it seems the better part of wisdom to look at the three papers separately.

Hutton and McNeill

As a case study in the almost overwhelming difficulties of the evaluation of the impact of social programs, the paper is enlightening on the one hand, but frightening on the other.

While the paper really adds little in a methodological sense, it demonstrates a well conceived application of appropriate methods to the special problems of evaluation of a public domain program of some complexity. The contribution, then, lies in its treatment of the special problems of such public program evaluation. By that, I mean, were this research associated with a test market for a new brand of some existing product class, or even a new product, the "end-result" (how much did we sell?) as well as what Hutton and McNeill call "diagnostic" and "formative" issues would surely have been included by most sophisticated industrial researchers.

The authors point to a few of the special problems of the research domain. For one, they say that there are special problems of researcher/decision maker interaction because, "The policy maker does not have adequate training to judge research methodology and policy makers are, at the very least, hesitantly supportive of evaluation." How serious a problem this is, and how much it differs from the interaction between researchers and, say, brand managers, probably varies. In any case, it's not really a methodological issue, but a political one. In short, while it may be a problem, it doesn't really need to be one.

They also key on a problem of the timing of both the placement of evaluation as coming late in the program design, and decisions needing to be taken before the results are available. This is also not a methodological problem, but one of planning. Third, the problem they point to of the "orientation" of evaluators being toward survey and single measure research is not methodological, it is really a problem of education.

Having been involved myself in a research where the users were in the public sector, I sympathize with the authors' problems, and am grateful for their strong argument that these kinds of problems need to be overcome. We should also be grateful for their public report of an example of a real program evaluation study which tries to overcome some of these problems. Their research, insofar as the "end-results" portion is concerned, is well designed and executed. The "diagnostic" part of the research is perhaps

as well as could feasibly be done in the field setting of such a quasi-experimental design. All alternative explanations, e.g., some inherent differences between New Englanders and New Yorkers, might have been more thoroughly explored, given of course, adequate time and resources.

Dardis and Stremel

The paper on risk/benefit analysis in relation to product safety is, on the one hand, both extremely interesting and methodologically "tight." On the other, it is sufficiently marred with unsupported (perhaps unsupportable?) assumptions that it is frightening if it were to be actually used for making public policy decisions. The authors argue irrefutably for both the need for a risk assessment methodology, and that such a methodology consider benefits as well as risks.

The methodology suggested in the paper is intriguing, and I hope it will stimulate considerable comment and debate, because the topic is one of such critical importance. Dardis and Stremel conclude that, "the proposed method for assessing risk appears feasible." Whether or not it will eventually emerge as meaningful and operational requires answers to a number of questions. Let's take as given that:

$$\text{Risk/Benefit} = P_f C_f + P_n C_n / V, \text{ and}$$

$$\frac{N_f C_f}{VQ} = \frac{N_n C_n}{VQ}$$

where each term is as defined in their paper.

Since both P_i are really N_i/Q_i , we need to examine the assumptions underlying this relationship. The value of Q is rooted in dollars spent on product i , (not actually quantity of items) during one year. This assumes each product has a life of one year (or that the durability of the products are equal). Our own experience leads us to question this assumption--nearly always my pants outlast my shirts and my wife's robes seem always to outlast her gowns. I have no solid empirical evidence which speaks to the differences in the clothing turnover rates, but the assumption of equal life seems questionable.

Since N_i is the incidence of consumers being burned, it would appear to include burns received in a fiery auto accident or home fire, neither of which seem to have anything at all to say about the safety of the garments the victim is wearing, though they are here lumped together with burns received which are so related.

The serious problem with V is that it treats benefits as value in exchange rather than value in use, and the footnote doesn't adequately deal with the problem. It is the use of a product, not its ownership, that exposes one to risk. One may own a large number of shirts, each one of which is used, say, once every two weeks, while one's robe is used almost daily. Again, we are without data, but the assumption built into the operationalization is open to serious question.

Finally, and perhaps of greatest concern is the proposed operational treatment of the cost constructs. Such an oversimplified, purely economic, approach is flawed on both logical and political grounds. The openly stated assumption is that the rankings of products is not invalidated by ignoring "pain and suffering" costs if they are proportionate to the sum of direct and indirect costs. Since one major component to economic costs is the present value of future

earnings, this view assumes that the widow of a \$100,000 per year ad agency executive is greater than that experienced by the widow of a \$25,000 per year consumer researcher, and infinitely more than the widower of an unpaid housewife. It is an assumption that is hard to swallow on logical grounds, and is one that the NAACP and NOW organizations are likely to see as untenable.

Perhaps what we have in this proposed methodology is a way of comparing the relative risks/benefits across products to society, or more coldly to the economy. However, the whole purpose behind product safety intervention is to protect individuals, not the economy. It seems to me that a major contribution of this paper is to make it clear to us that the methodology that should eventually be used is one that is not likely to rest on readily available secondary data sources. The contribution lies in identifying those constructs for which we will need data in order to do risk/benefit analyses. The challenge will be to develop the appropriate operationalizations, which will no doubt be extremely complex and arduous tasks. That is why I said earlier that I am pleased this paper is here, and that hopefully it will stimulate comment and debate leading to those badly needed research efforts.

Uusitalo and Uusitalo

The final paper calls us to task at a much more fundamental level. The authors argue cogently for researchers in our young discipline to slip the bonds of a philosophy of science with roots in the physical sciences. I can remember when it was next to impossible to study the philosophy of science without using texts written by physicists, and for someone who knew nothing about physics, that was no easy task.

The authors have prepared some truly challenging suggestions with which I find myself largely in agreement. To the extent that researchers in our discipline are wedded to the tenets of logical positivism, we are losing opportunities to explore approaches to research questions which might be very fruitful.

Uusitalo and Uusitalo also suggest that we have continued to be wedded also to research approaches that are "traditional" to either economics or psychology. If they were saying that we were guilty of borrowing theories and concepts from those mother discipline and applying them blindly to consumer behavior phenomena, our reaction could legitimately be "ho-hum." But that's not what I hear them saying. What I hear is something much more fundamental. I have a friend who is a very fine consumer psychologist. When faced with a research issue of interest to him, he will study it, analyze it, and conceptualize the research in terms of the results as they would appear in an ANOVA table. What I hear the authors saying is that there are other, some radically different, ways to conceptualize both the problem and the entire approach to the research the problem calls for. Let's open ourselves up to those alternative approaches.

I found the paper very difficult to read, and I think the reason is that it says so much in so few words. Had the authors had the opportunity to expand their arguments and illustrations to three or four times the paper's current length, it would no doubt be much clearer. I hope that this fact does not stand in the way of lively debate and response to these proddings. For instance, would following this advice lead us to a research tradition in consumer behavior which is even more eclectic than we have now? and would that be functional or dysfunctional? Many of us could take either side of that question and have a debate that would shed more light even than the "debates" between presidential candidates. While some may find these authors poking at their sacred cash cows, I think we should thank them for doing the poking.

AN EVALUATION OF SEX ROLE THEORIES: THE CLASH BETWEEN IDEALISM AND REALITY

Susan Hesselbart, Florida State University

Abstract

This paper describes how different sex role theories can create new or perpetuate old myths about women and men. Research evidence is used to evaluate how well these myths fit reality.

Many consequences of contemporary feminism are now part of our daily lives. As with other social changes, theories are devised to explain new realities. This paper explores some new and old theories about gender and society, and the potential these approaches have for gender mythology.

We have discarded old and created new myths about the sexes during the past 15 years. Largely discarded myths include: the dire consequences to the offspring of employed women; that only men enjoy sex; or that sex discrimination vanished years ago. New myths include: the spectre of men as "emotional cripples"; that successful employed women now acquire men's "stress diseases"; or that both sexes are now converging in role demands and behavior.

Social myths serve functions for those who hold and promulgate them. Myths can be used to block or to foster social change. By "mythology" I mean that either: (1) data exist to invalidate the myth; (2) at best biased data support the myth; or (3) the mythological structure is illogical or inconsistent.

I will examine sex role theories along two dimensions: (1) postulating many versus relatively few sex differences; and (2) the attribution of similarities or differences to mainly biology ("nature") or to culture ("nurture"). Although each dimension is a continuum, for simplicity here I will treat each as a dichotomy.

Major Approaches to Gender

The Extent of Sex Differences

Most past and current sex role theories assume that there are many sex differences in character and social position. These differences correspond to common stereotypes held by laypersons. Men are generally portrayed as dominant, aggressive and stoical while women are viewed as warm, nurturant, cooperative and sensitive. The social positions resulting from these perceived differences assign men to "instrumental" task and leadership roles, taking an active part in major societal organizations, e.g., the economic and political. Meanwhile, women play "expressive" roles emphasizing nurturance, e.g., nursing or child care. Some claim that these are universal patterns. Approaches postulating many sex differences include most sociobiologists (e.g., Barash 1977, Goldberg 1974, Wilson 1978); recent work on men's roles (see the collections in Lewis and Pleck 1979 or Pleck and Brannon 1978); functional sociologists (Parsons and Bales 1955); and those advocating "crossover" or convergent sex roles (e.g., Giele 1978).

A theorist could believe that these sex differences are mainly shaped by "nature" or "nurture". Virtually all theorists believe that women's part in reproduction powerfully influences sex roles. Beyond this, theories diverge widely in terms of how much can be attributed to biology and why biology is important.

Far fewer theorists see relatively small sex differences. Those in this category focus generally more on social roles than on personality characteristics. Not surprisingly, most favor "nurture" over "nature". They place the greatest emphasis on how social change--technical, demographic or ideological--influences both sexes (e.g., see Huber 1976).

Nature Versus Nurture

Sociobiologists believe that traits are selected for "evolutionary fitness". Those enabling an organism to survive long enough to reproduce are passed to offspring. Many of these theorists believe that females have more "investment" in offspring than males. Females bear and nurse their young and are certain that a child is theirs, whereas males cannot be certain. This "investment" difference puts males in a "seller's market" competing to be found attractive by females. Economic provision and protection are two assets males can use to increase their attractiveness. Thus traits such as aggression or mental acuity were "selected for" in men. Since humans have a long, dependent infancy, traits such as nurturance in caregivers increase offspring survival; thus these were "selected for" in women.

Sociobiologists do not explicitly say that "anatomy is destiny". Rather biological sex is a predisposition or potential. Nevertheless, sociobiologists are also those who claim near universality of certain patterns, e.g., men's economic productivity or women as caretakers of children.

"Nature" theorists believe that biology shapes individuals who then shape society. In contrast, "nurture" types see the role of the sexes in reproduction as shaping societal institutions which then shape personality. To survive, any society must keep fertility at least at replacement levels. Societies with high mortality also need high fertility. In labor-intensive societies, children are also an economic asset.

All this implies that in earlier societies women bore and attempted to rear many children. Survival past reproductive ages was rare for both sexes until this century. Average longevity is below age 50 for many nations today. Thus, until recently, women spent most of their adult lives as mothers. Women's productivity was constrained by maternal mobility.

Changes influencing both sexes occurred with the industrial revolution. Productive labor moved from the home to the factory. Given women's maternal constraints, this largely meant that men left the home to take part in an increasingly cash monetary system. Home crafts that formerly had been women's work also moved to the factory. At the same time, improved living conditions led to less mortality. Children were less often seen as economic assets.

These changes have been viewed as leading to the devaluation of women. Productive "women's work" shifted outside the home or (in the case of childbearing) was seen as less valuable. Wage-earners in a cash economy have more power than those with little cash income. Thus, men as a class have been hypothesized to gain power over women as a class.

This view suggests that sex differences respond to the condition of larger society and are mainly the result of socialization. During childhood, the sexes are trained to appropriately fit into adult roles. If men earn their living in an industrialized culture stressing achievement, then boys are trained to be competitive and ambitious. If women remain within the home tending to children and adult men, they need training instead in forms of emotional expressivity (e.g., Hoffman 1977, Huber 1976, Oakley 1974, Parsons and Bales 1955, Stockard and Johnson 1980.)

Many "nurture" theorists are especially interested in adult socialization. For example, Kanter (1977a) stresses how adults are shaped by occupational hierarchies within corporations. Persons of either sex low in the hierarchy behave in stereotypically "female" ways, for example, stressing sociability on the job or having low aspirations for promotion. If the job is a dead-end one, these may be realistic responses.

Those taking a "nurture" stance also stress how demands made by different institutions shape individual behavior. Role incompatible demands can occur among both sexes, but they are especially pronounced for women--particularly for mothers in the paid labor force. These women face two jobs, one outside the house and one within. This "role overload" means that women have less leisure time than men. For example, employed married men in 1975 averaged 36 hours a week leisure time compared with 32 hours a week for employed married women (U.S. Department of Commerce 1977: Table 10/1). These types of differential demands may produce "sex differences" rather than biology or personality traits.

Finally, a few theorists perceive relatively small sex differences but attribute these mainly to "nature". The most famous example is Maccoby and Jacklin's *The Psychology of Sex Differences* (1974, but also contrast Bardwick 1979 with Bardwick 1971). These perceived sex differences mainly center around greater male aggression.

Theories of Gender and Social Science Mythology

These theoretical approaches have implications for the creation and perpetuation of social science mythology. Such myths extend beyond the academic community. They can shape perceived social reality, and can define whether social change is possible or desirable, the kinds of changes sought and the benefits and costs of change. For example, the distinction that Freud made between vaginal and clitoral orgasm influenced research, therapy and women's self images for decades.

The Case of "Fear of Success"

A more recent case is the "motive to avoid success" (M_{as}), sometimes called "fear of success" (Horner 1968). M_{as} was conceptualized as a stable, internalized motive, learned early, held by bright women who had the greatest chances for success. High M_{as} women fear the consequences of succeeding in male-typed professions such as medicine. Social rejection is seen as the "reward" for these accomplishments. When high M_{as} women are placed in a competitive or male-typed achievement situation, they will tend to lower their performance.

"Myth" may seem too strong to assign to "fear of success". However, most data do not replicate the patterns that Horner predicted. Indeed reanalyses of Horner's dissertation data do not support her hypotheses. Tresemmer (1976) found that high M_{as} women in Horner's data had the highest performance scores of all women in five out of six conditions. With more sophisticated research designs, what was categorized as an internal motive among women may turn out to be negative stereotypes about achieving women held by both sexes (e.g., see Levine and Crumrine 1975, Zuckerman and Wheeler 1975).

Let's consider some implications of the "fear of success" example. It implies that women's actual performance is below men's because women are hampered by M_{as} . In an economy that claims that merit is the most important determinant of reward, this suggests that women's lower rewards (e.g., income) result from their own performance. It should be noted, however, that many studies have found that the identical performance is rated lower when a woman's name is attached rather than a man's.

The "fear of success" perspective also specifies the remedy: women must change through therapy; "assertiveness workshops"; remedial education; or self-help books teaching one office politics. Finally, hidden in "fear of success" and many other individual differences approaches to gender is the idea that major traits and motives are instilled in childhood and adolescence and are very difficult to change during adulthood. This means we will need to wait many years for a new, strong generation of women to grow up before substantial social change can occur. The generational change remedy contrasts with suggestions by those stressing social structure (e.g., Kanter 1977a 1977b) and the short-term changes that organization changes can produce.

"Fear of success" became popularized during the early 1970s. Ironically, better designed later research which does not support the M_{as} has remained cloistered in academia. I suspect that a social science approach stressing individual differences is probably more acceptable than a social structure approach to Americans, who are told that it is individual merit that "really counts". A structural approach is also obviously more costly to existing organizations than one stressing individual change. "Fear of success" serves the needs of business and government far better than potential myth stressing massive social change.

The Seeds of Mythology

Different approaches to gender and society have differential susceptibility to mythology. In the preceding example, I pointed out that an approach stressing individual differences will try to locate motives and traits to explain social phenomena such as the lower economic rewards to women than to men. In general, I believe that approaches that either (1) stress relatively large numbers of sex differences or (2) attribute most perceived sex differences to biology have the greatest probabilities of creating or perpetuating mythology. Theories assuming relatively few differences with a nurture approach may be less susceptible to myth creation.

This does not mean that there are no sex differences in reality or that biology plays no role in gender differentiation. For example, cross-cultural work strongly suggests that pregnancy (rather than physical strength or personality traits) has influenced the gender differentiation of labor. Jobs that are predominantly male involve mobility: hunting; fishing; soldiering. Women's productive labor is closer to home: domestic agriculture; infant care. With low birth rates, highly predictable contraception, medically safe abortion and safe substitutions for mother's milk products of the last century, it is not surprising that much of women's roles had centered around the care of young children. Cross-cultural and time series data also imply that greater physical activity (which does not necessarily mean aggression) is more characteristic of males than females (e.g., Boserup 1970, Huber 1976, Maccoby and Jacklin 1974, Oakley 1974, Stephens 1965).

Thus, a theory need not stress that there are no sex differences at all. What I am saying is that approaches that a priori assume large sex differences or that biology strongly influences sex differences in "personality" are more open to social science myth creation. A priori assumptions constrain which topics are studied, how topics are studied, and

how results are interpreted. Research using true "null hypotheses" would assume at first that there would not be sex differences. To confirm or reject such hypotheses, we need to ask similar research questions about both sexes. In most cases, this is exactly what has not been done and our research accordingly has often contributed to mythology.

Consider again "fear of success". Horner a priori emphasized the importance of personality traits over, say, social structure in academic and economic achievement. This meant that effects from, for example, the hierarchical organization of corporations and the roles of each sex within these were not considered. Horner also assumed that the achievement conflicts within women would be reflected in the stories college students told about Thematic Apperception Test cues. Given that TAT responses were assumed to reflect internal conflict (rather than social stereotypes), Horner saw little need to present cross-sex cues. Thus, women responded only to a female referent and men only to a male referent. When later research did use cross-sex cues, men reacted with higher "fear of success" imagery to successful women cues than women did--hardly confirming the M_{AS} as an internal personality motive.

Instrumental and Expressive Role Mythology

In another current myth, many functional sociologists assumed a priori that sex differences result in a certain evolution of social structure. Men's instrumental or task oriented role was to "represent" the family to the outside world. His employment sets the social status for the entire family. Women play an expressive, socio-emotional role within the home. This efficient division of labor prevented spouses "competing" with each other for status were both to have an engrossing "career" (Parsons 1942, Parsons and Bales, 1955). The nearly wholesale buying of this myth continued until the late 1960s.

The "expressive/instrumental" mythology resulted in the neglect of many research questions even though existing data did not support it. For example, Parsons himself (1942) noted that this differentiation was mainly true only of the upper middle class. Working class families usually needed both spouses in the labor market (but see Oppenheimer 1977 on middle and working class data). Even in 1979, only 30 percent of employed men were in professional or technical occupations, i.e., those with the greatest chances to become upper middle class. The majority of men have blue collar jobs which are neither "engrossing careers" nor the route to upper middle class status. Thus, relatively few Americans can afford the "expressive/instrumental" differentiation.

The functionalists' division of roles assumes that a man's top priority is his work and a woman's is her family. This is particularly mythical when applied to men. While the largest allotment of time for married men may be to their job, whenever the question has been asked, men rank their families as their highest priority and greatest source of satisfaction. Survey data have been consistent on this since at least the 1940s, and such results have been obtained from all strata of men (see a review in Hesselbart 1978 and also Pleck and Lang 1978). Because of the functionalist approach, we have largely ignored men's family roles and treated women's employment as a "social problem" (Feldberg and Glenn 1979, Lewis and Peck 1979).

Perhaps because social science has rarely explored the family for both sexes with parallel questions, functionalist though has precisely reversed the roles of each sex within the family. A correlated myth is that full-time housewives do not "work". Housework was hardly examined until the middle 1970s, and it is made clear that

the major role women play within the home is instrumental (Berk 1980, Glazer 1976, Oakley 1974, U.S. Department of Commerce 1977: Table 10/1). It is hard to define the average 44 hour housework week for fulltime housewives and the 25 hour housework week for wives in paid employment as "expressivity". It is also usually the wife who represents the family to institutions other than paid employment, e.g., churches, PTA meetings, kin contact, or tradespersons. In contrast, husbands enjoy more expressive family relations, such as playing with (rather than routine care for) children or recreation. In comparison with wives, married employed men spent only 10 hours a week in family care in 1975.

"Nature" and Myth Perpetuation

Those who postulate many sex differences due to biology often end up either apologists or strident defenders of the status quo. Many sociobiologists argue that most sex differences in social structure and social institutions are inevitable universals. Male dominance in politics, culture or the economy are proposed as true of virtually all societies. Man's greater physical strength and logical character cause male dominance (e.g., Barash 1977, Goldberg 1974, Wilson 1978).

In proclaiming these universals sociobiologists exaggerate cross-cultural similarities. These are used to justify the relative lack of change in Western gender stratification. Even if large changes were possible, they would be so costly, change would not be worth it. Sex differences appearing dysfunctional now (e.g., norms encouraging women to have large families) were adapted in "99 percent" of our evolutionary history and would take generations to change. These assertions ignore the great variability in women's contributions to society, especially in economic productivity (e.g., Aronoff and Crano 1975, Boserup 1970, Rosaldo and Lamphere 1974).

Sex differences treated as biological or universal can also betray a curious sense of history. For example, women's greater longevity is a phenomenon of the twentieth century and may be due more to living habits (e.g., lower accident rates) than to biology (Harrison 1978). Such "universal" assertions are also contradicted by research that finds that sex differences appearing in early childhood during the 1950s or 1960s now do not emerge until adolescence during the 1970s (e.g., male superiority in math skills, Maccoby and Jacklin 1974). Even the most determined sociobiologists do not argue for spontaneous mutation and evolution over a 20 year period.

Nurture, the "Feminine Principle" and "Role Crossover"

Those subscribing to many sex differences instilled by culture can fall prey to other mythologies. In particular, they are susceptible to the mythology of the "feminine principle". Women's lives and "women's traits" are seen as healthier, more flexible and more useful to modern society than traditionally "male traits". These theorists argue for and perceive a society in which both sexes are becoming more alike or are "crossing over" into roles or traits that embody the "best" of both gender stereotypes. Thus, women gain more interests in paid employment or politics while men become more familial oriented. The healthiest persons are those showing the strengths of both sexes: "masculine" competence and "feminine" nurturance (e.g., Bem 1975 or Giele 1978).

Derivations from the "feminine principle" caution women against becoming "too masculine". "Executive stress" diseases once confined to men now strike the women entering "men's professions" or managerial roles. Related to "executive stress" is the belief that men suffer because they are not allowed women's emotional expressivity. A

theme of "how my life as a man has been crippled because I am not allowed to cry" repeatedly appears in those theorizing about male roles. Yet another spin-off is the role of "society's housekeeper" a myth that is the societal analogue to women's Victorian housekeeper role. Once women share power more equally with men, women will clean up the "messes" that men, like little boys, leave behind; war; ecological waste; inflexible life timetables; and hierarchical organizations (e.g., Bardwick 1979, Ciele 1978, Kenzer 1978, Lewis 1978).

There is a common core to these myths. Researchers creating and perpetuating them hold highly stereotyped views about the sexes. Men really are much more aggressive, repressed and hierarchical than women; women really are much more nurturant, emotional and cooperative than men. The problem is that men and women neither describe themselves so differently nor behave so differently to support such myths (Maccoby and Jacklin 1974). What are sometimes called "sex differences" by those unfamiliar with research on social structure may result from playing superior and subordinate social roles (e.g., Henley 1977, Kanter 1977a). Unfortunately, occupying a subordinate social position does not imply greater mental health or flexibility but instead a possible legacy of self-doubt, anxiety or depression (Deaux 1976, Dohrenwend and Dohrenwend 1976, Pearlin and Johnson 1977). Finally, women's roles are changing more than those of men suggesting that "role crossover" is quite a way into the future.

How well are these recent myths supported? Let's first look at "executive stress". Relatively few women are employed as managers--six percent in 1979 (14 percent of employed men). Most, of course, are not at the upper echelons where ulcers and heart disease reputedly strike. Comparisons of health rates over time fail to suggest that women were starting to suffer from "executive stress" more during the late 1970s. Age-adjusted death rates on cause of death showed lower not higher ratios in 1976 than in 1970 for women on heart disease, cirrhosis of the liver and suicide. These ratios were slightly higher for women in 1976 than in 1970 on respiratory diseases, homicide and ulcers--but women's rates were under half of men's. Life expectancy tables favoring women's longevity increased by a tenth of a year between 1970 and 1977. Death rates for heart diseases among young adults fell steeply for both sexes between 1950 and 1977. Five percent fewer women in the early 1970s said they had had or were near a "nervous breakdown" than in the early 1960s (men's comments fell by less than one percent). Age adjusted rates of acute illness dropped on respiratory illness very slightly for both sexes and rates of digestive illness and injuries rose for both sexes between 1970 and 1977. Rates of increase were nearly identical for both sexes (U.S. Department of Commerce 1977: Tables 5/8 5/39, U.S. Department of Commerce 1979: Tables 2/1 2/3, U.S. Department of Commerce 1980: Tables 100 112).

What is apparent is that more young women are smoking cigarettes. In 1974 among 17 and 18 year olds, 31 percent of boys and 26 percent of girls smoked cigarettes. The rates were identical for boys in 1968 but only 19 percent of girls smoked then (U.S. Department of Commerce 1977: Table 5/27). This is taking its toll in converging rates of respiratory illness. Of course, the inferential leap from increased smoking among teenage girls to the dire prophecies of "executive stress" is large and other data here do not support this leap.

There is little systematic research examining how social reactions to emotional expression differ across subgroups or situations. However, it is interesting that so many recent theorists on men's roles seem to believe that women's crying receives social approval. Women's supposed greater emotionality is frequently cited as a

reason to exclude women from political office or executive suites--hardly a sign of social approval. Further, men may be allowed to express anger more frequently than women. Some psychologists (e.g., Chesler 1972) believe that anger turned inward instead of out can result in depression, and it is the case that women in the United States generally show a higher incidence of depression than men (e.g., Dohrenwend and Dohrenwend 1976).

The role of "society's housekeeper" is actually a partial resurrection of the "female purity" myth used in the early 1900s when women sought the right to vote. Women are portrayed as more humane, opposed to war or to the sufferings of poverty than men. In public opinion polls there is some substance to this myth, but sex differences are relative rather than polarized. For example, if 53 percent of women and 66 percent of men support a military draft (Gallup Poll February 1980), just how humane could we expect a government of women to be? Both sexes are usually on the same sides of the issue even when sex differences in public opinion occur (see "The Polls" in the Public Opinion Quarterly or the Gallup Opinion Index for more examples).

Despite stereotypes among both managers and laypersons, women do not seem to bring unique humane contributions to business either. Women and men in similar jobs have similar values and personalities (Kanter 1977a, Reif, et al. 1975) whether through self-selection or adult socialization. It is possible that with more women in management larger sex differences might appear, but this is really conjecture at this time.

If sociobiologists and functionalists seem to serve the status quo, many of those arguing that large sex differences occur because of culture hold out a beacon of social change to benefit everyone. If androgyny becomes more prevalent, both sexes will gain. However, evidence supporting the "feminine principle" myths does not seem any more compelling than those supporting "gender universals".

I suspect that the "feminine principle" is a lure to sweeten the problems of social change. Shifts in gender and society influence us all, especially if these shifts involve a rebalancing of social power. Scanzoni (1979) suggests that women's negotiations to change men's roles in marriage may involve self-interest and/or altruism to effect greater role sharing. For example, it can be argued that a wife's employment (which seems to give her more power in marriage) makes her "more attractive" or a "better companion" than if she is not employed. A similar phenomenon of a "self-interest sell" may be occurring among social scientists who want more gender equality.

Socio-technical Models and Social Change

Socio-technical and structural models of gender have been left until last because I believe these models offer the least fertile ground for myth creation. Most theories of gender, these included, are "middle range". A larger theory should be grounded in tentative assumptions about both sexes that allow the greatest possible comparisons about sex differences and similarities over time and space. Socio-technical and structural models at this time come closest to this goal but systematic research using "null hypotheses" is just beginning.

We must begin to study the roles of men as parents, spouses, friends or lovers and women's labor whether cash compensated or not. "The family" seems to be a high priority for both sexes but we don't yet know how or whether the familial concept differs for women and men (the data on housework and men's expressivity do give some clues here). Survey data do not suggest that men's interest in the family is a recent one. Changes

that we see now, such as child custody in divorce may reflect old values activated by the recent women's movement. The connections between women's and men's value priorities and changes in social structure need a good deal more research.

Socio-technical theories appear to better explain the dramatic shifts that have occurred in women's labor force participation and the implications of these shifts for other institutions. The rise of married women in the labor market and rising divorce rates preceeded rather than followed contemporary feminism. A combination of low birth rates in the 1930s, high birth rates in the 1950s and an economy increasing in the service sector drew married women into the job market. By 1979, 49 percent of all women, 57 percent of the married mothers of school age children and 42 percent of the married mothers of preschool children were in the labor force. Accordingly, changes occurred in attitudes toward the acceptability of employed wives.

With more permanence in the labor market and with legal changes against discrimination, women's academic aspirations have changed. In 1977, women received 46 percent of bachelor's degrees, 47 percent of master's degrees and 25 percent of doctorates compared with 42 percent, 40 percent and only 14 percent respectively in 1971. In 1977, 12 percent of women receiving bachelor's degrees majored in business, engineering or physical science, double the 1971 percentage. Nearly 20 percent of new physicians were women in 1977 and 23 percent of new lawyers compared with eight and five percent in 1971 (Digest of Educational Statistics 1972: Table 114 U.S. Department of Commerce 1980; Tables 281 282 283).

With greater chances for self support (even with 59 percent of a man's income), more women seem to be electing to stay single. In 1978, 48 percent of women in their early 20s and 18 percent of those in their late 20s had never married, compared with 28 and 11 percent in 1960. Live birth rates plummeted from 118 per thousand women aged 15 to 44 in 1960 to 68 in 1977. Of ever married women, 25 percent were still childless in their late 20s in 1978 compared with 13 percent in 1960 (U.S. Department of Commerce 1980; Tables 54 83 91). Thus, the demographic and socio-technical factors which helped bring married women into paid employment in the 1950s and 1960s have probably "spilled over" indirectly into other forces of social life.

These sociocultural issues have not been postulated to cause large changes in men's roles, and, indeed men's lives seem not to have changed as much as women's over the last 20 years. Husbands do not generally do more housework when wives are employed, although they may now be spending a little more time in the 1970s than 15 years earlier (Pleck 1979). While women opted for more "masculine" college majors in the 1970s, men did not "cross-over" into fields such as nursing or childcare (see Hesselbart and Bayer 1978). At least one recent study (Margolis 1979) suggests that the "corporation man" is still wedded to his company with its attendant geographical mobility. Possibly it is the changes in women's work roles juxtaposed with the lack of changes in men's family roles that have led young women and men to marry and have children less frequently.

Even changes in the roles of one sex, women, have brought about substantial debate, legal change and scholarly research. Perhaps it is a measure of the anxiety caused by changes in basic roles that has laid the ground for new social science mythology on gender and the revival of old myths. It seems time to stop creating mythology and to do basic, unbiased research that is decades overdue.

References

- Aronoff, J. and Crano, W (1975), "A Re-examination of the Cross-Cultural Principles of Task Segregation and Sex Role Differentiation in the Family," American Sociological Review, 40, 12-20.
- Barash, D.P. (1977), Sociobiology and Behavior, New York: Elsevier.
- Bardwick, J.M. (1971), Psychology of Women: A Study of Bio-Cultural Conflicts, New York: Harper and Row.
- Bardwick, J.M. (1979), In Transition: How Feminism, Sexual Liberation, and the Search for Self-Fulfillment Have Altered America, New York: Holt, Rinehard and Winston.
- Bem, S.L. (1975), "Sex Role Adaptability: One Consequence of Psychological Androgyny," Journal of Personality and Social Psychology, 31, 634-643.
- Berk, S.F., ed (1980), Women and Household Labor, Beverly Hills: Sage Publications.
- Boserup, E. (1970), Woman's Role in Economic Development, New York: St. Martin's Press.
- Chesler, P. (1972), Women and Madness, New York: Doubleday and Company.
- Deaux, K. (1976), The Behavior of Women and Men, Monterey, California: Brooks/Cole Publishing Company.
- Dohrenwend, B.P. and Dohrenwend, B.S. (1976), "Sex Differences and Psychiatric Disorders," American Journal of Sociology, 81, 1447-1454.
- Feldberg, R.L. and Glenn, E.N. (1979), "Male and Female: Job versus Gender Models in the Sociology of Work," Social Problems, 26, 524-538.
- Giele, J.Z. (1978), Women and the Future: Changing Sex Roles in Modern America, New York: The Free Press.
- Glazer, N. (1976), "Housework," Signs: Journal of Women in Culture and Society, 1, 905-922.
- Goldberg, S. (1974), The Inevitability of Patriarchy, New York: William Morrow and Company.
- Harrison, J. (1978), "Warning: The Male Sex Role May Be Dangerous to Your Health," Journal of Social Issues, 34, No. 1, 65-86.
- Henley, N.M. (1977), Body Politics: Power, Sex and Non-verbal Communication, Englewood Cliffs, N.J.: Prentice-Hall.
- Hesselbart, S. (1978), "Some Underemphasized Issues about Men, Women and Work," paper presented at the Annual Meetings of the American Sociological Association, San Francisco.
- Hesselbart, S. and Bayer, A.E. (1978), "Gender Equity and Postsecondary Education: Evolution, not Revolution," in Drew, D.E., ed., Increasing Student Development Options in College, San Francisco: Jossey-Bass, 39-59.
- Hoffman, L.W. (1977), "Changes in Family Roles, Socialization, and Sex Differences," American Psychologist, 32, 644-657.

- Horner, M. (1968), "Sex Differences in Achievement Motivation and Performance in Competitive and Non-Competitive Situations," Unpublished doctoral dissertation, University of Michigan.
- Huber, J. (1976), "Toward a Socio-Technological Theory of the Women's Movement," Social Problems, 23, 371-388.
- Kanter, R.M. (1977a), Men and Women of the Corporation, New York: Basic Books.
- Kanter, R.M. (1977b), Work and Family in the United States: A Critical Review and Agenda for Research and Policy, New York, Russell Sage Foundation.
- Kinzer, N.S. (1978), Stress and the American Woman, New York: Anchor Press/Doubleday.
- Lewis, R.A. (1978), "Emotional Intimacy among Men," Journal of Social Issues, 34, No. 1, 108-121.
- Lewis, R.A. and Pleck, J.H. (1979) eds., Men's Roles in the Family, special issue of The Family Coordinator, 28.
- Levine, A. and Crumrine, J. (1975), "Women and the Fear of Success: A Problem in Replication," American Journal of Sociology, 80, 964-974.
- Maccoby, E.E. and Jacklin, G.N. (1974) The Psychology of Sex Differences, Stanford, California: Stanford University Press.
- Margolis, D.R. (1979), The Managers: Corporate Life in America: Work, Family and Community, New York: William Morrow and Company.
- Oakley, A. (1974), Woman's Work: The Housewife, Past and Present, London: Penguin Book Ltd.
- Oppenheimer, V. (1977), "The Sociology of Women's Economic Role in the Family," American Sociological Review, 42, 387-406.
- Parsons, T. (1942), "Age and Sex in the Social Structure of the United States," American Sociological Review, 7, 604-616.
- Parsons, T. and Bales, R.F. (1955), Family, Socialization and Interaction Process, Glencoe, Illinois: The Free Press.
- Pearlin, L. and Johnson, J.S. (1977), "Marital Status, Life-Strains and Depression," American Sociological Review, 42, 704-715.
- Pleck, J.H. (1979), "Men's Family Work: Three Perspectives and Some New Data," The Family Coordinator, 28, 481-488.
- Pleck, J.H. and Brannon, R., eds. (1978), Male Roles and the Male Experience, Journal of Social Issues, 34, No. 1.
- Pleck, J.H. and Lang, L. (1978), Men's Family Role: Its Nature and Consequences, Wellesley, Massachusetts: Wellesley College Center for Research on Women.
- Reif, W.E., Newstrom, J.W., and Monczka, R.M. (1975), "Exploding Some Myths About Women Managers," California Management Review, 7, No. 4, 72-79.
- Rosaldo, M.Z. and Laphere, L., eds. (1974), Woman, Culture and Society, Stanford, California: Stanford University Press.
- Scanzoni, J. (1979), "Strategies for Changing Male Family Roles: Research and Practice Implications," The Family Coordinator, 28, 435-442.
- Stephens, W.N. (1963), The Family in Cross-Cultural Perspective, New York: Holt, Rinehart and Winston.
- Stockard, J. and Johnson, M.M. (1980), Sex Roles: Sex Inequality and Sex Role Development, Englewood Cliffs, New Jersey: Prentice-Hall.
- Tresemmer, D. (1976), "Do Women Fear Success?" Signs: Journal of Women in Culture and Society, 1, 863-874.
- U.S. Department of Commerce (1977), Social Indicators 1976, Washington, D.C.: Bureau of the Census.
- U.S. Department of Commerce (1979), A Statistical Portrait of Women in the United States: 1978, Current Population Reports, Special Studies, Series P-23, No. 100, Washington, D.C.: Bureau of the Census.
- U.S. Department of Commerce (1980), Statistical Abstract of the United States 1979, Washington, D.C.: Bureau of the Census.
- U.S. Department of Health, Education and Welfare (1972), Digest of Educational Statistics, Washington, D.C.: Education Division.
- Wilson, E.O. (1978), On Human Nature, Cambridge, Massachusetts: Harvard University Press.
- Zuckerman, M. and Wheeler, L. (1975), "To Dispel Fantasies about the Fantasy-Based Measure of Fear of Success," Psychological Bulletin, 82, 932-946.

HOUSEWIVES, BREADWINNERS, MOTHERS, AND FAMILY HEADS: THE CHANGING FAMILY ROLES OF WOMEN¹

Janet A. Kohen, Institute for Social Research, University of Michigan and San Diego State University

Abstract

Delayed age at marriage, decreased fertility, higher employment rates, and increased family headship indicate a major change in women's attitudes and relationships to family roles. Review of recent research findings suggests that women have redefined their relationship to the family so that their personal goals and interests can be met.

Only a generation ago, most young women planned their adult lives around marriage and full-time homemaking. Barbie dolls, Seventeen magazine, and Father Knows Best fed their dreams of a home in the suburbs, a white picket fence, and several loving children. Women's participation within the family has shifted dramatically since then. Women are delaying marriage; planning whether, when, and how often they will be mothers; entering the labor force before their children leave home, often before their children enter school; and, a great many are taking the responsibilities of heading their families for some part of their adult lives. They are participating in their families not only as wives and mothers but also as breadwinners and family heads.

In the process, family forms and domestic relationships have been transformed. New issues have arisen around marriage and family organization. Such questions as whether to live together before marriage, what should be included in the marriage contract, whether to have a second child, when the kids are old enough to stay by themselves after school, and, who does the laundry or changes the oil in the car are becoming common family decisions. The impact, however, extends beyond individual families to virtually every institution in the society. Volunteer work in schools and hospitals; fringe benefits of part-time jobs; social security, pensions, and insurance benefits; child custody and child support; even census classifications have come under scrutiny.

The response to these changes has not been entirely positive. Some social commentators fear that the family is failing. A closer examination reveals that it is not the family which is being threatened but the traditional ideas about how families should be organized. Families continue to function but women no longer participate in them in the same old ways. Delaying marriage, planning fertility, participating in the labor market, and heading families indicate women's increasing desire and ability to control the conditions of their lives, including the conditions of their relationship to their family. Such increasing control challenges pre-existing ideology of women's tie to wifehood and motherhood. Beliefs that women want and emotionally prefer self-sacrificing behavior, particularly for their husbands and children, or what Barbara Ehrenrich and Deidre English (1978) have labeled "masochistic motherhood," are no longer consistent with women's consciousness or behavior. Neither are theories which attribute a "culture of narcissism" to changes in the form of husband and wife roles (Lasch 1979). Instead, current statistics and recent research in the four major areas I've listed above suggest not selflessness, nor selfishness, but increasing autonomy of women.

Delayed Age of Marriage

Young women are approaching marriage today with a perspective that includes other roles besides family roles. Not long ago, women suspended self-commitments to their futures until marriage, expecting to build their lives according to their husband's interests and careers (Bardwick and Douvan 1972). Marriage occurred for some women even before high school graduation, although for most it occurred soon after. In comparison, young women are now waiting longer before they become wives. In 1977, 45 percent of young women between 20 and 24 years of age had never married. This represents an increase of 17 percent since 1960 (U.S. Bureau of the Census 1978). Instead of cooking, cleaning, and taking care of infants, almost one out of every two young women in their early twenties were accumulating advanced education and work experience for futures that most expect will include both family and employment. In a recent study by Cherlin (1980) enrollment in higher education and the expectation of being employed rather than a full-time housewife at age 35 were associated with delay of marriage.

Since most young women expect to marry, these findings demonstrate an increasing awareness of the eventualities of marriage--financial pressures, divorce, widowhood, and, 20 years or more of employment eligibility following their children's departure from home. Rather than a passive awareness though, these women's actions imply self-conscious planning both in relationship to the roles they expect within the family and within the labor market.

Other evidence suggests an equally rational approach to marriage among young women. In the 1976 replication study of Americans View Their Mental Health, a study with which I've been associated as a post-doctoral fellow over the past year, Elizabeth Douvan's analysis supports a significant shift in the way young women view marriage. In 1957, the year respondents were interviewed for the first study, single women were the most positive toward marriage of all marital status groups of women. They were even more positive than single men, a noteworthy difference since overall men were more positive about marriage than women. In comparison, the 1976 single women showed the greatest decline of any group of women in the number who placed a positive value on marriage. They were much more accepting of women who chose to remain single and were substantially more likely than their 1957 counterparts to view marriage as restrictive and burdensome (Douvan 1978). Instead of the idealistic and perhaps over-conformist stance of single women in 1957, the 1976 single women appeared to be more critical of the options provided by marriage.

The consequences of young women's shifts in marital timing and in attitudes toward marriage raise questions about these women's future marital roles. Delay of marriage could mean a swing to alternative family arrangements, or it could mean that marriage will be embraced with greater awareness of problems and greater commitment to their solution. Young women are bringing more psychological and social resources to family formation but the type of family that results in the future is not clear. Whatever the family arrangements that get negotiated, there are hints from the 1976 replication study of Americans View Their Mental Health that young women will take part in those negotiations with a greater sense of personal identity. More married women than married men viewed their marriages as composed of two separate people rather than being a couple and married women were also less

¹The work on this paper was funded by National Institute of Mental Health Grant MH-14618 to Elizabeth Douvan, Joseph Veroff, and Richard Kulka.

likely to view marriage as the basis for their social validity (Douvan, Veroff and Kulka in press).

Fertility

One of the major factors enabling women to modify their family roles has been their control over fertility. Young women can explore relationships and decide on the form they will take without the perils of pregnancy. Once a relationship is established they can plan their parenthood around a variety of events including hers as well as his employment career, family income needs, child care supports, and a variety of other marital and family specific events. Such events, hers, his, and theirs, have had an impact on fertility. Married women are currently having 1.8 children. This lowered fertility has been associated with increased formal education, planned employment, and actual employment (Stolzenberg and Waite 1977).

While on the surface lowered fertility might imply a rejection of motherhood for career, a more thorough examination indicates the opposite. According to the Bureau of the Census, lowered fertility appears to result from a reduction in family size rather than a reduction in the number of women who become parents. Instead of three or four children, women are now completing child-bearing with one or two children (U.S. Bureau of the Census 1976). While the replication study of Americans View Their Mental Health indicated a significant relaxation of pro-natal forms with an increasing acceptance of childless couples (Douvan, Veroff, and Kulka in press), and the number of young women who say that they expect to remain childless is increasing (U.S. Bureau of the Census 1977), still childless women constitute about 5 percent of the population of young women and the number of women who are still childless by the age of 45 has remained relatively constant since 1960 (U.S. Bureau of the Census 1979). The decline is more likely the result of the ability of women to control their fertility to match their desired family size rather than having to adapt to unwanted fertility as they did several decades ago. Even though the decline has been associated with women's employment, there is reason to argue that the increasing financial costs of raising children may stimulate both women's employment and their lowered fertility.

A recent estimate of the costs of the first year of a child's life is over \$3,000 (Ann Arbor News, July 6, 1980). Given both this initial cost and the continuing ones in raising children, women's wages become important contributions to the expenses of children both prior to birth and thereafter. According to the Bureau of Labor Statistics Studies, much of married women's income gets earmarked for the children whether it is basic survival needs, college educations, or certain luxuries. Thus fertility and employment are doubly tied. Having children increases the need to work; work tends to reduce fertility.

Furthermore, the mere ability to plan children may also promote both women's employment and their limited fertility. Having children when they are wanted makes each child more important and valuable than when children were simply a contingency of marital sexuality. In turn, the increased value of children increases the assessment of the goods and services that seem necessary for successful child-rearing. In a study of the value of children, Hoffman and Thornton report that American parents want children for the love and relationship they add to the parents' lives but they also found that these parents report financial cost as a predominant disadvantage of parenthood (Hoffman, Thornton and Manis 1978).

Equally important in fertility plans are the problems of managing and arranging parental responsibilities within the financial, social, and emotional demands of the family. For the most part, these problems have remained a women's issue. Thus, number of hours in the labor force and the

availability of child care have been associated with women's fertility behavior. For example, a recent study indicates that child care restraints are related to reduced fertility among mothers of young children who want to enter the labor market or increase their hours of employment (Presser and Baldwin 1980). Since time-use studies indicate that husbands' contribution to child care does not substantially increase when their wives enter the labor market (Robinson 1977; Berk and Berk 1979), employed wives' concern with the availability of child care seems to be a realistic consideration in their fertility plans. In most cases, mothers have more intimate awareness of their children's abilities than fathers. They, rather than their husbands, are more likely to experience the conflicts between employment to secure the children's needs and the quality and quantity of time necessary for rearing additional children.

From a research perspective, women's fertility behavior raises a number of questions. Studies of fertility have usually been within the field of demography. As a result, the mapping of fertility change and its correlates has been informative of large populations. Few studies have focused on fertility decision making that occurs within the context of family events. Partially this is because panel studies are needed and panel studies require commitments of time, energy, and money. Partially this reflects the recency of fertility control on a mass basis. Not enough time has elapsed to estimate the affects. Partially, too, this indicates a neglect or bias against studies which focus on women's issues. We know that most women, and men too, have plans for eventual family size. We don't know how these plans get negotiated between husband and wife, how spacing gets determined, what family events precede fertility decisions of when or whether to have a child. Is having a baby a way of leaving a routine and monotonous job? On the other hand, is getting a job promotion or receiving tenure a career plateau that makes the possibility of having a baby more likely? Is the experience of having the first child--the medical experience, the post-partum family organization, the availability of support following childbirth--an important experience in fertility decisions? Oakley's recent study of childbirth would seem to suggest the answer is yes (Oakley 1980). Her extraordinary research has located fertility decisions within the events that surrounded women's experience prior, during, and after having a child. Since women do control their fertility, hopefully, Oakley's research will provide a new model which focuses on women's experiences as the context for those decisions.

Employment

One decision that has become as commonplace in women's lives as fertility, particularly among young women, is employment. According to results from the 1976 replication of Americans View Their Mental Health, attitudes toward both housework and paid work have shifted since 1957. Housewives are less positive about housework and an increasing number plan to seek jobs. At the same time employed married women express a much higher level of job commitment, saying they would work even if they didn't need the money. Leisure is no less important to them than it is to full-time housewives but for many it is less important than work in fulfilling their major life values (Douvan, Veroff and Kulka in press).

These attitudes have been translated into labor force participation. For mothers of schoolage children, employment has become the norm. Even among mothers of preschoolers it has become increasingly common with 42 percent in the labor market (Hoffman 1979). Housework, childcare, and husband-wife interaction have been adjusted to the job demands of wives and mothers.

Time use studies have indicated that one solution to the multiple demands incurred when women enter the labor market is to put less time into housework (Berk and Berk 1979). Since the fewer hours put into housework by employed married women is not made up by contributions from other family members, the amount of housework which gets done within the family declines absolutely. Regardless of the decline though, a large portion of employed women's week still gets devoted to it. Besides getting up earlier to do housework before work and devoting weekends to laundry, shopping, and cleaning, working wives also spend part of their wages for time-saving foods and appliances, although rarely for paid household help. Thus, employment gets translated into lowered household production--less complicated meals, less ironing, less sewing--and higher household consumption--microwave ovens, dishwashers, and meals from fast food restaurants. Although to my knowledge we have no longitudinal studies on when, how, and what part of household production gets translated into household consumption, it seems evident that when family complaints arise during the process of household reorganization that accompanies entry into the job market, women assume the responsibility and adjust their lives accordingly.

While housework may lend itself to some reorganization, child care is more of a problem. Purchasing childcare replacements is limited both by availability and by costs. Working mothers often have to put together fragile packages of childcare from two or three sources. When these arrangements break down, as they invariably do, women assume the responsibility and adjust their lives to the crisis (Presser and Baldwin 1980). Despite these and other problems that employment brings into the lives of working mothers, the children rather than the mothers have received the most research attention.

Until the last two decades, researchers often hypothesized that being an employed mother was bad for children. As a result, the Nye and Hoffman volume, Employed Mother in America (1963), was a ground-breaking contribution to the study of the effects of employment on children. Their comprehensive review of research indicated that employment of mothers was not necessarily bad for children and even had benefits. Since then a variety of studies have indicated that children with employed mothers have more androgynous sex-role attitudes and that girls, in particular, have higher achievement aspirations than those from families without employed mothers. The implications for sons are less clear (Hoffman 1979).

If the children of employed mothers are somewhat different from those of full-time mothers, then the mothers' child-rearing activities must have changed as well. Obviously employed mothers spend less time around their children, although not necessarily less time in mother-child interaction (Hoffman 1977). What about the standards and expectations they hold for themselves and their children? Little research has been devoted to this question yet it seems more than likely that such change has occurred. If so, then some problems that have been attributed to the effects of maternal employment may, in fact, be the result of a lag between non-familial institutional expectations and those of parents. In my collaborative study of divorced mothers,² the mothers frequently felt that they expected a more mature and responsible role from their children than did the children's teachers and that classroom problems sometimes resulted. These mothers may or may not be correct in their evaluations but what this emphasizes is the intersection between changing family

roles and the responsiveness of other institutions to these changes.

A more common concern of researchers has been with the impact of women's employment on marital happiness. Loxley's (1980) recent analysis of the 1976 replication of Americans View Their Mental Health suggests no significant differences in overall marital happiness by wife's employment. However, when age is considered, younger employed women report lower marital happiness than full-time housewives. The complications introduced by balancing child-care and housework with employment apparently dampens marital satisfactions. Since women shoulder most of the burdens of adjusting family to work demands, such findings are not surprising. On the other hand, among older employed women who no longer have such burdens, employment apparently increases marital happiness (Douvan, Veroff and Kulka in press). Thus, in some cases, women's employment may give rise to family tensions, in others, contribute to the overall emotional climate of the family.

Women Heading Families

Nevertheless, expanding employment opportunities have often been cited as a reason for the increase in female headed families. Today, one out of every six families is being headed by a woman. Six out of ten divorces in 1975 involve children (Norton and Glick 1979) and Bane (1976) has projected that 40 percent of all children will spend some time before they are 18 in a female headed household. Almost 70 percent of the women who currently head these families were employed during their marriages, a rate much higher than among married wives (Bane and Weiss 1980).

For a variety of reasons, though, employment opportunities alone are unlikely to explain the rise in female headed families. First, both employment rates and divorce rates are higher among younger women. Second, women may enter the work force just prior to divorce in anticipation of heading their families. Third, women may leave marriage because their husband fails to earn money forcing them to hold a job despite their desire to be full-time mothers. In the divorced mother study mentioned earlier, the latter situation was common among young mothers. Having supported the family until their first child was born because their husband would or could not do so, they realized that the birth of their child did not change his earning ability. Nor, for that matter did he take on greater responsibilities by caring for the child. As a result, being a divorced mother meant the same responsibilities they had as a married mother but with one less person to care for.

Finally, the economic and social situation of divorced mothers who head their families can hardly be described as an inducement to leave marriage. The end of the marriage for the women in the divorced mother study meant a drop of over 50 percent in family income, residential mobility, and, reliance on used or second-hand goods to meet family needs (Kohen, Brown and Feldberg 1979). While welfare and child-support stipends are a resource for female headed families, usually they are inadequate to support a family. As a result, whether the mother wants to be employed or not, she comes to depend on employment to provide for her family. Consequently, almost 90 percent of divorced mothers are employed (Bane and Weiss 1980). Yet, many women who head their families receive such low wages that they also qualify for welfare assistance. Thus, the picture of a married wife getting divorced because of the self sufficiency engendered by being employed appears to be a myth.

Certainly, employment opportunities enable women to head their families, but it is unlikely that divorces would occur if there were not changes in role expectations within families that produce tensions and conflicts. Returning once again to the divorced mother study, the analysis

² This study was done by Boston Women's Research Center and supported by a grant from Russell Sage. Co-researchers include Carol Brown, Roslyn Feldberg, Elizabeth Mawry, and Ruth Brandwein.

indicates that women choose single parenthood to escape such problems as alcoholism, violence, and continual marital disputes. If employment plays a role, it may do so by encouraging a sense of competence and self worth that enables a woman to end a marriage and head a family when either hers or her children's needs are threatened. In the divorced mother study, the women often had trouble finding jobs and worried about making ends meet, but they valued the independence and control that resulted from heading their families. These women were less likely to be self-sacrificing victims of unhappy marriages than women were several decades ago. Loneliness, fatigue, and poverty were balanced with a new sense of self identity (Kohen in review).

Summary

Each of the major areas I've addressed, delaying marriage, planning children, taking on jobs, and heading families, indicate changes in the way women are relating to marriage, to family, and to themselves. Women no longer accept or expect the traditional family roles of their parents. Marriage and motherhood continue to be part of women's lives but women are remaking these family roles without giving up their self-identities. They are developing a sense of autonomy and are no longer uncritically responding to traditional patterns and institutions.

References

- Bane, Mary Joe (1976), Here to Stay: American Families in the Twentieth Century, New York: Basic Books.
- Bane, Mary Joe and Weiss, Robert S. (1980), "Alone Together," American Demographics, 2, 11-16.
- Bardwick, Judith M. and Douvan, Elizabeth (1972), "Ambivalence: The Socialization of Women," in Vivian Gornick and Barbara Moran (Eds.), Women in Sexist Society, New York: Signet.
- Berk, Richard A. and Berk, Sara Fenstermaker (1979), Labor and Leisure at Home: Content and Organization of the Household Day, Beverly Hills: Sage.
- Cherlin, Andrew (1980), "Postponing Marriage: The Influence of Young Women's Work Expectations," Journal of Marriage and the Family, 42, 355-366.
- Douvan, Elizabeth (1978), "The Marriage Role: 1957-1976," paper presented at American Psychology Association, Toronto.
- Douvan, Elizabeth, Veroff, Joseph, and Kulka, Richard (in press), The American Experience: A Self-Portrait Over Two Decades, New York: Basic Books.
- Ehrenreich, Barbara, and English, Deidre (1978), For Her Own Good: 150 Years of the Experts' Advice to Women, New York: Anchor Press.
- Hoffman, Lois Wladis (1977), "Changes in Family Roles, Socialization, and Sex Differences," American Psychologist, 32, 644-657.
- Hoffman, Lois Wladis, Thornton, Arland, and Manis, Jean D. (1978), "The Value of Children to Parents in the United States," Population: Behavioral, Social, and Environmental Issues, 1, 91-131.
- Hoffman, Lois Wladis (1979), "Maternal Employment: 1979," American Psychologist, 34, 859-865.
- Kohen, Janet, Brown, Carol, and Feldberg, Roslyn (1979), "Divorced Mothers: Costs and Benefits of Female Family Control," in George Levinger and Oliver C. Moles (Eds.), Divorce and Separation: Context, Causes, and Consequences, New York: Basic Books.
- Kohen, Janet A. (in review), "From Wife to Family Head: Transitions in Self Identity."
- Lasch, Christopher (1979), The Culture of Narcissism: American Life in an Age of Diminishing Expectations, New York: Warner Books.
- Loxley, Anne (1980), "On the Effects of Wives' Employment on Marital Adjustment and Companionship," Journal of Marriage and the Family, 42, 337-346.
- Norton, Arthur J. and Glick, Paul D. (1979), "Marital Instability in America: Past, Present, Future," in George Levinger and Oliver C. Moles (Eds.), Divorce and Separation: Context, Causes, and Consequences, New York: Basic Books.
- Nye, F. Ivan and Hoffman, Lois Wladis (1963), The Employed Mother in America, New York: Rand McNally.
- Presser, Harriet B. and Baldwin, Wendy (1980), "Child Care as a Constant on Employment: Prevalence, Correlates, and Bearing on the Work and Fertility Nexus," American Journal of Sociology, 85, 1202-1213.
- Robinson, John P. (1977), How Americans Use Time: A Social Psychological Analysis of Everyday Behavior, New York: Praeger.
- Stolzenberg, Ross M. and Waite, Linda J. (1977), "Age, Fertility Expectations, and Plans for Employment," American Sociological Review, 42, 769-783.
- United States Bureau of the Census (1976), "Fertility of American Women: June, 1975," Current Population Reports, Series P-20, No. 301, Washington, D.C.: U.S. Government Printing Office.
- United States Bureau of the Census (1977), "Fertility of American Women: June, 1976," Current Population Reports, Series P-20, No. 308, Washington, D.C.: U.S. Government Printing Office.
- United States Bureau of the Census (1978), "Marital Status and Living Arrangements: March 1977," Current Population Reports, Series P-20, No. 323, Washington, D.C.: U.S. Government Printing Office.
- United States Bureau of the Census (1979), Statistical Abstract of the United States, 1979, Washington, D.C.: U.S. Government Printing Office.

PSYCHOLOGICAL AND SOCIOLOGICAL PERSPECTIVES ON
WOMEN'S PAID AND UNPAID WORK CHOICES

Ruth B. Ekstrom, Educational Testing Service

Abstract

Working women are not a homogeneous group. The decision to work is related to situational, experiential, psychological, societal and opportunity factors. These factors interact in different ways for working class vs. middle class women and for wives vs. women who are the sole support for themselves and their families.

Introduction

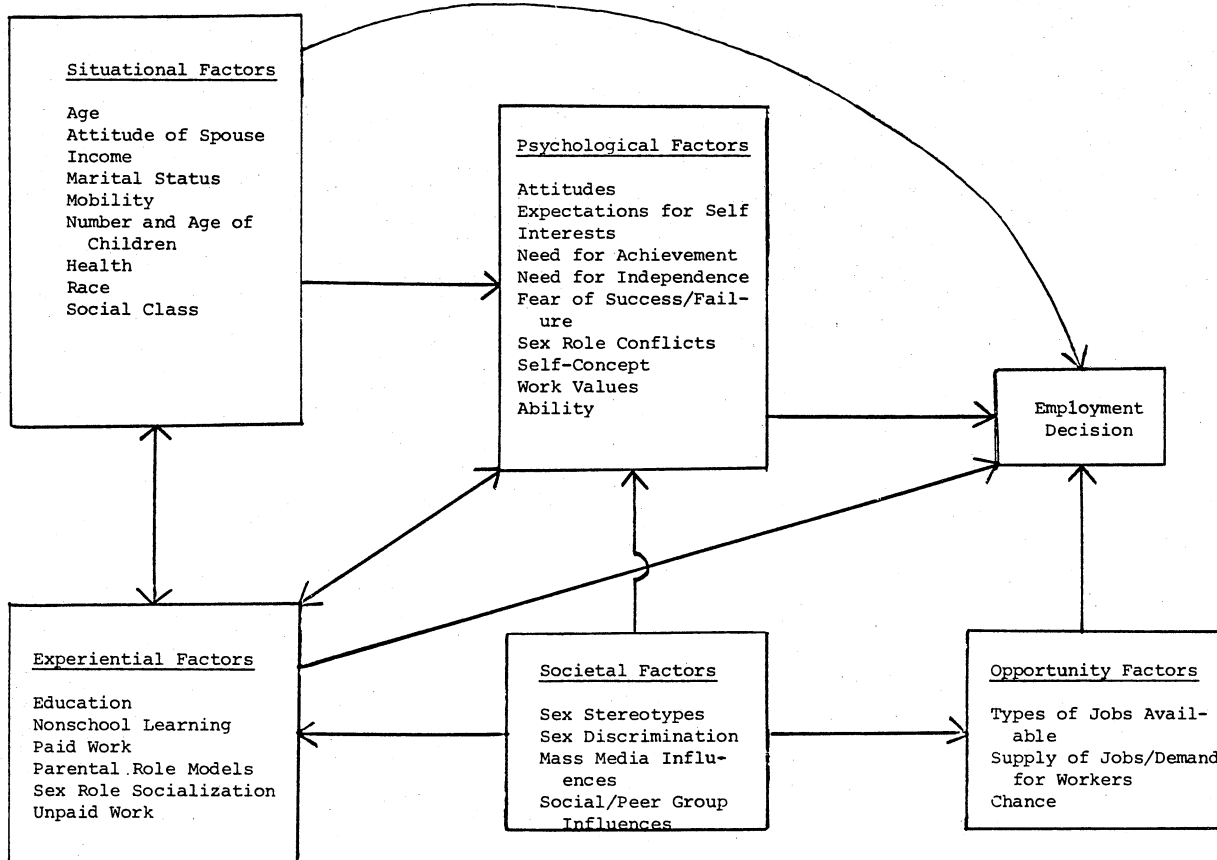
Today more women than ever before are working for pay outside of the home. Data for 1979 show that 51% of all females age 16 and over are in the labor force (the comparable figure for males is 77%). There has been a 10% increase in women's labor force participation rate in the past decade. As Hafstrom and Dunsing (1978) have pointed out, this affects not only the level of consumption but also the types of goods and services purchased. Therefore it is important to understand the factors that enter into women's employment decisions.

What Are the Factors Affecting Women's
Employment Decisions?

I hypothesize that there are five major factors which determine women's decisions about paid work (see model in Figure 1). Three of these factors are personal, specific to the woman herself. These include: (1) situational characteristics, such as age, race, social class, marital status, and number and age of children; (2) experiential characteristics, such as education, parental role models, sex role socialization, and previous paid work; and (3) psychological characteristics, such as ability, attitudes, expectations, interests, needs, and sex role conflicts. Each of these three personal factors can affect the employment decision directly. In addition, the situational and experiential factors affect the decision indirectly by modifying the psychological variables.

There are two other factors which affect the employment decision but which are external to the woman herself: (4) societal factors, such as sex discrimination, sex role stereotyping, mass media influences, and social/peer group

FIGURE 1
Model of the Factors Affecting Women's Employment Decisions



influences; and (5) opportunity factors, such as the types of jobs available, the supply of these jobs and the demand for workers to fill them, and, finally, chance--that is happening to hear about a job or being in the right place at the right time. The opportunity factor affects the employment decision directly while the societal factor affects it indirectly through the opportunity, experiential, and psychological factors.

A detailed discussion of the opportunity and societal factors is outside of the main scope of this paper but I would like to comment on them briefly. When the society at large accepts certain sex stereotypes (such as women lack mechanical aptitude) it not only limits the job opportunities available to women but it also affects women's feelings about themselves and leads them to believe that it is unfeminine or inappropriate to do certain kinds of work. This, in turn, leads women to narrow their job search to sex stereotyped occupations or, if they enter nontraditional occupations, to fear success and have sex role conflicts. Perceived support from both male and female peers is an important component of the societal factor and affects career aspirations (Angrist and Almqvist 1975; Parsons, Frieze and Ruble 1978). The availability of societal support through institutions, such as day care centers, women's groups, and other community institutions is also important in career decisions (Nye and Hoffman 1963).

It is important to point out that women's work choices are often not a single decision but a group of decisions. Many of the research studies have divided women into different experience groups or patterns of paid work. One common differentiation is between career-oriented and homemaking-oriented women; this may involve determining if the women are currently working and if they plan to work in the future. Thus, one can conceptualize women as falling into four groups: (1) those currently working who plan to continue; (2) those currently working who plan to leave the labor force; (3) those not currently working who plan to re-enter the labor force; and (4) those not currently working who plan to remain homemakers. Another common differentiation is between women who are homemakers, those in traditionally female occupations, and those in occupations that are nontraditional for women (pioneers). Still another differentiation is between women who are "career-oriented," as defined by intention to pursue a specific profession or occupation (or a group of hierarchically related occupations) and those who are "work-oriented" but not career oriented, as defined by intention to be in the labor force for much of their adult life but without a commitment to a specific occupation.

How Do Situational Characteristics of Women Relate to Employment Decisions?

Working women are not a homogeneous group. Situational characteristics of women are clearly related to differences in labor force participation.

Age cohort data on women's work is shown in Table 1. As can be seen, labor force participation increased between 1960 and 1978 in every age group except for women over age 65. The 1960 and 1970 pattern shows that women decreased their labor force participation during the child-bearing and childrearing years; this is no longer evident in the 1978 data.

One of the fastest growing groups of working women is mothers with children under the age of 14. In 1979, 41% of women with children under age 3 were in the labor force, as were 52% of women with children ages 3 through 5, and 62% of women with children ages 6 through 13.

Another rapidly increasing group in the labor force is women age 20-24. This is related to the trend toward

TABLE 1
Labor Force Participation Rates, by Age Cohorts
1960, 1970, and 1978

Age Cohort				Increase
	1960	1970	1978	1960-1970
20-24	46%	58%	69%	23%
25-34	36%	45%	63%	27%
35-44	43%	51%	62%	19%
45-54	50%	54%	58%	8%
55-64	37%	43%	41%	4%
65+	11%	10%	9%	-2%

later marriage. In 1970 approximately one-third of women in this age group were single; today approximately half are single.

Another rapidly growing group of working women, sometimes called displaced homemakers, are the middle-aged women who have been widowed or divorced. The data shows that the labor force participation rate is higher for divorced women (74%) and for single women (63%) than for married women (49%).

There are also racial differences in labor force participation. About half (51%) of all White women work for pay outside the home. This figure is slightly higher for Black women (53%) and slightly lower for Hispanic women (47%).

Most women make the decision to work for the same reason as men do--they need to provide for themselves and their families. Of all women in the work force in March 1978, 25% were self-supporting single individuals and 19% were self-supporting widows or divorced individuals; 18% were married to men whose 1977 income was below \$10,000, 15% were married to men whose income was between \$10,000 and \$14,999, and 23% were married to men whose income was \$15,000 or more.

The research on situational factors affecting women's employment reinforces these data. Miyahira (1976) found that women are most likely to work if they are single, if they have no children, and if financial need exists; they are least likely to work if they have children under age 6, if there is adequate income, and if the husband disapproves of the wife's working. Heckman (1978) found that the major variable leading a married woman to seek employment was the unemployment of her spouse; the major variable keeping women working in the home was the presence of children. Tang (1976) found that husband's positive attitude about wife's working, husband's low income, and few or no children were the major situational variables affecting women's employment.

Mobility, on first consideration, would appear to increase the probability of employment since migration can be to areas where the opportunity factors are better. This is true for single women. However, married women may lose job opportunities when they move. Data from a national longitudinal study (Parnes et al. 1975) found that migration led to the improvement of earnings for the family unit as a whole but earnings of the wife deteriorated as a result of the move.

There has been very little research on working class women for whom paid employment is more often a necessity than a choice. Studies by Walshok (1976) suggest that this group, which does not have the resources for the "elaborative" aspects of homemaking (such as, interior decorating or gourmet cooking), or a tradition of participating in unpaid work in voluntary organizations, may

find paid work a major source of personal satisfaction as well as providing additional income for family well-being. For the married middle class woman with adequate income paid employment is a choice; it is in this population that differences in the experiential and psychological factors have their greatest effect.

Stroud (1977) identified four patterns of work experience: (1) the career to homemaking pattern in which the woman leaves the work force after the birth of her children; (2) an unstable, intermittent pattern of labor force participation; (3) double-track women with high family orientation but who are in the labor force; and (4) work-committed women, with high career involvement who may or may not have families. She found that socioeconomic status significantly predicted career decisions in all groups. Family and personal characteristics were significantly related to work commitment for women who had attended college but there was no relationship between these variables and work commitment for women who had not attended college. College educated women with double-track or intermittent careers were unhappy in their marriages and in their role as homemakers.

How Do Experiential Differences Affect Women's Employment Decisions?

Education and other forms of job training have a large influence on employment decisions. As Mott and his colleagues (1978) have shown in their longitudinal study of the educational and labor market experience of young women that "women capable of earning higher salaries are more likely to maintain their work ties. Conversely, women with the least economic bargaining power--the less educated and the less skilled--are least likely to find and maintain employment at a reasonable wage level." The higher a woman's educational level and the greater the amount of occupational training, the more likely the woman is to be committed to a career rather than being home or work oriented.

The nature of women's unpaid work experience, especially their experience as homemakers, may lead them to seek paid employment. Early studies of homemakers (Weiss and Samuelson 1958) found them happy and content. More recently Ferree (1976) found that full time homemakers, compared to employed working-class women, are more dissatisfied with the way they spend their time (26% vs. 14%), feel that they have not had a fair opportunity in life (47% vs. 38%), and want their daughters to be different from them (35% vs. 22%). Ferree concludes that housework does not lead to a sense of competence, social connectedness, or self-determination to the same extent as does paid employment. Iglehart (1979) found a significant decrease in positive feelings about housework between 1957 and 1976. She also found a significant increase in the proportion of full-time housewives who plan to work at some time in the future. A study by Segre (1978) suggests that the happiest homemakers may be those who perceive that they have a choice between remaining at home and entering the work force; she also found that women who were part-time returnees to the work force were happier than either homemakers or full-time workers because they felt they had exercised a choice of position.

Several studies have indicated that there is a "chicken and egg" problem in studying the relationship between work experience and psychological variables, such as attitudes. For example, Macke, Hudis and Larrick (1978) found not only that attitudes affect the decision to work but also that work experience has a strong effect on attitudes. These authors believe that early sex role attitudes influence initial employment decisions and that the nature of this paid work experience affects later sex role attitudes and work decisions.

The early sex role socialization experience of young women affects their adult work decisions. Nickerson and Pitochelli (1978) describe how feminine sex role socialization often involves learning to be passive and dependent, what Lipman-Blumen (1972) calls "learned dependency." In later life this can lead to reduced feelings of control and reduced self-confidence when entering the labor force.

It has been generally acknowledged that parental role models, especially the mother, play a major part in the development of career aspirations in young women (Angrist and Almquist 1975, Nye and Hoffman 1963, O'Leary 1974). In general, daughters whose mothers worked during their childhood have higher career aspirations than daughters whose mothers were not in the labor force. More recent analyses have suggested that this relationship is somewhat more complex. Data from a national longitudinal study of working women (Mott 1978) shows that women who grew up in two parent families were more likely to work if they had a mother who worked during their childhood. This finding did not hold for women who grew up in single parent households. This suggests that when a mother makes the choice to work, she provides a positive model for her daughter but when the work decision is forced on the mother by circumstances, such as poverty or widowhood, it may be perceived by the daughter as a negative model. Parsons, Frieze and Ruble (1978) found that both mothers with careers and homemakers who were dissatisfied with their lives had daughters with high career aspirations but mothers who were satisfied with their role as housewives had daughters with lower aspirations.

How Do Psychological Factors Affect Women's Employment Decisions?

Much of the literature on the psychological factors affecting women's careers is based on research with college educated women. It seems likely that it is in this group, with a high investment in education and high income husbands, that ambivalence about paid work may be the greatest. However, it is unwise and unrealistic to over-generalize from college educated women to women in general.

Women often express conflicting feelings about their wish to enter the labor force and the fears they hold about the consequences their employment may have in their marriage and for their children. As the number of roles increases (from worker to wife and worker to mother and wife and worker), a woman feels increasing role conflicts. Home pressures appear to be the most important contributors to role conflicts (Hall and Gordon 1973). The absence of a supportive male, as is the case with displaced homemakers and other women who are self-supporting, increases the perception of role conflict and is a factor limiting interest in higher-level careers and inhibiting the movement of working women into positions of greater responsibility (Jerdee and Rosen 1976). O'Leary (1974) has suggested that at the point in the life cycle where women face multiple responsibilities, employing organizations should focus on helping women cope with role conflicts rather than on how to motivate them to achieve in the occupation.

Three psychological factors that inhibit women's occupational aspirations have been identified by DiSabatino (1976). These are sex role conflicts, poor self-esteem, and fear of failure.

The psychological literature shows that, in general, females have a poorer self-concept than males and that females of lower socioeconomic status may have poorer self-concept than middle class females. Thus, women are likely to have more doubts than men about their ability to compete in the work force successfully.

Mendelsohn (1979) compared the personality and characteristics of career-oriented and homemaking-oriented college

educated women. He separated these women into four groups: (1) those currently working who plan to continue working until retirement; (2) those currently working who plan to leave the labor force; (3) those not currently working who plan to re-enter the labor force; and (4) those not currently working who plan to remain homemakers. For all groups except the nonworking women who plan to remain at home, the opportunity to use skills was rated as the most important reason for working; this difference reached statistical significance between women now working who plan to continue and women not now working who do not plan to work. This finding and my own research with re-entry women (Ekstrom, Beier, Davis and Gruenberg, In press) shows that although homemakers have many skills they view themselves as having fewer skills than women in the labor force. The other significant difference between Mendelsohn's four groups was that women now working who plan to continue rate need for money more highly than women who are not now working but who plan to return to work.

Fear of success is a concept originated by Horner (1972). She has shown that women perceive a conflict between societal expectations for women to be noncompetitive and dependent and the competent, achievement-oriented behavior needed for success in many occupations. These findings have not been consistently replicated. Some researchers think that fear of success comes into play only when women are employed in occupations that are predominantly male.

Women choosing nontraditional careers tend to be more independent, more committed to their careers, and to hold more nontraditional sex role attitudes than are women who choose traditional occupations (Farmer 1979).

Walshok (1976) has identified the work values important to women in nontraditional blue collar occupations. These women are, first of all, concerned with the economic rewards of work, not only as a means of obtaining money needed for support and as fair compensation for contributions made, but also as a means of becoming self-sufficient and independent and of achieving an enhanced sense of self-worth. They also view paid work as a way of escaping the boredom of the home, of obtaining outside communication and friendship, of obtaining recognition and a sense of achievement, and of being challenged by the opportunity to learn new things and prove oneself capable.

The women in the Walshok study were, as a group, quite liberal in their sex role attitudes. For example, 93% felt that it was all right for women to compete with men for jobs; 93% also felt that most jobs can be done as well by women as by men. They showed little problem with sex role conflicts; 86% felt that a woman could have a career and still keep her femininity. These women also felt that husbands should take an active role in the family; 93% expected husbands to help in the kitchen and in doing other housework; 87% expected husbands to share actively in raising children. In contrast to this group, women who are homemakers have more traditional attitudes about sex roles (Segre, 1978).

A number of researchers have used Eyde's Work Values Scale to identify differences in the values held by working women and homemakers. Wolkon (1972) found that college educated women who were homemakers and who were in traditional occupations ranked independence as their highest value but women who were pioneering in nontraditional occupations ranked economic rewards as their highest value. In general, there were larger value differences between the traditional and pioneer working women than between the traditional working women and homemakers.

Summary

In summary, the vast majority of women make the decision

about whether or not to enter the work force on the basis of need and other situational factors, such as marital status, presence of children, and husband's attitudes about women and work. Opportunity factors not only affect the decision to work but also play a large part in the job choice of women with limited education and/or job training. College educated women appear to have higher expectations for their careers but feel greater sex role conflicts in their decision to enter the labor force. Women who elect the unpaid work of homemaker may do so because of traditional sex role attitudes, because they perceive high conflict between femininity and success, or because they feel they lack the skills needed in paid work. Married women in paid work, whether by choice or necessity, report problems arising out of their multiple roles as worker, wife and mother.

References

- Angrist, S. S. and Almquist, E. M. (1975), Careers and Contingencies, New York: Dunellen.
- DiSabatino, Marie (1976), "Psychological Factors Inhibiting Women's Occupational Aspirations and Vocational Choices: Implications for Counseling," Vocational Guidance Quarterly, 25, 43-48.
- Ekstrom, Ruth B., Beier, Juliet J., Davis, E. Leta, and Gruenberg, C. Brooke (In press), "The Career and Educational Counseling Implications of Women's Life Experience Learning," Personnel and Guidance Journal.
- Farmer, Helen (1979), Counseling Programs and Services for Women in Non-Traditional Occupations, Columbus, Ohio: Ohio State University.
- Ferree, Myra M. (1976), "Working-Class Jobs: Housework and Paid Work as Sources of Satisfaction," Social Problems, 23, 431-441.
- Hafstrom, Jeanne L. and Dunsing, Marilyn M. (1978), "Socioeconomic and Social-Psychological Influences on Reasons Wives Work," Journal of Consumer Research, 5, 169-175.
- Hall, Douglas T. and Gordon, Francine E. (1973), "The Career Choices of Married Women: Effects on Conflict, Role Behavior, and Satisfaction," Journal of Applied Psychology, 58, 42-48.
- Heckman, James J. (1978), "A Partial Study of Recent Research on the Labor Supply of Women," American Economic Review, 68, 200-207.
- Horner, Matina S. (1972), "Toward an Understanding of Achievement-Related Conflicts in Women," Journal of Social Issues, 28, 157-173.
- Iglehart, Alfreda P. (1979), Married Women and Work, Lexington, Massachusetts: Lexington Books.
- Jerde, Thomas H. and Rosen, Benson (1976), "Factors Influencing the Career Commitment of Women," Paper presented at the Symposium "Women Managers, How Are They Different?" American Psychological Association, Washington, D.C.
- Lipman-Blumen, Jean (1972), "How Ideology Shapes Women's Lives," Scientific American, 226, 34-42.
- Macke, Anne S., Hudis, Paula M. and Larrick, Don (1978), "Sex-Role Attitudes and Employment Among Women: Dynamic Models of Continuity and Change," In Women's Changing Roles at Home and on the Job, Washington, D.C.: National Commission for Manpower Policy.
- Mendelsohn, Robert A. A. (1979), The Personality and Attitudinal Characteristics of Career-Oriented and Home-

- making-Oriented Women, Detroit: Wayne State University.
- Miyahira, Sarah D. (1976), "College Women's Career Orientation as Related to Work Values and Background Factors," Unpublished doctoral dissertation, Ohio State University.
- Mott, Frank L. and others (1978), Years for Decision: A Longitudinal Study of the Educational and Labor Market Experience of Young Women, Washington, D.C.: U.S. Department of Labor.
- Nickerson, Eileen T. and Pitochelli, Elaine T. (1978), "Learned Helplessness and Depression in Married Women: Marriage as a Depressing Life Style for Women," Paper presented at the Symposium "Wife, Mother, Employee: Role Combination and Conflict," Washington, D.C.: Eastern Psychological Association.
- Nye, F. Ivan and Hoffman, Lois W. (1963), The Employed Mother in America, Chicago: Rand McNally, 1963.
- O'Leary, Virginia E. (1974), "Some Attitudinal Barriers to Occupational Aspirations in Women," Psychological Bulletin, 81, 809-826.
- Parnes, Herbert S. and others (1975), Dual Careers: A Longitudinal Analysis of the Labor Market Experience of Women, Columbus, Ohio: Ohio State University.
- Parsons, Jacquelynne E., Frieze, Irene H., and Ruble, Diane N. (1978), "Intrapsychic Factors Influencing Career Aspirations in College Women," Sex Roles, 4, 337-347.
- Segre, Jessica (1978), "Self-Concept and Depression: Mothers Returning to Work or Remaining at Home," Paper presented at the Symposium "Wife, Mother, Employee: Role Combination and Conflict," Washington, D.C.: Eastern Psychological Association.
- Stroud, Janice G. (1977), "Careers of Middle-Aged Women in Work and Family: Personal and Social Concomitants and Antecedents," Unpublished doctoral dissertation, University of California at Berkeley.
- Tang, Dena B. E. (1976), "Labor Force Participation of College-Educated American Women: The influence of Early Work Commitment and Later Situational Factors," Unpublished doctoral dissertation, Purdue University.
- Walshok, Mary L. (1976), "Occupational Values and Family Roles: Women in Blue-Collar and Service Occupations," The Urban and Social Change Review, 11, 12-20.
- Weiss, Robert S. and Samuelson, Nancy M. (1958), "Social Roles of American Women: Their Contribution to a Sense of Usefulness and Importance," Marriage and Family Living, 4, 358-366.
- Wolkon, Kenneth A. (1972), "Pioneer vs. Traditional: Two District Vocational Patterns of College Alumnae," Journal of Vocational Behavior, 2, 275-282.

WOMEN IN PAID WORK: SOME CONSEQUENCES AND QUESTIONS FOR
FAMILY INCOME AND EXPENDITURES

Hila Kahne, Wheaton College

Abstract

The paper examines consumer preference and expenditures in relation to women's market work and income. Working women's participation rates, family forms, and income patterns are analyzed. Identification of research needs emphasize consumer expenditure studies at moment of time and studies of market responsiveness to changes over time in working women's lives.

Now that the social revolution in women's roles has become a settled way of life, analysis has begun of the economic consequences of these changes both for individual lives and for the economic life of the society. This paper looks at one aspect of these consequences which result from changes in numbers, characteristics and income of women engaged in paid work. It is concerned with how consumer preferences and expenditures are altered by women's labor force participation and by the income earned.

I will begin my discussion by reviewing the changing trends in women's work force participation and their family characteristics occurring over the past decade. I will then present some data on income patterns that have resulted from the influx of women into paid work. The section will conclude with brief comments on expected future trends related to working women, which have consequences for consumer needs and demands. The final section discusses three broad questions which require research attention if we are to understand more fully the relationship between working women and consumer behavior.

Trends and Characteristics of Women in the Labor Force

World War II and its aftermath speeded up a process of women's entry into paid work earlier given impetus by the dramatic population shift from farm to city, a shift which made it feasible for women to take paid jobs in place of unpaid labor on the family farm (Waldman 1976). Between 1950 and 1965 the proportion of women in the civilian labor force increased from 29.6% to 35.2%. More than one-half of the net addition of 7.8 million women workers were accounted for by mature middle-aged women, most of whose children, if they had any, were grown.

Since 1965 the social revolution in women's paid work has moved to a different age group. In this period, it is young women, particularly young married women and mothers of pre-school children who have experienced the most marked shifts of entry into paid work. During the decade of the 1970s, 58.8% of the 20.2 million person increase in the civilian labor force was due to the influx of women with the result that by 1979 43.4 million women, more than double the 18.4 million in 1950, were working or looking for work. They represented 42.1% of the paid civilian labor force of 103 million persons (29.5% of the labor force of 62.2 million in 1950). Fifty-one percent of all women 16 years and over were in the labor force. Between 1970 and 1979 women aged 25-34 accounted for nearly one-half of this increase in the number of female workers, despite the fact that almost three-fourths were married with a husband present and had young children at home (Norwood and Waldman 1979). By the end of 1979, almost two-thirds of all women aged 25-34 and 54% of mothers in this age group were working or looking for work. Overall, between 1950 and 1978, although the labor force participation rate

of women increased by 50%, that of married women, husband present doubled, reaching 47.6% in 1978, and that of married women, husband present with children under age 6 increased by more than three times, being 41.6% in 1978. Just as for mature women in the earlier period, the rate changes were mostly due to continued rising participation rates of women with similar characteristics, not to demographic changes in the female population itself. Paid work discontinuities, associated with marriage and child rearing in the past, are today strikingly diminished as important causes of women's life cycle work pattern.

One of the most important consequences of these changes in women's labor force participation rates has been the increased prevalence of the dual-worker family, now 60% of all married couple families (36% in 1950). The group, now numbering 28.4 million persons, has increased by 3 million since 1970, due almost entirely to the increase in the number of working wives (Hayghe 1976).

But other important changes in family forms are also taking place. Since 1970, the increase in the number of never-married women has risen significantly, an outgrowth of the coming of age of the "baby boom" generation and of the greater prevalence of postponed marriages. The rise in their labor force participation rates between 1970 and 1978 was greater than that for women in any other marital status, reaching 60.5% in March 1978, with the result that single women now comprise about one-fourth of the working woman population, a proportion not reached since 1957 (Johnson 1979). At the same time, the number of divorced women in the labor force has doubled between March 1970 and March 1978 as a result of a dramatic rise in the divorce rate, from 2.5 per thousand population in 1965 to 5.2 per thousand population in May 1978. The labor force participation rate of divorced women is higher than that for women in any other marital status (Johnson 1979).

Largely due to the rising incidence of separation and divorce (and also somewhat due to an increased number of never-married women who head families) the past eight years has also seen a significant rise in the proportion of families maintained by women. One in seven families in 1979 fits this description; their numbers are up 48% from 1970. Their 1979 labor force participation rate is 59.3%. Their earnings are extremely low.

How can we relate these trends in family characteristics of women in the labor force to our interest in consumer behavior? We have shown that although the numbers and proportion of working women have increased markedly, the increase is different for different marital groups. In the 1970s, increases in labor force participation are being felt most strongly among young married women, among women with young children, and among single women. At the same time there are marked shifts in the relative prevalence of family forms. Although the norm among women workers continues to be that of married status, there has been a sharp increase in the number of never-married women, a doubling of numbers of women in divorced status and a rapid rise of families headed by women, about 60% of whom are in paid work. By March 1979 30.1 million children had mothers in the labor force, almost 20% more than in 1970. Family lifestyles and incomes interact in their effects on consumer behavior. We must be alert not only to this contemporary experience with evolving family forms and labor force participation, but also to the rewards of paid work for women.

Together they impact on available income and on the kinds of consumer demands for its use.

Income of Women in the Labor Force

The income derived from women's paid work reflects a number of influences including the economic reward structure for particular occupations, continuity of work histories, training and education, the degree of part-time and part-year work, and at times still, the extent of discrimination. Because of the wide variation in influence of these factors among working women, a variation which is not diminishing over time, earnings distribution among women shows considerable range, larger than that experienced by men. Large earnings differences between women and men workers also continue, even when both work full time, year round (Henle and Ryscavage 1980). For many women and their families, an adequate standard of living is still contingent upon the existence of another income in the family.

Women's employment, as we know, is concentrated in a relatively few occupations, many of which have low status, little career progression, and limited economic reward. It is true that in recent years the proportion of women in some male professional fields has been growing -- e.g. accountancy, medicine, law. Professional, managerial, and sales jobs have expanded for women in a number of industries -- in the Federal government, in transportation, public utilities, finance, and insurance (Davis 1980). A recent survey of over 250 corporations documented considerable progress in the kinds of training and promotion opportunities offered to women at higher managerial levels in the private sector, particularly in female intensive industries (Schaeffer 1978). There is some indication that, spurred on by economic need and supported by affirmative action policies, blue collar women are finding jobs in construction trades and as miners, although in each division, employment is still under 10% (Davis 1980). The employment of black women as private household workers has dropped dramatically and an increasing proportion have become employed in clerical work. Among white women, the proportion in white collar positions have risen somewhat, while that of operatives have declined (Young 1980). Yet, despite such advances in women's proportionate representation in higher paying and more career-oriented occupations, the numbers of women so involved are still not large. Women have yet to reflect in their employment the full range of occupations in which men engage and for the most part have yet to achieve vertical career mobility once entry on a career track has been gained (Reubens and Reubens 1978). Women's median full-time earnings continue to be about 60% those of men, a relationship that has prevailed for the past 20 years (Norwood and Waldman 1979). The data on occupational distribution and earnings explains why income in families headed by women, even when working, continues to be so low. Occupational disadvantage reinforced by earnings disadvantage related to lack of training or experience, or opportunities, or discontinuities, or discrimination, affect both aggregate and specific consumer expenditures that would otherwise be generated by earnings from the work that women do.

Working women who are family householders and those in dual-worker families provide an interesting contrast of the characteristics that prevail among working women. Families with female householders numbered 8.2 million in 1978, and constituted 14.4% of all families. About 60% were in the labor force. They are a group who have grown sharply in relative importance as the divorce rate has spiraled upward (an increase of 47.7% in the past eight years). They are more likely to be black than are other families. They are more likely to have dependent children under age 18. Their labor force participation rate (58.9%) is higher than that of married women living with husbands (47.6%). Among women who head families, labor force participation is highest for divorced women (77.6%) and lowest for widows (39.4%). Their level of education

is considerably less than that of married women, husband present, and indeed of working men, and this affects their occupational work and income. Except for a small movement into non-traditional occupations, the overwhelming majority of female family householders work in traditionally female occupations. Second workers in families headed by women are much less common than in families headed by men (31% and 56% in first quarter, 1977). In 1979, 53% of families where the husband was unemployed had at least one employed family member; only 17% of families maintained by an unemployed female householder, had an employed family member. Responsibility for family maintenance rests heavily on the shoulders of female family householders. What can we say about income of these families?

Because of low average earnings, lack of steady employment, and absence of supplementary income from other earnings or asset sources, median income of these families is low. In 1977, it was \$7,700, or only 44% of the \$17,720 total for husband/wife families. Women family heads earned about \$6,370 or \$8,850 if they worked at full-time jobs all year. (Men earned an average of \$13,600 or \$15,600 for full-time-year-round jobs.) Consistently during this decade, about one-third of all families headed by women had money income below the poverty level, compared with about 5% of husband/wife families. Consumer behavior for individuals in this family form is strongly influenced by the absence for many of even a basic minimum income.

Dual-worker families provide contrast in both their composition and income level. They are about 60% of all married-couple families; in one-half of such families, both the husband and wife have jobs. Such families tend to be not only with higher income, but also younger, more mobile, and of greater educational achievement than single-earner families (Rawlings 1978). Despite the fact that only about 17% of husband/wife dual-worker couples both hold full-time jobs (1976 figure), the pay of working wives contributes about one-fourth of the family income, a figure that has remained constant over the past two decades (Johnson 1979). Wives who worked year-round full time in 1977 (41.7% of working women in dual-worker families) had a median income of \$8,600 and contributed 38% to the family income. Median income for families with the husband as the only earner was \$15,796 in that year compared with \$20,723 when both husband and wife worked (Johnson 1979). In general, wives' earnings were lower than husbands' earnings, although in one of every three married couples in which the wife had earnings, they were in the same or a higher class interval than that of her husband (Rawlings 1978). The contribution of the second family earner to the family standard of living is significant. In 1954, 57% of all families with incomes of \$25,000 or more had two incomes; in 1974, the year of latest available data, in 80% of households with the same high real level of income there were two earners. In recent years a rising proportion of these higher income level households have come into being because of women's entry into paid work.

In terms of societal income distribution, the implications of the two-paycheck families are more complicated, for they cause a marked division in income between families with a single earner and those with two or more earners. In 1977, as in 1970, the income of husband/wife families when only the husband was an earner was about three-fourths that of families when both husband and wife were earners (Johnson 1979). Moreover, 60% of working wives who entered the labor force in the 1960s and 1970s came from families with husband's income in the upper middle and upper ranges (Ryscavage 1979). If this trend continues, the gap between above-average income families and below-average income families could widen further, increasing the inequality of income distribution and causing problems for maintenance of economic well being for those families still in a single earner mode.

There is no reason to believe that present trends will be reversed in the near future and in fact they may be reinforced by additional factors we have not discussed. Family forms may develop further variety -- we did not refer to forms of communal families or those in non-married family status, both two sex and single sex partners. Nor have we discussed the greater variability than in the past of movement into and out of family forms over the life cycle, a seemingly permanent pattern that also has a bearing on consumer behavior. Similarly, not only is the rate of growth of women's labor force participation expected to continue to increase, though perhaps at a slower rate, but the percent of women holding two jobs may continue to grow -- 2.3% of women workers in 1969, 3.5% in 1979. Estimates of labor force participation of women in 1990 range between 53.8%, if the number of children born to each woman turns upward again, to 60.4%, if participation continues to rise at its present rate of growth. It is expected that the proportion of the labor force accounted for by women may reach 45-46% of the total by 1990, increasing the potential influence of working women on consumer behavior. Married women are expected to account for a larger proportion of the female labor force by the end of the 1980s than today (due both to an older population and to continued increasing participation rates) and a sharp rise is expected to continue in the numbers of working mothers with pre-school children. By the end of the 1980s, two-thirds of all married women under fifty-five may be working, including over one-half of all mothers of young children (Smith 1979).

The increase in the number of households and the greater prevalence of the smaller household, partly related to the decisions of young working women to postpone marriage, will further influence consumption patterns. During the past decade, the number of households increased 26%, three times faster than the population growth. One-person households increased 40% between 1970 and 1976 compared with an 11% increase in multi-person households. There is some uncertainty about the rate of continued growth during the 1980s. Rising unemployment and inflation are beginning to cause elderly persons to move in with children and young married couples to reside with one set of parents. Estimates of household growth in the 1980s range from 16% to about 23%, depending on the assumptions made about life-style trends of marriage, divorce, fertility, and residence preference. The pattern of future household formation will further strengthen the influence which working women's decisions have on the consumption of goods and services.

Working Women: Implications for Consumer Behavior

Having described some of the recent changes in the characteristics and income of women who work for pay, we come to the critical question of what difference changes in women's labor force participation and income will make in consumer needs and expenditures? By asking the question in this way, we seek to focus particularly on those changes related either directly (through earnings, e.g.) or indirectly (through changes in living styles such as smaller households or in consumer needs associated with larger numbers of working women of middle age) to the numbers and characteristics of women in paid work.

I suggest three general areas which need research attention if we are to understand more fully the influence of the particular factor of working women on consumer behavior. Economic research is not yet extensive in this area. But the studies that exist not only provide important information about the relationship between income of working women and expenditures, but suggest a context for thinking about the issues raised and about next questions for investigation. As we review the evidence, we must keep in mind not only the changes as they affect individual family units but also the relative importance of those units in the societal structure. We must remember, too, that consumer needs can be translated into

effective demand either by spending of personal income or income generated by social programs.

1. What do we know and what do we need to know about the new directions in consumer needs and choices that reflect changes in the numbers and characteristics of working women and the family forms in which they live?

With the recent dramatic changes in family forms and in women's labor force participation it is important that there be available a descriptive comparative picture of the expenditure patterns of each of these family groups which are of continuing growing importance in the society. What are the budgetary allocations -- and needs -- of families with female householders? What are the needs and expenditures of single working woman households which have grown so rapidly in recent years? What is the consumer expenditure experience -- and unmet needs -- of families with working mothers, whether male- or female-headed? On the other hand, what can we learn if we hold family form constant and compare consumer behavior of working-woman families by age (life cycle phase), by race, by occupation or blue collar/white collar status?

Some of this information is provided in a recent paper by Clair Vickery (1979). In her discussion, which focuses particularly on husband/wife families with and without a working wife, she addresses two issues. First, how do family expenditure patterns differ if the wife is a homemaker from those of a family with the wife in the labor force? In the absence of a wife's paycheck, Vickery states, expenditures on clothing, transportation, recreation, and retirement are as much as 50% lower, and expenditures on the basics of food and shelter are slightly lower as are savings and financial assets. Except for work-related expenditures, these differences reflect the different expenditure patterns of a 30% difference in income. When a wife enters the labor market, there is both some replacement of a housewife's time with market produced goods and services (e.g. restaurant meals, laundry and dry cleaning services) and some additional work-related expenses.

The second issue studied by Vickery in her own research, relates to these work-related expenses. Using data from the 1972 Consumer Expenditure Survey, she analyzed spending patterns of middle-income families with similar household characteristics (number and age group of children, life cycle stage of family head) and assets, and with the same income, to see the differences in spending for families with and without a working wife. Her analysis revealed several items of working-wife expenditures not incurred by nonworking wives which reduced the equivalence of full income between the two kinds of families who had the same money income. Working wives spent substantially more for transportation, for social security and pensions, and for clothing than did families with non-working wives. These work-related expenses (about 14% of the wife's earnings before taxes), reduced the value of the family full income and may account for the lower estimated savings and market value of financial assets, and lower rates of home ownership found for a working wife than a homemaker family. Other expenditures did not vary much between the two kinds of families except that working-wife families seemed partly to balance the work-related expenditures by allocating a lesser proportion of income to shelter. Interestingly, child care expenses did not show up as a significant quantitative expenditure for the average working wife. Whenever possible, family members appear to care for children within the home.

Vickery has provided us with a beginning understanding of budgetary patterns that result from having a working wife. But as we have seen, not all working women are part of two-earner families. Nor, even if they are at the moment, are they necessarily continuously so throughout their lives. We need to study this. We also need studies of

budgetary allocations of working women -- white and minority -- in other family forms, reflecting a range of socioeconomic levels. We need to know more of the trade-offs involved in making consumer choices and of the unmet needs not registered by income allocations.

2. To what extent are consumer choices -- either particular purchases or amounts spent -- influenced by the fact of women's participation in paid work or by the amount of additional family income generated? Is there a different decision-making process about consumption -- and outcome -- when women are the income earners?

Theoretical conceptualization and empirical research on this question, again focuses particularly on wives' economic behavior. Economists have suggested a number of alternative hypotheses to explain purchases made with wives' earnings: Mincer suggests (1960) that working wives spend more on durable goods (a form of saving) than nonworking wives because their earnings have a transitory component which is likely to be saved; Galbraith states (1973) that they spend less on durable goods than nonworking wives because of less time available for consumer administration; Drucker claims (1976) that they use income for "extraordinary purchases" because their earnings are transitory (reported in Strober and Weinberg 1977). Strober and Weinberg (1977), using data from the 1968 Michigan Survey Research Center 1967-70 panel Survey of Consumer Finances, have compared purchases of consumer durables of TVs, dishwashers, washers, dryers, stoves and refrigerators, for a sample of married couple non-farm families under age 65 with and without working wives, holding income, net assets, life cycle stage and other factors constant. They found that size of family income and whether the family had moved recently, but not the wife's labor force behavior per se, were important factors in determining the purchase of specific goods and services. Moreover, income and assets, but not wife's labor force behavior were significant in determining how much was spent for such items. The findings do not imply that wives' earnings are unimportant; rather, they stress the fact that it is the income and not its source that is the determining factor in purchases. If family income, assets, and several other variables are held constant, there is no significant difference between working and nonworking wife families in family expenditures on durable goods. A more recent study by the same authors (1980) using 1977 data from a consumer marketing panel of husband/wife families designed by Needham, Harper and Steers, extends their investigation to consumer purchases of microwave ovens and freezers. Again, holding income and life cycle stage constant, they found no greater purchase of these "time saver" consumer durables among working wife than nonworking wife families, reaffirming their hypothesis that wife's income, like husband's income, is treated as a permanent component of family income. Even the time-saving quality of the items did not alter the expenditure preference for time-pressured working wives.

The findings, not inconsistent with those of Vickery, suggest the complexity as well as the need for currency in study of these budgetary issues. We need now to extend analyses to categories of nondurable consumer goods and services (where there is some evidence of higher expenditure for working-wife families) to identify these consuming patterns. We need to study families with working women in different socioeconomic groups and family settings where tastes and pressures of time may have different effects on the outcome. Experience of families with working women in different life cycle phases, of working women who experience changing family forms over time, and of working women in a non-continuous paid work pattern within one-family form must each be examined to test the consistency of findings in different settings.

3. How do economic conditions of unemployment and inflation affect consumer needs and expenditures of working women and their families? Are studies of consumer behavior based on 1972 experience outmoded by these recent conditions? Will the affluent buying behavior wrought by the two-earner family give way to a standard of living crisis when unemployment strikes one earning member? Is the seeming work flexibility, made possible for the two-earner family, over-ridden by the rigid financial requirements established by a standard of living that requires both incomes? How do female householders with lower income and fewer second earners cope with needs and expenditures when unemployment or inflation strikes the household? For these, and for other working-women groups, how are expenditures adjusted when there are uneven and significant price rises in some essential categories such as energy?

In this area I have no research answers; only questions. They are critical ones for us to begin to address.

We live in a world where women now share with men responsibility for the production of goods and services of the economy and where their incomes comprise permanent and critical components of family income. Working women live in a variety of family forms. Their standard of living ranges from one of poverty, particularly prevalent among elderly, widowed and female householders, to one of affluence, often existing because both husbands and wives work. Their consumer behavior is also influenced by their roles, values, and tastes, which reflect both societal movement over time and individual change over the life course.

Research and marketing analysis has only begun to inform our understanding of this aspect of women's economic behavior. We need to study further consumer expenditures in relation to age, race, life cycle phase, socioeconomic situation of working women so as to learn more about the potential market for such items as luxury condominiums for two-worker small families, microwave ovens for working wife/mothers, sophisticated maternity clothes for the woman professional, or a new range of services for the traveling woman business executive. We need more expenditure studies for working women in all family forms. We need to know how economic conditions impact on their choices.

But in addition to a concern with effective consumer demand at a point in time, we need a better grasp of the responsiveness of the market to the changes taking place over time in the lives of working women. These have both to do with changing social mores and changing social roles. They sometimes take place because of movement among family forms in which working women live over the course of their lives and sometimes have to do with changes in women's paid work status over time that occur within a single family form.

Needs may be covered by effective private demand. They may be unmet needs that require social expenditures. Both kinds should be analyzed in research inquiries.

If research can extend and enhance our understanding of these aspects of consumer behavior, we will perhaps get a more complete grasp of the totality of consumer needs in relation to the flow of individual lives. My hope is that all these aspects, and not only individual product preference, can be addressed by our studies, so that we will be better able to assure that improvement in the quality of our lives proceeds apace with increased sales and marketing returns.

References¹

- Davis, Howard (1980), "Employment Gains of Women By Industry, 1968-78," Monthly Labor Review, 6, 3-9.
- Hayghe, Howard (1976), "Families and the Rise of Working Wives - An Overview," Monthly Labor Review, 5, 12-19.
- Henle, Peter and Rycsavage, Paul (1980), "The Distribution of Earned Income Among Men and Women, 1958-77," Monthly Labor Review, 4, 3-9.
- Johnson, Beverly L. (1979), "Marital and Family Characteristics of Workers, 1979-78," Monthly Labor Review, 4, 49-52.
- Norwood, Janet L. and Waldman, Elizabeth (1979), "Women in the Labor Force: Some New Data Series," U.S. Dept. of Labor Bureau of Labor Statistics Report 575.(1980).
- Rawlings, Stephen (1978), "Perspectives on American Husbands and Wives," U.S. Dept. of Commerce, Special Studies Series P-23, No.77.
- Reubens, B.G. and Reubens, E.P. (1977), "Women Workers, Nontraditional Occupations and Full Employment," in U.S. Joint Economic Committee, American Women Workers in a Full Employment Economy, 103-126.
- Rycsavage, Paul (1979), "More Working Wives Have Husbands with 'Above-Average' Incomes," Monthly Labor Review, 6, 40-42.
- Schaeffer, Ruth G. (1978), "Improving Job Opportunities for Women -- From A U.S. Corporate Perspective, 1970-1975," The Conference Board.
- Smith, Ralph E., Editor (1979), The Subtle Revolution: Women at Work, Washington, D.C.: The Urban Institute.
- Strober, Myra H. and Weinberg, Charles B. (1980), "Strategies Used by Working and Nonworking Wives to Reduce Time Pressures," Journal of Consumer Research, 6, 338-348.
- Strober, Myra H. and Weinberg, Charles B. (1977), "Working Wives and Major Family Expenditures," Journal of Consumer Research, 4, 141-147.
- Vickery, Clair (1979), "Women's Economic Contribution to the Family," in Ralph Smith, Ed., The Subtle Revolution, Washington, D.C.: The Urban Institute.
- Waldman, Elizabeth and Klein, Deborah P. (1976), "Woman in the Labor Force: Secular Trends and the Impact of Recession," paper presented to the Society of Government Economists, Sixth Annual Conference.
- Young, Ann McDougall (1980), "Work Experience of the Population in 1978," Monthly Labor Review, 3, 43-47.

¹Unreferenced figures are from U.S. Department of Labor and U.S. Dept. of Commerce, Bureau of Census sources.

WOMEN'S CHANGING ROLES -- A CONSUMER BEHAVIOR PERSPECTIVE

Mary Lou Roberts, Boston University

Abstract

The multiple, and often conflicting, roles of women today pose significant challenges for marketers and researchers. This paper discusses actual and anticipated changes in women's consumer behavior on the basis of issues raised by studies of women's roles in non-marketing disciplines. Implications for research, marketing strategy, and public policy are discussed.

Change brought about by the continuing evolution in women's roles have affected, and will continue to affect, all aspects of our society. The purpose of this paper is to analyze the actual and anticipated effects of these changing roles on women's consumer behavior. The issues treated herein are those raised by the other five participants in this session. Underlying the multi-disciplinary approach taken in this session is a strong belief that we can better understand and anticipate the effects of women's changing roles on consumer behavior if we make use of knowledge available in the behavioral sciences and economics.

As an organizing framework, I will first treat issues directly observable in current demographic trends and then treat some interpersonal issues.

Four major factors which are evident in demographic data - delayed age of marriage, the decreasing fertility rate, women's employment, and women-headed families - have potential for causing important changes in women's consumer behavior. These factors have both individual and interactive effects.

Delayed Age of Marriage

One implication of this phenomenon is that women (and men, too) are spending more post-school years as single adults than have past generations. As young adult singles they may continue to live with their parents, live alone, live with one or more persons of the same (or with increasing frequency, opposite) gender, or live under the arrangement euphemistically described as co-habitation. Whichever living arrangement, or combination thereof, is chosen, delaying marriage implies a longer period of independence for the young adult.

The individuals who have experienced this protracted period of social and financial independence will bring to their eventual marriages a broad array of consumer skills. These consumer skills will be the result of both their own experience and of exposure to the life styles of other young adults including roommates, co-workers, and friends. Singlehood tends to force the acquisition of non-traditional consumer skills for both genders. For instance, the woman is likely to have purchased and maintained one or more cars and the man is likely to have prepared meals, maintained an apartment, and done the shopping associated with the performance of traditionally-female tasks.

Upon marriage, these persons are less likely to model their consumption patterns and decision-making processes upon the, probably traditional, patterns of their parents than are couples who lack this lengthy exposure to diverse and, often non-traditional, life styles prior to marriage. Both the nature of consumption-related decisions and the processes by which they are made will be affected. Kohen's statement that "young women will take part in

these negotiations with a greater sense of personal identity" is likely to apply in consumer decision-making as well as in family formation decisions. This sense of identity, reinforced by dual incomes, is likely to result in more autonomous decision-making by these couples.

Extensive pre-marital experience as single consumers may, over time, result in a more flexible approach to the division of household and other consumption-related tasks. Change is unlikely to be rapid, however. The most non-traditional attitudes and behavior should be exhibited by cohabiting couples, since they have chosen a relationship which is not sanctioned by a large proportion of our society. Still, a study of the division of household labor which included cohabiting couples (Stafford, Backman and Dibona 1977) found that even these couples were dividing tasks in the traditional, gender-linked manner. Even so, some change is becoming apparent. In earlier research dealing with meal preparation by husbands, Wortzel and I speculated that couples today seem increasingly inclined to perform at least the more creative tasks on the basis of personal preference as opposed to traditional gender roles (Robert and Wortzel 1979a). This is likely to become increasingly true as more young persons experience protracted periods of singlehood. While this is promising from the viewpoint of the demands on time of the working woman - and admittedly an over-optimistic one from the viewpoint of current time use research - it is a perplexing one for marketers. It has been difficult enough to select target segments, determine appeals, and accurately portray women as their roles have multiplied. Men's assumption of multiple, non-traditional roles will further complicate the situation. The best approach to take will be a cautious, research-based one, since change is not likely to occur simultaneously either across product categories or across market segments.

Since higher education is related to delayed marriage, another implication is that both partners will bring increased financial resources into the initial years of marriage. This will clearly allow for a more rapid accumulation of a stock of high-quality durable goods as well as for continued high expenditures on personal consumption goods, services, travel, and leisure pursuits.

Fertility

While the social implications of women's ability to control their fertility cannot easily be overstated, there are also important implications for consumer behavior.

One might start with the proposition that for the majority of American families having a child is now a conscious decision reached jointly by husband and wife. In her introduction to a special issue of the Psychology of Women Quarterly which deals with determinants of fertility, Russo states that "there is overwhelming scientific evidence that we value what we perceive ourselves to have freely chosen" (Russo 1979, p. 12). Therefore both necessity born of women's employment and positive attitudes toward nurturing of children should lead to greater male involvement in child rearing. Marketers should question whether fathers react similarly to mothers with respect to attributes of child care and entertainment products. Will they be more, less, or equally concerned about convenience in use, health and safety aspects, and educational benefits of products, for instance?

It seems clear that promotional activities for child-related products should reflect increasing male involve-

ment. Advertising themes, media selection and role portrayals may all be affected. Current television advertising for a new brand of disposable diapers can be cited as an example of portraying equal involvement of both parents in infant care.

The high educational level of these parents is likely to increase the size of the market segment which is extremely critical and demanding with regard to child-related products. Their expectations for quality in all types of products will be high. They will expect toys to provide developmental and learning experiences as well as entertainment. The increased financial resources they have accumulated as a result of delayed marriage, coupled with few children as a result of decreased family size, will allow sizeable expenditures on each child. The impact of large numbers of first births as well as many single-child households will increase the economic impact of this "echo baby boom" since higher per-capita expenditures are typical for first-born children (Cardozo and Haefner 1969).

Since many women may be delaying the birth of the first child until careers are well established, most of them are likely to return to the labor force very quickly, if they leave it at all. The resulting segment of well-educated and affluent two-worker families may choose to spend as little time as possible on activities that represent merely "custodial care" of the home or the children. Instead, they will try to devote as much time as possible to activities that provide rewarding family interaction. Products and services which promote interaction between parents and children that is both enjoyable and intellectually stimulating should be especially attractive.

Employment

In examining the effects of women's employment, Kohen points out that, while leisure is important to the working woman, it is frequently less fulfilling of them than is their work. In terms of their basic life priorities, this seems quite reasonable. Yet, it also seems reasonable to expect that changes in women's role-related attitudes will be reflected in their leisure pursuits. There is limited support in the literature for this hypothesis (Gentry and Doering undated). There is also the suggestion that women, like men, consider work and leisure to be inter-related and that women, even more than men, enjoy competitive leisure pursuits (Hawes, Blackwell and Talarzyk 1975).

In view of the limited amount of research available, some research questions seem to be in order. It seems clear that the scarce leisure time of the working woman should be highly valued. How does she approach the use of that leisure time? Does she see it as an end in itself or as an integral part of a well-rounded life, *i.e.*, as a way of keeping herself mentally and physically fit to cope with her work and personal activities? If she is married, does she reserve her leisure time for family activities or does she feel that she is entitled to some "time of her own?"

The theme of lack of time for the working woman to call her own recurs frequently in the more popular current literature on women's issues (see Bird 1974, for example). It seems that a majority of working wives correctly perceive themselves to be not only contributing to household income but also spending more time on household and child care than do other members of the family. They may, therefore, be very open to appeals related to doing something, either active or passive, for their own personal benefit and enjoyment.

In terms of the household consumer behavior that results from women's employment, marketers have more experience, but are not really much closer to developing useful

generalizations. The easy hypotheses have tended not to be confirmed by empirical data. Kahne points out that, although economists have hypothesized that purchase patterns for durable goods will be affected by women's employment, Strober and Weinburg (1977, 1980) have not been able to support this hypothesis with empirical data. The seemingly reasonable hypothesis that working women will tend to rely more on convenience foods than non-working women has fared similarly (Anderson 1971, Anderson 1972), although there is contradictory evidence (Editor and Publisher, 1972). Pointing out that sales of one convenience food, frozen vegetables, have actually declined in recent years, Wells sounds a warning that is widely applicable across product categories when he says, "Remember that frozen vegetables represent only one of a number of acceptable ways (emphasis mine) to shorten meal preparation time. The link between the desire to make meal preparation easier and the use of frozen vegetables is neither necessary nor direct (Marketing News 1980). Understanding the various possible approaches to task performance (see Roberts and Wortzel 1979b for an example), will shed considerable light on product and brand choice behavior. Again, a careful research-based approach is necessary, because there is no reason to believe that approaches to task performance will be consistent either across tasks within the same household or across households.

We should not overlook the fact that women's employment is creating for them another role -- customer for business goods and services. There is an easily observable tendency for marketers to fall back on the prevailing mythology about sex differences when dealing with women in business decision-making roles. Hesselbart sounds an important warning when she notes that "women and men in similar jobs have similar values and personalities." We should therefore not expect to find decision criteria and processes differing a great deal between male and female managers. Even in a situation where women do have a special set of problems (business travel is a good example), proprietary research has found no major difference in decision criteria between men and women. This type of woman, however, does appear to be very easily offended by two extremes of marketing behavior. At one extreme is blatant sexuality, usually in promotion, which is counter to her own role-appropriate behavior. At the other extreme is anything which bespeaks a patronizing attitude toward "the little woman." Actually, it may be that marketers of business goods and services have an easier basic task where women are concerned than do marketers of consumer goods and services; they should simply treat every role incumbent in the same manner.

Women-Headed Households

Both Elkstrom and Kahne point out that single women are disproportionately represented in the labor force and Kohen adds that one out of every six families is headed by a woman. Given current rates of divorce and of chosen singlehood, these trends seem likely to continue.

Expected differences in consumer behavior suggest that we should be looking at three different groups of currently single women:

Single women without families who are less than 55 years old (3 million in 1978).

Single women without families who are 55 years or older (7.3 million in 1978).

Single women who are family heads (5 million in 1978) (U.S. Department of Commerce 1978).

From 1970 to 1978, the most rapidly growing group of never-married women was the 25-34 age range (14% of all never-married women). Age alone makes them a prime

market for consumer goods, as does the fact that they are the best-educated of the women in this age bracket. In spite of the size of this market, only media catering to the young, single woman have published information which examines it specifically. It is clear that these women have a great deal of buying power, but how are they spending it? Are they buying condominiums or making other types of investments? Do they spend heavily on non-durable goods including clothing and personal care items? Are they especially heavy consumers of travel and entertainment services? The answer to all of these questions is undoubtedly "yes" but we need to understand the budgetary allocations of this segment. We also need to understand their shopping attitudes and habits. Are they part of the shrinking group of consumers who regard shopping as a pleasurable activity, perhaps because their free time is not constrained by family responsibilities? Are their shopping attitudes product specific - i.e. they like to shop for clothing and other personal items but dislike to shop for food? Do they mix shopping and entertainment? The basic hypothesis would be that the consumption patterns of this segment differ as a result of their relatively high level of "truly discretionary" income and that their shopping attitudes and behavior also differ from those of other women as a result of their life styles.

The segment of women living alone aged 55 and older presents a very different picture. They are less well educated and have lower incomes than other groups of employed women. In addition, women aged 55 to 64 are less likely to be employed than are any other group under age 65, although it is uncertain whether or not this is true for the single women in this age group.

Because elderly consumers make up a large portion of the market for many goods and services, there are several studies available. However, they tend not to cross-classify their data so that female responses can be distinguished from those of male respondents. One study (Martin 1976) studied fashion shopping behavior among young, middle-aged and elderly women. He found numerous differences between the three age groups with elderly women using more information from both media and sales clerks, shopping fewer stores, and shopping in specialty boutiques less frequently than either of the other two groups. Although there was little inter-group difference, elderly women perceived themselves as being fashion conscious and said they enjoyed shopping. While it is true that one might expect fashion-related behavior to differ widely between generations, different shopping patterns as well as the different information processing patterns discussed by Phillips and Sternthal (1977) suggest that different marketing strategies would be appropriate for the elderly female consumer. Single elderly women, an apparently unstudied group, should continue to increase in importance as a market segment, since a woman's life expectancy continues to exceed a man's by almost eight years and the differential is predicted to increase. (U.S. Department of Commerce 1978, p. 11).

The final group of single women is those who are family heads. As Kahne points out, the vast majority of these families are composed of children, a few include a male who is not defined as the household head, and still others are composed of other adult family members. While there are other similarities, the single most unifying aspect of these families is their low income level. It is not surprising that Kohen's research has found them to be an especially good market for used goods.

The low level of discretionary income does not make this a particularly appealing market target for most goods. The exception is products specifically designed to stretch the consumer's dollar, such as generic products in the supermarket. Even here, we may be missing out, because a recent study of the market for generics cites greater usage by the upscale customer (Murphy and Laczniak 1979).

This highlights the need for consumer information programs which will effectively reach this and other low-income segments. While this is a difficult task, it is one that deserves the attention of both marketers and researchers.

There is, however, another aspect of women-headed families that may prove to be of even greater long-run significance to consumer behavior. It is the consumer socialization of children from woman-headed households. If the alarming projection quoted by Kohen that 40% of all children will spend at least a portion of their formative years in such a household turns out to be true, we may see differences in socialization practices and the resulting behavior of adolescents and young adults.

It is clear that a great deal of learning about sex-role-appropriate behavior occurs in early childhood. However, the literature on single-parent families headed by women (see Herzog and Sudia 1973 for a literature review) indicates that, while absence of the father tends to delay this learning, other role models are used so that by young adulthood there is no discernable difference in internalization of sex roles between children from two-parent and single-parent families. It is not absence of the father per se that causes developmental and social problems in the children of single-parent families. Rather, it is the stress and fatigue engendered by the long hours, lack of dependable child care, and the low salaries of most of these women that lead to problems in the family.

The research to date does not seem to provide strong evidence to suggest that children from woman-headed families will be more sex-role egalitarian than children from two-parent families. It is more important to note that sex role stereotyping is less prevalent among children of all working mothers (Del Boca and Ashmore 1980), so maternal employment appears to be the key factor.

Kohen's finding that divorced mothers expect a more mature and responsible role from their children does suggest differences in consumption-related behavior in single-parent families with older children and teenagers. These children may shop for themselves and do routine shopping for the family at an earlier age. Time pressure in all families with a working mother, and especially in those with a single parent, suggests that more research attention should be given to assumption of household tasks and to shopping behavior by older children and teenagers.

Household Task-Sharing and Decision-Making

When considering the effect of changing sex roles on consumer behavior, one of the first concerns of marketers is changing patterns of household task performance. This concern seems to be based on the premise, which to the best of my knowledge has yet to be empirically tested, that persons who are involved in the performance of household tasks will be more influential in product and brand choice for task-related products.

There is ample evidence that both men and women hold increasingly egalitarian attitudes towards women's roles, and as a result, feel that household tasks should be shared. Unfortunately, egalitarian attitudes have not been readily translated into wholesale male assumption of household tasks (Lopata et al. 1980, p. 128). However, there is limited evidence that males are performing some traditionally female tasks (Roberts and Wortzel 1979a). When one of these tasks is grocery shopping, the results, in terms of brand choice, may be quite different than if the wife had done the shopping herself (Newsweek 1979, Progressive Grocer 1980). With the dramatic increase in two-wage-earner families in recent years, the topic of task sharing has relevance for marketers of a wide range of goods and services.

Various theories have been advanced to explain male-female roles in household decision making. Chief among them are the relative power, cultural role expectations, relative investment (Davis 1976) and time-available (Stafford and Dibona 1977) hypotheses. While all provide useful insights, none appears to provide consistent explanatory power with respect to household task sharing (See Wortzel and Roberts forthcoming for an extensive review of the literature on household task sharing).

Both observation and descriptive writing on the subject of household task sharing suggest that there are a variety of processes used in arriving at a, hopefully satisfactory, allocation of household tasks. They range from the marital contract which may specify duties in great detail, through recommended processes for arriving at allocations (see Seiden 1980, for an example) to the ad hoc arrangements which probably exist in the majority of households. The strategies used to allocate household tasks should provide insight into a broad range of family decision-making processes and are therefore well worth study.

At the risk of oversimplifying a complex situation, I would like to propose a paradigm which may be useful in studying household task allocation (Figure 1).

		Standard Relative to Task Outcome	
Attitude Toward Performing the Task			
	High	Low	
Positive	Individual likely to perform task	Individual likely to perform task 'when convenient'	
Negative	Individual likely to try to shift task performance to another person	Individual unlikely to care if task performed at all	

Figure 1.

As traditional roles become blurred, attitudes - whether actually positive toward task performance or just less negative than the partner's - are likely to become increasingly important. At the same time, perfectionist standards toward housework may be less important to women, especially working women, who have other inputs to their evaluation of their own self-worth. It is quite possible that a husband may have higher standards than his wife with regard to one or more specific task outcomes. This paradigm, then, applies separately to husband and wife and also to children old enough to assume general household tasks. Task allocation will be the result of the mix of attitudes and standards present within the family.

Other things equal, a positive attitude coupled with high standards should predict that the individual will perform the task. If two individuals in the family fall into this cell, they may share the task, perhaps on a time-available basis. The person(s) who perform(s) this task may actually prefer the type of product that requires more effort (materials to cook or bake 'from scratch' or non self-polishing floor wax, for example). On the other hand a person with a negative attitude but high standard, if (s)he must perform the task, would look for a product which achieves good results with minimal effort. The individual who falls into the positive attitude/low standard cell may be relatively indifferent to the attributes of the product used. The cell which couples negative attitudes with low standards, especially if it is prevalent throughout the family, suggests that the task may not be performed at all. There are two key warnings for marketers: not to assume that standards are uniformly high across the population and not to assume that they are gender-linked.

Hesselbart's distinction between expressive and instrumental tasks is also worth noting. Research has found that child care is the task husbands are most likely to share (Lopata 1971). Recent interest has centered on other activities such as cooking, which can also be expressive in nature. If husbands are most willing to share tasks which have expressive aspects and if women are less eager to find recognition or self-fulfillment tasks that are essentially instrumental, it suggests a more pragmatic approach to the promotion of products used in routine tasks.

It is also worth returning briefly to the problem of power versus time in the two-wage-earner household. If power is conferred by contribution to the family's income, then working women should indeed have more power in family decision-making. It is easy to hypothesize that women will exercise this power with the result being more joint decision-making. However, we must realize that time spent in income-producing activities is, for most women as well as men, time not spent at home. The trade-off between more power and less time may result in less assumption or sharing of previously male-dominated decisions than one would at first suspect. In fact, one study found more male participation in shopping and financial decisions that had previously been handled solely by wives (Pralle 1980).

Investigating how couples handle two incomes can also give useful insights into more specific consumption decisions. A study of more than 15,000 couples conducted by Family Circle magazine in 1976 and 1977 and quoted by Bird (1979, pp. 155-164) identifies four types of couples based on the manner in which they handle the wife's income:

- Pin money couples in which the wife's income is hers to save or spend as she chooses
- Earmarker couples who designate the wife's income for a specific purpose (children's college expenses, down payment on a home, etc.)
- Pooler couples who combine both incomes, usually in a joint checking account, and spend it without regard to its source
- Bargainer couples who treat income as the property of the person who earns it and negotiate the division of household expenses.

Preferred decision-making strategies are implicit in these categorizations. In addition, the freedom, or lack thereof, which the wife has in spending "her" money is likely to affect both product and brand choice behavior.

Working-Class Women

One large segment of women has been almost entirely overlooked in our discussion thus far. That segment is the working class woman. This is not surprising. Ekstrom notes that there has been little research on employment decisions of the working class woman. This is true of research on working class women in general with the classic studies (Rainwater 1959, Komarovsky 1962) predating the women's movement and therefore having limited applicability today. Major studies through 1974 are listed in an annotated bibliography (Samuels 1975), and a journalistic-type study which contains useful insights was published in 1977 (Howe 1977). However, research which explicitly deals with the consumer behavior of working-class women is virtually non-existent. The most recent study was published in 1973 (Social Research Inc. 1973).

Some of the salient findings about working-class women are:

- o Working-class women have been slower than middle-class women in breaking with traditional roles, and have tended to react negatively toward changes which appear to threaten established values.
- o For ethnic working-class women, especially, religious institutions tend to reinforce traditional values.
- o The working-class woman tends to be part of a tightly-knit social group composed primarily of female kin.
- o Although the necessity of working has expanded the social horizons of many working-class women, the traditional female jobs which most of them hold are not intrinsically rewarding.
- o These expanded horizons are reflected in the decreasing propensity of working-class women to define family responsibilities, especially child care, as the central focus of their lives. (Peters and Samuels 1976).
- o Working class husbands and wives tend to adhere to traditional household roles and to engage in sex-segregated leisure activities, even if wives work.
- o Wives' employment is often viewed as threatening by the working class husband whose ability to provide for the family is often the major source of his sense of self-worth (Samuels 1975).
- o While feelings of lack of self-worth and competency in dealing with the world outside the home have decreased among working-class women, they have decreased more slowly than have the same feelings among middle-class women. The result is an increasing disparity in feelings of self worth and self-confidence between the two social class groupings.
- o Working-class women are more likely than their middle-class counterparts to feel that their adult life is better than their childhood. Since part of the perceived improvement is due to acquisition of desired material goods, working-class women are more positive toward business in general and specific products, and toward the media in general and advertising in particular, than are middle-class women. (Social Research, Inc. 1973).

Taken together, these points emphasize the fact that most of our data about women come from middle-class respondents and is based on middle-class values. Working-class women make up an important market segment for many products and brands. In addition, they may react to some types of advertising appeals very differently from middle-class women (Johnson and Satow 1978). If we could believe that they are soon likely to adopt middle-class attitudes, values and patterns of behavior, their absence from most consumer behavior studies would not be an important omission. However, the S.R.I. study (1973) also indicates that working-class emphasis on upward mobility into the middle class appears for many to be giving way to an emphasis on achieving desired goals within the working-class milieu. The attention of researchers to the working-class market is therefore important and overdue.

Implications

For Research

The opportunities for research that are inherent in issues surrounding women's changing roles are so numerous that I shall make no attempt to enumerate them. However, there are some general issues that should be considered.

One of these is indiscriminate borrowing of concepts from other disciplines. The patterns of work experience described by Ekstrom are a good example of potential use and misuse. These patterns - career to homemaking, intermittent labor force participation, double-track women, and work-committed women - would seem intuitively to have potential in studies of family financial decisions. However, I see no reason to believe that they would be useful predictors of product or brand choice decisions.

A related problem is indiscriminate borrowing of methodology. Several years ago Green and Cunningham (1975) used Arnott's Feminism scale in a study of husband-wife decision-making and, citing this work, marketing researchers have continued to use it. This scale, which is about ten years old, is composed of ten attitudinal statements. Although the scale appears reliable, attitudes of the general populace continue to shift toward the "liberated" end of the scale, with the result being that the scale does not discriminate well. Since we know that role-related behaviors are not changing as rapidly as are related attitudes, role-orientation scales which have behavioral components should provide more ability to discriminate.

Another possibility is that we should move away from the idea than general role orientations will be good predictors of consumer behavior. Perhaps the concept of androgyny, infrequently used in marketing research to date, will be more useful.

The best solution of all, however, may be to focus on task and product-specific attitudes. They should not only be more powerful predictors of related behavior but will also enable us to avoid the troublesome problems of defining concepts such as "traditional" and "contemporary." Asking parallel questions of all family members will help us to understand the real nature of decision making within the family group.

And finally, all research, whether basic or applied, should recognize the inherent bias toward the middle class view of the world that is apparent in most of our work. Removing this bias will help us to understand the diversity that exists among American women today.

For Marketing

Those who market goods and services to women will be well advised to look beyond the "outputs" of consumer decision-making -- product and brand choice -- and to investigate in depth the processes by which those final decisions are made. The life style choices available to women today are many and varied. Products and services are sought both to implement these life style choices and to reflect them. Understanding how products and services fit into chosen life styles is a necessity if they are to be marketed effectively.

On the basis of research previously cited, it seems clear that we should expect to find few major differences in the purchase of consumer durable goods between families with working and non-working wives. Differences, however, may exist in buying behavior relative to frequently purchased goods and services, which have received less research attention to date. We should look for variations in product quality expected or preferred, in shopping patterns including outlet choices and timing of purchases, and in attitudes and family practices with regard to shopping.

In developing, as well as marketing, products and services we should look at the incompatible demands which often result from women's multiple roles. Any working wife or mother could assure us that these role-incompatible demands are legion. Take, for example, services of all kinds. Every provider of services -- from the washer

repairman to the orthodontist -- prefers to work from nine to five on weekdays. Obtaining services at these times is highly incompatible with the demands of most employers. Offering services at times and places convenient to the consumer could be a major coup. The same type of reasoning can be applied to the development and distribution of products which ease the time pressure brought about by the demands of multiple roles.

As marketers we should also look more clearly at the budgetary allocations and decision-making process of non-traditional families, especially the dual-income and the woman-headed family, two rapidly-growing groups. In this respect, we should also carefully evaluate the composition of households in the U.S. over the coming years. The number of households is expected to increase rapidly. Will the increase come primarily from affluent singles and dual-income families or will a large part of it be less-than-affluent woman-headed households? The demand for all types of goods and services, including consumer durables, will be greatly affected by the composition of households.

In effect, marketers should not look for easy answers to marketing strategy problems based on whether women work or not. The diversity of working, as well as non-working, women is too great for generalizations that will prove useful across broad groups of people and of products. Approaches centered around life styles, decision processes and task orientations should replace a simple working/non-working dichotomy.

For Public Policy

The papers in this session have also raised numerous issues which have public policy implications. Neither time nor the objectives of this session allows for a full discussion of these important issues, many of which relate to the job satisfaction and income-producing ability of women. Only a few can be highlighted here.

Ekstrom noted that numerous women lack the necessary skills to obtain employment that pays well enough to compensate for their absence from the home. Women in families where husbands or other adults are present may be able to generate worthwhile income by adjusting hours so that the husband or other adult can assume child care and other household tasks. Such arrangements can be economically and personally viable, but the behavioral science literature frequently refers to the strains on the marital relationship caused by such arrangements. The group affected most by lack of skills and low income is single mothers. If businesses can upgrade the skills of these women and assist them in finding satisfactory, low-cost child care, they will increase the level of effective demand emanating from a significant segment of our population. In this context, businesspeople may wish to consider the implications of Kahne's statement that consumer needs can be translated into effective demand either through spending of personal income or through social programs.

The possibility of an increasing gap between the incomes of one- and two-wage earner families is interesting, although its implications are less clear. It does seem worthy of continuing attention, however, both from the standpoint of marketing and of employee relations and for its potential in generating public policy issues.

Conclusion

It seems that, both as marketers and as researchers, we have overcome the notion that all women are or should be homemakers, even though some of the mythology concerning sex differences still pervades many of our activities. We have also come a long way toward understanding that working women do not make up a single, homogenous group,

nor are all non-working women alike. It is to be hoped that this session has taken us one step further toward a genuine understanding of women's multiple roles. That step would be the realization that, while change in women's role-related attitudes and behavior is pervasive, it is not occurring with equal speed on all the same dimensions throughout our society.

Only when we, as marketers and researchers, recognize the diversity that exists among women and their families in the way they interpret their work, family, and consumer roles will we be able to effectively cope with the changes that are occurring.

References¹

- Bird, Caroline (1979), The Two Paycheck Marriage, New York: Pocket Books.
- Davis, Harry L. (1976), "Decision Making Within the Household," Journal of Consumer Research, 2, 241-260.
- Del Boca, Bella M. and Ashmore, Richard D. (1980), "Sex Stereotypes Through the Life Cycle," in Ladd Wheeler, ed., Review of Personality and Social Psychology: 1, Beverly Hills: Sage Publications, 163-192.
- Herzog, Elizabeth and Sudia, Cecelia E. (1973), "Children in Fatherless Families" in Bettye M. Caldwell and Henry N. Ricciuti, ed., Child Development Research, Vol. 3, Chicago: The University of Chicago Press, 141-232.
- Howe, Louise Kapp (1977), Pink Collar Workers, New York: Avon Books.
- Lopata, Helena Z. (1971), Occupation: Housewife, New York: Oxford University Press.
- _____, Barnewolt, Debra and Norr, Kathleen (1980), "Spouses' Contributions to Each Other's Roles," in Fran Pepitone-Rockwell, ed., Dual-Career Couples, Beverly Hills: Sage Publications, 111-142.
- Roberts, Mary Lou and Wortzel, Lawrence H. (1979a) "Husbands Who Prepare Dinner: A Test of Competing Theories of Marital Role Allocations," in Jerry L. Olsen, ed., Advances in consumer Research, 7, 669-671.
- _____, (1979b), "New Life-Style Determinants of Women's Food Shopping Behavior" Journal of Marketing, 43, 28-39.
- Samuels, Victoria (1975) "Nowhere To Be Found, A Literature Review and Annotated Bibliography on White Working Class Women," Institute on Pluralism and Group Identity, Working Paper Series, Number 13.
- U.S. Department of Commerce, Bureau of The Census (1980), "A Statistical Portrait of Women in the United States: 1978."
- Wortzel, Lawrence H. and Roberts, Mary Lou (forthcoming), "The Cooking, Food Shopping and Eating-Out Behavior of Husbands and Wives."

¹Due to lack of space only a portion of the references cited are listed here. A complete list of references is available from the author.

SEPARATING AGE, COHORT AND PERIOD EFFECTS IN CONSUMER BEHAVIOR

Joseph O. Rentz, University of Virginia,¹
 Fred D. Reynolds, University of Georgia²

Abstract

Traditional research designs for the study of consumer aging confound the effects of age, cohort and period. Cohort analysis is a group of methods designed to separate the three effects. Problems with conventional designs are discussed and cohort analysis methods critically reviewed.

Introduction

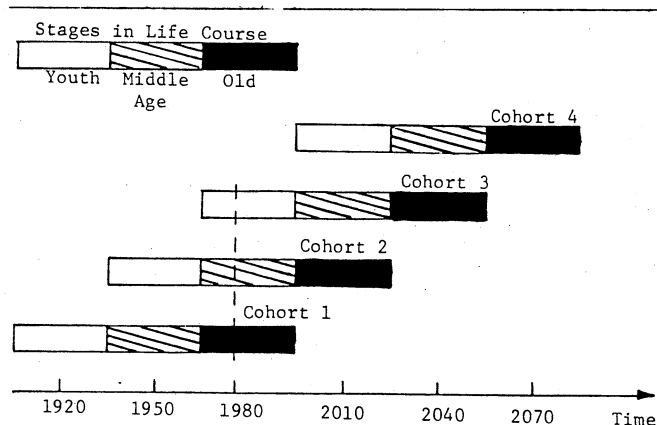
Our purpose in this paper is to point out to consumer behavior researchers some advances in research designs and methods for the study of aging. Generally called cohort analysis, these methods attempt to separate the effects of age, cohort and period. Use of the methods over the past decade has resulted in an increased understanding of the aging process. Researchers have found that some phenomena once thought to be related primarily to the aging process are instead related primarily to the cohort membership of the groups studied.

In this paper we critically review the cohort analysis methods. First the relationships between age, cohort and period are discussed and some shortcomings of traditional research designs for the study of aging are pointed out.

Relationships Between Age, Cohort And Period

Figure 1 graphically illustrates the relationships between age, cohort, and period. The figure represents the concomitant processes of aging and succession of four selected cohorts and reveals cross-sectional differences among age classes at any given point in time (Riley, Johnson and Foner 1972).

FIGURE 1
 Aging, Cohort Succession And Age Differences
 Source: Riley, Johnson and Foner (1972)



Johnson and Foner 1972). Each cohort is comprised of people born at approximately the same time. People within each cohort age over time.

Specifically, a cohort is an aggregate of individuals within a given population who experienced the same event(s) within the same time interval (Ryder 1965). In Figure 1 the defining event is birth although numerous types of cohorts can be identified including marriage cohorts, divorce cohorts, retirement cohorts and education cohorts. The word is typically used without a modifier to refer to birth cohorts. Cohort succession refers to the immutable process whereby cohorts of people are born, live out their life courses and die, being replaced by other cohorts (Riley 1973).

Cohorts can be expected to differ from one another on a variety of dimensions for a variety of reasons. Each cohort occupies a unique interval in time. Since society is constantly changing over time, each cohort experiences unique events or experiences events differently because each cohort is at a different stage in the life course when the events occur. Cohorts also differ in initial size and composition (Ryder 1965). The problems faced by members of the post-war baby boom cohorts -- crowded schools and limited job opportunities, for example -- are examples of the effects initial size can have on cohorts. Thus we might expect cohorts to differ from one another in many respects.

Age differences in the population at a given time are revealed by a cross section of the population at that time -- see the vertical line at 1980 in Figure 1. These cross-sectional differences can be important in their own right since they have implications for current marketing practice, public policy, etc. However from a developmental perspective, the important issue is what accounts for the current differences.

Figure 1 shows that there are two interpretations for cross-sectional differences in the age classes (Riley 1973). The "aging interpretation" is that the 1980 age class differences are due to the development (aging) of consumers. This is the interpretation that is nearly always given by consumer researchers to age differences. The alternative interpretation, the "cohort interpretation," is that the 1980 age class differences are due to the different cohort memberships of the age classes. If the cohort interpretation is correct then the characteristic of interest will remain relatively stable over the life courses of people in the cohorts and the characteristics of the age classes in the future (say in the year 2010 in Figure 1) will change as new cohorts occupy each of the age classes. For example, elderly consumers in the future will be different from today's elderly consumers. When aging is producing the age class differences, the characteristic of interest will change over the life courses of people in the cohorts and the characteristics of the age classes in the future will be similar to the characteristics of the current age classes. For example, the elderly of the future will be similar to the elderly of today.

The clearest example of a variable which is related to cohort membership rather than aging is education. At any point in time, elderly people report having completed fewer

¹Joseph O. Rentz is Assistant Professor of Marketing, McIntire School of Commerce, University of Virginia.

²Fred D. Reynolds is Professor of Marketing, University of Georgia.

years of formal education, on average, than young people. Clearly the number of years of education completed does not decline as an individual ages. Rather, age class differences in education are a function of cohort membership and we can expect the average educational level of the elderly, for example, to increase as more educated cohorts age into the elderly age classes and replace less educated cohorts.

Cohort membership accounts for a significant portion of age class differences in political party affiliation (Cutler 1970; Glenn 1972; Glenn and Hefner 1972; Knoke and Hout 1974), cognitive performance (Baltes and Reinert 1969, Schaie and Strother 1968a; Schaie and Strother 1968b; Schaie and Parham 1977), and consumption of a product class (Rentz 1980) as well as other variables.

Clearly the effects of aging are confounded with the effects of cohort membership in the cross-sectional design. Unless one has evidence that cohort effects are minimal, errors of interpretation can occur by attributing age class differences to the aging process.

A second research design, the longitudinal design, has also been employed by consumer researchers to study aging. For example, in Figure 1, one might measure a sample of individuals in a particular cohort at two or more points in time. The same individuals might be followed over time or an independent sample from the cohort might be taken at each time point. The longitudinal design also has weaknesses. In longitudinal comparisons the effects of aging are confounded with "period effects." Period effects are environmental influences which affect aging cohorts at one point in time (between two times of measurement). They include marketing efforts, public policy, and social change in general. Because it is impossible to employ control groups who do not age, one cannot separate period effects and age effects in longitudinal comparisons. Thus age effects and period effects are confounded in longitudinal comparisons.

A third research design, the time-lag design, is sometimes used to study differences among cohorts (or among generations). The same age class is compared at two or more points in time. For example, in Figure 1, youth in 1980 might be compared to youth in 1950. Differences are typically attributed to differences in cohorts (or generations.) However the comparisons are made over time so cohort effects are confounded with period effects in time-lag comparisons.

The confounding of age, cohort, and period effects in empirical measures led Baltes (1968, p. 153) to conclude: ". . . the conventional methods are in no way adequate research designs for the assessment of age effects."

How, then can consumer researchers separate the effects of age, cohort and period? A group of methods which fall under the general heading "cohort analysis" has been developed in an attempt to separate the effects of the three variables. The method can be used in studies of aging, in inter-cohort-studies, and in the forecasting of consumer behaviors. Reynolds and Rentz (1980) illustrate the use of cohort analysis in forecasting changes in women's roles. In the next section we discuss the basics of cohort analysis and then we discuss several cohort analysis methodologies.

Cohort Analysis: Separating Age, Cohort, And Period

Cohort analysis refers to any study in which one or more cohorts are measured on some variable at two or more points in time (Glenn 1977). Either panel data or cross-sectional data may be used. Cohort analysis normally begins with the construction of a "standard cohort table" (Glenn 1977). Tables 1, 2 and 3 are examples of cohort tables and illustrate the pattern of variation resulting from "pure" age, cohort and period effects. Normally a series of cross-sectional studies comprise the columns of the table.

TABLE 1
Cohort Table Showing Hypothetical Data (Percentages)
In Which All Variation Is Due To Age Effects
Source: Adapted from Glenn (1977)

Age	Time of Measurement (Period)			
	1950	1960	1970	1980
20-29	40	40	40	40
30-39	45	45	45	45
40-49	50	50	50	50
50-59	55	55	55	55
60-69	60	60	60	60
70-79	65	65	65	65

The tables are constructed so that the interval between any two times of measurement (periods) is the same as the intervals used to delineate age classes. In such a table cohorts are represented in the diagonals of the table. Trends in an aging cohort can be traced by reading down the diagonals of the table as illustrated for one cohort by the underlined measures in Table 1. Nine cohorts are represented in each of the cohort tables presented here.

TABLE 2
Cohort Table Showing Hypothetical Data (Percentages)
In Which All Variation Is Due To Cohort Effects
Source: Adapted from Glenn (1977)

Age	Time of Measurement (Period)			
	1950	1960	1970	1980
20-29	50	40	30	20
30-39	60	50	40	30
40-49	70	60	50	40
50-59	80	70	60	50
60-69	90	80	70	60
70-79	100	90	80	70

TABLE 3
Cohort Table Showing Hypothetical Data (Percentages)
In Which All Variation Is Due To Period Effects
Source: Adapted from Glenn (1977)

Age	Time of Measurement (Period)			
	1950	1960	1970	1980
20-29	70	60	50	40
30-39	70	60	50	40
40-49	70	60	50	40
50-59	70	60	50	40
60-69	70	60	50	40
70-79	70	60	50	40

Unfortunately, the interpretation of a cohort table is not straightforward. We have seen that age and cohort effects are confounded in cross-sectional comparisons (comparisons within any column in Table 1), age and period effects are confounded in longitudinal comparisons (comparisons within any diagonal in Table 1), and period and cohort effects are confounded in time-lag comparisons (comparisons within any row of Table 1). Comparisons between mean values for rows, columns, or diagonals does not eliminate the confounding. For example the difference between the means of the first two rows in Table 1 is the same as the mean of the four cross-sectional differences between the two rows. In each of the four differences, age and cohort effects are confounded so the mean of the four differences also contains confounded effects.

To further complicate matters, there is a formal linear

dependency between each of the independent variables of interest (age, cohort, and period) and the other two. That is, when age, cohort and period are indexed by chronological age, year of birth, and year of measurement respectively, then age equals period minus cohort, cohort equals period minus age and period equals cohort plus age.

Mason, et al. (1973) present a lucid discussion of the implications of the linear dependency and show that parameter estimation becomes impossible in most statistical models. Thus traditional techniques such as analysis of variance or regression are not appropriate without modification. The nonindependence of the variables is a case of the "identification problem" discussed by Blalock (1966). Identification of the operative variables is possible only through a priori assumptions.

Cohort Analysis Methods

Several methods have been proposed for separating age, period and cohort effects including "visual inspection" methods, methods to correct for period effects (Agnello 1973; Crittenden 1962; Klecka 1971), Schaie's (1965) sequential strategies, Baltes (1968) bifactorial model, the constrained multiple classification method developed by Mason, et al. (1973), and an approach developed by Palmore (1978). We prefer the methods of Mason, et al. and Palmore although there is by no means a consensus in the literature concerning which method is most appropriate.

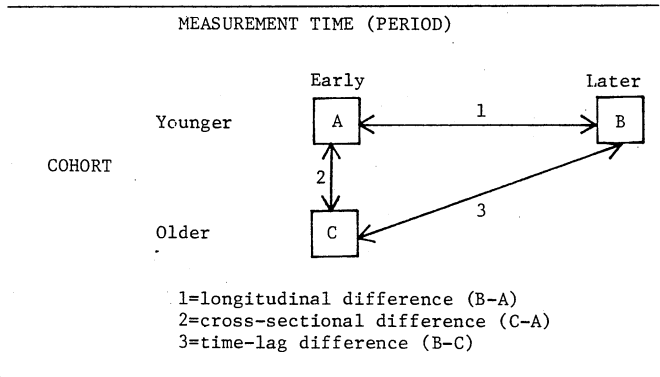
In this section we critically review four of the methods which we feel merit the consideration of consumer behavior researchers. For a complete review, see Rentz (1980). Palmore's method is presented first because his approach clarifies much of the conceptual and terminological confusion which exists in the literature.

Palmore's Method

Palmore (1978) points out that much confusion surrounding the age-cohort-period problem has arisen from a failure to distinguish between three necessary levels of analysis: computing observable differences, inferring effects, and determining the theoretical causes of the effects. Cohort analysis involves inferring effects from observable differences. The causes of the effects can only be decided on the basis of evidence outside the cohort table.

Observable differences are those differences which are calculated from empirical data. Three types of differences are observable (Figure 2). Cross-sectional differences (cell C - cell A) are differences between younger and older cohorts (and younger and older age groups) at one point in time.

FIGURE 2
Observable Differences
Source: Palmore, 1978



Longitudinal differences (cell B - cell A) are differences between early and later measurements on the same cohort (at different ages). Time-lag differences (cell B - cell A) are differences between two groups of the same age, one from the older cohort measured at the early time and one from the younger cohort measured at the later time. Computation of the observable differences is not a difficult process once relevant data are collected.

Palmore's second level of analysis is inferred effects. Age, cohort and period effects are inferred from the three observable differences. Each observable difference is composed of the effects of two and only two variables:

$$\begin{aligned} \text{Longitudinal Differences} &= \text{Age} + \text{Period Effects} \\ \text{Cross-Sectional Differences} &= \text{Age} + \text{Cohort Effects} \\ \text{Time-Lag Differences} &= \text{Period} - \text{Cohort Effects} \end{aligned}$$

The effects of age, period, and cohort can only be inferred from the observable differences. Since any one difference is composed of two effects one cannot equate any one effect with any one difference without further analysis or assumptions. For example, if one has other evidence or makes the assumption that cohort effects are zero, then cross-sectional differences can be equated to age effects.

One triad of data of the form shown in Figure 2 is the minimum number of cells necessary to compute the three observable differences. Palmore works with one triad of data at a time, inferring effects from observable differences in each triad. Each of the differences is tested for statistical significance. Three patterns of significant differences are possible: no significant differences, two significant differences and three significant differences. Because the three differences are related (the longitudinal difference is the sum of the cross-sectional and time-lag differences) one cannot obtain one significant difference except when sample sizes are markedly different.

We will illustrate Palmore's inference process using a pattern of two significant differences. The following data illustrate the pattern:

		Period	
		Early	Later
Cohort	Younger	10	20
	Older	20	

Here the longitudinal difference is ten, the cross-sectional difference is ten and the time-lag difference is zero. There are three possible interpretations of this pattern.

The first interpretation is that the age effect is ten and cohort and period effects are zero:

$$\begin{aligned} \text{Longitudinal Difference} &= 10 = \text{Age} (10) + \text{Period} (0) \\ \text{Cross-Sectional Difference} &= 10 = \text{Age} (10) + \text{Cohort} (0) \\ \text{Time-Lag Difference} &= 0 = \text{Period} (0) - \text{Cohort} (0) \end{aligned}$$

The second interpretation is that period and cohort effects are exactly counterbalancing one another in the time-lag difference and age effects are zero:

$$\begin{aligned} \text{Longitudinal Difference} &= 10 = \text{Age} (0) + \text{Period} (10) \\ \text{Cross-Sectional Difference} &= 10 = \text{Age} (0) + \text{Cohort} (10) \\ \text{Time-Lag Difference} &= 0 = \text{Period} (10) - \text{Cohort} (10) \end{aligned}$$

The third interpretation is that a complex combination of all three effects is producing the differences. Palmore argues that in the absence of other evidence the principle of parsimony favors the first interpretation over the other two. We have found that by making a priori assumptions about the directionality of two or three effects one can often rule out one or two interpretations. For example if one assumes that period and cohort effects are in opposite

directions then the second and third interpretations in the preceding example can be ruled out since period and cohort effects could not counterbalance one another in the time-lag difference.

Palmore offers similar interpretative guidelines for the patterns of no significant differences and three significant differences. Once the effects are inferred in each triad of the cohort table one can summarize effects over groups of triads.

Palmore notes that the major advantage of his method is that it is not necessary to assume some systematic pattern of effects. Some methods require the assumption that effects are additive or that the effects are unidirectional and positive. Palmore's method provides a way of determining whether systematic patterns exist and what they are. However the method also has weaknesses. Often directionality assumptions do not allow a choice of alternative interpretations and one is forced to rely on the principle of parsimony. Also the method is oriented toward finding one main effect and works best when no differences or two differences are significant. The typical situation may be that all three effects are operative and three differences are significant.

Schaie's Sequential Methods

Recognizing that only two of the three independent variables of interest (age, cohort, and period) can be varied independently, Schaie (1965) proposes three bifactorial designs in which different combinations of the three variables are represented two at a time. His cohort-sequential method is an age by cohort design, his time-sequential method is an age by period design and his cross-sequential method is a cohort by period design. The analysis of variance is used as the test of significance in each case.

Schaie recognizes the confounding of the variables in each of his three methods and does not advocate using any one of the methods alone to estimate age, period, and cohort effects. Rather he advocates using two or more of the methods in conjunction. He presents a set of decision rules for inferring effects using the time-sequential and cross-sequential methods in conjunction or the cohort-sequential and cross-sequential methods in conjunction.

Schaie's decision rules require the assumption that age, period, and cohort effects are unidirectional and positive, that is, "an incremental progression exists from a younger to an older age, from an earlier to a later measurement time, and from a younger to an older generation" (Schaie 1965, p. 103).

Essentially, the estimates provided by the analysis of variance are estimates of either mean longitudinal, mean cross-sectional or mean time-lag differences. Schaie's decision rules follow a logic similar to Palmore's. For example the estimate of the period effect from the cross-sequential method is actually an estimate of the mean longitudinal difference (which contains age and period effects). One of Schaie's decision rules states that if this estimate is zero then variation in the table is a result of cohort effects. Since both age and period effects are assumed to be positive, (or zero) if the mean longitudinal difference is zero then both age and period effects must be zero and all variation must be due to cohort effects. Similar logic can be employed to verify Schaie's remaining decision rules.

Variations of Schaie's methodology have been employed by others and especially by Riley, Johnson, and Foner (1972). Their "period analysis" is the same as Schaie's cross-sequential strategy, their "cohort analysis" is the same as Schaie's cohort-sequential strategy and their "cross-sequential analysis" is the same as Schaie's cross-

sequential strategy. They do not employ Schaie's decision rules but rather emphasize the need to make assumptions and use outside evidence in order to untangle age, cohort, and period effects.

The usefulness of Schaie's methods is limited by the necessity to assume that all effects are unidirectional and positive. However, Schaie's general approach of specifying the type of slope assumed for each of the three effects and developing decision rules based on the type of slope assumed is certainly valid when theory or outside evidence supports the assumptions.

Constrained Multiple Classification Method

Mason, et al. (1973) develop an approach to estimating age, cohort and period effects using multiple classification analysis. Multiple classification analysis is essentially a multiple regression analysis using dummy variables (Andrews, Morgan, and Sonquist 1967).

Mason, et al. employ a model of the following form:

$$Y_{ij} = \mu + B_i + \gamma_j + \delta_k + \epsilon_{ij}$$

where the effect of the i -th age group is given by B_i , the effect of the j -th period by γ_j , the effect of the k -th cohort by δ_k ; where μ is the grand mean of the dependent variable and where ϵ_{ij} is the random error. They show that the parameters of a model of the above form are not estimable because only one cohort exists for any given age and period (i.e., any one variable is linearly dependent on the other two). Mason, et al. also show that by constraining the above model such that any two ages, periods, or cohorts have identical effect parameters, linear dependency is eliminated and the parameters of the model are estimable. Mason, et al. proceed to show that when only two ages or two cohorts or two periods are constrained to be equal, the fit of the model to the data does not vary but the estimates do vary with different equality constraints. Thus one cannot judge the suitability of the many possible constraints. To remedy such a situation, Mason, et al. advise that the analyst place constraints on more than one dimension. For example, we have used the method by fitting three models with different sets of constraints: one with two ages and two cohorts set equal, one with two cohorts and two periods set equal and one with two ages and two periods set equal. The model that best fits the data is chosen as providing the best estimates of the effects.

The method of Mason, et al. is appealing since the assumptions involved appear to be minimal compared to other methods. Most investigators would be untroubled by assuming two ages, periods, or cohorts are equal. In many cases strong a priori reasons might exist for making the assumptions. In any case the model and assumptions are clearly specified. Another advantage of the method is that other independent variables can be included in the analysis along with age, cohort and period. (See, for example, Knoke and Hout (1974) and Hout and Knoke (1975). However, Glenn (1976; 1977) argues against the constrained multiple classification method because the model assumes that age, period, and cohort effects are additive, i.e., do not interact. The additivity assumption implies that age effects are the same for all cohorts and periods, that cohort effects are the same for all ages and periods, and that period effects are the same for all ages and cohorts. Glenn presents examples of theoretical and empirical evidence which suggest that the additivity assumption is unrealistic. For example, theory and empirical evidence suggest that attitudes tend to become less responsive to change as the individual ages. Thus Glenn maintains that period effects should not, as the additive model assumes, be the same for all age groups. Glenn concludes that for many additional and behavioral dependent variables, the

additivity assumption is unrealistic.

We feel that, given the alternative methods, the additivity assumption is not unrealistic. Rather than reject the method we attempt to incorporate likely interactions in our interpretation of the effects.

Visual Inspection Method

The "visual inspection method," advocated primarily by Glenn (1977)³ centers around the researcher being familiar with the patterns of variation produced by age effects only (Table 1), cohort effects only (Table 2), and period effects only (Table 3). The researcher should then be able to recognize the effects which produce the pattern found in his own data. However, Glenn stresses that there are two explanations for the pattern of variation in any cohort table. For example, the pattern shown in Table 1 could have been produced by age effects alone. There is no variation across periods or across cohorts in the table. However, the pattern also could have been produced by a combination of period and cohort effects. An increase in the percentages with each successive period combined with an offsetting decrease in the percentages with each successive cohort would produce the pattern of variation in Table 1. To further complicate interpretation, the pattern of variation also might result from a complex combination of all three effects.

Glenn maintains that the correct interpretation cannot be made by examination of the cohort table alone. He advocates the use of theory and evidence from outside the cohort table to choose the correct interpretation. For example, if one can assume on the basis of theory or other evidence that period and cohort effects result from the same causal influences, then period and cohort effects would tend to reinforce one another in the cohort table and one could rule out the interpretation of period and cohort effects offsetting each other in Table 1.

Glenn offers other interpretive aids to the visual inspector of cohort tables and concludes that by relying on theory and evidence from outside the cohort table, "there usually is only one plausible explanation for a pattern of variation predicted by pure age, cohort, or period effects" (Glenn 1977, p. 52).

The visual inspection method requires that outside evidence be employed when interpreting a cohort table. Usually assumptions about the direction or magnitude of the three effects must be made. The necessity to use outside evidence and/or assumptions is not unique to the visual inspection method. However the visual comparison between one cohort table and another is unique to this method and is not analytically appealing. When reasonable assumptions can be made, more rigorous methods of cohort analysis are available.

Summary

The validity of the various cohort analysis methodologies depends on the validity of the assumptions necessary for their use. While we prefer the Palmore method and the constrained multiple classification method, we urge consumer researchers to make their own judgments based on the nature of the problem at hand and the assumptions required for the use of the methods.

While cohort analysis has its limitations, it has much to offer to the student of consumer aging. Recognition of

³The term "visual inspection method" is the authors' and is not employed by Glenn.

the confounding of age, cohort and period in empirical measures and attempts to separate the effects using cohort analysis should produce more accurate descriptions and theories of consumer aging.

A Long-Term Solution to the Age-Cohort-Period Confounding Problem

The confounding of age, cohort, and period effects in empirical data is the result of the way researchers measure the three variables. Age is measured as chronological age, i.e., the number of years since birth. Cohort is measured as year of birth and period is measured as the year of measurement. The three operational measurements are formally dependent.

Each of the three measures is only a surrogate for more fundamental and meaningful variables. For example, chronological age is a surrogate measure of more meaningful dimensions of aging: biological, psychological, and social age (Glenn 1977; Maddox and Wiley 1976).

Maddox and Wiley (1976) and Mason, et al. (1976) point out that if we could measure these more fundamental and meaningful variables then the confounding problem would not exist. For example, Maddox and Wiley point out that replacing "period," indexed by time of measurement, with appropriate measures of environmental variables would be sufficient to remove the confounding since "there is no reason to suppose that age, cohort, and 'environment' are linearly dependent" (Maddox and Wiley 1976, p. 21). Similarly, cohort, indexed by date of birth, could be replaced by direct measures of environments which impinge on successive cohorts, and age, indexed by number of years since birth, could be replaced by direct measures of biological, psychological, and social age.

Unfortunately measures of biological, psychological, and social age are not well developed. Also, our understanding of the environmental conditions contributing to cohort and period effects is limited for many types of dependent variables. Because of these limitations, the direct measurement of the three basic dimensions of aging and the direct measurement of environments is not practical in most instances (Rosow 1978). However, as our understanding and measurement capabilities increase, it is certainly desirable to replace measures of age, cohort, and period with the more basic variables which they index. As Mason, et al. (1976, p. 905) point out, "The replacement of proxies by the variables they index is a universal goal of research." Thus a long-term solution to the age-cohort-period confounding problem is to replace the confounded variables by the unconfounded variables they index.

References

- Agnello, Thomas J. (1973), "Aging and the Sense of Political Powerlessness," Public Opinion Quarterly, 37, 251-259.
- Baltes, Paul B. (1968), "Longitudinal and Cross-Sectional Sequences in the Study of Age and Generation Effects," Human Development, 11, 145-171.
- _____ and Reinert, Guenther (1969), "Cohort Effects in Cognitive Development of Children as Revealed by Cross-Sectional Sequences," Developmental Psychology, 1, 169-77.
- Blalock, Hubert M. (1966), "The Identification Problem and Theory Building: The Case of Status Inconsistency," American Sociological Review, 31, 52-61.
- Crittenden, John (1962), "Aging and Party Affiliation," Public Opinion Quarterly, 26, 648-57.

Cutler, Neal E. (1970), "Generation, Maturation, and Party Affiliation: A Cohort Analysis," Public Opinion Quarterly, 33, 583-588.

Glenn, Norval D. (1972), "Sources of the Shift to Political Independence: Some Evidence from a Cohort Analysis," Social Science Quarterly, Dec., 494-519.

_____, (1976), "Cohort Analysts Futile Quest," American Sociological Review, 41, 900-905.

_____, (1977), Cohort Analysis, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-001. Beverly Hills and London: Sage Publications.

_____ and Hefner, Ted (1972), "Further Evidence on Aging and Party Identification," The Public Opinion Quarterly, 36, 35-44.

Hout, Michael and Knoke, David (1975), "Change in Voting Turnout, 1952-1972," Public Opinion Quarterly, 39, 52-68.

Klecka, William R. (1971), "Applying Political Generations to the Study of Political Behavior: A Cohort Analysis," Public Opinion Quarterly, 35, 358-373.

Knoke, David and Hout, Michael (1974), "Social and Demographic Factors in American Political Party Affiliation, 1952-1972," American Sociological Review, 39, 700-713.

Maddox, George and Wiley, James (1976), "Scope, Concepts, and Methods in the Study of Aging," in R. Binstock and E. Shanas (eds.) Handbook of Aging and the Social Sciences, New York: Van Nostrand Reinhold.

Mason, Karen O., Mason William M., Winsborough, H.H. and Poole, W. Kenneth (1973), "Some Methodological Issues in Cohort Analysis of Archival Data," American Sociological Review, 38, 242-258.

Palmore, Erdman (1978), "When can Age, Period and Cohort be Separated?," Social Forces, 57, 282-295.

Reynolds, Fred D. and Rentz, Joseph O. (1980), "Cohort Analysis: An Aid to Strategic Planning," Working Paper, University of Georgia.

Rentz, Joseph O. (1980), "The Impact of Aging, Period and Cohort Membership on Consumption of a Product Class," unpublished Ph.D. dissertation, University of Georgia.

Riley, Matilda White (1973), "Aging and Cohort Succession: Interpretations and Misinterpretations," Public Opinion Quarterly, 37, 35-49.

_____, Johnson, Mailyn and Foner, Anne (1972), Aging and Society: A Sociology of Age Stratification, New York: Russell Sage Foundation.

Rosow, Irving (1978), "What is a Cohort and Why?," Human Development, 21, 65-75.

Ryder, Norman B. (1965), "The Cohort as a Concept in the Study of Social Change," American Sociological Review, 30, 843-861.

Schaie, K. Warner (1965), "A General Model for the Study of Developmental Problems," Psychological Bulletin, 64, 92-107.

_____, and Parham, Iris A. (1977), "Cohort-Sequential Analysis of Adult Intellectual Development," Developmental Psychology, 13, 649-53.

_____, and Strother, Charles R. (1968a) "The Effect of Time and Cohort Differences on the Interpretation of Age Changes in Cognitive Behavior," Multivariate Behavioral Research, 3, 259-293.

_____, and Strother, C.R. (1968b), "A Cross- Sequential Study of Age Changes in Cognitive Behavior," Psychological Bulletin, 70, 671-680.

COGNITIVE AGE: A NONCHRONOLOGICAL AGE VARIABLE

Benny Barak, Rutgers--The State University of New Jersey
Leon G. Schiffman, Baruch College (CUNY)

Abstract

The determination and measurement of self perceived age as an alternative to chronological age has received almost no attention in consumer behavior and marketing research. This paper discusses a newly developed self-perceived age measure entitled "cognitive age", and presents some results concerning its reliability and its response patterns.

Introduction

While demographic variables have been a mainstay of marketing and consumer behavior research, this group of variables is typically selected and operationally defined quite automatically and usually without much imagination. Moreover, there has been a general lack of attention given to the development of new forms of demographic and "demographic-like" variables. Within a consumer behavior context, Roscoe, LeClaire and Schiffman (1977) have been sensitive to this problem when they suggested the need to refine existing demographics and develop new ones. In particular, they proposed that the age variable, the variable of concern in this paper, should be broadened so that it reflects such age-related factors as: the age of the household, age at birth of first child, age of siblings, birth order, age at first awareness of a product (or brand), age at first trial of a product and perceived age (i.e., youthfulness).

In the spirit of the Roscoe, LeClaire and Schiffman (1977) recommendations, this paper is concerned with the need for both refined and new measures of age, especially non-chronological measures of age. More specifically, this paper will endeavor to: (1) set out some of the major limitations of chronological age, (2) review the major types of nonchronological age variables, (3) propose a new perceived age variable, "cognitive age," and present some initial results pertaining to its reliability and response patterns in comparison to chronological age, and (4) offer our thinking on the future use of this age variable in various types of research, especially consumer behavior research.

The Limitations of Chronological Age

Chronological age is usually defined as either the number of years lived (Hendricks and Hendricks 1976), or as the distance from birth (Jarvik 1975). As a demographic variable, chronological age stands out from all other variables in terms of frequency of its use. In consumer behavior research it is often employed in descriptive consumer behavior studies, or in efforts to segment consumer markets.

Despite its great popularity, the use of chronological age is problematic for researchers interested in age-related research, particularly research that examines the attitudinal or behavioral patterns of the elderly. More precisely, chronological age does not lend itself well to functioning as a dependent variable; that is, it is exceedingly difficult to justify employing almost any behavioral variable of interest to consumer researchers as a predictor of chronological age. Stated differently, the unique antecedent character of chronological age restricts

its usefulness to being employed as a predictor variable.

Still further, from the perspective of the present paper, and consumer behavior research in general, the overriding shortcoming of chronological age would seem to be that it does not take into account the fact that people frequently perceive themselves to be at an age other than their birth age, and that this self-perceived or cognitive age seems to influence purchase behavior. There have been few references in the marketing and consumer behavior literature to the influence of self-perceived age on consumer behavior. A noteworthy exception is the Ford Motor Company's eventual recognition of the importance of self-perceived age in positioning its Mustang automobile:

The car was designed to appeal to young people who wanted an inexpensive sporty automobile. Ford found to its surprise that the car was being purchased by all age groups. It then realized that its target market was not the chronological young, but those who were psychologically young (Kotler 1976, p. 147).

Following this line of thinking, it might be expected that consumers would tend to consume many products according to their perceived age, and not according to their chronological age. This suggests that an individual's identity (and behavior) may depend, as much, if not more, on perceived or felt age than upon chronological age. Thus, a flexible and versatile perceived age measure would provide consumer researchers, marketers and public policy makers with an attractive alternative to relying on a chronological measure of age; and more importantly, it might provide greater insights into the patterns of aging and the consumer behavior of the elderly.

Nonchronological Age Variables

Gerontological researchers have suggested a variety of non-chronological age variables. The three broad categories of nonchronological age which are most frequently championed are briefly considered here, i.e., biological age, social age and social-psychological age.

Biological Age

Biological age is an estimate of an individual's present position with respect to his or her potential life-span (Birren and Renner 1977; Jarvik 1975). The measurement of biological age is difficult to accomplish and tends to take the approach recommended by Bell (1972); namely, the measurement of biochemical age through assays of blood serum and urine. Moreover, in discussing biological age, Bromley (1974) has pointed out that the body's organs are made up of different types of cells, so that an estimate of the biological age of any particular organ is extremely difficult to establish. In addition, the overall effectiveness of a human body is determined by the least efficient part of the system required to keep the body functioning and this varies among humans.

Social Age

Social age is the age of an individual as defined in terms of social roles and habits (Birren and Tenner 1977). It implies that age expresses an individual's place in the

social structure; which is indexed by such variables as socioeconomic status, occupation, education, race and sex (Bengston, Kasschau and Ragan 1977).

Social aging is also concerned with the different roles a person takes while passing through the life cycle. There is a continuous role change that takes place and the patterned sequencing of these roles reflects some of the changes in an individual's life (Blau 1973). Tied to this role alteration is a subjective perception of appropriate and inappropriate age-specific, society-determined norms that are an integral part of the varied roles (Bengston, Kasschau and Ragan 1977).

Social-Psychological Age

Three major types of social-psychological age measures have received special attention: (1) subjective age, (2) personal age, and (3) other-perceived age.

Subjective Age

Subjective (or identity) age measures an individual's self-perception in terms of reference age groups, i.e., "middle-aged," "elderly," or "old" (Blau 1956, 1973; Peters 1971; Rosow 1967, 1974; Ward 1977). It subjectively establishes how a person feels about such reference age groups.

Several conclusions can be drawn from the findings of research examining subjective age:

1. The majority of elderly have a strong tendency to see themselves as considerably younger than their chronological age (Blau 1956, 1973; Peters 1971; Rosow 1967, 1974).
2. The self-identification with a younger age group varies in terms of social class standing (Bengston, Kasschau and Ragan 1977; Peters 1971; Rosow 1967).
3. Women are more sensitive to the negative stereotypes associated with "elderly" and "old," and tend to see their age differently from their male counterparts (Bengston, Kasschau and Ragan 1977; Peters 1971).
4. Loss of critical roles and status also have a differential effect on subjective age perception; specifically, Neugarten (1977) suggests that especially "off-schedule" crises causing adaptational problems bring about change in the subjective perception of age.
5. Elderly who perceive themselves as younger are more likely to be innovative (Blau 1973).
6. Elderly voters who perceive themselves to be younger tend to have a more liberal and less traditional outlook on life (Bengston and Cutler 1976).
7. Those who perceive themselves to be younger are more likely to have had more education than those who perceive themselves as older (Rosow 1967, 1974; Peter 1971).
8. Finally, research indicates that subjective age is related to subjective well-being (i.e., life satisfaction or morale) and self-confidence (Bengston, Kasschau and Ragan 1977; Peters 1971).

A major problem with the subjective age construct has been the ambiguity surrounding how it has been defined and measured. Subjective age has been measured through self-

rating scales, expressed in terms of some form of nominal age-reference categorization (i.e., "young," "middle-aged," "elderly," "old"). This kind of rating is suspect since one is never sure just what such a nominal age-reference categorization means to the respondents in terms of commonly accepted units such as years.

Personal Age

Personal age is a different type of self-perceived age than subjective age. It is established by means of a self-report of an individual's age perceptions measured in terms of units of years. As operationalized by Kastenbaum, Derbin, Sabatini and Artt (1972), personal age consists of four major age dimensions: (1) feel-age (how old a person feels), (2) look-age (how old a person looks), (3) do-age (how involved a person is in doing "things" favored by members of a certain age group), and (4) interest-age (how similar a person's interests are to members of a certain age group).

In addition to the four age dimensions, respondents are also required to answer questions dealing with their individual feelings about age in general and how the four personal age dimensions compare with the respondents' chronological age. Thus, personal age is measured as a basic part of a complex procedure designed to investigate people's feelings about age. In the format employed by Kastenbaum, Derbin, Sabatini and Artt (1972), the personal age measure requires lengthy personal interviews that do not lend themselves to the types of survey research normally conducted by consumer and marketing researchers. Even though Kastenbaum's personal age measure has these limitations, it nevertheless provides a cornerstone-framework for the development of the cognitive age measure to be reported on here.

Other-Perceived Age

This final social-psychological age measure is concerned with the subjective evaluation of the age status of an individual as assessed by others. The measure is especially appropriate for examining stereotyping of age groupings and seems to be largely based on perceived physical looks and the perceived social roles of the individual(s) being observed (Lawrence 1974). This type of perceived age, while not as yet dealt with in the consumer behavior or communication-mass media literature, would seem especially ripe for exploring how consumers' perception of another person's age (e.g., a model in an advertisement) interacts with the product or brand, the usage situation, and other relevant environmental or situational factors.

An Exploration of Cognitive Age

In this section we will: (1) discuss how we operationally defined our nonchronological age measure, (2) describe the sample of elderly consumers that served as the subjects for the initial examination of our self-perceived age measure, (3) present the results of our effort to examine the reliability of the measure, and (4) contrast some of the basic characteristics of our survey results to the findings reported by Kastenbaum, Derbin, Sabatini and Artt (1972).

The Measurement and Analysis of Cognitive Age

Our self-perceived age measure, which we have labelled "cognitive age," was operationally defined in terms of four questions which were designed to correspond to the four dimensions of personal age suggested by Kastenbaum, Derbin, Sabatini and Artt (1972). The introductory statement which is either read by an interviewer, or read directly by the respondent, and the questions and response

mode for the cognitive age variable are displayed in the Appendix. To prepare survey results for analysis, responses to each of the four age dimensions (i.e., feel-age, look-age, do-age and interest-age) can either be scored separately, or an overall or composite score can be derived. As a first step, each respondent's score, for each of the four dimensions, is assigned a midpoint value (e.g., a response of "50's" was recoded to be "55"). This procedure provided an opportunity to set a numerical value (in terms of years) for a person's cognitive age (either for each of the four dimensions or a composite score), and it also allowed us to compare cognitive age and chronological age. In our preliminary research the composite score for each respondent was formed by the simple average of the midpoint values of the four age dimensions.

The Sample

The initial data gathered on the cognitive age measure was collected as part of a study that questioned some 324 elderly consumers who met the following qualifications: (1) they were female, (2) they reported their exact chronological age, (3) they were chronologically 55 years or older, and (4) they responded to all four cognitive age questions.

The respondents who were all residents of the counties that comprised a major Northeastern city, were all personally interviewed by specially trained interviewers. The questionnaire, which required about 45 minutes to be administered, focused on hair care product behavior and attitudes. Some of the major consumer behavior and consumer-related variables measured were: venturesomeness, self-confidence, dogmatism, opinion leadership, life satisfaction, club membership, a range of media habit questions, a battery of product/brand usage and experience questions, and a selected number of demographic items.

Estimation of Reliability

The cognitive age variable was subjected to three widely employed measures of reliability: test-retest reliability, Guttman's Lambda Test, and a split-half reliability test. The test-retest reliability was based on data collected during two interview sessions (separated by a three week period), conducted with the same small sample of 15 respondents. The resulting test-retest coefficient was .88. The Guttman Lambda and Spearman-Brown split-half reliability tests were respectively .86 and .85. On the basis of these three reliability estimates, it seems reasonable to suggest that the cognitive age scale has good internal consistency and is reliable.

Cognitive Age: Patterns and Regularities

In proposing that a nonchronological age variable such as cognitive age be employed along with, or in certain cases as a substitute for the traditional chronological age variable, it seems critical to attempt to ascertain the extent to which the cognitive and chronological age variables are measuring unique aspects of age. To this end, we elected to compare our findings with those reported by Kastenbaum, Derbin, Sabatini and Artt (1972) who undertook the original exploratory work which examined the four age dimensions.

Table 1 presents the percentage of our respondents who reported their chronological age in response to the four cognitive age questions. As Kastenbaum, Derbin, Sabatini and Artt (1972) found with their data, our results indicate also only a moderate degree of correspondence between the four cognitive age dimensions and chronological age. Specifically, none of the 16 percentages reported in Table 1 reach the 40 percent agreement mark. The greatest degree of agreement was 44 percent (i.e., look-age at chronological age 50's) and the least was 16 percent (i.e.,

interest-age at chronological age 80's).

TABLE 1

Decade	Percentage of Respondents Expressing Agreement Between Chronological and Cognitive Age			
	Feel-Age	Look-Age	Do-Age	Interest-Age
50's	42%	44%	28%	32%
60's	26%	35%	22%	18%
70's	35%	33%	17%	22%
80's	26%	32%	21%	16%

These results reveal that for the majority of our elderly subjects their chronological age does not adequately correspond to their perceived age as reflected by any of the four cognitive age dimensions. More importantly, the results can be interpreted as suggesting that the cognitive dimensions of age capture distinctly different aspects of age than is reflected by chronological age.

An important point that needs to be determined is the basic directionality of the decline in mean percentage of agreements that occur with advancing decades. What we want to know here is the extent to which our elderly subjects tend to view themselves as younger or older than their chronological age. In this regard, Table 2 shows the percentage of respondents who perceived themselves to be at a younger age grouping than their chronological age for each of the four dimensions of our cognitive age variable. The results reveal that our elderly respondents are considerably more likely to identify their age related feelings and actions with a younger age group than the one which is consistent with their chronological age.

TABLE 2

Decade	Percentage of Respondents Expressing A Cognitive Age That is Younger Than Their Chronological Age					Mean %
	Feel-Age	Look-Age	Do-Age	Interest-Age		
50's	54%	52%	69%	66%	60%	
60's	67%	63%	77%	71%	70%	
70's	63%	66%	80%	77%	72%	
80's	74%	68%	79%	84%	76%	

Also, the results indicate (i.e., the difference between the results in Tables 1 and 2) that the percentage of respondents who identify with an older age group than the one corresponding to their own chronological age was exceedingly small (zero for interest-age at chronological age 70's, and at the high end only five percent for feel-age at chronological age 60's).

Still further, the mean percentages appearing in the last column of Table 2 reveal that as our respondents' chronological age increases, they are more likely to identify with a younger cognitive age grouping.

It is also important to look at the degree of consistency between the four age dimensions that constitute the cognitive age variable. Table 3 reveals that the number of

cases in which each of the four cognitive age dimensions received the same decade-response as one of the other cognitive age dimensions ranges between a low of 45 percent (for interest/look age) to a high of 57 percent (for do/interest age). Performing a similar type of analysis, Kastenbaum, Derbin, Sabatini and Artt (1972) found somewhat lower levels of agreement (i.e., all were less than 50 percent -- the range was 28 percent for look/interest age and 49 percent for feel/do age). Nevertheless, the extent of inter-dimension agreement in our data is still quite low, suggesting that the four dimensions tend to reflect separate aspects of this perceived age variable.

TABLE 3

Consistency Percentage Among Cognitive Age Dimensions

	Feel-Age	Look-Age	Do-Age	Interest-Age	Mean %
Feel-Age		54%	55%	50%	53%
Look-Age	54%		48%	45%	33%
Do-Age	55%	48%		57%	53%
Interest-Age	50%	45%	57%		51%

Finally, the mean percentages in the last column of Table 3 reveal that when our respondents reported their perceived look-age, they were providing a response that had only a 33 percent mean correspondence with the other three cognitive age dimensions. In contrast, the other three cognitive age dimensions each have mean percentage agreement scores that clustered slightly above 50 percent. That the look-age had the lowest level of mean percentage agreement is consistent with the Kastenbaum, Derbin, Sabatini and Artt (1972) results. They found a 36 percent mean agreement score for look-age and the other three items that make up their personal age variable.

These results indicate that the four dimensions of our cognitive age variable are capturing aspects of age that are not adequately reflected in an individual's chronological age. Moreover, the analysis reveals the importance of examining each of the four cognitive age dimensions, for the composite measure that combines the dimensions is likely to mask some of the differences reflected by each of dimensions. Still further, the four age dimensions related to each other and to chronological age in a manner quite similar to patterns found by Kastenbaum and his associates (1972).

Conclusions and Recommendations

It is our feeling that the cognitive age variable, as well as other nonchronological age measures, will enrich the process of studying the impact of age on consumer behavior (and vice versa). Also, it is likely that nonchronological age variables will provide information not generally possible when chronological age is related to consumer behavior.

The present results also provide additional support for the increasingly popular view that the elderly should not be viewed as if they are a single group or market segment, with uniform attitudes and behavior. For instance, a consumer who is in her sixties might perceive of herself as being in her forties and identify with role models of that age. In such a case, the possibility exists that she really belongs to a different target market than indicated by her chronological age; that is, if she perceives herself to be younger, she actually might belong to a younger target market, even though her chronological age is older.

It would seem desirable that research be undertaken to examine the relationship between cognitive age (and chronological age) and a cross-section of different product categories (some likely to be age sensitive and some not) to see if the product category makes a difference in terms of age-related appeals and age-related role models.

Of course, the cognitive age variable should be employed in studies that do not solely focus on the elderly; that is, it should be used in studies that concentrate on other age groups, as well as a cross-section of age groups. For example, this research could include investigations of age perception among teenagers as compared to the elderly. (Teenagers might like to perceive of themselves as being older than their chronological age, while the elderly perceive of themselves as being younger than their chronological age.)

As a final suggestion, the cognitive age measure should lend itself to cross-cultural consumer behavior studies. Specifically, studies that explore the scope and nature of cultural differences in the perception of age, and how these differences influence various aspects of consumer behavior. It would also be desirable to assess the effect of subcultural and situational factors on the response to the cognitive age measure.

Appendix

Instruction and Format for the Self-Perceived Age Measure

Most people seem to have other 'ages' besides their official or 'date of birth' age. The questions which follow have been developed to find out about your 'unofficial' age. Please specify which age group you FEEL you really belong to: twenties, thirties, forties, fifties, sixties, seventies, or eighties.

	20's	30's	40's	50's	60's	70's	80's
I <u>feel</u> as though I am in my --	_____	_____	_____	_____	_____	_____	_____
I <u>look</u> as though I am in my --	_____	_____	_____	_____	_____	_____	_____
I <u>do</u> most things as though I were in my --	_____	_____	_____	_____	_____	_____	_____
My <u>interests</u> are mostly those of a person in her --	_____	_____	_____	_____	_____	_____	_____

References

Bell, Benjamin (1972), "Significance of Functional Age For Interdisciplinary and Longitudinal Research in Aging," Aging and Human Development, 3, 145-147.

Bengston, Vern L. and Cutler, Neal E. (1977), "Generations and Intergenerational Relations: Perspectives on Age Groups and Social Change," in Robert H. Binstock and Ethel Shanas, Eds., Handbook of Aging and the Social Sciences, N.Y.: Van Nostrand Reinhold Company, 130-159.

_____, Kasschau, Patricia L. and Ragan, Pauline K. (1977), "The Impact of Social Structure on Aging Individuals," in James E. Birren and K. Warner Schaie, eds., Handbook of the Psychology of Aging, N.Y.: Van Nostrand Reinhold Company, 327-359.

Birren, James E. and Renner, V. Jayne (1977), "Research on the Psychology of Aging: Principles and Experimentation," in James E. Birren and K. Warner Schaie, eds., Handbook of the Psychology of Aging, N.Y.: Van Nostrand Reinhold Company, 3-38.

Blau, Zena S. (1956), "Changes in Status and Age Identification," American Sociological Review, 20, 198-202.

_____ (1973), Old Age in A Changing Society, N.Y.: Franklin Watts, Inc.

Bromley, Dennis B. (1974), The Psychology of Human Aging, 2nd ed. Harmondsworth, Middlesex, England: Penguin Books, Inc.

Hendricks, C. Davis and Hendricks, Jon (1936), "Concepts of Time and Temporal Construction Among the Aged, with Implications for Research," in Jaber F. Gubrium, ed., Time, Roles and Self in Old Age, N.Y.: Human Sciences Press, 13-49.

Jarvik, Lissy F. (1975), "Thought on the Psychobiology of Aging," American Psychologist, 30, 576-583.

Kastenbaum, Robert, Derbin, Valerie, Sabatini, Paul and Artt, Steven (1972), "'The Ages of Me' Toward Personal and Interpersonal Definitions of Functional Aging," Aging and Human Development, 3, 197-211.

Kotler, Philip (1976), Marketing Management, Analysis, Planning and Control, 3rd ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Lawrence, J., "The Effect of Perceived Age on Initial Impressions and Normative Role Expectations," International Journal of Aging and Human Development, 5, 369-391.

Neugarten, Bernice L. (1977), "Personality and Aging," in James E. Birren and K. Warner Schaie, eds., Handbook of the Psychology of Aging, N.Y.: Van Nostrand Reinhold Company, 626-649.

Peters, G.R. (1971), "Self-Conceptions of the Aged, Age Identification and Aging," The Gerontologist, 11, 69-73.

Roscoe, Jr., A. Marvin, LeClaire, Jr., Arthur and Schiffman, Leon G. (1977), "Theory and Management Applications of Demographics in Buyer Behavior," in Arch G. Woodside, Jagdish Sheth and Peter D. Bennet, eds., Consumer and Industrial Buying Behavior, N.Y.: Elsevier North-Holland, Inc., 67-76.

Rosow, Irving (1967), Social Integration of the Aged, N.Y.: The Free Press.

_____ (1974), Socialization to Old Age, Berkeley, CA.: University of California Press.

Ward, Russel A. (1977), "The Impact of Subjective Age and Stigma on Older Persons," Journal of Gerontology, 32, 227-232.

LEISURE AND THE QOL CONSTRUCT:
A REVIEW AND SOME MODEST PROPOSALS

Lynette S. Unger, Miami University
Jerome B. Kernan, University of Cincinnati

Abstract

Consumer researchers have demonstrated an interest in the leisure construct both as a result of the broadened concept of consumer behavior and a specific interest in quality of life (QOL) assessment, with which leisure seems so closely to be associated. This paper reviews the voluminous literature of leisure, emphasizing construct definition and measurement, and develops a proposal for enhancing our understanding of this construct and hence that of QOL as well.

Background

Since the adoption of the broadened concept of consumer behavior (Jacoby 1976, Zaltman and Sternthal 1975), consumer researchers have addressed domains that extend considerably beyond those of traditional concern to marketing. Leisure is one of the domains considered germane under this expanded concept. The leisure experience is frequently associated with the acquisition of products/services and even more often requires that the person invoke a choice process--he/she must choose from among activities, uses of time or money budget alternatives. Perhaps most appealingly, leisure can generate satisfaction--a criterion most familiar to consumer researchers. While the for-profit leisure industry has long commanded marketers' attention, leisure's demonstrated contribution to life quality encourages research attention for social purposes as well.

The leisure domain is generally recognized as a significant component of life satisfaction across the total American population and particularly among certain segments. Leisure's importance to life quality has been noted frequently in non-empirical research across many disciplines: in psychology (Neulinger 1974), sociology (Dumazedier 1974, Riesman et al. 1950), philosophy (de Grazia 1962), anthropology (Huizinga 1950), communications (Stephenson 1967), and economics (Morgan 1962), and most empirical studies of life quality include measures of the leisure domain (Andrews and Withey 1976, Campbell et al. 1976, Cantril 1965, Canadian Minister of Supply and Services 1977, George 1979, Liu 1973, 1976, U.S. Department of Commerce 1977).

More importantly, in two recent subjective studies of life quality, satisfaction in the leisure domain was found to contribute significantly to global well-being. Andrews and Withey (1976) found that satisfaction with "how much fun you are having," a leisure domain item established by earlier perceptual mapping, was the strongest contributor to global satisfaction among twelve selected predictors. In a second subjective study (Campbell et al. 1976), satisfaction with "the ways you spend your spare time," designated as the leisure domain by the authors, was the individual domain which most strongly correlated with an index of global well-being. In multiple regression, the leisure domain was ninth in importance among twelve selected predictors.

Among certain segments of the population, leisure's contribution to perceived happiness is underscored. Because the elderly have more leisure time, satisfaction in this domain seems to have a greater impact on overall perceived well-being. The gerontological literature largely indi-

cates that there is a positive relationship between leisure participation and global satisfaction (Barfield and Morgan 1978, Edwards and Klemmack 1973, Graney 1975, Hepner 1969, Knapp 1976, Larson 1978, Lemon et al. 1972, Markides and Martin 1979, Szewczuk 1966, Toseland and Rasch 1978, Wolk and Telleen 1976). Finally, Neulinger (1974) has noted other specific groups of people with leisure-related problems (alcoholics, drug addicts and the urban poor) that seriously impair life quality.

Current trends also augur well for the continuing importance of the leisure domain. Innovative changes in work time such as flexitime, shared jobs, the four day/forty-hour week and three-day weekends will create larger blocks of time for leisure activities (Clawson and Knetsch 1966, Poor 1971, Roberts 1971). Total leisure spending (an estimated \$160 billion in 1977) has proved itself largely resistant to energy shortages and recessions, enjoying an average eleven percent annual growth rate over the past decade (Commerce America 1977). Demographic and social developments also point to growth. The sixty-five and older segment will increase dramatically over the next twenty years, the feminist movement will continue to expand leisure horizons for women, and Americans' heightened interest in health and physical fitness, now well beyond the fad stage, promises continued leisure growth.

Because it has been shown to contribute significantly to overall life satisfaction, accurate assessment of leisure quality becomes essential. Several researchers have warned against "rushing to the data" without sound constructs and measures (Churchill 1979, Reynolds and Barksdale 1978). The purpose of this paper, therefore, is to review the leisure literature, emphasizing construct definition and measurement. Based on that review a set of properties is proposed that adequate leisure domain/QOL measures would seem to require. Finally, some "next steps" toward achieving those measures are discussed.

Measurement of the Leisure Domain

Leisure has generally been measured in three ways: as time, as expenditures, or as activities. These are also used as surrogate measures of leisure quality, both as domain-specific items in QOL studies and in empirical leisure studies by psychologists, sociologists and recreation scientists.

Measuring Leisure as Time

Leisure is often equated with leisure time, and there are two common objective measures. In one method, leisure time is defined as time free from paid work (total time - work time = leisure time), a viewpoint rooted in the microeconomic labor-leisure analysis (Henderson and Quandt 1971, Robbins 1930). Trends in aggregate working conditions such as average workweek length, number of paid vacation days or holidays and length of average working life are used to trace longitudinal changes in the amount of leisure available. While the workweek has declined very little since World War II, researchers generally conclude that Americans are enjoying greater "blocks" of leisure time as a result of increased vacation/holiday time (de Grazia 1962, Moore 1971, Neulinger 1974, Wilensky 1961).

The second objective measure is the time diary or time budget. With variation among studies, respondents record their activities during short time intervals over the course of one day, and these data are aggregated. Alternatively, subjects might be asked simply to estimate how much time in a day or week they spend engaged in particular activities. The methodological problems of time diaries and budgets have been widely discussed (Bishop et al. 1975, Ennis 1968, Foote 1961, 1966, Holman and Venkatesan 1979, Szybillo et al. 1979). In these studies leisure time is generally defined as that time remaining after hours devoted to work and nondiscretionary activities are subtracted from total time. This is based on the tripartite model of time (total time - (work + nondiscretionary time) = leisure time), which has been described by a number of researchers (Bell 1975, Brightbill 1963, de Grazia 1962, Hendrix et al. 1979, Jacoby et al. 1976, Parker 1971, Voss and Blackwell 1975) and is based on a modification of the traditional labor-leisure analysis (Becker 1965, Mincer 1963, Voss 1967). Many early time diary studies were conducted during the 1930s, when the Depression's "forced leisure" generated interest in the use of free time (Lundberg et al. 1934, National Recreation Association 1934, Sorokin and Berger 1939). Renewed interest in leisure study during the 1960s and 1970s resulted in a number of new studies (Chapin 1974, Hawes 1977, Robinson 1977, Szalai 1972). In addition, most of the recent objective QOL studies contain at least one measure of leisure quality expressed as amount of discretionary time (Canadian Minister of Supply and Services 1977, Terleckyj 1975, U.S. Department of Commerce 1977).

There are three conceptual problems with measuring leisure as leisure time. First, it is often difficult to segregate work, non-discretionary and leisure time components (Bell 1975, Dumazedier 1974, Moore 1971, Murphy 1974, Schary 1971). There are many gray areas where the characteristics of each type of activity overlap. Where, for instance, does one classify time spent gambling or commuting or caring for children? Second, a priori classification of time spent pursuing a specific activity as "leisure" may be inappropriate. An activity typically categorized as a leisure pursuit may not be perceived at all as leisure by some individuals. Playing golf with friends, for example, may be a thoroughly enjoyable pastime for one person and a terrible chore for another. Even within the same person, a particular activity may shift from leisure to non-leisure, depending on situational context. A third problem with using leisure time as a surrogate measure of leisure is the question of direction. A "more is better" normative stance is implied when quite the opposite may be true, as in the case of too much leisure time among the unemployed, youth or the elderly. Inherent in all three problems is the fact that leisure has existential elements which extend beyond time constraints. Time diary researchers frequently suggested that subjective qualities differentiated leisure from other nonwork activities. Chapin (1974) and Szalai (1972) proposed categorizing activities according to their position along an obligatory/discretionary continuum. Robinson (1977) pointed to three criteria to segregate leisure from other nonpaid activities: enjoyment, discretion and spontaneity. Extensive empirical testing of subjective leisure elements was beyond the scope of these large-scale time diary studies, however.

Measuring Leisure as Expenditures

Researchers measuring leisure expenditures generally combine secondary data sources such as aggregate leisure sales data (U.S. Department of Commerce, Bureau of Census 1972a, 1972b, U.S. Department of Commerce, Bureau of Economic Analysis 1977) and aggregate leisure consumption data (U.S. Department of Labor, Bureau of Labor Statistics 1976, 1978a, 1978b). Several authors have used objective leisure indicators of this type (Canadian Minister of Supply and Services 1977, de Grazia 1962, Ennis 1968, Fisk

1963, Kaplan 1960). Measuring leisure as expenditures presents two problems similar to those encountered regarding leisure time. First, it is difficult to segregate leisure from nonleisure spending as, for example, in the case of gasoline (Eisenpreis 1971, Fisk 1963). Second, the validity problem reappears. One cannot assume direction, that more leisure spending means higher quality leisure.

Measuring Leisure as Activity Participation

Leisure is often defined as participation in recreational activities. This view of leisure grew out of the "recreation movement" of the 19th and early 20th centuries. The "classic" theories of play, prominent during this time, suggested that recreation served certain functions such as restoration or release of surplus energy (Ellis 1971). Because recreation was often viewed as a cure for social problems, it acquired a certain moralistic flavor. People were encouraged to recreate because it was good for them (Arnold 1980). There are two objective measures of activity participation. The first measure aggregates individual participation rates, association memberships, institutional attendance records or number of leisure facilities as gauges of leisure quality. Some researchers have indicated that these measures may underestimate actual participation (Ennis 1968). Indicators of this sort, however, are commonly used as leisure measures in QOL studies (Canadian Minister of Supply and Services 1977, Liu 1976, Terleckyj 1975, U.S. Department of Commerce 1977).

A second group of activity-based studies is generally not used in quality of life research, yet its prominence in the leisure literature warrants discussion. The theoretical basis is the same: activity participation frequency is used as a surrogate measure of activity enjoyment or leisure quality. In this body of research, however, activity participation frequency is used as a criterion variable. Four major types of predictors are used: socioeconomic variables, early leisure behavior, demographics or personality variables.

Socioeconomic Variables. Studies of this type use either occupational prestige/social class or occupational role to predict participation in leisure activities. Some researchers have found that differences in leisure activity participation are explained by differences in social class or occupational status (Clarke 1956, Dowell 1967, Reissman 1954, Romsa 1973, Settle et al. 1979, White 1955). This is based on the theory that leisure activities are economically determined. Others, by contrast, found these variables did not predict leisure activity (Burdge 1969, Defee, et al. 1974, White 1975). In a slightly different approach, some researchers have linked occupational role with activity participation (Bishop and Ikeda 1970, Gerstl 1961). Inconsistencies in findings using socioeconomic predictors have been explained as conceptual problems (Gerstl 1961) or resulting from the changing U.S. socioeconomic system (Settle et al. 1979).

Early Leisure Behavior. In leisure socialization studies, leisure activity participation early in life is used to predict frequency in later life. With few contradictions, studies have found that childhood leisure behavior can be used to predict activities pursued in adulthood (Burch 1969, Sofranko and Nolan 1972, Yoesting and Burkhead 1973), and that early adulthood activity is linked to later adult behavior (Kelly 1974, 1977). In studies attempting to relate pre-retirement leisure behavior with post-retirement activities, results are conflicting. Atchley (1971), Bultena and Wood (1970), Dye et al. (1973), Lambing (1972), McPherson and Guppy (1979), Peppers (1976), Pfeiffer and Davis (1971) and Szewczuk (1966) found post-retirement activity was strongly related to pre-retirement leisure. Others, however, found overall activity declined with age, as a result of health or mobility problems (Barron 1961,

Palmore and Cleveland 1976, Palmore and Luikart 1972).

Demographic Variables. In addition to age-related differences, other demographics have also been used to predict activity participation frequency. Sexual differences in leisure participation have been noted, usually as a part of larger studies (Duncan 1978). Family life cycle stage has also been used as a predictor (Landon and Locander 1979). Rural/urban residence effects on leisure activity have been demonstrated by Hendee (1969) and Knopp (1972). Some studies grouping several demographic and socioeconomic variables have isolated significant predictors of activity participation frequency (Tatham and Dornoff 1971), while others have concluded that demographics are not helpful in grouping leisuers (Romsa and Girling 1976).

Personality Variables. This genre of studies uses personality variables to predict leisure activity participation frequency. The theory of this research is that people tend to favor those activities which satisfy certain needs (Hawes 1978, London et al. 1977). Examples of this type of study include: McKechnie (1974), Procter (1962), Tinsley et al. (1977) and Weller and Brasch (1979). The amount of variance in participation frequency explained by personality varies widely across these studies.

In using leisure participation frequency as a measure of leisure quality, construct validity again comes into question. While the inconsistent findings in the four types of empirical activity-based studies may be due to theoretical problems peculiar to each area, it is also likely that the criterion variable, participation, lacks validity as a surrogate leisure measure. Some authors have stated (Meyersohn 1972, Murphy et al. 1973) that this measure assumes intrinsic enjoyment of an activity when situational/extrinsic motivation might better explain participation. Indeed, researchers have shown that situational variables such as antecedent conditions (Belk 1975, Markides and Martin 1979, Witt and Bishop 1970) and social situation (Cheek et al. 1976, Field and O'Leary 1973) can be used to explain participation.

Overall, the questionable construct validity of the three "objective" measures of leisure--time, expenditures, and activity participation--suggests a need for more subjective measures of leisure quality. While many authors have called for psychological investigation of the leisure experience (Mannell 1980, Murphy et al. 1973), objective measures remain the most widely used. The development of subjective measures of leisure domain life quality has begun in two directions: in QOL studies and in the empirical leisure literature. Among most QOL studies, the bulk of the "subjective" items are actually based on objective indicators. In the gerontological literature, for example, leisure activity is measured almost universally in objective terms: number of social contacts, organization affiliations, activity participation (Edwards and Klemmack 1973, Graney 1975, Knapp 1976, Larson 1978, Lemon et al. 1972, Wolk and Telleen 1976). Out of six "subjective" leisure indicators in the Andrews and Withey (1976) study, three are time-based: "the amount of relaxation in your life;" "your chances for relaxation - even for a short time;" "the amount of time you have for doing the things you want to do;" and one is activity-based: "the way you spend your spare time, your nonworking activities." Two measures, however, approach a subjective concept of leisure: "how much fun you are having," and "the amount of fun and enjoyment you have." Importantly, one of these items was also the leading predictor of global QOL, as discussed earlier. In a second direction, several empirical leisure studies have attempted to identify subjective experiences enjoyed during participation in specific activities (Hawes 1978, Hawes et al. 1975, Tinsley et al. 1977). These, however, are not designed to measure overall leisure quality. In summary, development of a subjective measure of leisure is needed, one that transcends specific activities and approaches leisure from a number of existential

dimensions.

Toward An Expanded QOL Measure: Subjective Properties of the Leisure Domain

As Churchill (1979) has stated, the first step in measurement development is to specify the domain of the construct in question. The literature should be reviewed, and the various dimensions of the domain should be assembled. Most definitional discussions of leisure compare it to three closely-related concepts: free time, recreation and play (Miller and Robinson 1963). As discussed, free time and recreation correspond to "leisure as time" and "leisure as activities," objectively measurable dimensions of leisure. Play, like leisure, however, is frequently described in more subjective terms. Some authors distinguish the two concepts (de Grazia 1962, Pieper 1952). Leisure is described as being restful or contemplative and is limited to adults while play is generally seen as the activity of children and animals and is characterized by its apparently "useless" nature. Conversely, other authors have suggested that it is impossible to segregate the subjective properties of leisure from those of play, and that the concepts are inextricably intertwined (Arnold 1980, Ellis 1971, 1973, Kaplan 1975). Some researchers use the terms synonymously, implying that leisure is simply the play of adults (Berlyne 1969, Smith 1980). All the major subjective properties of leisure described below have also been used to describe play, attesting further to the similarity of the two constructs. Because the concepts are so closely related, it appears appropriate to include both leisure and play in the construct domain.

There appear to be several major subjective properties of leisure discussed in the literature: intrinsic satisfaction, perceived freedom, high involvement, arousal, feelings of mastery and spontaneity.

Intrinsic Satisfaction

Most leisure definitions suggest that the leisure experience offers pleasure or gratification. Leisure is pursued for its own sake and is an end in itself rather than a utilitarian means to an end (Berlyne 1969, Caillois 1961, de Grazia 1962, Douglass 1960, Kaplan 1975, Kerr 1962, Lee 1929, Piaget 1962, Riesman 1963, Schary 1971, Smith 1980, Stephenson 1967). Because of its intrinsic rewards, leisure offers a means of self-expression and affirmation (Dumazedier 1974, Farina 1969, Murphy et al. 1973, Pieper 1952).

Perceived Freedom

Leisure is often described as "free," something one does voluntarily, without coercion or obligation (Caillois 1961, Douglass 1960, Ennis 1968, Huizinga 1950, Kaplan 1975, Stephenson 1967). At least two authors have maintained that perceived freedom is the single precondition of subjective leisure (Bregha 1980, Neulinger 1974). As noted earlier, many time-oriented researchers recognized this subjective leisure element and proposed an obligatory/discretionary continuum to segregate leisure from other nonwork activities (Chapin 1974, Hendrix et al. 1979, Robinson 1977, Szalai 1972).

High Involvement

Some have suggested that true leisure means high involvement or total absorption in an activity. When we leisure, we become so involved that we enter a microcosm distinct from daily life (Anderson 1961, Berlyne 1969, Douglass 1960, Dumazedier 1974, Erikson 1950, Foote 1966, Huizinga 1950, Kaplan 1975, Lee 1929, Piaget 1962, Riesman 1963, Stephenson 1967). Some have described this experience as a fantastic escape from reality while others simply call it an interlude from the ordinary. While the micro-

cosm may be relatively free from social control, it is governed by its own rules and order (Caillois 1961).

Arousal

Several researchers indicated arousal or tension are present in leisure (Berlyne 1969, Ellis 1971, Lee 1929, Murphy et al. 1973, Piaget 1962, Riesman 1963). Novelty-seeking, exploration and risk taking behavior have been noted. Also related to arousal is the repetitious or ritualistic nature of leisure activity (Huizinga 1950, Stephenson 1967).

Mastery

Feelings of mastery have been identified as another property of the subjective leisure experience. One has the opportunity to test one's self (Murphy et al. 1973) or conquer the environment in some way (Berlyne 1969, Erikson 1950, Piaget 1962, Riesman et al. 1950, Riesman 1963).

Spontaneity

Finally, several authors have noted leisure's spontaneous nature (Lee 1929, Piaget 1962, Riesman 1963, Robinson 1977).

It seems desirable to relate currently-used objective measures and the proposed subjective properties. Churchill (1979) stressed that bringing together widely varying definitions of a construct was a way to synthesize current knowledge and to avoid future confusion. Old measures of a construct should be considered along with new viewpoints in order to avoid "throwing the baby out with the bath-water." Importantly, QOL researchers also maintained that, while subjective measures may offer improved construct validity in measuring satisfaction, they must be tied in some way to objective conditions in order to contribute to policy making (Schneider 1976). Obviously, in terms of measurement development and policy application, it is necessary to relate the objective conditions and subjective properties of leisure.

The two-dimensional matrix in the Exhibit provides a framework for relating objective measures and subjective properties. The rows of the matrix represent the major objective indicators of leisure quality. The $E_{i,j}$ s equal the general evaluation of the objective conditions across subjective properties. For example $E_{2,1}$ would indicate how much amount of expenditures contributes to leisure quality. The columns show the major subjective properties of leisure, and the $E_{i,j}$ s indicate the evaluation of subjective properties across objective conditions. $E_{2,2}$ would measure, for example, how perceived freedom across all objective leisure conditions contributes to overall leisure quality. The grand mean ($E_{..}$) would provide a general evaluation of leisure quality "as-a-whole." It is important to note that the lists of properties and objective conditions shown are by no means complete, and that a general evaluation of leisure quality would optimally include as many measures as possible. Finally, on a more specific level, the $E_{i,j}$ s would measure leisure quality under a particular objective condition with respect to a particular subjective property. The relationship between each objective condition and subjective property could then be determined.

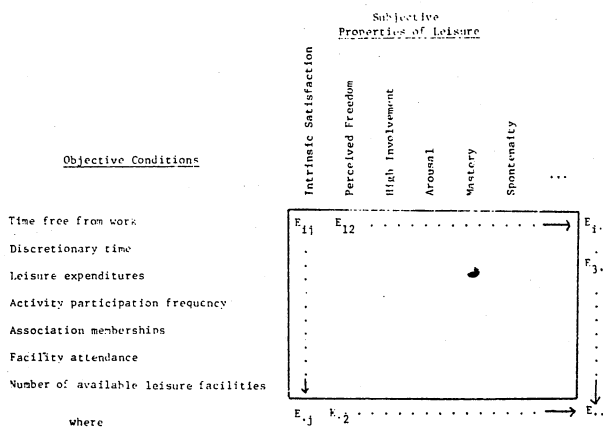
Next Steps

Next steps include the development and integration of subjective and objective leisure measures and the exploration of personal and situational antecedents of the subjective leisure experience. The development of subjective leisure measures should be continued, following the steps Churchill (1979) has enumerated. This would entail a more extended review of the leisure literature to add other properties not included above. The construct domain should be adequately sampled and items generated for each dimension. Following data collection within the proposed framework, the measures should be purified and reduced to a parsimonious set, and reliability should be assessed. Convergent validity of the measure might be evaluated by correlating an overall leisure quality measure developed from the various items with existing one-item leisure domain satisfaction measures such as those used in subjective QOL studies. Discriminant validity could be assessed by correlating the overall leisure quality measure with QOL study measures of other, theoretically unrelated domains (satisfaction with safety, home or money, for example). In assessing construct validity, one might expect the overall leisure domain satisfaction measure to contribute significantly to global satisfaction, as demonstrated in past subjective QOL studies. The matrix discussed above might be used to relate subjective properties with objective measures for policy purposes. This could be done using simple Pearsonian correlations or canonical correlation. Further, analysis of variance or multiple regression could be used to assess the subjective properties and objective measures as predictors of leisure domain satisfaction. Research could be conducted on an individual level of analysis or between communities, as done by Schneider (1976).

Further investigation of the antecedents of subjective leisure is warranted, paralleling similar studies of objective measures. As discussed above, socioeconomic variables, early leisure behavior, demographics and personality variables have been related to objective leisure measures, albeit often weakly. These same personal variables might be tested as predictors of subjective leisure properties as well. Further, the subjective leisure experience appears to be highly situational, varying across different experiences with the same activity or even within the same person. This situational quality of leisure also has been discussed by several researchers (Huizinga 1950, Stephenson 1967). Some have begun to investigate the effects of situational variables such as antecedent conditions or social surroundings on objective measures (Field and O'Leary 1973, Witt and Bishop 1970). Their effects on subjective properties might also be

EXHIBIT*

Dissagregation of the Leisure Domain



- $E_{i,j}$ = Affective evaluation (response) under a particular objective condition with respect to a particular subjective property
- $F_{i,j}$ = General evaluative response to an objective condition (across properties)
- $E_{.j}$ = General evaluative response to a property (across objective conditions)
- $E_{..}$ = General evaluative response to all objective conditions over all properties; i.e., to leisure as a whole

*After Andrews and Withey (1976)

explored.

Summary

Since there are many reasons for assessing leisure, there are many measures for it, each with its rationale and proponents. There seems little merit to arguing for subjective over objective assessment, particularly since the benefits of each can be combined. It should be recognized, however, that such a combination, although useful, is merely an initial step toward the systematic explication of the leisure domain and our ultimate understanding of its contribution to QOL evaluation.

References

Because of space limitations, this paper's references (which number 137) must be obtained from the lead author: Lynette S. Unger, Department of Marketing Management, Miami University, Oxford, OH 45056.

THE CONSUMER AND THE HEALTH CARE PROCESS

Lynn Langmeyer, Wright State University
George Miaoulis, Wright State University

Abstract

A review of perceived past and current directions in health care delivery research is presented to set the stage for a paper on a suggested future direction in this critical area. Consumer data on attitudes toward health care is presented and discussed in order to reveal a relatively unexplored direction within which consumer research can make unique and important contributions: the interpersonal dimension of the health care consumption process. The paper concludes with suggestions for specific research of both a theoretical and practical nature.

Introduction

The focus of this paper is consumer behavior issues of health care consumption. It is generally accepted that many consumers are dissatisfied with the quality of health care they receive and, at the same time, feel unable to remedy the situation. In this paper, we suggest that the health care provider - patient interaction, the interpersonal dimension of the system, has often been neglected as an important source of these dissatisfactions. At the same time, the interpersonal dimension represents an important starting point for dealing with these dissatisfactions. The objective of the paper is to provoke consideration and exploration of health care consumption behavior and not, at this point, to provide a detailed analysis of consumer dissatisfaction with the health care process.

Health Care Delivery -- Where's the Consumer?

Zaltman and Vertinsky (1971) mention, rather briefly, in their article "Health Service Marketing: A Suggested Model," that one of the "perceived barriers to taking proper health action is the psychological distance between patient and physician." This distance, they suggest, "may cause patients to ignore the treatment prescribed by the physician." Health care professionals, particularly nurses, have been aware for many years of the inequality inherent in the physician-patient relationship (See M. Karmar 1972) and the negative effect of this imbalance on the patient. However, the physician component of the profession has changed little in the past ten years with respect to the interpersonal dimension, and little consumer-oriented research has been done in this critical area. Indeed, although the interpersonal dimension has been a relatively unexplored issue of health care research, we can find hints of its importance in health care literature.

For example, Hochbaum (1969) discussing consumer participation in health planning states:

We must recognize and accept the fact that it is the consumers of health services who are the final and proper judges of what kinds of services they want, how they want them delivered, what form they should take and in what setting they should be provided.

It is only on the medical and technical details that the health professions have any exclusive right to make decisions.

To further emphasize, support and underscore this issue, we found the majority of specific dissatisfactions that

the Wright State study panel members experience with health care (aside from cost) can be traced to what Zaltman and Vertinsky call "psychological distance," what Kramer calls "power imbalance" and what Hochbaum calls "inequity." As marketers, if we are to assist health care professionals in addressing these dissatisfactions, then we must begin to understand, explain, and eventually model general health care consumer behavior. We believe the first step in this research process is to examine the issues of health care consumption behavior with emphasis on the interpersonal dimension. We will identify and examine these issues by first discussing previous directions in health care literature and research, which suggest or consider but do not thoroughly explore consumption behavior and then, in the balance of the paper, directly address the issue of consumer attitudes and the consumption process.

Identified Research Issues (Directions?)

The literature and research in the area of the consumer and health care appears to address four major issues and one minor issue. These research issues are:

- 1) consumer participation in the planning and delivery of health care
- 2) consumer satisfaction with delivery systems
- 3) health care decision models
- 4) health promotion
- 5) health care facility location

Each of these is summarized below from the perspective of the role of and impact on the consumer.

Planning and Delivery

The passage of Public Laws 88-164, and 93-64 in 1963, 1966, and 1974 respectively, require that representatives of the consumers of health care facilities be participants in the planning and operating councils of such facilities. The conceptual literature and empirical research in this area, with respect to the consumer, focuses on issues such as, "The Whys and Why Nots of Consumer Participation" (Thomson 1973), "Community Health Planning or Who Will Control the Health Care System" (Brown 1972), "Consumer Participation in Health Planning: Toward Conceptual Clarification" (Hochbaum 1968) and "The Merits of Using Experts or Consumers as Members of Planning Groups: A Field Experiment in Health Planning" (Nutt 1976).

These articles have appeared in public health journals or management journals and emphasize the organizational ramifications and problems concomitant with consumer participation in the planning and delivery of health care. The discussions and research results are useful for educational, skill training and general managerial purposes. They are, however, relatively unuseful for exploring patient health care consumption behavior.

The issue of consumption behavior is not completely ignored. Nutt (1976), in discussing the role of consumers as members of planning groups makes the following observations: "people should have a role in articulating their needs"....yet we are "co-opting health consumers by giving them a voice in the decision process and not fundamentally altering the decision process itself (e.g. the criteria selected or criteria weightings used to make

decisions)." We would add that additional co-opting occurs when health consumers are given no role or voice in evaluating how well articulated needs are being met which raises the issue of consumer satisfaction with delivery systems.

Consumer Satisfaction Issues

The second major theme, appearing at this time almost exclusively in public health journals, concerns consumer satisfaction/dissatisfaction with specific types of delivery systems -- HMOs (Pope 1978), prepaid group practice (Weinerman 1954; Bashshur, et.al. 1967; Tessler and Mechanic 1975), health plan coverage (Gerst, et.al. 1969), etc. The results of these studies are somewhat confusing, perhaps, as Kramer (1972) suggests, due to the lack of differentiation between the "outcome of medical care as contrasted to the activities involved in obtaining or being given that care." She raises the issue of the interpersonal dynamics in health care consumption and Weinerman (1964) reports "disappointment with the degree of personal interest shown by the doctor" attributing this patient perception to conflict between lay and professional concepts of proper care.

This interpersonal theme is further suggested by others. Pope (1978) states his respondents are "more satisfied with the technical aspect of HMOs than the interpersonal components;" Friedman and Di Matteo (1979) mention "power differential," "the patient as an object," "the impact of health care professionals on patient's reactions," "satisfaction comes from treatment as a person," and "the needs and values of patients determines 'best' procedures;" O'Connor (1978) states "accurate and complete information about patients' needs and satisfactions" must be available. Weinerman (1964) summarizes these ideas in his statement, "the most neglected and least appreciated element in medical care planning and evaluation is overall perceptions of patients...the reactions of the consumer to medical care."

Generally, despite these selected comments, the satisfaction/dissatisfaction health care studies are concerned with the implications for health care organizations and not health care consumption itself. For example, Swan and Carroll (1980) conducted an extensive literature review to explore the possible conceptual and research contributions the patient satisfaction literature might provide for research on consumer satisfaction. It is clear from the reviewed research that "physician-patient interactions" are strongly related to patient satisfaction. However, explanations as to why this relationship exists or how it operates in health care consumption are not offered in the literature reviewed.

Decision Models

The contribution of consumer behaviorists to health care marketing, in the form of health care decision models, comprises the third major research theme. These studies have appeared in marketing journals and conference proceedings and have concentrated on modelling action/no action decisions with respect to illness (Zaltman and Vertinsky 1971) or exploring the demographic and psychographic factors associated with the decision to seek medical care (Wortzel 1976). The emphasis appears to be on social marketing and segmentation issues of health care rather than on identification and exploration of consumption behavior issues. As such these articles contain important implications and suggestions for preventive health care behavior changes. However, we feel it is also important to follow the consumer in this process beyond the initial action/no action state. Indeed, it is possible that action/no action decisions are tied inextricably to prior experiences in health care situations, particularly the interpersonal aspects of those situations,

and the existing models do not demonstrate this orientation.

Health Promotion

The fourth area, health promotion, addresses health educators attempts to increase utilization of health services, and to stimulate changes in health behaviors, and changes in health attitudes, knowledge, and values. The focus of Keyes (1972), Rosenstack (1966 and 1974) and most recently Green (1979) has been on those activities designed to increase consumer participation in the health care system and to reinforce consumers' positive health behaviors. Of the five areas being reviewed, health promotion comes closest to identifying and responding to consumption behavior issues. However, health promotion currently fails to consider the consumer's interpersonal needs as an integral element of the promotion and reinforcement process. Some recent research by Green (1979) and Miaoulis (1979) suggests that health educators are beginning to address this issue and we are suggesting that consumer behavior researchers do the same. There is substantial opportunity for marketers to assist in the development process as health educators have not historically recognized the importance of consumer decision processes.

Facility Location

The final theme, rather undeveloped at the present time and therefore somewhat minor, is the use of mathematical models to determine health care facility location. Parker and Srinivasan (1976) report on a modelling approach which incorporates consumer preferences in planning rural primary health care facilities. Their study indicates that the method, although elaborate and time consuming, has "substantial reliability and predictive validity" as well as viability. This aspect of health care delivery, however, is only tangentially related to our field of interest.

Consumer Attitudes

Consumption Process

The specific theme conspicuously absent (aside from passing mentions here and there) in the previous discussion of health care research (which is by no means meant to be exhaustive) concerns issues surrounding consumer health care behavior subsequent to participating in the location, planning, and operating of a health care delivery system and the decision to use it. In other words -- what factors affect the behavior of a consumer, negatively and positively, as he or she is in the process of consuming general medical care and/or local hospital services?

Imagine that you are a consumer in the examining room of your physician's office. You have just been informed that the results of the series of tests you underwent last week indicate surgery is necessary. Your physician says, "No, it is not really serious, but should be taken care of as soon as possible." This typical exchange between consumer and physician leads to some important consumption issues:

1. Does the consumer participate in the surgery decision making process?
2. Does the consumer question the physician as to the benefits of the surgery and its outcomes? Unfortunately, there is no money back guarantee or trial size.
3. If the consumer seeks a second opinion, how are opinions compared with respect to quality, cost, etc.?

4. If the consumer decides to undergo surgery, does he or she have the authority to determine the who, when, where and under what conditions it will be performed?
5. To summarize: Is the consumer a participant in the consumption process?

Now imagine that you have decided to undergo surgery and are seated in the waiting room of "your" hospital, filling out insurance forms since that is the necessary first step in your consumer/patient role. How do you feel? Is there a knowledgeable person available with whom you can share your feelings of anxiety and powerlessness? Are you being treated like "the patient" or "a case"? This scenario certainly suggests you have not been a consumer in the traditional usage of that term -- involved decision maker.

Illustrative Data

To illustrate the consumer's lack of involvement in the health care consumption process, we draw upon a 1978 Wright State University consumer panel study designed to assess the overall quality of medical care and hospital services in the Dayton, Ohio SMSA.* While the study was not specifically designed to analyze the health care consumption process, several open-ended questions illustrate the consumer lacks of involvement, and warrant reporting here.

The data from the panel study suggests that consumers are feeling helpless, depersonalized, uninformed and scared. Panelists were asked to respond to the question, "If you could change one thing about the general level of hospital services, what would it be?" The following are selected verbatim comments directed explicitly at the interpersonal component of services:

- Let the patient know more, explain more.
- Take more time to be sympathetic and explain what is happening and why.
- Improve education of personnel, especially to be more compassionate and devoted to patients.
- More caring doctors.
- Quicker more efficient personal attention.
- More communication between physician and patient.
- More individualized care.
- More personal concern.
- Be more aware of the family in regard to the truth of the conditions of the patient.
- More empathy with patients.
- Medical professionals should look past their income levels and more into what they're being paid to do.
- Quality of 'bedside' manner of professionals.
- More personalized service. Patients are scared humans, not just bodies.
- More concern, more interest in people than money.

The following summary to the question, "If you could change one thing about the general level of hospital services, what would it be?", indicates clearly the consumers' need for improving communication between the health care providers and the patients.

Suggested Improvements	Response
Cost containment	25%
Better communications with patients	20
Emergency room services	15
Better preventive medicine	15
Better food services	15
All other comments	10

*The Wright State University Consumer Panel is a randomly selected demographically cross sectional sample of 1,000 households in the Dayton, Ohio SMSA.

The results suggest that patients feel they are not active participants in the health care consumption process. They have little or no control and power within the system, and their expressed dissatisfactions (aside from cost) are with interpersonal and other non-technical aspects of the health care system.

Quality of Health Care

It is also evident from panel responses that quality health care is defined in terms of both technical aspects and interpersonal aspects. When asked, "How would you define quality care within a hospital?," panel members gave considerable attention to the non-technical, interpersonal dimension:

Quality of Hospital Care	Response
Technical aspects	
Qualified staff	24%
Efficient staff	16
Non-Technical aspects	
Patient concern	22
Patient care	9
Friendly personnel	7
All other comments	22

Panel members were asked a similar question regarding their perception of quality care by physicians. Their responses to "How would you define quality care by a physician?," also reflect concern for both the technical and interpersonal aspects of health care.

Quality of Physician Care	Response
Technical aspects	
Diagnosis	25%
Medical knowledge	15
Proper treatment	11
Non-Technical aspects	
Patient interest	17
Personal communication	11
Physician availability	4
All other comments	17

The interpersonal dimensions of health care are equally as important, and perhaps more important to consumers. To fully participate in the consumption process, and to have the opportunity to assess issues such as quality care, consumers require interpersonal attention and communication. The technical and non-technical aspects of health care cannot continue to be treated as separate elements of the delivery system. Illness is an inherently anxious condition and hospitals are scary places. If consumers' health care are to be met satisfactorily then we must design human systems to meet them and not simply buildings and support facilities.

Implications

The data strongly suggests that the non-technical, interpersonal dimension of health care consumption, generally ignored or given cursory attention in health care consumer behavior studies, must be considered and studied regardless of the "major themes" being investigated. The issues of lack of active participation, lack of communication, feelings of helplessness, powerlessness and depersonalization not only influence but perhaps determine health care behavior.

If, as consumer behavior researchers, we are to contribute to the understanding and explaining of health care behavior, and provide genuinely useful information to health care professionals, then we must address and systematically investigate the issues of the consumer and the health care consumption process. To this end, we recommend that the following research be undertaken:

- Examination of the formal and informal communications networks in various health care settings.
- Patient oriented research to identify and investigate the role inequities of the interpersonal patient physician relationship.
- Physician oriented research to identify mechanism through which physician (and medical students) can be taught to emphasize the interpersonal dimensions of health care delivery.
- Broader conceptualization and integration of the health care process through more interdisciplinary research.
- In-depth research on consumer motivations and attitudes toward and benefits sought from health care interactions.

The research studies outlined above will contribute to understanding consumer attitudes and the health care consumption process.

References

- Bashshur, R. L., Metzner, C.A., and Worden, C. (1967), "Consumer Satisfaction With Group Practice -- The CHA Case," American Journal of Public Health, 57, 1991-9.
- Brown, Douglas R. (1972), "Community Health Planning or Who Will Control the Health Care System?," American Journal of Public Health, 62, No. 10, 1336-39.
- DiMatteo, M. Robin (1979), "A Social Psychological Analysis of Physician-Patient Rapport: Toward a Science of the Art of Medicine," Journal of Social Issues, 35, No. 1, 12-33.
- Friedman, Howard S. and DiMatteo, M. Robin (1979), "Health Care As An Interpersonal Process," Journal of Social Issues, 35, No. 1, 1-11.
- Gerst, A., Rogson, L. and Hetherington, R. (1969), "Patterns of Satisfaction with Health Plan Coverage: A Conceptual Approach," Inquiry, 6, 37-51.
- Green, Lawrence W. (1979), "How to Evaluate Health Promotion," Hospitals, 106-108.
- Harris, Reuben T. (1978), "Improving Patient Satisfaction Through Action Research," Journal of Applied Behavioral Science, 14, No. 3, 383-402.
- Hochbaum, G. M. (1969), "Consumer Participation in Health Planning: Toward Conceptual Clarification," American Journal of Public Health, 59, No. 9, 1698-1705.
- Keyes, Lyn Ford L. (1972), "Health Education in Perspective: An Overview," Health Education Monographs, 31.
- Kotler, Philip and Zaltman, Gerald (1971), "Social Marketing: An Approach to Planned Social Change," Journal of Marketing, 35, 3-12.
- Kramer, Marlene (1972), "The Consumer's Influence on Health Care," Nursing Outlook, 20, No. 9, 547-78.
- Miaoulis, George (1979), "Good Samaritan Hospital and Health Care Centers: Conceptual Analysis of Consumer Panel Study," Unpublished Working Paper, Wright State University.
- _____ (1979), "Marketing and Health Education," Keynote address presented at Ohio Health Education Institute, Kenyon, Ohio.
- Nutt, Paul C. (1976), "The Merits of Using Experts or Consumers as Members of Planning Groups: A Field Experiment in Health Planning," Academy of Management Journal, 19, No. 3, 378-94.
- O'Connor, John T. (1978), "The Radically New Medical Care Industry," Journal of Applied Behavioral Science, 14, No. 3, 275-305.
- Paap, Warren R. (1978), "Consumer-Based Boards of Health Centers: Structural Problems in Achieving Effective Control," American Journal of Public Health, 68, No. 6, 578-82.
- Parker, Barnett R. and Srinivasan, V. (1976), "A Consumer Preference Approach to the Planning of Rural Health-Care Facilities," Operations Research, 24, No. 5, 991-1025.
- Pope, Clyde R. (1978), "Consumer Satisfaction in a Health Maintenance Organization," Journal of Health and Social Behavior, 19, 291-303.
- Rosenstock, Irwin (1966), "Why People Use Health Services," Milbank Memorial Fund Quarterly, 44, 194-227.
- _____ (1974), "The Health Belief Model and Preventive Health Behavior," Health Education Monographs, 2, 354-386.
- Swan, John E. and Carroll, Maxwell G. (1979), "Patient Satisfaction: An Overview of Research -- 1965 to 1978," Paper presented at the Fourth Annual Conference on Consumer Satisfaction/Dissatisfaction and Complaining Behavior, Indiana University, Bloomington, Indiana, October 4-5, 1979.
- Tessler, R. and Mechanic, D. (1975), "Consumer Satisfaction With Prepaid Group Practice," Journal of Health and Social Behavior, 16, 95-113.
- Thomson, Ruth (1973), "The Whys and Why Nots of Consumer Participation," Community Mental Health Journal, 9, No. 2, 143-50.
- Weinerman, E. Richard (1964), "Patients' Perceptions of Group Medical Care," American Journal of Public Health, 54, No. 6, 880-89.
- Wortzel, Lawrence H. (1976), "The Behavior of the Health Care Consumer: A Selective Review," Proceedings, Association for Consumer Research, Sixth Annual Conference, 295-301.
- Zaltman, Gerald and Vertinsky, Ilan (1971), "Health Service Marketing: A Suggested Model," Journal of Marketing, 35, 19-27.

ATTITUDES TOWARD PUBLIC POLICY ALTERNATIVES TO REDUCE AIR POLLUTION*

David A. Aaker, University of California, Berkeley
Richard P. Bagozzi, Massachusetts Institute of Technology

Research into public concern with environmental quality has mainly attempted to identify causes or determinants of environmental concern. Although there have been exceptions (Koenig 1975, Kinnear, et. al. 1974), the evidence suggests that concern is associated with higher levels of education and socioeconomic status, younger age groups, urban dwellers and nonminorities (see, e.g., Erskine 1972, Tognacci, et. al. 1972). With respect to political orientation, the evidence indicates that Democrats, liberals and those less politically alienated are more concerned with the environment (see, e.g., Tognacci, et. al. 1972, Dunlap 1975, Koenig 1975, Buttell and Flinn 1978). There is also evidence that concern is related to perceived consumer effectiveness in influencing environmental quality (Kinnear, et. al. 1974).

This article will attempt to extend this research stream in several directions. First, the study will focus upon a specific domain of concern: concern for air pollution. Past research has tended to use a much broader definition of concern which would generate constructs like concern for the overall environment, ecology, or pollution. When the constructs consist of multi-item scales, the individual items are often quite diverse (e.g., Weigel and Weigel 1978). The problem with the use of such a general concern construct is that people can hold a variety of attitudes and beliefs toward the many components of the environment. If these components are combined, especially when they are combined in an unspecified way, the meaning of, and attitudes toward, the resulting construct may be unstable. Second, and perhaps most important, the research will study not only the antecedents of concern but also the impact of concern and other constructs upon the advocacy of a set of public policy alternatives such as closing freeways or adding additional pollution control equipment on automobiles. Previous research with few exceptions has been content to consider generalized concern as the sole dependent variable of interest. However, concern is more properly treated as an intervening variable. Inquiry should be extended to the link between concern and various public policy alternatives. The practical question is whether concern becomes translated into support for public policy alternatives that will be costly, inconvenient, or both. It is this link that is really important to understand the process by which an issue such as pollution gets addressed in our system of government.

Among the few studies that have looked beyond concern in order to investigate possible action that might be taken are several which use a "willingness to pay" construct (cf National Wildlife Federation 1969 and 1970, Schuller and Ervin 1972). Although such a construct is somewhat artificial and therefore less useful than more realistic public policy alternatives, the results of these studies are suggestive. Basically, they indicate that, although a majority of people are willing to pay something to ameliorate problems like pollution, the amounts are likely to be quite modest. Thus, we might hypothesize that a

concern for problems of pollution may not necessarily be transmitted into support for public policy programs if those programs have associated costs. More particularly, the link between concern and support for public policy alternatives should be related to the costs of those alternatives for the respondents. A natural alternate hypothesis is that concern will indeed generate support for public policy alternatives.

Third, this study will introduce a variety of constructs which normally have not been considered in past investigations but which will act as important control variables, add useful exogenous variables, and introduce intervening variables enriching the model building process. The determination of the link between concern and support for public policy alternatives is potentially confounded by many factors which can be controlled. Thus, in this study, a liberal-conservative construct is included to control for individuals' inclinations toward governmental solutions; and a miles driven construct is included to control for each respondent's attachment and reliance upon the automobile. Useful exogeneous variables included are objective measures of actual pollution levels. The few studies that have used pollution variables have found that these can influence attitudes (Degroot et. al. 1966). Several endogenous variables have the potential to increase our understanding of the attitude formation process. One such variable is the attitude toward the efforts of automobile companies to improve air pollution and another is beliefs as to the change of air pollution. Instead of the conventional two-way analysis, these variables are explored in the context of a structural equation model.

In sum, this study hypothesizes that whether people will advocate certain policy recommendations affecting themselves, industry, and government will depend upon (1) their beliefs that the problem in their immediate environment is getting worse, (2) their concern as to the seriousness of the perceived problem, (3) their political orientation, (4) their demographic characteristics, (5) their use of the automobile, and (6) the amount of actual pollution in their region. The particular sequence of relationships among these variables is developed below.

Method

In May of 1973 (before the OPEC oil embargo), a total of 521 respondents were interviewed in their homes as part of the regular California Poll conducted by the Field Research Corporation. A cluster sampling design was used with over 100 randomly selected starting points for each cluster. Interviewing was conducted late afternoons and evenings on weekdays and all day on the weekend. Up to four call-backs were employed. One adult per household was selected systematically to provide a representative age and sex distribution. The sample was divided on a random basis into an analysis sample (n=348) and a validation sample (n=173).

Exhibit 1 describes the eight endogenous variables and the eight exogenous variables developed from questions in the survey. The first five variables can perhaps be considered behavioral predispositions or policy variables in that they reflect respondents' positions on public policy legislative alternatives to combat air pollution. These include the installation of anti-pollution devices, gasoline rationing,

*The authors are grateful to the Field Research Corporation which originally collected the data used in this study, the University of California State Data Program which provided access to it, and the Institute of Business & Economic Research at the University of California which supported the study: Of course, these organizations are not responsible for the analysis or interpretation of data appearing herein.

EXHIBIT 1

VARIABLES EMPLOYED IN THE RESEARCH

Endogeneous Variables

1. DEVICE All cars and trucks to install anti-pollution devices that could cost up to 300 dollars. (Four point disagree (1) to agree (4) scale.)
2. RATION Ration gasoline. (Four point disagree - agree scale.)
3. FREEWAY Sum of two questions proposing to reduce freeway driving by closing of freeways or slowing freeway construction drastically. (Four point disagree - agree scale.)
4. TRANSIT Sum of two questions proposing to buy busses with gasoline tax and set aside exclusive bus lanes and to build a rail rapid transit system. (Four point disagree - agree scale.)
5. AIR Sum of three questions on whether the auto air pollution standards are too strict. (Three point scale from too strict (1) to too lenient (3) and a two point scale from easing up some (1) to standards should be met (2).)
6. CONCERN How serious is the air pollution problem? (Not at all serious (1) to extremely serious (4).)
7. BELIEF In the last year has air pollution become worse? (Not as bad (1) to worse (3).)
8. AUTCO Are auto companies doing all they can to reduce the air pollution problem? (Doing all they can (1) to could be doing more (2).)

Exogenous Variables

9. MILES Approximate number of miles driven in an average year. (Five point scale.)
10. LIBERAL Do you consider yourself a conservative or a liberal. (Five point scale from strongly conservative (1) to strongly liberal (5).)
11. NO2 The amount of NO2 air pollution in the area in which the respondent lives. (The state was divided into ten areas.)
12. OZONE The amount of ozone air pollution in the respondent's area.
13. DUST The amount of dust air pollution in the respondent's area.
14. INC Family income. (Seven point scale.)
15. AGE Age. (Twelve point scale; 18-20, 21-24, 25-29, etc.)
16. ED Education. (Seven point scale.)

restricting freeway use, encouraging rapid transit, and reducing automobile air pollution standards. Three alternatives were among those being seriously considered by state and local governmental units throughout the country who were under a mandate from the Clear Air Act of 1970 to submit plans for meeting the Environmental Protection Agency's air quality standards.

The CONCERN variable, a central variable in previous research, is based upon a question asking how serious is the air pollution problem. The BELIEF variable indicates the degree to which the respondent thinks that pollution has become worse during the past year. Finally, the AUTCO variable reflects the respondent's judgment as to whether the automobile companies have done all that they could to reduce air pollution.

The first exogenous variable represents the miles driven by the respondent, and the second reflects a liberal/conservative self disclosure. It was felt that these variables should explain people's position with respect to public policy issues. Someone who drives more than average might feel that legislation is unnecessary or inconvenient. Further, a liberal orientation should lead toward a tendency to look to the government for solutions. These variables are not only interesting in themselves, but serve as important control variables when trying to determine the predictive ability of demographic variables.

The next three variables reflect the main types of air pollution in California. One interest is to see if one seems to be more dominant in the analysis than the other. The final three variables are the demographic variables of income, age, and education. The highest correlation among the exogenous variables was .34 (excluding the correlations among the three types of air pollution measures).

A Recursive Model

Stepwise regression was run upon the analysis sample for each of the endogenous variables. A recursive structure was used. The BELIEF regression used only the exogenous variables. The CONCERN regression also included the BELIEF variable as an independent variable. In addition to the exogenous variables the AUTCO regression included the CONCERN and BELIEF variables and the remaining "behavioral" predisposition variables included the BELIEF, AUTCO and CONCERN variables as independent variables. The recursive structure is shown in Figure A.

The results of the stepwise regressions are shown on the top line of each row in Table 1. Each variable coefficient reported has a beta weight of close to or exceeding 0.10 and each was at least significant at the 0.10 level. Also shown are the mean values for each variable and the R² values.

The resulting model obtained from the stepwise regression was fitted to the validation sample using ordinary least squares. The results are presented on the second line of each row in Table 1. The numbers in parentheses are the t-values of the validation run. The results are summarized in Figure A. Only paths are included that have a beta weight of at least 0.15 on one of the data sets and were "confirmed" by the other. Stepwise regression with a validation sample is appropriate here since the prior theory is strong enough to support the model structure but is not strong enough to provide a priori specification of the paths. In fact, a purpose of the research is to identify paths. The low correlations among the exogenous variables and the use of a validation sample reduce the classic dangers of stepwise regression.

TABLE 1

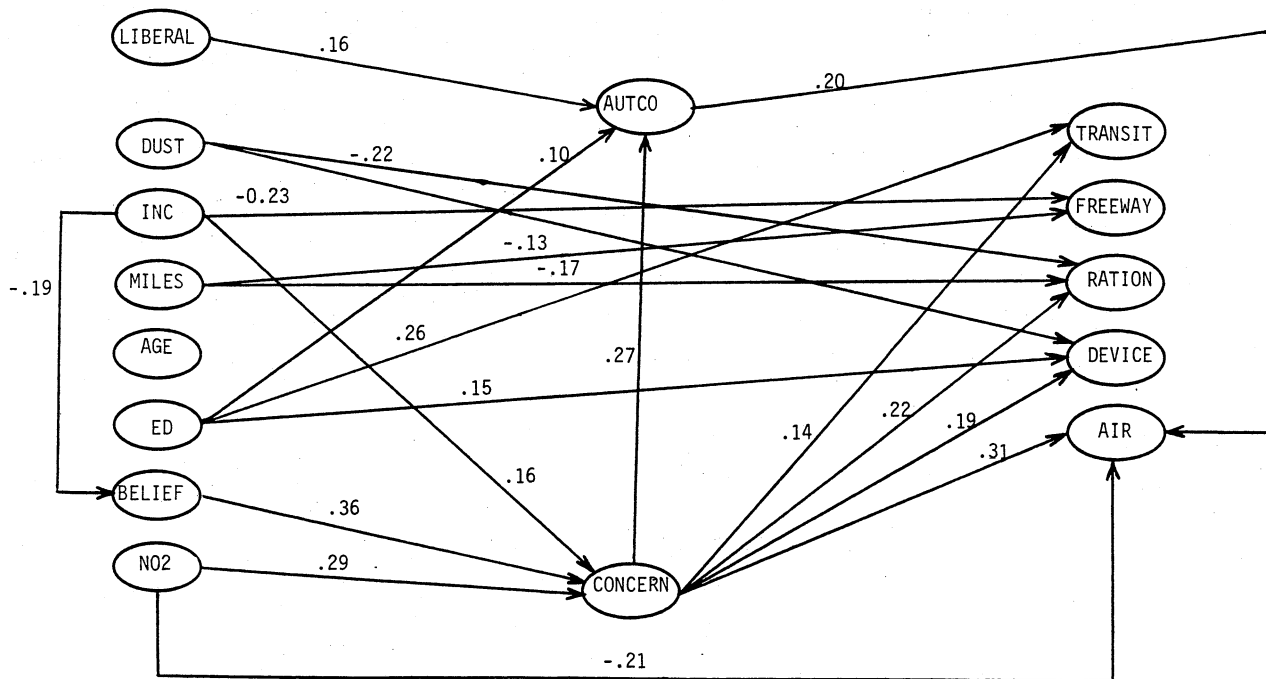
THE ANALYSIS AND VALIDATION RESULTS: BETA WEIGHTS

	CONCERN	BELIEF	AUTCO	MILES	LIBERAL	NO2	OZONE	DUST	INC	AGE	ED	Mean	Std. Dev.	Adj. R ²
DEVICE	.189 .213 (2.7) ^a			-.088 -.110 (1.4)				-.155 -.152 (1.9)			+.147 +.137	2.58	1.02	.071 .049
RATION	.220 .089 (1.1)			-.173 -.180 (2.2)	+.105 .095 (1.2)			-.219 -.090 (1.1)		+.103 .065 (.8)	+.093 .096 (1.2)	2.26	1.09	.123 .020
FREEWAY	.106 .120 (1.6)			-.125 -.152 (1.9)	+.130 .071 (.9)			-.229 -.184 (2.2)			+.091 .127 (1.6)	2.07 ^b	.87 ^b	.102 .064
TRANSIT	.143 .097 (1.3)				+.096 .030 (.4)						+.260 .176 (2.3)	2.50 ^b	.88 ^b	.102 .026
AIR	+.314 .247 (3.0)		.200 .256 (3.5)	-.119 -.098 (1.3)	.095 .008 (.1)	-.210 -.101 (1.3)		.104 .030 (.4)				5.56	1.13	.202 .124
CONCERN		.358 .273 (7.0)				+.292 .436 (6.4)		+.164 .048 (.7)				2.43	1.14	.211 .222
BELIEF								-.191 -.194 (2.5)	-.130 .079 (1.0)			1.62	.69	.042 .029
AUTCO	+.268 +.244 (3.4)				.160 .052 (.7)			.094 .005 (.1)			.097 .161 (2.1)	1.77	.40	.127 .072

^at-values for the validation sample.

^bScale divided by 2 so that it can be compared to Device and Ration.

FIGURE A



Discussion

The Means

The means on the first four variables are directly comparable. A 2.5 rating indicates a neutral position. Two of the proposals, DEVICE and TRANSIT, obtained a neutral score and the other two were regarded negatively.¹ Actually, the TRANSIT variable included a bus system proposal with exclusive bus lanes which was not supported (2.2) and a rail rapid transit system which was supported (2.8). During this time some exclusive bus lanes were tried in Los Angeles and proved to be unpopular. About 40 percent of the probability sample was from the Los Angeles area. Thus, the anti-pollution device and the rail rapid transit were judged neutral or favorable and the rest were judged unfavorably. Clearly, attacks on freeways and gasoline rationing are not popular among Californians who are heavily dependent on the automobile for transportation.

With respect to the AIR variable, respondents felt that the automobile standards were a bit too lenient but a small majority felt that they would still be relaxed if car performance were affected. The CONCERN variable indicated that respondents exhibited attitudes between feeling that the pollution problem was very serious (2) and somewhat serious (3). They further felt that the problem was (a) worse than it had been in the past (BELIEF) and (b) by a wide margin that the automobile companies could have been doing more (AUTCO).

The Intervening Variables

BELIEF. BELIEF is a cognitive construct indicating whether air pollution is perceived to be worsening. It is hypothesized to be influenced by the pollution variables. In fact, the effects of the pollution variables were actually negative, although they were too small to be included in the table. Two qualifications need to be mentioned. First, the pollution variables did not reflect the change in pollution, only the absolute magnitude. Further, with minor exceptions, the areas had about the same levels of pollution level. Thus, the pollution variables are confounded with the "Los Angeles area effect." This interpretation problem is probably inevitable when dealing with pollution on an area basis as opposed to a neighborhood basis.

Those believing that air pollution has worsened tended to be younger as expected, but with lower income, which was unexpected. Perhaps the income variable served as a surrogate for neighborhood pollution levels. In any case, the R^2 values were relatively low.

CONCERN. The CONCERN construct has been the focus of many of the reported studies and is a central intervening construct here. Note that the R^2 value for both data sets is over 0.20, an impressive level considering that disaggregative data are involved and that the dependent variable consists of a four point scale. If the respondents had been grouped, the R^2 would be much higher. Bass, et. al. (1968) provide a discussion why R^2 values are low for disaggregated data, and Aaker (1972) provides an illustration of what effect grouping can have on R^2 .

The two most significant coefficients have the expected sign. The pollution variable is large and positive as expected. Again, it should be noted that this variable

¹The differences between the means of the first four variables are statistically significant except for the difference involving DEVICE and TRANSIT.

is somewhat confounded by the fact that to some extent it is an indicator variable for the Los Angeles area. However, it seems safe to assume that a causal link should and does exist between pollution levels and CONCERN.

The BELIEF variable also has a large beta weight in both samples. Thus, there is evidence supporting a causal link between a perception that pollution is increasing and CONCERN. It could be argued that BELIEF is simply another measure of CONCERN and therefore what is observed is that two indicators of the same construct are correlated. Such a conclusion is regarded as not persuasive for two reasons. First, BELIEF is a perceptual or cognitive construct, while CONCERN is affective in content. Further, an attempt was made to consider BELIEF and CONCERN as indicators of a latent or unobservable variable in a recursive structure similar to that of Figure A. The resulting model had an extremely poor fit to the data indicating that such an approach was simply not compatible with this data. For a discussion of the use and testing of unobservable variables in structural equation models, see Aaker and Bagozzi (1979).

It should be noted that NO₂ was only marginally superior to DUST and OZONE, so the evidence is very weak indeed that NO₂ is a more relevant pollution variable. In fact, the three variables have intercorrelations of .70, .86, and .93.

INCOME had a positive coefficient, although it was not significant in the validation sample. Still, it is of interest that income was included in the regression equation when such control variables as LIBERAL, MILES, AGE, and ED were also available.

AUTCO. The most significant explanatory variable in the AUTCO equation is CONCERN. The R^2 value with only CONCERN was nearly 0.09. Note also that the link between CONCERN and AUTCO is higher than the link between CONCERN and any of the behavioral predisposition variables except AIR. Thus, it seems that a concern about air pollution indeed seems to result in a tendency to "blame" the automobile companies for not solving the problem. The AIR variable also relates to this judgment since it involves the legislated automobile exhaust standards which would impose an automobile company generated solution. The other behavioral predisposition variables, on the other hand, involve more of a sacrifice on the part of the motorists.

The ED variable is positive, as expected, and the INC variable is also positive, though its influence disappears in the validation sample. The relatively weak performance of the demographic variables may, in part, be due to the existence of the LIBERAL variable which was significantly positive. To some extent, demographic variables might be simple surrogates for a liberal/conservative orientation.

The Behavior Predisposition Variables

Air. Among behavior predisposition variables, the AIR equation generated the highest R^2 value: 0.20 for the analysis sample and 0.12 for the validation sample. Again, it should be noted that these levels of R^2 values are relatively high considering disaggregative data are used, the fact that the relationships involved are nontrivial, and the findings of previous studies (cf., Sharma, Kivlin, and Fliegel 1975).

A tendency to advocate stricter automobile exhaust standards is explained in large part by CONCERN (concern about the air pollution problem) and by AUTCO (the belief that automobile companies should have been doing more about the problem.)

A third significant variable is a pollution variable, NO₂. However, the pollution variable has the wrong sign! Note

that pollution also appears with a negative sign in the DEVICE AND RATION equations. The explanation for this finding is that the pollution variable is acting as a surrogate for the Los Angeles area and that other characteristics of residents of the Los Angeles area dominate or override the influence of pollution itself. This is somewhat surprising in view of the fact that MILES (a control variable for miles driven) is also in the equation. Thus, the fact that Los Angeles area residents drive more is not the characteristic that is causing the negative coefficient of the pollution variable. Note also that the coefficient of the MILES variable is negative as expected.

The remaining two explanatory variables, LIBERAL and INC, are in the expected direction but are not confirmed in the validation sample. The failure of the demographic variables to have more explanatory value is somewhat surprising.

The Remaining Behavior Predisposition Variables

The equations for the four other "behavioral" variables had smaller R^2 values on the order of 0.10, and these values decreased in the validation sample. In the case of the TRANSIT and RATION variables, the drop in R^2 was substantial. Still there are some worthwhile observations that emerge even for these two equations.

Overall the most significant explanatory variable is CONCERN. This finding is expected. In fact, it is somewhat surprising that the CONCERN variable did not have more explanatory power in the validation sample. Note that the BELIEF variable did not appear in any of the equations. This fact lends support to the hypothesized recursive structure. BELIEF seems to influence the behavior variables indirectly only through CONCERN and not directly.

The pollution variables appeared in the DEVICE and RATION equations (they were essentially zero in the others), but again with a negative coefficient. The most plausible interpretation is probably that the Los Angeles area residents do not support performance degrading automobile emission devices and gasoline rationing even after controlling for miles driven.

ED was consistently a substantial explanatory variable with the expected positive sign. Of interest is the fact that the other demographic variables failed to contribute. An exception is the large negative INC coefficient in the FREEWAY equation. The higher income people are opposed to radical efforts to cut down freeways whereas the lower income people are more disposed to such a move. This finding is unexpected. The hypothesis was that income would be associated with concern and with support of the policy alternatives. Thus, the direct negative links between income and other behavioral predisposition variables are somewhat surprising.

MILES and LIBERAL both made modest contributions in three of the equations. As expected, a willingness to support three proposals to restrict (RATION and FREEWAY) or affect (DEVICE) driving are negatively associated with miles driven. The support of the more radical government imposed solutions (RATION, FREEWAY and TRANSIT) is positively associated with LIBERAL.

Combining the Behavioral Predisposition Variables

It can be hypothesized that the five behavioral predisposition variables can be considered as indicators of a single underlying construct. Thus the model with a single unobservable "behavioral" construct was used with various combinations of the five behavioral predisposi-

tion variables as indicators. The results of this analysis (not shown) supported the counter hypothesis that the five variables are sufficiently different in the context of the structural model that they should not be considered indicators of a single unobservable. Any combination that included the AIR variable had exceptional poor fits as measured by a chi square value that is a by-product of the maximum likelihood estimation that was used (Aaker and Bagozzi 1979). When AIR was excluded, the fit was still unsatisfactory and, in fact, would improve dramatically each time the number of indicators was reduced. Only when one indicator was used was the fit extremely good. Hence, a separate recursive model for each of the behavioral predisposition variables was judged to be the most realistic model of the situation.

Conclusions

The results lead to the following summary comments and conclusions. First, education was found to have a direct although modest impact upon the policy variables. Income had a modest indirect effect through the CONCERN variable. However, the impact of these variables was perhaps smaller than expected, and the reason can be traced probably to the inclusion of the various control and intervening variables. Of note is the lack of impact of the AGE variable.

Miles driven had a direct impact on the automobile-associated public policy alternatives. The liberal/conservative orientation variable had a direct impact upon the more radical alternatives involving substantial governmental actions.

The pollution variable did seem to have an impact upon the CONCERN variable. However, its direct impact upon the variables was negative due to other characteristics of Los Angeles area residents for which the pollution variables was acting as a proxy.

The hypothesis that the extent to which concern about air pollution generates support for public policy alternatives depends upon the cost of those alternatives is supported in this research. Concern about air pollution had a greater effect upon the AUTCO construct than upon the behavioral predisposition variables. This indicates that people were more ready to blame the automobile companies than they were to face certain alternatives involving sacrifices. Among the policy alternatives, the one with the strongest link to CONCERN was the proposal that the automobile air pollution standards should be stricter, a proposal that deflects the problem to the automobile companies and only indirectly affects the respondents. The DEVICE alternative, which also had a relatively strong link to CONCERN, was certainly much less extreme than the others. Of interest is the fact that TRANSIT, which was rated about as desirable as DEVICE, had a weaker link to CONCERN. Thus, the alternatives represented in the TRANSIT construct, although relatively palatable, were regarded positively in large part for reasons other than the fact that they represent a possible partial solution to the air pollution problem.

References

- Aaker, D.A. (1972), "A Measure of Grand Acceptance," Journal of Marketing Research, 9, 160-67.
- _____ and Bagozzi, R.P. (1979), "Unobservable Variables in Structural Equation Models," Journal of Marketing Research, 16, 147-158.

- Bass, F.M., Tigert, D.J., and Lonsdale, R.T. (1968), "Market Segmentation: Group Versus Individual Behavior," Journal of Marketing Research, 5, 264-270.
- Buttel, F.H. and Flinn, W.L. (1978), "The Politics of Environmental Concern: The Impacts of Party Identification and Political Ideology on Environmental Attitudes," Environment and Behavior, 10, 17-36.
- Degroot, I., Loring, W., Rihm, A., Samuaels, S.W., and Winkelstein, W. (1966), "People and Air Pollution: A Study of Attitudes in Buffalo, N.Y.," Journal of Air Pollution Control Association, 16, 245-247.
- Dunlap, R.E. (1975), "The Impact of Political Orientation on Environmental Attitudes and Actions," Environment and Behavior, 7, 428-454.
- Erskine, H. (1972), "The Polls: Pollution and Its Cost," Public Opinion Quarterly, 36, 120-135.
- Hummel, C.F., Levitt, L., and Loomis, R.J. (1978), "Perceptions of the Energy Crisis: Who is Blamed and How do Citizens React to Environmental-Lifestyle Trade-offs?" Environment and Behavior, 10, 37-88.
- Kinnear, T.C., Taylor, M.R. and Ahmed, S.A. (1974), "Ecologically Concerned Consumers: Who are They?" Journal of Marketing, 38, 20-24.
- Koenig, D.J. (1975), "Additional Research on Environmental Activism," Environment and Behavior, 7, 472-485.
- National Wildlife Federation (1970), A Study of the Attitudes of the American Public toward Improvement of the Natural Environment, Washington, D.C.: The Federation.
- _____ (1969), The U.S. Public Considers Its Environment, Washington, D.C.: The Federation.
- Reizenstein, R.C., Hills, G.E., and Philpot, J.W. (1974), "Willingness to Pay for Control of Air Pollution: A Demographic Analysis," 1974 Combined AMA Proceedings, Chicago: American Marketing Association, 323-328.
- Schuller, C.R. and Ervin, O.L. (1972), Public Attitudes and Regional Environmental Problems: Distributions and Correlates in a Southeastern City, Oak Ridge, Tenn: Oak Ridge National Laboratory.
- Sharma, N.C., Kivlin, J.E., and Fliegel, F.C. (1975), "Environmental Pollution: Is There Enough Public Concern to Lead to Action?" Environment and Behavior, 7, 455-471.
- Stalker, W.W. and Robison, C.B. (1967), "A Method for Using Air Pollution Measurements and Public Opinion to Establish Ambient Air Quality Standards," Journal of the Air Pollution Control Association, 17, 142-144.
- Tognacci, L.N., Weigel, R.H., Wideen, M.F., and Vernon, D.T.A. (1972), "Environmental Quality: How Universal is Public Concern?" Environment and Behavior, 4, 73-86.
- Weigel, R.H. and Weigel, J. (1978), "Environmental Concern: The Development of a Measure," Environment and Behavior, 10, 3-15.

CONCEPTUAL AND METHODOLOGICAL ISSUES IN
RESEARCH ON THE QUALITY OF LIFE

William K. Zikmund, Oklahoma State University

Abstract

This paper discussed three papers on consumer research and the quality of life. Issues of theoretical development, research design and data analysis are considered.

Introduction

At the outset, it should be stated that each of these papers is conscientiously trying to investigate a relatively unresearched topic. While they may be grouped together under the heading, Quality of Life, the diversity of these papers illustrates that there are numerous dimensions and aspects of this issue from a consumer standpoint. Each of the three papers deals with the very distinct topic.

Two of the papers are empirical. Some standard cautions should be given. There are limits to the inferences that can be made when no measure of reliability or validity is given. This seems doubly important in quality of life research because one of the major empirical works indicates that quality of life measures had very low stability coefficients (Campbell, Converse and Rodgers 1976). Both empirical papers suffer because of the lack of these measurers. A second deficiency, typical of conference papers, is that the data appear to have been collected in a limited geographical region (i.e., in the state where the researchers reside). These limitations should be considered by those reading the papers.

The Papers

Unger and Kernan

The authors provide a comprehensive, if not exhaustive, review of the literature. Initially, a quick scan of the paper brings an awareness that the paper's references are absent. When I observe this in conference papers, I am somewhat skeptical. I often jump to the conclusion that the authors have not been concise in their presentation of the material. However, after receiving a copy of the bibliography I found that its length exceeds that of the paper. In this case, the proceedings space limitations should have been relaxed. The authors should be commended for their extensive search.

Some minor aspects of the literature review require clarification. When discussing the value of leisure for the various market segments, there seems to be a confusion between the perceived happiness construct and perceived satisfaction construct (c.f. Campbell, et. al. 1976, p. 32-36). A distinction is usually made between these two constructs in quality of life research. Second, although this issue might be the subject of a second review, no comments are made about the reliability and validity of existing measurers. It has already been suggested that this caution is more than the generalized statement of good research technique to provide true measurement because measures of the quality of life (Campbell, et. al. 1976) have lacked stability.

Beyond reviewing the literature, the authors' goal was to develop a subjective measure of leisure quality; "one that transcends specific activities and approaches leisure from a number of existential dimension." This subjective existential perspective seems to allow for a philosophical definition of leisure, somewhat similar to the classical Greek ideal of leisure as exposed by De Grazia (1964). According to De Grazia, "Leisure is not fully realizable, and hence an ideal, not alone in idea. Leisure refers to a state of being free from the necessity to labor, a condition of man, which few desire and fewer achieve" ... "Leisure is the state of being in which activity is performed for its own sake or as its own end." This view may be that of a philosopher, rather than a psychologist, sociologist or consumer researcher, but it does illustrate that the concept of leisure is somewhat elusive when subjective interpretations are allowed. To clarify the concept of leisure the authors attempt to identify the subjective properties of leisure: intrinsic satisfaction, perceived freedom, high involvement, arousal, mastery, and spontaneity. The question arises does this help us more precisely define leisure. This, to those of you who read Pirsig's (1974) Zen and The Art of Motorcycle Maintenance, may be reminiscent of Phaedrus' exploration into the meaning of the term Quality. "As he became involved in finding quality he singled out aspects of Quality such as unity, vividness, authority, economy, sensitivity, clarity, emphasis, flow, suspense, brilliance, precision, proportion, depth and so on; kept each of these as poorly defined as quality itself ..." And of course at one time before his madness he wrote, "But Even Though Quality Cannot Be Defined You Know What Quality Is!"

Take for example, "perceived freedom." Is it freedom from work? Or as Dumazebier (1967) suggest is "Contemporary Leisure ... Defined by contrast not just to one's job, but to all of the ordinary necessities and obligations of existence?" Is freedom any more concrete than leisure? Doesn't a subjective obligatory/discretionary continuum have the same definitional problems as leisure? Consider that the functions of leisure according to one view are relaxation, entertainment and personal development. It is not clear what empirical rules would be utilized so we can be confident that we should classify relaxation into "arousal" (its absence) or into intrinsic satisfaction. Of course, there may be some hope if the rules for specifying the subjective properties are clearly identified. However, they are not in this proposal.

It seems to me an inherent problem of a subjective measure is similar to the problems of objective measures. Of particular relevance is the situational content of leisure. The authors recognize that objective leisure measurers and subjective leisure experience appear to be highly situational. It is difficult to see how they will tackle this problem even though it is a recognized problem. In sum, the authors are dealing with an exceedingly difficult measurement problem, and while they shed some light on the issues they do not solve the problem.

The authors focus on a timely topic that is important to all of us.

The authors' begin the paper by identifying five research issues. The first three deal with understanding the behavior of health care consumers and the discussion of a normative prescription of the role consumers should have in health care planning and administration. For the most part, a satisfactory review of key articles is given. The authors do state that their literature review was not meant to be exhaustive, and they should not be severely criticized for this. However, it should be pointed out that there was an over-emphasis on the physician-patient interaction with almost a complete absence of the discussion of other health care alternatives. While there is a brief mentioning of health maintenance organizations, there is no review of the literature on physician assistant, nurse practitioners, or other forms of health care delivery. Research in this area (e.g., Zikmund and Miller 1979) indicates that there is a similarity between the underlying cognitive factor consumers hold toward nurse practitioners and those they hold towards physicians. Further, a clear-cut interpersonal dimension has been identified. The next two research issues reviewed deal with two marketing mix variables: health promotion and facility location (distribution). One might question if pricing issues (e.g., pricing strategies of denturists, medicaid/medicare, etc.) and other marketing mix questions should not be of concern to those interested in health care/satisfaction consumption issues. Only one citation from the literature related to facility location is given. Undoubtedly this issue has been dealt within proprietary research and as a regular business matter that has not generated substantial interests in the theoretical literature in recent years.

The main strength of this paper is to provide selected responses from several open ended statements. While the data base is composed of 1,000 Dayton, Ohio residents, there is no breakdown of the data by market segments to take advantage of this large sample size. One wonders whether dissatisfaction lies with a particular group, such as the aged, poor, or those who received healthcare in a particular location within Dayton? Further there is not any investigation of the different medical contexts that may create satisfaction or dissatisfaction. It is most surprising that no inferential statistical analysis is presented in the paper. One wonders why not. Surely the data permit some reasonably sophisticated analysis.

Aaker and Bagozzi

The paper by Aaker and Bagozzi has several strengths. The paper is well written. It reads with clarity in both the conceptual and methodological sections. The data analysis also has some strong points. The research divides the sample into an analysis sample and a validation sample. Certainly a very sound methodological technique; too often omitted in consumer research papers.

In general, the use of stepwise regression seems appropriate to identify paths. Of course, the normal caveat that the widespread use of assuming ordinal scales have interval properties is violated. For example, the 4 point agree/disagree scale seems to be ordinal. Shouldn't the authors know that "despite the widespread treatment of the total (Likert) score as an interval scale, there is no evidence it has no more than ordinal properties" (Aaker and Day 1980, p. 186).

Some questions may arise about the data base. The survey was conducted in 1973, and is obviously somewhat dated. As the authors suggest, things may have changed since the OPEC embargo. Things have changed! Car pooling has increased. People didn't even know what Kalaka was in 1973. BART (Bay Area Transportation) has been completed and many other things have changed including the OPEC oil embargo. Certainly attitudes towards gasoline rationing, restricting freeway use have changed dramatically, especially in California where the sample was taken. And, of course, California isn't like anywhere else on this planet. The sampling question is of particular relevance in this paper. To the authors' credit, "objective measurers of actual pollution levels were included in the study." This is important especially in L.A. where in the autumn the air turns brown and falls to the ground?

However, the paper addresses theoretical issues and perhaps these issues can be minimized. Lets turn our attention to this topic. It appears to me that this study is an example of theorizing with an existing data base in mind. Certainly other constructs (not examined here) could be considered as factors influencing attitudes towards air pollution. Further, a more important conceptual issue is that this study seems to assume that air pollution is an exclusive function of the automobile. Automobile companies are not the sole "villians." Steel mills, oil refineries, and other factories have for years been portrayed with large black clouds flowing from their smoke stacks. The domain of "concern" in this study apparently is a variable that should be labeled "concern for air pollution generated by automobiles." A person living next to an oil refinery may have a great concern for air pollution, but not attribute this to an automobile company. To make a public policy decision based on this logic and under these circumstances would be unwarranted.

Summary

Like most discussants I have mentioned the papers' strengths and dwelled on their shortcomings. I assume the readers will realize that each of these papers deserves some merit. Because of the diversity of these three papers it is difficult to summarize them taken together and make a universal recommendation. One common characteristic is the fact that quality of life constructs, whether it is pollution concerns, health care satisfaction, or leisure, all are relatively subjective concepts that will vary within situational contexts. This presents difficulty to researchers wishing to measure these concepts. Certainly measurement is the main issue to be addressed in future research.

References

- Aaker, D.A. and Day, G.S. (1980), Marketing Research, New York: John Wiley and Sons.
- Campbell, A., Converse, P.E. and Rodgers, W.L. (1976), The Quality of American Life, New York: Russell Sage Foundation.
- Pirsig, R.M. (1974), Zen and the Art of Motorcycle Maintenance: An Inquiry into Value, Toronto: Bantam Books.
- Zikmund, W.G. and Miller, S.J. (1979), "A Factor Analysis of Attitudes Toward Nurse Practitioners," Research In Nursing and Health, 85-90.

TRADE-OFFS IN ATTRIBUTE LEVELS MADE BY ECOLOGICALLY CONCERNED
AND UNCONCERNED CONSUMERS WHEN BUYING DETERGENTS¹

Karl E. Henion, University of Texas at Austin²
Russell Gregory, University of Texas at Austin
Mona A. Clee, University of Texas at Austin

Abstract

Environmental concern is less intense than it was a decade ago. Today when selecting products, consumers may be giving much greater weight to price than to ecological attributes of product. Whether this suspected trend holds for ecologically as well as non-ecologically concerned buyer segments is the principal focus of this paper.

Introduction

"During the past fourteen months we have seen a virtual storm of public protest and journalistic attention devoted to the nation's ecology. One of the prime targets of the environmental hue and cry has obviously been the business community and the role that it plays in our deteriorating environment." Times change. Written almost ten years ago, that opening paragraph of an article on ecological factors and buying behavior (Herberger and Buchanan 1971, p. 644) seems strangely out-of-step with public opinion today. No longer does environmental concern rank particularly high in national priorities (Jaroslarosky 1980). In the early and middle 1970's it did, and national policy favored a strong, regulatory approach to problems like environmental protection (Nixon 1973). As that decade drew to a close, national sentiment seemed to reflect greater sympathy for the problems of business, especially the regulatory burden it carries, which is widely considered to be an important cause of inflation (DeMuth 1980). And inflation seems to be on everybody's mind. More and more, consumers are shopping price and they seem to be less interested in buying products that are ecologically benign than they were at the start of the decade (Marketing News 1979).

This turn of events is critical to the credibility of a concept that has been referred to in the marketing literature as ecological marketing (Henion 1976). Both a suggested national policy alternative to increased regulation (Henion 1978) and itself a sub-discipline of social marketing, this concept has one central idea. It is that environmental protection and resource conservation can be better advanced through less regulation by the public sector and more enterprise in the private sector. This idea, in turn, is based on the premise that the ecologically concerned consumer is a legitimate but largely untapped market segment -- one that is identifiable, accessible and measurable (Henion 1976). Such a

¹The authors gratefully acknowledge the financial assistance of the Sperry and Hutchinson Company whose grant in ecological marketing helped make this study possible. The authors also express their appreciation to Professor Mark Alpert for many ideas for improving the analysis and discussion sections; also, to three blind reviewers for numerous helpful suggestions.

²Karl E. Henion is Professor of Marketing; Russell Gregory is a doctoral student in the Department of Marketing; and Mona A. Clee is Assistant Instructor of Marketing.

segment is a potential market for products that are environmentally benign (EN-products, for short).

But, clearly, no customer -- whatever his or her degree of ecological consciousness -- will buy EN-products on the basis of their ecological impact alone. The consumer obviously buys a collection of attributes. In the case of a detergent, the attributes would probably include price, cleaning ability and -- for some shoppers, such as the ecologically concerned consumer -- ecological impact. Just as for other products, a bundle of attributes is sought; and the EN-product bought will depend on trade-offs the consumer is willing to make among various levels of such attributes.

The Problem

What trade-offs in attribute levels then do consumers make today when choosing products that are particularly relevant from an ecological perspective? Only two studies in the marketing literature have addressed this question.

The earliest study, which took up the issue in indirect fashion, was based on a simulated shopping experiment whose data today are ten years old (Herberger and Buchanan 1971). In that study price and ecological information were experimentally manipulated for four brands of detergents. Ecological information was found to influence brand preferences across all three price levels studied. The effect was in favor of low-phosphate brands at the expense of high-phosphate brands. Yet, in the context of the extent of the change in preference, change was not as great when phosphate-free detergents were priced disadvantageously. Less conclusive results were reported for soft drinks.

In the other study, Kerin and Peterson (1974) considered the subject of trade-offs more directly. Responding to mailed questionnaires, housewives rank-ordered detergent and soft drink product configurations, each made up one of two levels of price, brand and ecological impact. Regardless of high or low level of environmental consciousness, respondents compromised ecological compatibility first, price second and most favorite brand third.

The aim of the present study was to make a current assessment of attribute level trade-offs for an EN-product. Former assessments reflected in the studies mentioned are dated and limited. For example, neither study reported results based on observed behavior. Also, in the earlier study ecologically unconcerned consumers were not distinguished from concerned consumers -- an important distinction. Although such a distinction was made in the later study, no information was supplied about either the scale used to measure ecological consciousness or its reliability and validity. Moreover, in the 1974 study, the only price levels studied were those below and above prevailing prices. Consequently, overlooked was the realistic and important situation in which products are priced the same.

In the present study the expectation was that ecologically concerned consumers would value attribute levels

of an EN-product differently from unconcerned consumers. This prediction is consistent with key findings in much of the ecological marketing literature. Numerous studies (e.g. 's, Anderson, Henion and Cox 1974, Kinnear, Taylor and Ahmed 1974, Webster 1975) have shown that these two segments exist and that consumers comprising them are inclined to believe and behave in different ways when buying products that are particularly ecologically relevant.

Method

As in the previous two studies, laundry detergents were also the **product** focus of the present study. The field setting was five stores of a supermarket chain situated in a medium-size city in the Southwest. They served shoppers living in middle to upper-middle class neighborhoods. Each store carried the same line of about 25 brands of powdered and liquid detergents. Several of them were phosphate-free: the powdered detergents Arm & Hammer, Purex, Trend, Topco Orange and Ivory Snow; and the liquid detergents ERA and Dynamo.

Trained interviewers, stating that they were conducting a survey on consumer preferences using a technique called conjoint analysis, administered a pre-tested five-minute questionnaire to a total of 188 in-store shoppers, mostly as they filed by the detergent aisles of the stores. Of this number 110 were observed and recorded, prior to the interview, as actually taking a particular brand of detergent from the shelf and selecting it for purchase. The three-part questionnaire was administered to the respondents in the following order: (a) trade-off tables of paired attribute levels; (b) an index of ecological concern (IEC) developed by Kinnear and Taylor (1973); and (c) several demographic questions, which are of no direct concern to the present study. Evidence of construct validity of the IEC has been reported by the above investigators (1973, pp. 192-193); additional evidence is provided elsewhere (Henion and Wilson 1976).

Data were collected by the trade-off method instead of the full-profile method (Green and Srinivasan 1978) for several reasons. Knowledge of the relative importance of the attributes was necessary for testing one of the hypotheses of the study; and this can be easily done by comparing utility ranges based on the information that respondents provide directly in the trade-off method (Johnson 1974). Also, the experience of one of the authors gained from other field experiments using both data collection methods resulted in a decided preference for the trade-off method because it was easier to apply and explain to busy shoppers in supermarkets. The issue of interactions, which cannot be estimated under this method, was not considered a potential problem; the attributes of the present study were viewed as independent and non-redundant and hence orthogonal.

Another issue that should be addressed arises from self-perception theory. According to that theory, some persons who bought low-phosphate detergents conceivably might have labeled themselves as ECCs and developed and registered an attitude on the IEC consistent with such behavior, both of which may have had low reliability if measured on a later occasion. However, disguiser steps were taken so that subjects would not be self-conscious about the fact that the ecological aspect of their buying behavior was being monitored. Moreover, the limited time shoppers were willing to take for the interviews did not warrant embedding the IEC in a lengthy battery of other measures.

The interviewer first explained to the respondent how to fill out a trade-off table in which pairs of attribute levels were to be rank-ordered; then the interviewer

had the respondent practice on two tables containing attribute pairs for unrelated products. These tables helped disguise the ecological interest of the study. Next, three 3X3 trade-off tables for detergents were separately presented. (Since only one of the three attributes was ecologically related, the disguise was further protected.) The three detergent attributes (and their three levels) were price (15¢ more than, 15¢ less than and same price as usual); cleaning power (somewhat cleaner, equally clean, and somewhat less clean); and phosphate content (high, average, and low). Choice of these attributes served the main purpose of the study, namely, to clarify their previous ambiguous role in the literature of ecological marketing. Other detergent attributes, such as softness or odor, while relevant for different purposes such as new product development, were not relevant for this study.

Hypotheses

Motivated by the studies of Herberger and Buchanan (1971) and Kerin and Peterson (1974), the following predictions were made:

- H₁: A low phosphate detergent has more utility for ECCs than for non-ECCs; and, conversely, a high phosphate detergent has more disutility for ECCs than for non-ECCs;
- H₂: A detergent's phosphate content will be considered by ECCs to be more important, and by non-ECCs to be less important, than other attributes.
- H₃: An EN-Product that is higher priced and/or less functional than its environmentally less desirable substitute will have less overall disutility for ECCs than for non-ECCs, e.g., a low-phosphate detergent priced higher and with less cleaning power than a high-phosphate detergent.

Results

From the sample of 110 respondents who were observed buying a detergent, 18 ecologically concerned and 29 unconcerned consumers were identified, as well as 63 neutral or inconsistent consumers. Identification of the first two groups (ECC and non-ECC) was based on two criteria: observed purchase of a detergent and score on the IEC, whose range was from 0 to 25 points. If shoppers bought a phosphate-free detergent and also scored above the third quartile (15) on the IEC, they were classified as ECCs. If they bought a phosphate detergent and scored below the first quartile (7), they were classified as non-ECCs. Interquartile scorers were considered neutral whichever detergent type was bought; there were only six respondents whose scores and purchase behavior were inconsistent.

From the rankings of attribute level pairs made by each of the 110 identified respondents, a parametric utility for each attribute level was estimated using conjoint analysis. The analysis was based on an additive model and was carried out using \$FTROFF (Nehls, Seaman and Montgomery 1976). This algorithm rests on the assumption that the dependent variable is, at most, ordinally scaled.

\$FTROFF prints out a badness-of-fit measure, ϕ , for each respondent. It is based on the extent to which the sums of the estimated utilities for each pair of intersecting attribute levels in a trade-off table have rank orders dissimilar to the input ranking that the respondent made of the sample pairs. This measure, which accounts for

both number and size of errors arising from the revealed dissimilarities, accumulates such errors across all of the tables that the respondent fills out. The average ϕ for the ECC group was .054; for the non-ECC group, .085. The average proportion of correct comparisons of utility sums in each cell for the ECC group was 87.5%; for the non-ECC group, 84.3%.

For each attribute level, an average utility was calculated from the utilities derived from the conjoint analysis of ranks at the individual level. The results for the consumers in the ECC and non-ECC groups are presented in Table 1. The algorithm forces the algebraic sum of each respondent's utilities for the three levels of each attribute to equal zero; hence the sum of the utility averages for each attribute at the group level approximates zero.

The utilities for the nine attribute levels (3X3) were cast as independent variables first in a multiple discriminant analysis, with the ECC vs. non-ECC vs. neutral and inconsistent group membership (utilities not shown) as the criterion variable. The results were not significant using the whole spectrum of subjects. However, the theoretical interest was in the ECC and non-ECC groups. Hence, the indifferent and inconsistent group was dropped from the model and a simple discriminant analysis was next run using the other two groups. Forty percent of the variance in the resulting discriminant function was explained by the group variable. The canonical correlation coefficient was .63; the Wilks Lambda of .60 was significant at .014 ($\chi^2 = 20.71$ with 9 d.f.).

The jackknife technique was used in cross-validating the results of a parallel stepwise discriminant analysis, owing to the small size of the sample. Correct classification and α -probabilities for the two groups were: ECC (78%, $p < .02$) and non-ECC (55%, $p < .28$). The rest of the paper focuses on findings for these two groups.

TABLE 1

Average Utilities For Three Levels of Three Attributes For Two Groups of Consumers

Attribute and level	Average utility (\bar{u})	
	Consumer group	
	Ecologically concerned (n=18)	Non-Ecologically concerned (n=29)
<u>Price</u>		
15¢ more	-.3398	-.3885
Same	.0501	.0748
15¢ less	.2891	.3075
<u>Cleaning Power</u>		
High	.2844	.3474
Average	.0866	.1225
Low	-.3710	-.4692
<u>Phosphate content</u>		
High	-.4669	-.2356
Average	.0569	.0394
Low	.4109	.1962

Within each attribute for each consumer group in Table 1, the pairs of correlated means (i.e., those in the vertical direction) were significantly different for each of the three possible pairs of attribute levels (t's had $p < .001$ or better in all cases but two, which had $p < .02$ instead: one in the ECC group, d.f. = 17, and one in the non-ECC group, d.f. = 28).

The direction and size of the utility means show that persons in both groups place higher utility on lower price and on higher cleaning power. This result one would expect to find. Of special interest, however, is the finding that not only ECCs but also non-ECCs place higher utility on low phosphate content than on average or high levels.

We turn next to a comparison of uncorrelated means across the two groups (i.e., those in the horizontal direction of Table 1). While there are 9 possible comparisons, only 6 are particularly relevant since values of the utilities for the 3 middle levels for the attributes will tend toward zero, owing to the logic of the \$FTROFF algorithm. Of the six comparisons, three revealed differences in mean pairs that were significantly different. The other three were not significant, although they were in the expected direction; the probability that three pairs would randomly be arrayed in the expected directions, given that each of them is equal likely to have been reversed, is .125, which approaches some statistical significance. Overall, strong support was found for H_1 .

The disutility of high phosphate content was significantly higher for ECCs than for non-ECCs ($p < .001$, d.f. = 45); and the utility of low phosphate content was significantly higher for ECCs than for non-ECCs ($p < .003$, d.f. = 45). In a similar vein, low cleaning power had less disutility for ECCs than for non-ECCs ($p < .05$, d.f. = 45); and, while not statistically significant ($p < .27$) high cleaning power had less utility for ECCs than for non-ECCs. Unfortunately, none of the price comparisons produced significant differences. It seems that price qua price (i.e., when not compared with other attributes) is about equally important to both groups of consumers. Nevertheless, numerically, the values of the means are in the expected direction. High price had slightly less disutility for ECCs than for non-ECCs; low price, slightly less utility.

Ranges

The relative importance of a set of attributes can be indexed by computing the range of utility values across the levels of an attribute and comparing it with the corresponding ranges for the other attributes. The greater the range the greater the importance. For each respondent the range of utilities for each attribute was calculated and the means of the ranges for the three attributes for the two groups are presented in Table 2.

The rank order of the attributes according to mean utility ranges is quite different for the two consumer groups. As mentioned earlier, low phosphate content had positive utility for both groups. However, ECCs considered this attribute to be of first importance (with the highest range of .9021), cleaning power next and price last. But non-ECCs considered phosphate content to be of least importance (with the lowest range of .5628), placing price and cleaning power ahead of it.

TABLE 2
Mean Utility Ranges for Three Attributes
For Two Groups of Consumers

Attributes	Consumer Group					
	Ecologically concerned ^a (n=18 ECCs)			Ecologically unconcerned ^b (n=29 non-ECCs)		
	Rank	Mean utility	%	Rank	Mean utility	%
Price	3	.6324	(28)	2	.7152	(34)
Cleaning Power	2	.7204	(32)	1	.8542	(40)
Phosphate content	1	.9021	(40)	3	.5628	(26)
		2.2549	(100)		2.1322	(100)

^aF = 3.39, p < .05

^bF = 5.74, p < .005

The one-way ANOVA of the mean ranges for each consumer group produced significant F-values. For the ECC group, the results of t-tests for the three pairs of correlated means were: no significance between price and cleaning power (p < .20); approaching significance between cleaning power and phosphate content (p < .10); and significance between price and phosphate content (p < .02). For the non-ECC group, the results for the same pairs were, respectively; approaching significance (p < .10); significance (p < .05); and significance (p < .005). Consequently, good support was found for H₂.

Amplifying these findings were the results of t-tests for two of the three pairs of uncorrelated means (i.e., those in the horizontal direction). For two attributes -- phosphate content and cleaning power -- significance (p < .005 and p < .056, respectively) was found between the ECC and non-ECC groups, but not for price.

Product Simulations

For each respondent, a composite utility was calculated for each product simulation or configuration. A configuration was made up of 3 attribute levels -- one from each of the 3 attributes. The utilities of the 3 levels were summed to form a respondent's composite utility for the configuration. The means of the respondents' composite utilities (\bar{u}_c) for the 27 possible product configurations are shown in Table 3 for the ECCs and the non-ECCs.

Direction signs, - 0 +, in Table 3 indicate whether a level for an attribute is desirable relative to the attribute's other two levels. These signs were employed in the following manner. Product configurations have been listed in Table 3 in descending order according to the sum total of the advantages or benefits that various attribute's levels in a configuration are judged to have. Thus, the first entry in the list is a detergent that has a low (L) price (i.e., 15 cents less), high (H) cleaning power and low (L) phosphate content. This configuration, symbolized by the ordered triple (LHL), is also represented by an ordered triple of signs (+ + +). Here a + indicates the most desirable level of an attribute; a 0, a middle level; and a -, the least desirable level. Let a + and a - be interpreted as +1 and -1, respectively. If equal weights (1 : 1 : 1) are then arbitrarily used as attribute weights for each sign in the triple, then the algebraic sum of the signs indexes the utility of a configuration. Hence, the first triple yields an index

TABLE 3
Means of Composite Utilities, Attribute Levels
And Benefit Signs For 27 Possible Product
Configurations For Two Consumer Groups

Con- figur- ation number	Attribute Levels						Util- Index	\bar{u}_c	
	Attributes							Consumer group	
	P	CP	PC	P	CP	PC		Ecolog- ically con- cerned	Ecolog- ically uncon- cerned
1.	L	H	L	+	+	+	3	.984440	.850984
2.	L	H	M	+	+	0	2	.630497	.694243
3.	L	A	L	+	0	+	2	.786597	.626088
4.	S	H	L	0	+	+	2	.745397	.618288
5.	L	A	M	+	0	0	1	.432654	.469347
6.	S	H	M	0	+	0	1	.391454	.461547
7.	S	A	L	0	0	+	1	.547554	.393393
8.	L	H	H	+	+	-	1	.106660	.419263
9.	L	L	L	+	-	+	1	.329018	.034434
10.	H	H	L	-	+	+	1	.355521	.154993
11.	L	A	H	+	0	-	0	-.091183	.194368
12.	L	L	M	+	-	0	0	-.024925	-.122307
13.	S	H	H	0	+	-	0	-.132383	.186567
14.	S	A	M	0	0	0	0	.193611	.236652
15.	S	L	L	0	-	+	0	.089975	-.198262
16.	H	H	M	-	+	0	0	.001578	-.001748
17.	H	A	L	-	0	+	0	.157677	-.069903
18.	L	L	H	+	-	-	-1	-.548762	-.397287
19.	H	H	H	-	+	-	-1	-.522259	-.276728
20.	H	L	L	-	-	+	-1	-.299901	-.661557
21.	S	A	H	0	0	-	-1	-.330226	-.038328
22.	H	A	M	-	0	0	-1	-.196266	-.226644
23.	S	L	M	0	-	0	-1	-.263698	-.355003
24.	S	L	H	0	-	-	-2	-.787805	-.629983
25.	H	A	H	-	0	-	-2	-.720103	-.501623
26.	H	L	M	-	-	0	-2	-.653844	-.818298
27.	H	L	H	-	-	-	-3	-1.177681	-1.093278

P = Price L = Low H = High
CP = Cleaning Power A = Average M = Medium
PC = Phosphate Content S = Same

of +3. Expectedly, for configuration 1 the \bar{u}_c obtained for each consumer group has the highest value; for configuration 27, the lowest value.

Abstracted from Table 3 and shown separately in Tables 4 and 5 are the mean composite utilities, respectively, for the nine detergent configurations with low phosphate content and the nine with high phosphate -- results on which H₃ principally depend.

Based on the composite utilities, support was found for this hypothesis. Considered first will be the "and" part of the "and/or" statement of H₃. For ECCs the mean composite utility (\bar{u}_c) for configuration 20 was -.299901 but only -.330226 for item 21, although the difference in the correlated means was not significant; whereas, for non-ECCs the corresponding \bar{u}_c 's were -.661557 and -.038328 (p < .001). Thus, a low-phosphate detergent priced higher (by one level) and with less (by one level) cleaning power than a detergent with a high level of phosphate content was considered by ECCs to have more utility (right direction); by non-ECCs, to have significantly less. However, these reversals did not hold if both the disadvantages of higher price and less cleaning power were at the extreme

TABLE 4

Means (\bar{u}_c) Of Composite Utilities For Nine Product Configurations With Low Phosphate Content

Con-figuration number	Attri-bute levels			Attri-bute levels			Util-ity index	\bar{u}_c	
	Attributes							Consumer group	
	P	CP	PC	P	CP	PC		Ecolog-ically con-cerned	Ecolog-ically uncon-cerned
1	L	H	L	+	+	+	3	.984440	.850984
3	L	A	L	+	0	+	2	.786597	.626088
4	S	H	L	0	+	+	2	.745397	.618288
7	S	A	L	0	0	+	1	.547554	.393393
9	L	L	L	+	-	+	1	.329018	.034434
10	H	H	L	-	+	+	1	.355521	.154993
15	S	L	L	0	-	+	0	.089975	-.198262
17	H	A	L	-	0	+	0	.157677	-.069903
20	H	L	L	-	-	+	-1	-.299901	-.661557

TABLE 5

Means (\bar{u}_c) Of Composite Utilities For Nine Product Configurations With High Phosphate Content

Con-figuration number	Attri-bute levels			Attri-bute levels			Util-ity index	\bar{u}_c	
	Attributes							Consumer group	
	P	CP	PC	P	CP	PC		Ecolog-ically con-cerned	Ecolog-ically uncon-cerned
8	L	H	H	+	+	-	1	.106660	.419263
11	L	A	H	+	0	-	0	-.091183	.194368
13	S	H	H	0	+	-	0	-.132383	.186567
18	L	L	H	+	-	-	-1	-.548762	-.397287
19	H	H	H	-	+	-	-1	-.522259	-.276728
21	S	A	H	0	0	-	-1	-.330226	-.038328
24	S	L	H	0	-	-	-2	-.787805	-.629983
25	H	A	H	-	0	-	-2	-.720103	-.501623
27	H	L	H	-	-	-	-3	-1.177681	-1.093278

levels, i.e., by two levels: for ECCs the utilities were -.299901 (again configuration 20) vs. .106660 (configuration 8) with $p < .01$; for non-ECCs, -.661557 vs. .419263 with $p < .001$. These last results however, are based on extremely strict criteria. The ECC is being asked to sacrifice for ecology not only price and cleaning power (function) but also to the extent of the most disadvantageous levels of these attributes.

It is noteworthy that in Table 4 the composite utility mean for the ECC group is higher than that for the non-ECC group for each of the nine product configurations with low phosphate content. The reverse is true in Table 5: the \bar{u}_c is lower for the nine with high phosphate content. The probability of either of these outcomes is less than .002 (binomial test). Obviously the ECC places greater value on phosphate content being at a lower level than does the non-ECC.

The utilities were next examined under criteria that were less strict, namely, the "or" part of the "and/or" in H_3 . Rather than both disadvantages, only one disadvantage of the other were considered -- again at one level and two levels. The results, which are broken out from Table 3 and summarized in Table 6, provide strong support for H_3 . In the case of these somewhat less unattractive configurations -- namely, a low-phosphate detergent with either a higher price (whether 1 or 2 levels) or less cleaning power (also whether 1 or 2 levels) -- the EN-product was still considered by ECCs to have higher utility than its environmentally less desirable substitute; by non-ECCs, lower utility. In six of the eight possible pairings of the \bar{u}_c 's of the EN-product with those of the substitutes, significant differences in correlated means were found ($p < .01$ in four cases and $p < .05$ in two); in the other two pairings -- both in the non-ECC group -- the one with configuration 25 approached significance ($p < .08$) while the other with 24 was not significant.

TABLE 6

Means (\bar{u}_c) Of Composite Utilities: Partially Disadvantaged EN-Product Detergent vs. Four Highest Phosphate Product Configurations For Two Consumer Groups

Product type and configuration for EN-product	Disad-vantage level	Direction signs of attribute levels			\bar{u}_c	
		Attributes			Consumer group	
		Price	Clean-ing power	Phos-phate con-tent	Ecologi-cally con-cerned	Ecologi-cally uncon-cerned
EN-pro-duct						
20		-	-	+	-.299901	-.661557
Highest phosphate product						
24	1	0	-	-	-.787805	-.629983
25	1	-	0	-	-.720103	-.501623
18	2	+	-	-	-.548762	-.397287
19	2	-	+	-	-.522259	-.276728

Discussion

These results, based on concomitant measures of observed behavior and attitude, help buttress the case of ecological marketing. Today inflation is driving consumers more and more to shop price, and yet concerned consumers in this study are at least implying -- by the trade-offs they registered -- that they would be willing to subordinate price to ecology. Unconcerned consumers, however, place price and function ahead of ecology. Therefore, our findings contradict those of Kerin and Peterson (1974) who reported both consumer groups as placing price

ahead of ecology. The strength of the present findings is our identification of the ecologically concerned consumer is more valid than theirs. One caveat is that these findings are applicable to consumers with middle to upper-middle socio-cultural background.

Given this caveat, these findings have relevance for a long-standing, unresolved issue in ecological marketing research. If consumers show greater concern for price, as is expected, will sales of an EN-product be adversely affected (Henion 1976)? That depends. Is a particular EN-product less or more expensive than the product for which it is a more ecologically desirable substitute? (Obviously the critical condition is when it is more expensive.) What is the cross-elasticity of demand for the two products? How do patterns of price inelasticity of an EN-product differ for ecologically concerned and unconcerned consumers? Based on the earlier studies, the presumption is that the patterns are the same for both groups of consumers, namely, little if any elasticity (and further that substantial cross-elasticity exists for an EN-product and its functional counterpart which is ecologically insulting). However, the present assessment of attribute levels trade-offs -- which were made by one sample of respondents for one EN-product and analyzed using a decompositional model of consumer preference -- suggest a different conclusion. Concerned consumers might possibly be willing to pay a premium for an EN-product, even during times of high inflation, although unconcerned consumers would probably not -- based on the results of the range analysis.

However, the external validity of the conjoint measurements taken is still an open question, making predictions of the degree to which consumers actually would trade-off higher prices against lower phosphate content quite another matter. As with previous studies, the design of this study may have increased the salience of the area. For we have presumed that buyers perceive phosphate levels as salient; further that such salience was probably augmented by labelling the levels for respondents in the trade-off tables. In the real world attribute levels compete for buyer recognition with other cues that may be more dominant, e.g., price; and in the case of detergents usually, phosphate content is merely registered rather than highlighted (although it was highlighted on the package of one of the six phosphate-free detergents in the present study). External validity would be more assured if phosphate content had been labeled prominently as it was in one study (Henion 1972); surveying instruments would then not differ much from the information available to buyers when shopping. At any rate, results from the present study can justifiably be compared with those of prior studies, since whatever cue bias exists in the present study was present in the others, too.

Accordingly, field experiments need to be conducted in which prices of EN-products are experimentally manipulated with phosphate content highlighted. If inelasticity of EN-products were found for the ecologically concerned consumer, then the suspicion of ecological lip service would be largely groundless for this group. Meanwhile, confirmation that concerned consumers value attribute levels of an EN-product differently from unconcerned consumers are encouraging from the policy perspective of ecological marketing, coming as it does five years after the 1974 study with its negative results.

References

- Anderson, W. Thomas, Henion, Karl E., and Cox, Eli P. (1974), "Socially vs. Ecologically Responsible Consumers," American Marketing Association Combined Conference Proceedings, 36, 304-311.
- Demuth, Christopher C. (1980), "Constraining Regulatory Costs, Part One: The White House Review Programming," Regulation, (January/February), 13-26.
- Green, Paul E. and Srinivasan, V. (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," Journal of Consumer Research, 5, 103-123.
- Henion, Karl E. (1972), "The Effect of Ecologically Relevant Information on Detergent Sales," Journal of Marketing Research, 9, 10-14.
- Henion, Karl E. (1976), Ecological Marketing, Columbus, Ohio: Grid.
- Henion, Karl E. (1978), "Ecological Marketing: From Normative to Descriptive Model?" in Future Directions in Marketing, Johan Arndt, George Fisk, and Kjell Gronhaug, eds., Boston: Marketing Science Institute, 301-311.
- Henion, Karl E., and Wilson, William H. (1976), "The Ecologically Concerned Consumer and Locus of Control," in Ecological Marketing, Karl E. Henion and Thomas C. Kinnear, eds., Chicago: American Marketing Association.
- Herberger, Roy A. and Buchanan Dodds I. (1971), "The Impact of Concern for Ecological Factors on Consumer Attitudes and Buying Behavior," American Marketing Association Combined Conference Proceedings, 33, 644-646.
- Jaroslovsky, Rich (1980), "Douglas Castle's Balancing Act," The Wall Street Journal, (April 11), 18.
- Johnson, Richard M. (1974), "Tradeoff Analysis of Consumer Values," Journal of Marketing Research, 11, 121-214.
- Kerin, Roger A. and Peterson, Robert A. (1974), "Selected Insights of Ecologically Responsible Behavior," American Institute for Decision Sciences Proceedings, 6, 33.
- Kinnear, Thomas C. and Taylor, James R. (1973), "The Effect of Ecological Concern on Brand Perceptions," Journal of Marketing Research, 10, 191-197.
- Kinnear, Thomas C., Taylor, James R. and Ahmed, Sadrudin A. (1974), "Ecologically Concerned Consumers: Who Are They?," Journal of Marketing, 38, 20-24.
- Marketing News (1979), "Permanent Inflation Is Key Factor in Future of Marketing," (September 7), 3-4.
- Nehls, Lyle, Seamen, Bruce, and Montgomery, David B. (1976), "A PLI Program for Trade-off Analysis," Cambridge, MA: Marketing Science Institute.
- Nixon, Richard M. (1973), "The President's Message," Environmental Quality: The Fourth Annual Report of the Council on Environmental Quality, Washington, D.C.: U.S. Govt. Printing Office, vii.
- Webster, Fred E., Jr., (1975), "Determining the Characteristics of the Socially Conscious Consumer," Journal of Consumer Research, 2, 188-196.

A CAUSAL PATH ANALYSIS OF ECOLOGICAL BEHAVIOR RELATING TO MARKETING

Lawrence A. Crosby, University of Nebraska
James D. Gill, University of Nebraska

Abstract

A study of voter preference on the Michigan "Bottle Bill" provides new insights into the characteristics of the Ecologically Concerned Consumer and strategies for marketing to this consumer segment. A path model is developed to account for complex interrelationships among the predictors of ecological concern and behavior. The model is based on the environmental behavior literature and psychological theory. The survey results are found to be generally consistent with the model.

Introduction

The decade of the 1970's was a turbulent period in the history of the American marketplace. Major shifts in the economic, political and social climates significantly impacted marketers. During this period serious questions were raised about marketing practices and especially those affecting the natural environment. Many felt that "ecological concern" might pose a threat to marketing by stimulating the passage of more regulation. However, a few astute marketers argued that it was possible to convert this threat into an opportunity through Ecological Marketing. The premise of Ecological Marketing is that a segment of consumers exist who are motivated to buy environmentally safe products and who will respond to an ecological appeal (Henion 1976).

Research was undertaken in an attempt to identify the ecologically concerned consumer segment (see e.g., Anderson and Cunningham 1972, Kinnear, Taylor, and Ahmed 1974, Webster 1975, Murphy, Kangun, and Locander 1978). This research focused on the demographic, personality, and attitudinal correlates of ecological concern that might be relevant to the design of effective communications strategies. Unfortunately, the findings were not very encouraging. It appeared that the ecologically concerned consumer (ECC) could be identified but that the segment was small and hard to reach.

Consumer research pertaining to Ecological Marketing has been hampered somewhat by two problems. First, only limited phenomena exist for study because few consumers presently engage in ecological consumption behavior. This has forced researchers to increasingly rely on experimental choice situations that necessarily require some sacrifice in external validity (Murphy 1978, Lepisto 1979, Arndt and Helgesen 1979). The second problem concerns the lack of an appropriate methodology to account for complex interrelationships between the correlates of ecological concern. Multicollinearity among the predictors has generally been viewed as a methodological defect rather than an important source of information about underlying causal influences.

The Research Application

While consumer demand for ecological products may be limited (at present), consumers seem more willing to support the regulation of marketing activity affecting the environment. This is particularly evident in the case of the beverage container problem. Consumers have generally shunned returnable containers in favor of the convenience of throwaways, even though throwaways waste resources and

contribute to solid waste and litter. However, in apparent contradiction of this marketplace behavior, consumers in eight states have supported the passage of mandatory deposit laws (the states are Oregon, Vermont, Iowa, Maine, South Dakota, Connecticut, Michigan, and Delaware).

The present study focuses on the 1976 beverage container deposit law referendum in Michigan. In that election, Michigan voters approved by a substantial margin (64% to 36%) a proposal to require refundable cash deposits for soft drink and beer containers. This election was selected for study because: (1) individuals were acting in their roles as both voters and consumers, (2) a large number of persons were engaging in ecological behavior by supporting the law, and (3) the successful passage of the law was expected to provide some guidance for the marketing of ecological products.

The Methodological Approach

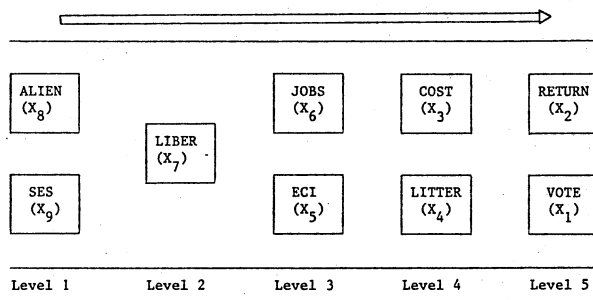
A generalized causal model was developed and tested for its ability to explain and predict deposit law voting preferences. It is suggested that the approach taken to developing and operationalizing the model can be applied to the study of various ecological behaviors. Although the model does not include all possible causal influences, it does include variables representing the most important classes of predictors. Furthermore, the model reflects causal influences among the predictors as well as those between the predictors and the criterion. Thus, both the direct and indirect effects of the predictors on the criterion can be assessed.

The Causal Model

The a priori causal model (see Figure 1) incorporates five levels of variables. The most basic level involves individual differences in demographics and personality. Included at Level 1 are the constructs of socio-economic status (SES) and alienation (ALIEN). The second level contains a political orientation measure of liberalism-conservatism (LIBER) which may reflect important differences in values. Level 3 contains general attitudes or opinion clusters relating to ecological concern (ECI) and concern about the unemployment consequences of environmental protection (JOBS). Level 4 contains more specific environmental and economic attitudes with particular relevance to the deposit law issue. Among these are concern about throwaway containers as a source of litter (LITTER) and concern about the effects of a deposit law on beverage prices (COST). The fifth level incorporates the behavior related variables. Included at Level 5 are variables dealing with voting preferences on the deposit law (VOTE) and previous habits regarding the purchase of returnable containers (RETURN). The criterion variable is VOTE.

The specific causal paths are not shown in the a priori model (Figure 1) because this would create undue complexity in the diagram. Basically, the a priori model assumes that causal influences move from left to right in the diagram and that each lower level variable has a causal path connecting it to each higher level variable. In addition, causal paths are assumed to exist from SES → ALIEN and from RETURN → VOTE while unanalyzed paths are assumed to exist between JOBS ↔ ECI and between COST ↔ LITTER.

FIGURE 1
A Priori Model



Model Justification and Hypotheses

The causal model was formulated on the basis of theoretical considerations and knowledge of the environmental behavior literature. The purpose of this section is to justify the choice of variables and their level of inclusion in the model.

Numerous researchers have tried to predict ecological concern from demographic and/or personality variables (Anderson and Cunningham 1972, Kinnear, Taylor, and Ahmed 1974, Webster 1975, Henion and Wilson 1976, Murphy, Kangun, and Locander 1978, Murphy 1978, Lepisto 1979). Demographically, the ECC's tend to be white and have higher incomes, higher status occupations, and higher socioeconomic status although the findings are not very consistent. Socio-economic status (SES) was included in the present model to determine how this variable influences an entire complex of environmental attitudes and behaviors, not just limited to ecological concern. It was felt this might uncover any countervailing effects tending to suppress the influence of SES.

Socioeconomic status was positioned at Level 1 as a basic causal influence in the model. It is assumed that individuals of similar social status have shared many of the same learning experiences and that these experiences serve to shape personality, values, attitudes, and behavior. For example, it is predicted the socio-economic status has a negative influence on feelings of alienation (ALIEN). Those who receive the rewards and respect of society and maintain positions of power and influence are less likely to experience feelings of powerlessness, meaninglessness, and social isolation.

Previous studies have produced contradictory findings about the relationship of alienation to ecological concern and behavior. Anderson, Henion, and Cox (1974) found the ECC to be more alienated than the non-ECC person. On the other hand, Nelson (1974) found the ECC to be less alienated. Anderson and Cunningham (1972) found that socially responsible persons were less alienated. It seems reasonable that alienation would exert some influence on environmental attitudes and behavior since alienated individuals tend to feel powerless in helping to solve social problems. Therefore, alienation (ALIEN) was included in the model with the expectation that it would relate negatively to the environmental variables.

The "alienated voter hypothesis" provides another reason for including a measure of alienation. Several studies in the political science literature have found that alienation is associated with political negativism and a tendency to vote against referendum issues like the "bottle ban" (Horton and Thompson 1962, Stone 1965, Aberbach 1969).

Alienation was also positioned at Level 1, as a basic causal influence in the model. Personality is thought to be a determinant of a person's political orientation (Adorno, et. al. 1950) and attitudes (Katz 1960). Also, personality is often tied to the concept of "consistent responses to the world of stimuli surrounding the individual" (Kassarjian 1971) suggesting a link to behavior.

There is some evidence in the literature of marketing (Webster 1975, Mayer 1976) and environmental behavior (Tognacci et. al. 1972, Constantini and Hanf 1972, Dunlap 1975, Koenig 1975, Buttel and Flinn 1976) to suggest that liberals are more ecologically concerned than conservatives. Such a relationship might be expected since most environmental programs involve some form of government intervention. Since this was also true for the Michigan deposit law, it justified including a measure of political orientation in the model (LIBER).

The LIBER variable was positioned prior to the other attitudinal variables at Level 2 on the basis of its "centrality." A belief, attitude, or opinion is said to possess centrality if it lies close to the person's values (Bem 1970). Also according to Bem "labels like liberal and conservative usually enable us to predict many of the individual's attitudes because these two terms refer to broad underlying values which are shared by large segments of the population." Since individuals attempt to maintain consistency with central values and attitudes, this implies a causal flow in the direction of decreasing centrality.

There is evidence that people can be arrayed in terms of the strength of their ecological concern (Maloney, Ward, and Braucht 1975) and that they exhibit consistency between their ecological attitudes and behaviors (Kinnear and Taylor 1974, Antril and Bennett 1979). Although the link between ecological concern and voting behavior has not been established, this seems to be a logical extension of these findings and is the reason for including ECI in the model.

It is clear academically that environmental protection can have serious economic consequences in terms of inflation, slowed GNP growth, reduced investment in productive facilities, and unemployment. It is suspected that members of the general public also see this environment-economic tradeoff to some extent, especially on such concrete matters as unemployment. A variable labeled JOBS was included in the model to represent an influence of economic concern that might run counter to the influence of environmental concern.

The variables ECI and JOBS were positioned at Level 3, indicating moderate centrality. These general attitudes and opinions are thought to represent the individual's orientation and affinity for certain social goals which serve to guide preferences and choices in specific situations. As such, these general attitudes may also be indicative of some of the individual's underlying values and motives.

The variables at Level 4, parallel those at Level 3, but with greater specificity to the question of the deposit law. Whereas the variable JOBS reflects a macroeconomic concern about environmental protection, COST reflects a microeconomic concern about the deposit law, namely its effect on prices. Likewise, the variable ECI reflects a concern about the environment in its many aspects while LITTER taps only a small part of that, namely concern about beer and soft drink container litter. It is expected that the attitudinal variables at Levels 3 and 4 will be associated. It is assumed that the causal flow is from the more general to the more specific attitudes (from 3 to 4). Finally, it is recognized that other potentially relevant variables dealing with the impact of the deposit law could have been included at level 4 (at the expense, of course, of greater model complexity).

The behavioral variables are represented at Level 5 in the model. They include not only the criterion VOTE but also the prior use of returnables (RETURN). Since both of these can be construed as ecological behaviors, an important question is whether they are influenced by the same variables in the same fashion. It is also assumed that the related consumption variable RETURN would causally influence VOTE since consumers are likely to resist changes in their consumption habits for convenience type goods like beverages. A good deal of support for this assertion exists within learning theory as applied to marketing (Howard and Sheth 1969).

Procedure

Data

The data utilized in this study were obtained from a telephone survey of 306 voting age adults in the State of Michigan. Interviews were conducted between October 23-30, 1976 with the last interview obtained three days before the election. Respondents were selected using a random digit dialing method similar to Waksberg's Two-Stage Procedure (Frankel and Frankel 1977). A randomization procedure was used to select from multiple eligible adults within the same household. The calling was done from a central location. The interview lasted about 25 minutes.

Interviews were completed with 61% of those who were contacted and determined to be qualified. The sample matched population estimates in terms of sex and area of residence. Some differences were noted on education, age, and income but not of sufficient magnitude to seriously affect the relationships under investigation. The survey results predicted the election outcome with a high degree of accuracy.

Measures

Both single items and multi-item indices were used to measure the theoretical constructs. Theoretically related items were combined whenever it was possible to form an index having satisfactory reliability for this type of research.

Socioeconomic status (SES) was measured as a three factor index including income, education, and occupation. As was the case for all the indices, each component received an equal weight. Cronbach's Alpha for SES was .48.

The alienation index (ALIEN) measures the extent to which feelings of powerlessness, meaninglessness, and social isolation are experienced by the respondent. High scores on the index indicate more alienation. This basic approach to measuring the construct is generally consistent with other research operationalizations (see: Aberbach 1969; Stone 1965). Cronbach's Alpha for ALIEN was .48.

A single item (LIBER) was used to measure the respondent's political orientation, with special reference to the role of government in economic affairs (i.e., whether government should force all products that pollute off the market). Because a large portion of regulatory activity in the 1970's concerned the environment, this seemed to be a meaningful way to ascertain feelings about government involvement. It is recognized that other dimensions of liberalism-conservatism exist. However, it is also true that the organization of political idea elements begins to breakdown as one moves from more to less sophisticated publics (Converse 1970), suggesting that the attitude domain should be more narrowly defined as was done here.

An ecological concern index (ECI) was formed to measure the respondent's concern about protecting the environment. The starting point for this index was an earlier version of the ECI developed by Kinnear and Taylor (1973), although

the current version is strictly attitudinal in nature. Higher ECI scores indicate more concern about the environment. Cronbach's Alpha for this index is .60.

A single item (JOBS) was used to measure the respondent's concern about the macroeconomic impact of environmental protection. The JOBS variable focuses on unemployment which is a concrete manifestation of the economic/environment tradeoff. Higher JOBS scores indicate more concern about the effects of environmental protection on unemployment.

A single item (COST) was used to measure the respondent's concern about a microeconomic impact of the deposit law, namely its effect on beverage prices. Higher COST scores indicate more concern about the adverse price effects of the law.

Concern about the degree to which beverage containers contribute to pollution and litter, an opinion highly specific to the deposit law, was measured by the LITTER index. Higher LITTER scores reflect the respondent's perception that container litter is a major problem. Cronbach's Alpha for this index is .72.

Prior use of returnables (RETURN) was determined by asking respondents whether they usually purchased soft drinks and beer in returnable bottles, non-returnable bottles, or cans. Responses to this question were coded: usually returnables = 1 and usually throwaways = 0.

To measure voting preference regarding the deposit law (VOTE), respondents were exposed to a statement of the law as it would appear on the ballot, and then asked "if the election was held today, would you vote yes or no on the deposit law?" Responses to this question were coded: yes = 1 and no = 0.

Analysis and Results

The causal model was analyzed by the method of structural equation analysis. Path analysis is a useful method to determine whether the data are consistent with the a priori model. However, consistency of the data with the model does not provide proof of the theory, but only lends support to it (Kerlinger and Pedhazzer 1973, p. 307).

The structural equations and path coefficients were estimated by ordinary least squares regression analysis. While this method of analysis assumes all variables are measured on interval scales, multiple regression is still the recommended procedure when ordinal scales are used, as in this study (Bohrnstedt and Carten 1971). Other assumptions underlying the application of path analysis appeared to have been met (see Kerlinger and Pedhazzer 1973, p. 309).

Table 1 shows how the a priori models were tested, trimmed, and how the path coefficients were estimated. The a priori models in this study were simply the regression models relating each variable to those variables preceding it in the causal flow of Figure 1. Table 1 shows the regression coefficients in unstandardized and standardized form. The standardized coefficients are referred to as "path coefficients."

Goldberger's (1970) approach to model trimming was used and a significance criterion of $p < .05$ was set as suggested by Duncan (1975). The trimming involves a reestimation of the regression equations with the predictors found non-significant in the first attempt eliminated. The regression and path coefficients for the trimmed models also appear in Table 1. A comparison of the coefficients of determination (R^2) for any of the a priori models and trimmed models indicates no significant loss in explanatory power resulting from theory trimming. Figure 2 is a diagram of the trimmed model showing the causal paths found

TABLE 1
Path Analysis Results

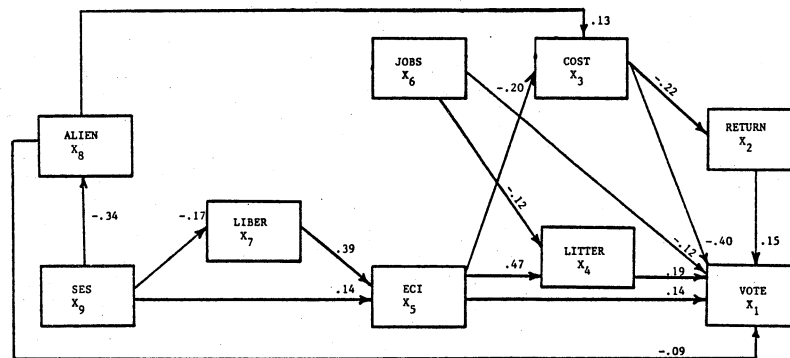
Model	Dependent Variable ^a	Explanatory Variable	Regression Coefficients for A Priori Models	Path Coefficients ^b for A Priori Models	Adj. R ² for A Priori Models	Regression Coefficients for Trimmed Models	Path Coefficients ^b for Trimmed Models	Adj. R ² for Trimmed Models								
VOTE	X ₁	X ₂	-.1583055	-.14757	R ₁ ² = 39.6	-.1588628	-.14809	R ₁ ² = 40.0								
		X ₃	-.1928743	-.39739		-.1940395	-.39979									
		X ₄	.09774416	.19681		.09602233	.19334									
		X ₅	.06726837	.13744		.06752978	.13797									
		X ₆	-.05938182	-.12089		-.05954105	-.12122									
		X ₇	-.0024070	-.00491 ^c		-----	-----									
		X ₈	-.03746405	-.07469		-.04278523	-.08530									
		X ₉	.01514523	.03018 ^c		-----	-----									
		Constant	0.5156688	-----		0.5154562	-----									
		RETURN	X ₂	X ₃		-.08820313	-.19495		R ₂ ² = 3.0	-.09845886	-.21762	R ₂ ² = 4.4				
X ₄	.01756845			.03795 ^c	-----	-----										
X ₅	.01759873			.03857 ^c	-----	-----										
X ₆	-.01127589			-.02463 ^c	-----	-----										
X ₇	-.0137542			-.03010 ^c	-----	-----										
X ₈	-.01655811			-.03541 ^c	-----	-----										
X ₉	.0039343			.00842 ^c	-----	-----										
Constant	0.3043634			-----	0.3052361	-----										
COST	X ₃			X ₅	-.2025349	-.20084	R ₃ ² = 6.0	-.2052948		-.20358	R ₃ ² = 4.2					
				X ₆	.09497975	.09385 ^c		-----		-----						
		X ₇	.02941970	.02913 ^c	-----	-----										
		X ₈	.1014887	.09820	.1357198	.13132										
		X ₉	-.07280656	-.07042 ^c	-----	-----										
		Constant	.1216202	-----	0.009639076	-----										
		LITTER	X ₄	X ₅	.4939858	.50126		R ₄ ² = 25.0	.4625380	.46935		R ₄ ² = 23.1				
				X ₆	-.1214614	-.12281			-.1155132	-.11660						
				X ₇	.08028185	.08134 ^c			-----	-----						
				X ₈	-.07009473	-.06940 ^c			-----	-----						
X ₉	-.1175786			-.11637 ^c	-----	-----										
Constant	0.006899822			-----	0.008397753	-----										
ECI	X ₅			X ₇	.3982107	.39760	R ₅ ² = 15.1		.3917214	.39112	R ₅ ² = 13.1					
				X ₈	-.08994626	-.08777 ^c			-----	-----						
				X ₉	.1113758	.10863			.1403550	.13690						
				Constant	0.01096278	-----			0.01452596	-----						
		JOBS	X ₆	X ₇	-.127128	-.12739 ^c		R ₆ ² = 1.0	-----	-----		-----				
				X ₈	.06562392	.06426 ^c			-----	-----						
				X ₉	-.04617223	-.04520 ^c			-----	-----						
				Constant	-.0.2303418	-----			-----	-----						
				LIBER	X ₇	X ₈			-.08238968	-.08052 ^c			R ₇ ² = 2.9	-----	-----	R ₇ ² = 2.9
						X ₉			-.1554383	-.15184				-.1830711	-.17883	
Constant	0.1130676					-----	0.008091028		-----							
ALIEN	X ₈					X ₉	-.3353914		-.33526	R ₈ ² = 10.9	-.3353914			-.33526	R ₈ ² = 10.9	
						Constant	0.3903081		-----		0.03903081			-----		

^aDefinition of Variables: X₁ = VOTE; X₂ = RETURN; X₃ = COST; X₄ = LITTER; X₅ = ECI; X₆ = JOBS; X₇ = LIBER; X₈ = ALIEN; X₉ = SES

^bThe path coefficients are the standardized regression coefficients of the regression estimate.

^cp > .05

FIGURE 2
Trimmed Model



to be significant.

From an analysis of Figure 2, structural equations were developed that could be used to decompose the total association between the variables. Due to the complexity of the model no attempt was made to specify the nature of the noncausal association (i.e., to distinguish between spurious association and unanalyzed correlation). By substituting the path coefficients of the trimmed models appearing in Table 1, into the structural equations, it was possible to decompose the total association between each variable and those that causally precede it. These results are summarized in Table 2. This table shows the decomposition of the total association (i.e., the simple correlation) into direct, indirect, total causal, and noncausal association.

TABLE 2
The Components of Total Association

Model	(1) Observed ^a Correlation	PATH ANALYSIS			
		(2) Direct Effect	(3) Indirect Effect	(4+3) Total Causal Effect	(5=1-4) Noncausal Association
VOTE	r ₂₁ ^a .27936	.14809	-----	.14809	.13127
	r ₃₁ ^a -.51973	-.39979	-.0322273458	-.4320173458	-.0877126542
	r ₄₁ ^a .36542	.19334	-----	.19334	.17208
	r ₅₁ ^a .35084	.13797	.1786942203	.3166642203	.0341757797
	r ₆₁ ^a -.23432	-.12122	-.022582112	-.143802112	-.090517888
	r ₇₁ ^a .07920	-----	.1238537097	.1238537097	-.0446537097
	r ₈₁ ^a -.18915	-.08530	-.0567325178	-.1420325178	-.0471174822
	r ₉₁ ^a .12068	-----	.0688203797	.0688203797	.0518596203
	RETURN	r ₃₂ ^a -.21762	-.21762	-----	-.21762
r ₄₂ ^a .09079		-----	-----	-----	.09079
r ₅₂ ^a .09342		-----	.0443030796	.0443030796	.0491169204
r ₆₂ ^a -.05907		-----	-----	-----	-.05907
r ₇₂ ^a -.00881		-----	.0173278204	.0173278204	-.0261378204
r ₈₂ ^a -.07818		-----	-.0285778584	-.0285778584	-.0496021416
r ₉₂ ^a .05243	-----	-.001681716	-.001681716	.054411716	
COST	r ₅₃ ^a -.21303	-.20358	-----	-.20358	-.00945
	r ₆₃ ^a .12359	-----	-----	-----	.12359
	r ₇₃ ^a -.02941	-----	-.0796242096	-.0796242096	.0502142096
	r ₈₃ ^a .14597	.13132	-----	.13132	.01465
	r ₉₃ ^a -.12612	-----	-.0576572478	-.0576572478	-.0684627522
LITTER	r ₅₄ ^a .48315	.46935	-----	.46935	.0138
	r ₆₄ ^a -.17223	-.11680	-----	-.11680	-.05543
	r ₇₄ ^a .12775	-----	.183572172	.183572172	-.055822172
	r ₈₄ ^a -.08483	-----	-----	-----	-.08483
r ₉₄ ^a -.03960	-----	.0314258035	.0314258035	-.0710258035	
ECI	r ₇₅ ^a .36664	.39112	-----	.39112	-.02448
	r ₈₅ ^a -.07193	-----	-----	-----	-.07193
	r ₉₅ ^a .06695	.13690	-.0699439896	.0669560104	0.0
JOBS	r ₇₆ ^a -.11086	-----	-----	-----	-.11086
	r ₈₆ ^a .06267	-----	-----	-----	.06267
	r ₉₆ ^a -.04396	-----	-----	-----	-.04396
LIBER	r ₈₇ ^a .13142	-----	-----	-----	.13142
	r ₉₇ ^a -.17883	-.17883	-----	-.17883	0.0
ALIEN	r ₉₈ ^a -.33526	-.33526	-----	-.33526	0.0

^a r₃₄ or r₄₃ = -.15973

r₅₆ or r₆₅ = -.11811

The VOTE Model

The test of the VOTE model indicated that RETURN, COST, LITTER, ECI, JOBS, and ALIEN all had significant direct effects on VOTE (p < .01) and, in the trimmed model, accounted for 40% of the variance in VOTE. The valences of the beta coefficients indicated that supporters of the deposit law were more litter conscious, were less concerned about the potential impact on either beverage costs or

unemployment, were more ecologically concerned, and were less alienated. Supporters were also more likely to have been previously purchasing returnables.

The magnitudes of the direct effects on VOTE generally decline as the causal variables become more remote. The variables with the largest direct influence on VOTE were assigned either to Level 4 (COST = -.40 and LITTER = .19) or Level 5 (RETURN = .15). Direct effects of less magnitude were obtained for the variables at Level 3 (ECI = .14 and JOBS = -.12), Level 2 (LIBER = N.S.), and Level 1 (SES = N.S. and ALIEN = -.09). These findings are consistent with the nature of the causal flow postulated in the a priori model.

An examination of the components of total association (Table 2) indicates that all predictors had some causal impact on VOTE. Several variables assumed greater prominence when total effects were computed because they had substantial indirect effects. This was particularly true for ECI which was second highest in terms of total effect (= .32). This finding provides additional support for the consistency of ecological attitudes and behaviors. Two of the variables that lacked significant direct effects were found to have notable indirect effects. This was the case for LIBER (= .12) and SES (= .07). The negative total association between VOTE and ALIEN (= -.14) supports the "alienated voter hypotheses."

Other Important Findings

Few of the variables appeared to exert much influence on the voluntary use of returnable containers (RETURN) and only 4.4% of the variance in RETURN was accounted for by the trimmed model. The total causal effect of ECI on RETURN (= .04) was much less than on VOTE (= .32). Apparently, returnables were being purchased for reasons other than ecological concern (e.g., habit, brand loyalty, etc.).

Contrary to the model, concern about the price effects of the law (COST) was unrelated to concern about the potential effects of environmental protection on unemployment. Instead, the general attitude of ECI was found to have a relatively strong direct influence on both of the specific attitudes: COST (= -.20) and LITTER (= .47). As might be expected, concern about the effects of the deposit law on prices was greater among alienated consumers and those of lower SES.

The general attitude of ECI was found to be related to LIBER and SES. As hypothesized by Mayer (1976), ecologically concerned consumers appeared to be more liberal. As he suggests, holding these opinions may offer the liberal a chance to express his or her progressive and humanitarian values. Ecological concern can be viewed as a relatively central attitude mediating these values.

A positive direct effect (= .14) was found to link SES to ECI. At the same time, however, those with higher SES scores tended to be more conservative. Greater conservatism implies less ecological concern. The total association linking SES to ECI was reduced (to .07) as the result of this countervailing tendencies. As expected, consumers of lower socioeconomic status were more alienated.

Interpretation

It should be re-emphasized that path analysis is not capable of proving causality but serves mainly to identify the logical consequences of the causal assumptions that are made. While the data in this study were generally consistent with the model, consistency itself does not establish causality. The data may also be consistent with other models not considered. Path analysis was found to be useful, however, in uncovering associations that would have been overlooked using conventional methods.

The substantive findings indicate that consumers perceived voting for the deposit law as a rather clear expression of their ecological concern. This points to a successful marketing effort on the part of the environmentalists. Prior to the election, however, consumers did not seem to associate buying returnables with ecological concern. Either that, or returnables were found to be unacceptable for some other reason (e.g., non-availability of favorite brands). This suggests that Ecological Marketing efforts by industry to sell returnables were either ineffective or non-existent prior to the election.

The success of the deposit law indicates there is sufficient ecological concern among the population to undertake Ecological Marketing programs. However, in marketing the "bottle bill" to Michigan consumers environmentalists did not rely solely on its ecological appeal. As part of their strategy, they successfully attached support of the law with the specific benefit of reducing litter. Also, they successfully overcame consumers' concern about the adverse effects of the law on prices and the economy.

Conclusion

Through the method of path analysis additional insight can be gained into the motivational basis of ecological behavior. The technique offers several advantages and should be used in future research dealing with the ecologically concerned consumer. One important advantage is that it forces researchers to make explicit their causal assumptions. A causal model employing demographic, psychological, and behavioral variables seems to have application to the study of the ECC. With additional elaboration, refinement, and testing a general model of ecological consumption behavior might be developed.

References

- Aberbach, Joel D. (1969), "Alienation and Political Behavior," American Political Science Review, 63, 86-99.
- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., and Stanford, R. N. (1950), The Authoritarian Personality, New York: Harper.
- Anderson, W. Thomas Jr., and Cunningham, William H. (1972), "The Socially Conscious Consumer," Journal of Marketing, 36, 23-31.
- _____, Henion, Karl E., and Cox, Eli P. III (1974), "Socially vs. Ecologically Responsible Consumers," American Marketing Association Conference Proceedings.
- Antril, John H. and Bennett, Peter D. (1979), "Construction and Validation of a Scale to Measure Socially Responsible Consumption Behavior," in The Conserver Society, ed. Karl E. Henion and Thomas C. Kinnear, Chicago: American Marketing Association, 51-68.
- Arndt, Johan and Helgesen, Thorolf (1979), "Marketing of Ecologically Compatible Products: Public Policy Issues and a Market Test," in 1979 Educators' Conference Proceedings, ed. Neil Beckweith et. al., Chicago: American Marketing Association, 594-597.
- Bem, Daryl J. (1970), Beliefs, Attitudes, and Human Affairs, Belmont, California: Brooks/Cole Publishing Company.
- Bohrnstedt, G. W. and Carten, T. M. (1971), "Robustness in Regression Analysis" in Sociological Methodology 1971, H. L. Costner, ed. San Francisco: Jossey-Bass, 118-46.
- Buttel, Frederick H., and Flinn, William L. (1976), "Environmental Politics: The Structuring of Partisan and Ideological Cleavages in Mass Environmental Attitudes," Sociological Quarterly.
- Constantini, Edmond, and Hanf, Kenneth (1972), "Environmental Concern and Lake Tahoe: A Study of Elite Perceptions, Backgrounds, and Attitudes," Environment and Behavior, 4, 209-242.
- Converse, Philip E. (1970), "The Nature of Belief Systems in Mass Publics," in William J. Crotty, ed., Public Opinion and Politics, New York: Holt, Rinehart, and Winston, pp. 129-155.
- Duncan, Otis Dudley (1975), Introduction to Structural Equation Models, New York: Academic Press.
- Dunlap, Riley E. (1975), "The Impact of Political Orientation on Environmental Attitudes and Actions," Environment and Behavior, 7, 428-454.
- Frankel, Martin R., and Frankel, Lester R. (1977), "Some Recent Developments in Sampling Design," Journal of Marketing Research, 24, 280-93.
- Goldberger, A. S. (1970), "On Boudon's Method of Linear Causal Analysis" American Sociological Review, 35, 97-101.
- Henion, Karl E., and Wilson, William H. (1976), "The Ecologically Concerned Consumer and Locus of Control," in Ecological Marketing, ed. Karl E. Henion and Thomas C. Kinnear, Chicago: American Marketing Association.
- _____, (1976), Ecological Marketing, Columbus, Ohio: Grid, Inc.
- Horton, John E. and Thompson, Wayne E. (1962), "Powerlessness and Political Negativism: A Study of Defeated Local Referendums," American Journal of Sociology, 67, 485-494.
- Howard, John A. and Sheth, Jagdish N. (1969), The Theory of Buyer Behavior, New York: Wiley.
- Kassarjian, Harold H. (1971), "Personality and Consumer Behavior," Journal of Marketing Research, 8, 409-18.
- Katz, D. (1960), "The Functional Approach to the Study of Attitudes," Public Opinion Quarterly, 24, 163-204.
- Kerlinger, Fred N., and Pedhazur, Elazar J. (1973), Multiple Regression in Behavioral Research, New York: Holt, Rinehart, and Winston, Inc.
- Kinnear, Thomas C., and Taylor, James R. (1973), "The Effect of Ecological Concern on Brand Perceptions," Journal of Marketing Research, 10, 191-198.
- _____, _____ (1974), "Identifying Ecological Buying Segments Through Attitude Measurement," Journal of Business Administration, 3, 33-43.
- _____, _____, and Ahmed, Sadrudin A. (1974), "Ecologically Concerned Consumers: Who Are They?" Journal of Marketing, 38, 20-24.
- Koenig, D. J. (1975), "Additional Research on Environmental Activism," Environment and Behavior, 7, 472-485.
- Lepisto, Lawrence R. (1979), "Environmental Product Attributes and Product Preference," in Proceedings, ed. Robert M. Hopkins, and Alfred G. Toma, Lafayette, Louisiana: Southern Marketing Association.
- Maloney, Michael P., and Braucht, G. Nicholas (1975), "A Revised Scale for the Measurement of Ecological Attitudes and Knowledge," American Psychologist, 787-90.

Mayer, Robert N. (1976), "The Socially Conscious Consumer - Another Look at the Data," Journal of Consumer Research, 3, 113-115.

Murphy, Patrick E. (1978), "Environmentally Concerned Consumers: Demographic Dimensions," in Proceedings of the 1978 Educators' Conference, ed. Subhash C. Jain, Chicago: American Marketing Association.

_____, Kangun, Norman, and Locander, William B. (1978), "Environmentally Concerned Consumers -- Racial Variations," Journal of Marketing, 42, 61-66.

Nelson, James E. (1974), "An Empirical Investigation of the Nature and Incidence of Ecologically Responsible Consumption of Housewives," Unpublished doctoral dissertation, University of Minnesota.

Stone, Clarence N. (1965), "Local Referendums: An Alternative to the Alienated-Voter Model," Public Opinion Quarterly, 29, 199-222.

Tognacci, Louis N. (1972), "Environmental Quality: How Universal is Public Concern?" Environment and Behavior, 4, 73-86.

Webster, Frederick E., Jr. (1975), "Determining the Characteristics of the Socially Conscious Consumer," Journal of Consumer Research, 2, 188-196.

THE EVOLUTION OF DISTRIBUTION CHANNELS FOR SOLAR PRODUCTS:
CONSUMER DECISION MAKING IN PERSPECTIVE

Jerald W. Blakely, University of Oregon
Scott M. Smith, University of Oregon

Abstract

Consumer acceptance of solar products is examined from a distribution channel perspective. The current focus on the consumer as the central decision maker for solar products is criticized. New channels of distribution are presented, along with the variables that will facilitate their adoption. Future directions for solar consumer research and public policy decisions are given.

Introduction

The need to develop alternative energy sources has been widely accepted in this country. Solar energy is one of the alternatives being developed and its potential contribution is highly touted. The possible utilization of solar energy for heating and cooling of public, commercial, and residential buildings has gained considerable attention and support. Information on the application and viability of solar energy for heating and cooling buildings is being distributed by both public and private sources in an attempt to generate awareness and interest. Recognizing that information alone will not result in widespread adoption, both Federal and state governments are offering financial incentives to encourage interested consumers to incorporate solar options into their homes. To date, efforts directed toward increasing the rate of adoption of solar as an energy source have been relatively ineffective, thus blunting progress toward energy independence. Solar advocates have become concerned and frustrated about the sluggishness of the adoption process. This problem would be better understood and efforts more effective if it were viewed not as an isolated case, but rather placed in the larger context of diffusion of any new technology or innovation.

When a new technology or innovative product is developed, the basic question of how to get it diffused and accepted in the marketplace must be answered. The diffusion of technological innovations is often retarded by distribution problems. Difficulties are encountered in establishing new distribution channels or adapting to or changing existing channels. A product may be so different that established distribution channels or patterns do not exist, or the product class may have established channels but the new technology may be so different as to not be accepted by the channel members. The array of solar options, with its many variations, encounters both these difficulties.

The present discussion will examine some of the previously ignored problems encountered in the diffusion of solar products. These problems will be discussed in the larger contexts of technological innovations in general and heating and cooling products in particular. The discussion will focus on distribution problems and will further consider the implications of these problems for consumer decision making and public policy decisions.

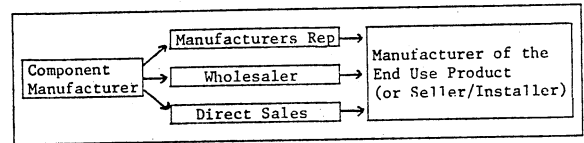
Distribution Channels

Technological innovations generally fall into one of two basic product classes: consumer or industrial goods. Many components used in the manufacture of consumer durable goods are classified as "industrial" products. The normal distribution channel for industrial goods is shown in Figure 1-A.

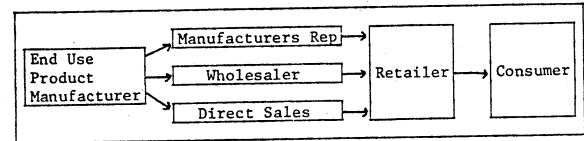
There may be several component manufacturers in succession prior to the end use product, which duplicates and extends the channel. The end use product has its own distribution channel which is shown in Figure 1-B.

FIGURE 1
Traditional Distribution Channels

A. Industrial Products



B. End Use Products



It is important to consider the differences in purchase decision making between the manufacturer of an end use product who buys and assembles components, and a consumer who buys the completed product. Consumer evaluation of durable goods typically centers on the end use product which flows from the product manufacturer to the consumer as diagramed in Figure 1-B. Consumer decision making research would therefore be evaluated as constrained to the end use products model. While this is appropriate for consumer durable goods, it will be argued here that solar products are not an end use product, but rather an industrial good following distribution channel 1-A.

Distribution Channels for Residential Heating and Cooling Systems

The traditional distribution channels for residential heating and cooling equipment are more similar to those used for industrial products than for consumer durable goods. It may appear at first glance that the seller/installer is the retailer selling end use products, but he is more than that. In the case of residential heating and cooling, the seller/installer plays a critical decision making role. The installer is a "packager," or final manufacturer of the system. He is the system designer. He takes a number of components, frequently from different manufacturers, and assembles them into a system. Thus, it is not a final product which flows through the channel to the installer, but rather a group of components. Therefore, the installer could be thought of as the final manufacturer (of the heating and cooling system), while the other members of the distribution channel are industrial goods suppliers.

The installer could also be viewed as another industrial goods supplier, since the heating and cooling system is another component in the final end use product, which is the home. The consumer does not purchase the heating and cooling system as an end use product, and in the technical

sense, the seller/installer is a manufacturer more than a conventional retailer. Therefore, the relationship of the consumer to the seller/installer is not the same as the normal consumer-retailer relationship for consumer durable goods.

Consumers as Solar Decision Makers

The ultimate consumer of technologically complex products has traditionally had little or no input into the selection decision for components that make up the product. This holds for many products across industries including home computer systems, audio components, and home building. The decision as to the type of system components and their relative advantage has typically been made by the packager/manufacturer. Given the complexity and technical nature of such systems and related decisions, it should be recognized that most consumers have inadequate knowledge and experience for making such decisions.

It is important to recognize the difference between heating and cooling systems and other durable goods with respect to consumer use patterns and decision processes. For example, with automobiles and major home appliances consumers have considerable "hands on" experience and well developed use patterns, resulting in sets of salient attributes and evaluative criteria. What consumers lack in technical knowledge is not so critical because other attributes such as psychological appeal and user benefits tend to be adequate for decision making in the absence of technical knowledge. For residential heating and cooling, consumers have little experience with product attributes other than comfort level and costs of operation and maintenance. Unlike other durables there are no substitutes for technical factors which consumers can use for making decisions. The majority of heating and cooling equipment sales go into new construction. It is therefore the builder who typically makes the equipment purchase decision. The builder's choice of equipment is strongly influenced by the installer. The installer serves as a "final manufacturer," of the system. Installers exhibit control over equipment selection by assembling components from different equipment manufacturers into a functional system. The design of the home and general mode of construction do provide parameters for system design. However, the installer still has considerable latitude in component selection and system design even when the requirements for heat gain and loss and air flow are considered.

Homeowners are rarely involved in heating and cooling system decisions except when they are confronted with replacement decisions which arise from equipment failure. Until recently, equipment failure was the only real motivation for shifting fuel source. There was seldom reason to consider replacing systems which were performing well. Homeowners have normally not seriously considered changing fuel source in the replacement decision because of the high initial cost of replacement. Now the high cost of oil is causing many homeowners to consider other fuel sources.

Even in replacement decisions, homeowners have traditionally deferred to the experts in making decisions. System decisions are rather technical in nature, involving degree days, heat gain and heat loss calculations, comparison of relative efficiencies, air flows, and equipment design features, to name a few. The primary consideration for the homeowner has been price and soliciting bids from several installers. Homeowners have been forced to rely on the installer's assurance that the replacement system would offer the same level of performance as the previous one. In some cases brand name recognition becomes an influential decision variable for consumers in the absence of technical knowledge.

The decision process for heating and cooling systems is a complex one, involving multiple decision makers and deci-

sion roles. The consumer's role has been minimized by a lack of technical knowledge and competence to judge the critical decision variables. Yet, in the case of solar options, which are merely an alternative method of heating and cooling, the majority of information and incentives are being directed at the consumer. The consumer is being asked to become the principal actor in evaluating products and making purchase decisions for residential solar systems. This is being done without adequate educational preparation of consumers.

Solar Distribution: A Mismatch of Channel and Product

Channels for distribution and sale of residential heating and cooling systems are well established and functional. However, these traditional channels have been little utilized for the distribution of solar space heating products. Of twenty-five solar installers listed in the 1978 Oregon Solar Directory, only five were also sellers of conventional heating systems. Conversely, this also means that only five of the hundreds of Oregon heating and cooling contractors were selling solar equipment. Solar installers are typically new, small firms specializing in solar equipment.

The selection and/or development of distribution channels for solar products is a sticky problem. There is no single "product," but rather a vast array of different products many of which are distributed in different channels. The basic differentiation of passive and active systems exposes the difficulty in identifying appropriate channels of distribution for the divergent products.

Many solar proponents consider passive solar systems to be the most feasible and desirable alternatives to conventional fuel systems. However, formal distribution channels for passive systems do not exist. Passive systems are largely design concepts which do not require distribution of physical products. They do require dissemination of designs, engineering specifications, performance data, etc. to architects, engineers, and builders. Because this is not a normal physical product supplied by manufacturers for profit, distribution becomes quite difficult to organize and coordinate. It is not yet clear how this can effectively be done on a large scale.

In contrast, active solar systems are mechanical products, for which distribution channels are available. However, unlike most other new heating and cooling products, solar equipment has not come from standard industry suppliers. When central air conditioning and heat pumps were introduced, they were distributed through the same channels as other heating and cooling products. This was largely due to the fact that the manufacturers were the same as for existing products. When technological innovations come from manufacturers who are new to an industry, these new manufacturers do not have established relationships with existing channels of distribution. So the distribution difficulties being encountered with solar should come as no surprise; it is common to technological innovations.

Carrying the problem a step further, the different forms of active solar products may even require different channels. For example, water transfer systems may require plumber/installers, air transfer systems may require ventilation/installers, and photovoltaic cells may require electrician/installers. Such a pattern could preclude one channel of distribution from supplying all types of solar heating and cooling systems. It can further create problems for equipment suppliers, who may be forced to build and maintain relationships with multiple channels, which can be very costly. It can also create problems for installers who may have to build relationships with multiple suppliers. Finally, such a pattern could severely dilute the kind of concentrated effort which is generally needed to launch a new technology. Another major difference between solar and conventional heating and cooling equipment is

that of sizing, installation and maintenance. Existing installers are presently ill-equipped to handle these activities. A substantial investment in training and equipment is required to adequately prepare for handling these new and different products.

Non-Acceptance by the Channel

Solar options are merely another way of heating and cooling buildings. It therefore seems logical that regular heating and cooling contractors would be best equipped to handle sales and distribution. However, it appears that they have been generally less than enthusiastic to date. That is not to say they have no interest at all. From a profit perspective, market demand is insufficient to warrant the required investment in training and preparation. Solar equipment will not be marketed by distributors of conventional heating and cooling equipment until such time that it becomes profitable for them to do so.

However, the blade cuts both ways. Francis DeWinter, Chairman of the American Section of the International Solar Energy Society (1978), commented that "designers balk at marketing their products through existing networks."¹ The apparent lack of interest shown by the established channels has caused these channel members to be viewed as competitors and anti-solar, simply because they are carrying on business as usual. Thus, they are labeled as outsiders, and producers of solar components or systems look elsewhere for distribution. As solar equipment manufacturers are learning, it is very difficult and expensive to build new distribution channels from the ground up. If the installers are not existing businesses with years of experience and a good local reputation, they may have difficulty gaining either consumer or designer/builder acceptance. Thus, distribution of solar systems may be the classic "chicken and egg" situation. Conventional distribution channels may not handle them until they are successful and the systems may not be successful until they can be processed by existing channels.

Future Distribution Patterns for Solar Options

Distribution channels evolve just as consumer adoption evolves over time. Distribution channels for a new technology often change as part of an evolutionary process. The introductory stage of the product life cycle for solar products has been characterized by a select group of solar innovators. These innovators are generally homeowners who have designed and built their own solar systems. They are motivated more by their interest in the technology and lifestyle than by economic justification for the investment.

Solar distribution channels have evolved from a state of non-existence, where the consumer innovators designed and assembled their own systems to the current state where many solar specialty firms now sell and install solar products. These firms are the channel innovators. Like the consumer innovators, they are motivated more by their interest in furthering the cause of solar technology than by a profit motive. The widespread lack of business experience, operating capital, and economic rationale will result in a high mortality rate for this group of channel innovators.

Just as the mass market, or majority of homeowners, will differ in characteristics from solar innovators, so will the eventual seller/installer differ from the retailer innovators. There will be heavy capital requirements for inventories, installation and test equipment, and education for sizing, installation, and maintenance. This invest-

ment may be compounded by the increasing sophistication and diversity of solar products. Therefore, it may become increasingly difficult for new firms to assume the role of the seller/installer, and thereby force solar options to be distributed through existing channels for heating and cooling products.

There are differing views as to the future development of solar distribution channels. Some believe that new channels must be created because of the lack of interest exhibited by members of conventional channels. One scenario is that the "solar system packager" is the best distribution alternative, but this is merely a solar specialist who operates in the same manner as conventional heating and cooling contractors.

The more prevalent view reported in the literature is that conventional channels will be the primary outlet. This seems logical and there are numerous signs of such development:

1. Johnson (1976), in discussing policy options for solar, contended that the regular operation of the national building industry is the most reasonable method for integrating solar options.
2. Overtures toward the solar industry are being made by conventional channel members. Trade publications in the traditional heating and cooling industries are increasingly reporting on the solar industry.
 - a. Air Conditioning, Heating, and Refrigeration News is the major weekly information source for this industry. It runs many news reports as well as frequent articles on solar activities.
 - b. The primary trade publication for the oil heating industry has been for years Fuel Oil & Oil Heat. This monthly periodical not only includes regular articles and reports on solar systems, but it recently added and Solar Systems to the title of the publication. This appears to be a sign of things to come.
 - c. Many of the articles which appear in these and other industry publications are focused on the interface between these conventional channels and the solar industry. For example, fuel oil suppliers and installers are being encouraged to become involved in selling and installing solar systems. It is claimed that they are compatible and that the dealers' future livelihood may be dependent on such a transition.
3. One recent article predicted that if solar is to achieve the market share being predicted over the next five to ten years, conventional distribution channels will have to be utilized in order for solar to become a mass market item (Air Conditioning, Heating and Refrigeration News 1977).

Regardless of whether new or existing firms eventually assume the principal role of seller/installers, the traditional decision making structure for heating and cooling systems — including solar options — can be expected to remain unchanged. That is, the majority of the purchase decisions are made by builders, and the majority of the sizing and configuration decisions are made by the seller/installers. Therefore, the current focus of marketing research on the homeowner as the decision maker is both inadequate and short sighted.

¹ Cited in "Solar Conference Hunts for Keys to Future Growth," Engineering News Record, Vol. 201, No. 9, September 7, 1978, p. 10.

Facilitating Distribution to the Builder and Seller/Installer

If solar heating and cooling is to meet growth targets, technology may have to take a back seat to marketing and industry organization. Channel members must see demand creation, manufacturer support, and profit opportunity if they are to adopt new technologies and incur the accompanying expenses. There are a number of facilitating factors which are critical to converting existing channel members to solar.

1. Involvement of conventional heating and cooling contractors could be enhanced by the development of packaged systems which could be more easily sized and installed by contractors.
2. Development of industry standards for comparison of solar system performance would dramatically reduce the frustration of explaining systems features and performance to prospective buyers, who are presently confused by conflicting information.
3. A substantial improvement in promotion and effective dissemination of technical information is needed. For the most part, solar manufacturers are underfunded and cannot afford adequate promotion. As yet, total industry sales are insufficient to provide the needed funds.

Directions for Consumer/Market Research

Three areas for future research appear promising using solar products. Decision making research has to date focused on the homeowner/consumer of the solar product. Future decision making is expected to shift from the homeowner to the builder and finally to the seller/installer, following the same pattern which presently exists for conventional heating and cooling products. The study of solar products offers a unique opportunity for research. Major product innovations that cause shifts in consumer decision making processes are rare. Suggested topics for study include causes of shifts in decision makers, and differences in the information acquisition process that occur at various points in the product life cycle for different decision makers. As consumer researchers we must explicitly recognize the different decision making segments and decision making roles as they shift over time.

Solar products are in search of a channel for distribution. As such, a longitudinal analysis of evolving distribution channels for a series of high technology innovations is possible. Worthy topics for investigation might include the modeling of variables that facilitate and enable shifts in the predominant channel decision maker. Channel conflict is another topic that is interesting in the study of traditional vs. new channel members.

Industrial marketing is a third area of research related to channel decision making. Multiple decision makers in the homeowner — builder — solar seller/installer triad permit the analysis of interaction both within and between channels distributing solar products.

Directions for Public Policy Decisions

It was pointed out earlier that homeowners/consumers have been the focus of promotion and incentives for solar products. This runs counter to the distribution and decision patterns for conventional heating and cooling products. Promotion and incentives for conventional heating and cooling products are primarily directed from the equipment manufacturer to seller/installers and builders, much the same as the pattern for industrial goods. In contrast to industrial promotion, consumer durable goods promotion is usually through or tied in with seller-installers. Solar,

promotion and incentives have come primarily from government and private sources — rather than existing channel members.

We have suggested that as solar products mature along the product life cycle, alternative distribution channels for solar products will necessarily unfold. This change in distribution, along with the associated shift in decision makers dictates that public policy programs and research be segmented. Information and incentive programs must eventually be shifted from the homeowner, and tailored for the designer/architects, builders, and retail solar sales and installation firms. Lastly, information and programs must be developed for conventional heating and cooling, plumbing, and electrical contractor channels as cost effective air, liquid and photovoltaic systems are developed. Again, both information and incentive programs are needed for the specific decision making segments as they develop to meet the changes in product technology and market demand.

A higher level of government involvement could accelerate the solar adoption process. For example:

1. Federal and state governments could carry the burden of "institutional" selling, or promotion of solar systems in general. Then the producers would only have to promote their own products, as the product class would be presold by public funds. This approach could, however, encounter opposition from competing industries.
2. A very strong impetus could be provided by government mandation of energy conservation or solar options. Two California communities have recently provided such a boost. The city of Davis has a performance standard for energy conservation which has resulted in the incorporation of solar options in new construction. San Diego County has mandated solar water heating in all new homes constructed in unincorporated areas of the county. These activities will obviously result in demand creation which will heighten the interest and involvement of conventional channel members.

Conclusions

When viewed in the larger context of developing technologies, it is obvious that the problems confronting residential solar heating and cooling are not unique. These adoption and diffusion problems are characteristic of most new technologies. When looking at the history of residential heating and air conditioning, it can be seen that similar problems have been previously encountered with new products in the same industry.

One of the primary causes of the slow rate of adoption of solar heating and cooling is inadequate and ineffective distribution of the products. This is often a problem with new technologies, but it is compounded by some unique aspects of solar technologies. The diverse array of solar products and complexity of application decisions create difficulties for both seller and buyer. There are many suppliers of solar products, most of whom are undercapitalized and are not traditional suppliers to the heating and cooling industry. A lack of product standards and slow accommodation of building codes pose problems for seller/installers. These factors combine to deter the type of demand pull and facilitation which normally help to develop functional and effective distribution channels. Solar heating and cooling products are presently at a very difficult stage in the evolution of distribution channels. Many of the current seller/installers are "innovators" who will eventually give way to more traditional heating and cooling operations. In the opinion of these writers, the conventional heating and cooling distribution channels are well equipped to handle solar products and they or similar outlets will eventually assume the role.

A separate but related problem is that solar products are being treated erroneously as consumer durable goods and marketed directly to consumers. This is in stark contrast to traditional distribution and marketing patterns for conventional heating and cooling products. They are treated more as industrial goods with technical decisions made by installers and most purchase decisions made by home builders. Prospective solar consumers are being asked to make product evolution and application decisions which they have previously not made for this product class. Research, information, and incentives for solar products have been directed at consumers. These efforts have, however, failed to adequately consider the critical decision roles played by the seller/installer and builder. In the future, both marketing research and marketing strategy development must address the multiple decision makers and their interactive decision roles if they are to be meaningful and effective.

References

Johnson, Augustus C. (1976), "Commercialization of Solar Heating and Cooling of Buildings," in Sharing the Sun: Solar Technology in the Seventies, Joint Conference of the American Section of the International Solar Energy Society and the Solar Energy Society of Canada, Inc., Winnipeg, Manitoba, Canada, Vol. 10, 35-39.

Oregon Solar Energy Directory (1978), University of Oregon Solar Energy Center.

"Solar Conference Hunts for Keys to Future Growth," (September 7, 1978), Engineering News Record, Vol. 201, 9, 10.

"Solar Industry Needs to Straighten Out Messy Distribution" (September 12, 1977), Air Conditioning, Heating and Refrigeration News.

PROVIDING REINFORCERS FOR ENVIRONMENTALLY UNCONCERNED CONSUMERS

Michael L. Rothschild, University of Wisconsin

Abstract

The marketing concept and behavioral learning theories both emphasize the need for positive reinforcement of behavior if repeat behavior is desired. In nonbusiness situations there is often a situation where there is individual cost and societal benefit but no individual benefit. The papers in this session show examples of this type of case in the area of ecological/environmental marketing and also suggest methods for dealing with this issue.

Discussion

Important new areas have been entered into by marketers and consumer behavior researchers in the past few years. These include public sector, nonprofit and public policy arenas.

In the public and nonprofit areas one deals with a complex set of issues including intangible products and prices, which translate to difficult to perceive benefits and costs. In public policy issues one tries to aid policymakers who may need to have an impact upon the knowledge and/or behavior of the general public. The area of environmental and/or ecological concern touches on both of these thrusts (i.e. public/nonprofit and policy) as can clearly be seen in the three papers presented in this session.

Two other recent papers consider issues parallel to those of environmental concern from a consumer behavior perspective. These papers deal with marketing communications in nonbusiness situations (Rothschild 1979) and a behavioral learning theory perspective of marketing (Rothschild and Gaidis 1980). Both relate to the three papers of this session; the three environmental papers serve as examples of the points made in the two theoretical papers.

As the tools of marketing develop uses in these areas of societal thrust, one can see a weakening of the marketing concept. The marketing concept offers the philosophy that success is achieved through meeting the needs of consumers. A parallel model is offered by behavioral learning theory and states that the probability of any future behavior is increased when that behavior has been positively reinforced in the past. These two models are essentially identical.

As marketers attempt to meet societal needs, though, they may lose sight of the exhortation to meet individual needs as well. Individual behavior which is not reinforced will not be maintained in the long run. Examples of this paradox abound in nonbusiness cases; the three papers presented here provide just such cases. It is clear that there are societal benefits to be derived from appropriate individual behavior; it is far less clear that the costs associated with the individual behavior will be offset by individual benefits. These cases are, then, removed from the *quid pro quo* type of exchange which typifies private sector marketing. The three papers provide insights into dealing with these unbalanced types of exchanges.

One can also consider the policy implications of unbalanced exchange situations. If there is a societal benefit and a societal need then perhaps policymakers should be called upon to provide sufficient incentives for behavior so that the individual cost benefit relationship is perceived favorably. From this perspective policymakers as well as

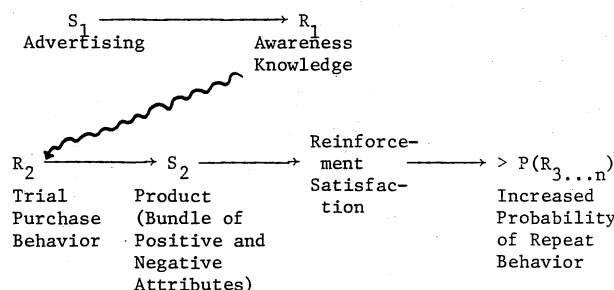
marketers are concerned with environmental issues.

Parenthetically, in the Henion, Gregory and Clee paper, reference is made to an earlier Henion paper (1978). In this paper Henion suggests that private enterprise is a viable alternative to governmental regulation in the area of environmental issues. That would seem to be a difficult assignment for a firm operating for profit. Perhaps sufficient individual reinforcers are only available through government subsidy; perhaps no individual firm or person has sufficient incentive to provide these reinforcers. An alternative to positive reinforcement which is available to government is brute force of law. Legislation may, by itself, provide adequate incentive for behavior. Again, the papers speak to this issue as well.

Figure 1 shows a model which combines education with positive reinforcement. Verbal learning theories can contribute to the top set of relationships; behavioral learning theories are necessary to the lower set of relationships. The three papers provide insight into application of this behavioral model.

Henion, Gregory and Clee discuss trade off analysis with respect to nonphosphate detergents and present a nicely done piece of research. An analysis of tradeoffs implies a search for reinforcers. One can see from their data that there are some individuals who will accept non-phosphates with their accompanying higher prices and lower cleaning power.

Figure 1
Learning Theory in a Marketing Context



Source: Rothschild and Gaidis 1980.

Given real world market shares it is clear, though, that most consumers do not see sufficient individual reinforcement in this product. Low phosphate detergents clearly provide a societal benefit but not an individual one. Low price and clean clothes help individuals; most individuals opt for these personal reinforcers.

The Henion, Gregory and Clee paper presents a useful method to show how consumers value the possible tradeoffs across attributes. Such a method would be useful in testing specific reinforcers and may also show that industry is unable to provide the necessary reinforcers. In this case the issue would become a policy one if society had a sufficient need for low phosphate detergents.

Crosby and Gill use path analysis to examine voting behavior with respect to a returnable bottle referendum in Michigan. Again a useful method is used well in providing

insights as to motivations and reinforcers.

In this case the events were more policy related since voting on the proposed policy was about to occur. The main issue of the paper, though, concerns which reinforcers to use to motivate the voter. In this paper, as in the previous one, one sees that the group of environmentally concerned citizens is quite small in the absence of reinforcers and that the private sector was unable to provide sufficient individual benefit to gain appropriate behavior. The referendum was a test of societal needs. Passage of the referendum was a mandate to policymakers to provide sufficient incentives and/or legal sanctions.

The data show that perception of cost is the main driving variable of vote; this is consistent with behavioral learning theory. Where there is less concern with cost, more votes result; where the perception of cost is low, more votes result. Secondly, behavioral variables are more important than attitudinal or SES variables. Again, individuals behave based on cost benefit relationships and reinforcers.

Later in their paper, Crosby and Gill state that the marketers of the bill marketed the benefits of reduced litter and minimized the perceived cost. Here a societal benefit (reduced litter) was transformed to an individual benefit and the individual benefit was minimized.

The third paper, by Blakely and Smith discusses the distribution of solar related products. Once again there is a case of insufficient reinforcement for consumers; demand for solar products is still quite low.

Blakely and Smith feel that a change in distribution systems (from a consumer directed to a builder/contractor directed system) would help to stimulate demand. Unfortunately the authors have no compelling argument as to why builders or contractors would be more reinforced than were consumers. These authors do, though, suggest a system of reinforcers. They point out the current lack of reinforcers and recommend a system of incentives and legislation. By doing so, they again bring a product class into the realm of public policy.

The three papers show commonality when considered within the framework of using reinforcers for nontraditional products. All too often private sector marketing tools are blindly applied to new situations. The profession is slowly learning that some nonbusiness problems are quite difficult and require new or modified marketing paradigms. The methods of path analysis and trade off analysis provide useful insights for dealing with these problems.

Researchers and practitioners in nonbusiness situations are urged to consider behavioral learning theory and the notion of reinforcement in dealing with these situations. Without reinforcers there will not be long-run behavior. The three papers in this session provide applications of this model.

References

- Henion, K. E. (1978), "Ecological Marketing: From Normative to Descriptive Model?" in Future Directions in Marketing, J. Arndt, G. Fisk and K. Gronhaug, Eds., Boston: Marketing Science Institute, 301-311.
- Rothschild, M. L. (1979), "Marketing Communications in Nonbusiness Situations or Why Its So Hard to Sell Brotherhood Like Soap," Journal of Marketing, 43, 2, 11-20.
- Rothschild, M. L. and Gaidis, W.C. (1980), "Behavioral Learning Theory: Its Relevance to Marketing and Promotions," University of Wisconsin Working Paper Series.

INTEGRATION OF ECONOMIC GEOGRAPHY AND SOCIAL PSYCHOLOGICAL MODELS OF
PATRONAGE BEHAVIOR

Robert F. Lusch, University of Oklahoma

Abstract

This article develops a model of patronage behavior which incorporates key concepts from geography, social psychology and economics. The model is shown to be rich in explanatory power but its predictive power is not assessed.

Introduction

The study of retail patronage behavior is not new. Over fifty years ago W.J. Reilly (1929) in his now classic monograph developed a gravity based model of retail patronage behavior. This model postulated that a household's attraction to a community for shopping purposes was positively related to the size of the community and inversely related to the distance to the community. Some thirty years later David L. Huff (1962) transformed Reilly's law into an intrurban patronage model. Huff's model suggested that a household's attraction to a store was positively related to the size of the store (i.e., square feet) and inversely related to the distance from the household's place of origin to the store. At about the same point in time Bob R. Holdren (1960) in an elegant model of the structure of retail markets postulated that the buyer evaluates the store offering on three dimensions: (1) the distance the buyer is from the store; (2) the level of prices at the store; and (3) the level of non-price offer variation. All three of these authors viewed patronage behavior from an economic geographic perspective. That is patronage behavior was viewed as being determined by the structure of physical space. This spatial orientation to patronage behavior continues to be pursued. For example, recent work by David MacKay (1973), Dennis Lord (1975) and William Young (1975) is reflective of the spatial research orientation toward patronage behavior.

On the other hand some early pioneers in patronage behavior research attempted to explain shopping behavior in terms of social-psychological concepts. Utilization of these concepts resulted in researchers treating shopping behavior as being unaffected by physical space. Notable early examples of this approach include Gregory Stone's (1954) work on categorizing shopping orientations used to evaluate stores. Stone's shopper orientations fell into five categories: (1) economic, (2) personalizing, (3) ethical, (4) apathetic, and (5) a residual category of unique or indeterminate criteria. Another early example includes the work of George Fisk (1961-1962) in which the consumer's image of the store was postulated as the major latent guiding force in determining patronage behavior. The social psychological approach to explaining patronage behavior continues to receive attention. Some relatively recent examples include the work of Darden (1980), Granzin and Painter (1976), Monroe and Gultinan (1975), James et al. (1976), Miller (1976), and Marks (1976).

As far as the author is aware the only attempt to integrate the geography and social psychology approaches to studying patronage behavior is in the area of what has been labeled cognitive mapping. In this stream of research the consumer's perceptions or cognitions of store size and/or distance are utilized in traditional geographic patronage models rather than real distances or actual store sizes. This approach is relatively recent and includes the research of Maze (1974), Cadwallader (1975) and MacKay and Olshansky (1975). Findings from this stream of research suggest that cognitive or perceptual measures are better

predictors of behavior than physical or real measures.

The purpose of this paper is to provide further integration of the geography and social psychological approaches to the study and explanation of patronage behavior. In addition useful concepts from microeconomics are integrated into the proposed patronage model. Portions of the model are taken from a theory of retail structure recently developed by Ingene and Lusch (forthcoming).

Theoretical Development

As in microeconomic theory we begin by assuming that the consumer (i.e., shopper) is rational. It is however recognized that not all consumers have the same rationality. What may be rational from one shopper's perspective may not be rational from someone else's perspective, but, nonetheless rationality as the individual defines it does exist.

In addition it is assumed that shoppers don't have perfect information. Therefore shoppers must make decisions in the face of uncertainty.

The fundamental underlying proposition upon which the model is based is that shoppers will make rational shopping decisions so as to maximize their net consumer surplus. Consumer surplus is a concept borrowed from economics. Consumer surplus occurs when goods are sold at a price lower than the maximal demand price a consumer would be willing to pay for them. It is because goods are offered by retailers below the consumer's maximal demand price that consumers are enticed into expending effort to shop for goods.

When the price the consumer must pay for the product at the store plus the costs associated with acquiring the product, such as the cost of traveling to and from the store, exceeds the consumer's maximal demand price shopping will become uneconomic. The preceding implies that there are certain costs incurred in shopping. When these costs are subtracted from consumer surplus one arrives at net consumer surplus. It is net consumer surplus that the consumer attempts to maximize.

The Patronage Model

The model that is being proposed of patronage behavior can be formalized as follows:

$$E(W_j) = P_j (P_m - P_e) q - 2rt' - [(2r/m) + s]t'' - t''' \quad (1)$$

where:

- $E(W_j)$ = expected net consumer's surplus from shopping at store j,
- P_j = probability that if store j is patronized that the product being sought is in stock,
- P_m = maximal demand price per unit given quantity desired,
- P_e = expected price per unit,
- q = quantity desired,
- r = distance in miles from store to household departure point, and therefore $2r$ is equal to round trip distance,
- t' = cost per mile of travel,
- m = speed of travel in miles per hour,
- s = expected in-store shopping time,
- t'' = opportunity cost per hour of time,
- t''' = psychic costs (benefits) of shopping trip evaluated in dollar terms.

Appreciation and understanding of this store patronage model can be facilitated by an elaboration of its individual components.

The first term (P_j) is included in order to capture the degree of uncertainty that confronts the shopper. Once a consumer recognizes a need for an item it must assess the set of stores which may have the item and estimate the probability that the item is in stock at each of the respective stores. This probability is the P_j term and it captures uncertainty on the part of the shopper. In reality the product the consumer plans to shop for either is or is not in stock at the store. That is if the consumer knew the truth then P_j would be assigned a value of 0.0 or 1.0. But the consumer does not know reality (i.e., certainty) and therefore P_j is its estimate of the likelihood that the item will be in stock at the store.

The term P_j is itself a function of several managerial and behavioral variables that are not explicitly included in the model. Most important from a managerial and behavioral perspective is the information the prospective shopper has available. This information is a function of the quantity and quality of promotion on the part of the retailer and also of the information processing capabilities of the shopper. Theory and concepts from social psychology can therefore be used to empirically study the determinants of P_j .

The second term, the maximal demand price (P_m) captures the consumer's desire, need or want for the item and the consumer's ability to acquire the item. As the consumer's purchasing power increases the maximal demand price will rise. Purchasing power may increase due to rising income or more readily available credit. Thus the retailer can have some influence over P_m via its credit policies. In addition the retailer could advertise or promote the item to have certain unique and strong traits or benefits which could result in the consumer wanting or desiring the item more and therefore shifting P_m upwards. Behaviorally oriented theories of advertising could be helpful in shedding light on this matter.

The expected price per unit (P_e) reflects what the consumer believes the retailer will charge for the item being sought. Once again this term suggests the shopper faces uncertainty. A shopper seldom knows a priori what the precise price of an item will be. Therefore the shopper must estimate the price. As you might well expect the retailer's promotional efforts--especially its advertising--can aid the shopper in better estimating prices. In addition the process of learning will become important in establishing (P_e). Everytime the consumer shops at a store it learns more about its pricing policies.

The quantity (q) the shopper plans to purchase on any given shopping trip is basically an inventory decision if the item is a nondurable rapidly consummable item such as food. A household will tend to purchase nondurables in larger quantities when its income is higher; the cost of travel is high; time pressures are great; storage is available and household size is larger. Most durable items (e.g., cars, furniture, appliances) are purchased one at a time. If more than one is purchased on a single trip it is usually due to income or wealth factors.

Combining together the four variables we have discussed (P_j , P_m , P_e , q) will allow us to formalize expected consumer surplus. Expected consumer surplus is equal to $[P_j(P_m - P_e)q]$.

From this the costs of shopping must be subtracted. The major variables influencing the costs of shopping will next be discussed.

In traditional economic geography models the most important variable is (r) or the distance from the store to the shoppers point of departure--usually its residence or place of employment. Naturally this variable is a function of where

the retailer locates its store and the location of households or businesses surrounding it. But (r) is also a function of city planning of road networks. The straight line distance between the store and shopper may be short but if travel is to occur on existing roadways the distance could be considerably greater. From a behavioral perspective (r) is a function of learning, information availability and information processing. The more times one travels to the store the more they will learn the shortest route. Or in terms of a first trip the more the information the consumer has on how to get to a store the better it will be able to travel the shortest route and thus minimize shopping costs.

Travel costs per mile are also important in determining patronage behavior. In the model the cost per mile of travel is labeled (t'). This important variable is influenced by the availability and cost of public transportation and if private transportation is used the cost of operating an auto. Costs of operating an automobile will vary by region of the country due to varying rates for insurance, gasoline, and taxes in different parts of the country. The total round trip costs of traveling to the store are represented in the patronage model by $2rt'$.

The costs of shopping not only include travel costs but also the implicit cost of time. As families are increasingly facing a scarcity of time the opportunity cost of that time rises. In the model the opportunity cost per hour of time is labeled t'' . This variable is a function of the alternative opportunities an individual has for leisure or work. One would expect that some of the correlates of t'' would include: income, lifestyle, and social class.

The total amount of time spent on the shopping trip can be divided into two components: (1) travel time; and (2) in-store shopping time. Travel time is equal to $(2r/m)$ in the patronage model, where m is average speed of travel in miles per hour. In most part speed of travel is a function of the type of transportation used, speed laws, road networks, and road congestion. In-store shopping time is labeled s in the model. This time variable is a function of a variety of factors which include: (1) the amount of information the shopper has on where the item(s) is located in the store; (2) in-store congestion; (3) store layout; (4) degree of sales assistance available; and (5) merchandise availability.

The cost of time per hour (t'') when multiplied by the amount of time spent on the shopping trip $(2r/m)+s$ will determine the total cost of time devoted to the shopping trip.

The final parameter in the patronage model is t''' or the psychic costs or benefits of the shopping trip. The shopping trip cannot only cost in terms of direct monetary travel costs and the opportunity cost of time devoted to the shopping trip but also in terms of psychic costs. For example let us assume that in order for a shopper to travel from its home in the suburbs to a large downtown department store that it must maneuver its auto over 32 miles of congested freeway and then through 12 blocks of ghetto neighborhoods. In addition to the time and travel involved this trip may be very nerve wrecking and tension creating thus resulting in certain psychic costs.

In principle psychic benefits can also accrue during a shopping trip. That is the shopping trip may be viewed as an enjoyable experience and thus of value in and of itself. For instance assume that it has been a long cold Winter and with the arrival of Spring the consumer desires to load his family into the car and drive to a regional shopping center forty miles away in order to browse in the shopping mall and look at the new Spring fashions. This shopping trip may actually be an enjoyable experience thus creating psychic benefits vs. psychic costs.

Explanatory Power

The traditional patronage models in economic geography, such as the Huff model, have good predictive power but are lacking in explanatory richness. This is not the case with the proposed patronage model. Not only can the proposed model explain the impact on patronage of changes in behavioral, environmental and geographic variables but it is also capable of explaining well accepted theories such as central place theory, Reilly's law and the Huff model.

Behavioral Variables

There are at least several behavioral variables which can fit well into the model and increase its potential explanatory power. Some of these variables include information processing, attitudes and consumer certainty.

Lets briefly look at information processing. As a person obtains more and better information about a store they will be better able to make shopping decisions which will increase their net consumer surplus. Increased net consumer surplus via better informed decision-making will lead to consumers being more store loyal.

A second behavioral variable is customer attitude toward the store. If attitudes become more favorable then the store should become a more enjoyable place to shop which should reduce the psychic costs (t'') of patronizing the store. This in turn could increase the distance (r) a shopper would be willing to travel to visit a store.

A third behavioral variable regards the degree of certainty the consumer has that the product being shopped for is in stock. As certainty increases the number of patrons a store can potentially attract increases because the distance a consumer would be willing to travel (r) rises. At the same time existing patrons will become more loyal because their net consumer surplus will rise.

Environmental Variables

In terms of environmental variables we could investigate the impact of changes in things such as transportation costs and speed of travel. As transportation costs rise or the average speed of travel declines several things could happen. Consumers may purchase in larger quantities so they will not need to make as many shopping trips; they may try to gather more information to increase their certainty as to whether the product is in stock; they may decide to patronize more stores closer to their residences or places of work; or they may try to reduce the amount of time involved devoted to in-store shopping by more carefully planning their shopping trips.

Geography Concepts and Models

The proposed patronage model is not only capable of explaining the impact of changes in behavioral and environmental variables on patronage behavior it also helps to explain central place theory, Reilly's law and the Huff model.

Central place theory (Christaller 1966) tells us that there is a hierarchy of communities each of which is progressively larger and performs a larger group of retail functions representing an increased variety of product availability. The smallest communities perform the most basic functions and the larger perform more specialized functions. In addition each successively larger community performs all lower-order basic functions. The larger communities are spaced farther apart and are bordered by smaller communities. In terms of the proposed patronage model this suggests that the larger the community the greater the probability that an item can be found in a store within the community. The larger the community therefore the greater will be the distance (r in the model) people will be willing to travel to the community. This is in principle what cen-

tral place theory concludes.

Both Reilly's law and the Huff model postulate that patronage is inversely related to distance and positively related to size of store or size of community. In the proposed patronage model distance (r) is also incorporated in such a fashion that an increase in distance reduces net consumer surplus and therefore the likelihood of a patron shopping at the store declines. Size of store is not directly incorporated in the proposed patronage model. Importantly, however, what size of store is a surrogate for is included in the model. Size of store itself is not an attraction force but rather the increased width and depth of merchandise assortments which accompany large stores is the attraction force. When these assortments increase then the probability that a shopper will find the item(s) it is seeking will increase which will in turn increase consumer surplus. Increased consumer surplus will increase a store's ability to draw patrons. Thus we see that what the size variable is reflective of is more directly incorporated in the proposed patronage model. And importantly it allows for the same end conclusion to be drawn as the size variable in the Reilly and Huff frameworks.

Discussion

The patronage model as it has been presented is not complete in all respects. Further work and refinement is needed. For example the model assumes that only one product is being shopped for. This is unrealistic. To make the model more realistic multiple product searches should be included. If the shopping trip is a multi-purpose single stop trip then the model can be extended by summing across items to be purchased. However if the trip is a multi-purpose multi-stop trip then extension of the model is considerably more complex.

In addition the model could become richer if it not only explicitly included demand related variables but also supply and competitive variables. This would enable the retail manager to manipulate certain managerial variables in order to determine a profit maximizing strategy.

Finally the model is in need of rigorous testing of both its individual components and its total composition. It is believed that all of the variables in the model can be measured directly except the psychic costs or benefits of shopping. Since many of the correlates of the major constructs in the model are social-psychological in nature, indirect measures may be necessary for them.

To better understand the relationship between patronage behavior and each of the variables in the model several approaches could be used. First, calculus could be used to establish the impact of a change in each variable on net consumer surplus. Second, simulation could be used to investigate the magnitude of these effects. Third, empirical research could be used to test hypotheses derived from the two preceding approaches.

Conclusion

A model of patronage behavior has been presented which incorporates concepts from economics, social-psychology and geography. The model appears to be rich in explanatory power. With additional refinement its explanatory power should be further improved. Its predictive power, however, still remains to be assessed.

References

Cadwallader, Martin (1975), "A Behavioral Model of Consumer Spatial Decision Making," *Economic Geography*, 51, 339-349.

Christaller, Walter (1966), Central Places in Southern Germany, Englewood Cliffs, NJ: Prentice-Hall, translated by C.W. Baskin from the 1933 German edition.

Darden, William R. (1980), "A Patronage Model of Consumer Behavior," in Competitive Structure in Retail Markets: The Department Store Perspective, Ronald W. Stampfl and Elizabeth Hirschman eds., Chicago: American Marketing Association, 43-52.

Fisk, George (1961-1962), "A Conceptual Model for Studying Customer Image," Journal of Retailing, 37, 1-8, 51.

Granzin, Kent L. and Painter, John J. (1976), "Store Attributes and Shopper's Attributes As Determinants of Retail Store Patronage," paper presented at the Southwestern Marketing Association Conference.

Huff, David L. (1962), Determination of Intra-Urban Retail Trade Areas, Los Angeles, California: Real Estate Research Program at the University of California-Los Angeles.

Ingene, Charles A. and Lusch, Robert F. (forthcoming), "A Model of Retail Structure," Research in Marketing.

James, Don L., Durand, Richard M. and Dreves, Robert A. (1976), "The Use of a Multi-Attribute Attitude Model in a Store Image Study," Journal of Retailing, 52, 23-32.

Lord, Dennis J. (1975), "Locational Shifts In Supermarket Patronage," Professional Geography, 27, 310-314.

MacKay, David B. (1973), "Spatial Measurement of Retail Store Demand," Journal of Marketing Research, 10, 447-453.

_____ and Olshavsky, Richard W. (1975), "Cognitive Maps of Retail Locations: An Investigation of Some Basic Issues," Journal of Consumer Research, 2, 197-205.

Marks, Ronald B. (1976), "Operationalizing the Concept of Store Image," Journal of Retailing, 52, 37-46.

Maze, Edward M. (1974), "Determining Shopper Movement Patterns by Cognitive Maps," Journal of Retailing, 50, 43-48.

Miller, John A. (1976), "Store Satisfaction and Aspiration Theory," Journal of Retailing, 52, 65-84.

Monroe, Kent B. and Gultiman, Joseph P. (1975), "A Path-Analytic Exploration of Retail Patronage Influences," Journal of Consumer Research, 2, 19-28.

Reilly, W.J. (1929), Methods for Study in Retail Relationships, Austin, TX: Bureau of Business Research at the University of Texas--Austin.

Stone, Gregory P. (1954), "City Shoppers And Urban Identification: Observations On The Social Psychology Of City Life," The American Journal of Sociology, 60, 36-45.

Young, William J. (1975), "Distance Decay Values And Shopping Center Size," Professional Geography, 27, 304-309.

WOMEN'S SELF-ASCRIBED OCCUPATIONAL STATUS
AND RETAIL PATRONAGE

Elizabeth C. Hirschman, New York University

Introduction

Since the early 1970's there has been a steady research stream concerned with the increasing occupational diversity of women and the relationship of this occupational diversity to women's behavior as consumers (e.g. Bartos 1977; Lazer and Smallwood 1977; McCall 1977). Investigations conducted at the outset of this stream of research tended to dichotomize women into only two groups: employed and unemployed (e.g. Anderson 1972; Douglas 1976; Strober and Weinberg 1977).

Later studies, however, utilized a three member typology: housewives, "working" women and "career" women; where "working" women were defined as those who viewed their employment as a job and worked primarily for financial reasons, and "career" women were those who viewed their employment as a career and worked primarily for self-fulfillment (e.g. Joyce and Guiltinan 1978; Hafstrom and Dunsing 1978; Bartos 1977).

Based on empirical findings that support the validity of a three-member occupational typology for women, Bartos (1977 p. 37) writes: ... We must go beyond a simplistic comparison of working women with housewives if we are to understand these two kinds of markets. The contrast found in the ... behavior of career women and the "just-a-job" working women is dramatic evidence of the need to go beyond the perspective of (employed) women as a monolithic group."

Purpose

The present research has two purposes. The first is to explore the overlap between a woman's self-ascribed perception of her occupational status (i.e. housewife, working woman, career woman) and her externally-ascribed occupational status. For example, professional-managerial, sales-clerical and unemployed occupational categories are three externally-ascribed status categories into which many women are often grouped by researchers (Miller, Schooler, Kohn and Miller 1979; Greenberg, Straight, Hassenger, and Roska 1978; Tumin 1967). In consumer research on such externally ascribed occupational status groupings, professionally employed women often are implicitly assumed to view themselves as "career" women, while women having sales-clerical jobs are frequently assumed to hold "working" women self-perceptions, and 'unemployed' women assumed to view themselves as housewives (e.g. Joyce and Guiltinan 1978). One important research hypothesis, therefore, is whether this implied congruence between externally-ascribed and self-ascribed occupational status actually is present for women.

If this congruence is found, the second objective of the research will be to examine differences among the self-ascribed status groupings of women in terms of retail patronage variables. In this regard, the product area of apparel is especially important, because one of the primary ways an individual displays status to others is through apparel (Belk 1978, Holman 1980). Four areas of potential differences among Housewives, Working Women and Career Women will be investigated: (1) types of stores shopped for apparel, (2) times at which apparel shopping is conducted, (3) the dimensional structure of store attributes relevant to apparel purchases and (4) information source importance in apparel purchases.

Internal and External Status

It is important to gain an understanding of the occupational status concept and the perspectives from which it may be viewed. The term status refers to a social position; social positions may be defined along many dimensions. Wealth, educational attainment, sex, race, religion and occupation are some of the more common dimensions for defining status in Western society (Tumin 1967). Within a given society, some social positions are normatively valued more/less highly than others, and hence may be ordered hierarchically in terms of their prestige or social value. This normative, hierarchical ordering gives rise to status ranking.

One of the most important sources of status ranking in U.S. society is that of occupation (Tumin 1967). It is because of the predominance of occupation as a social ranking dimension and women's traditionally unemployed condition that women were often said to occupy a "low-status" position in society. As women acquired jobs outside the home, they altered their status and provided a means for differentiating themselves. That is, by acquiring occupations other than that of housewife, women generated greater within-group status heterogeneity. Occupational status became an additional dimension that could be used by researchers and by women, themselves, for differentiation.

Two perspectives may be used for studying women along this "new" differentiating dimension of occupational status: one which is externally-assigned and one which is internal to the individual. An example of an externally-assigned status system is that of labeling certain jobs as professional-managerial, sales-clerical, skilled labor and so forth. Such a labeling system is used as a social convention to ascribe status to occupations. These externally-ascribed statuses subsequently are often normatively ranked as to social desirability; for example, professional-managerial occupations are generally viewed as higher in social value than those based upon those based upon skilled labor (Tumin 1967).

A second perspective is to ask women to ascribe a status to themselves based upon self-perceptions of their occupations. Here, instead of labeling herself as (say) a 'sales-clerical worker', the woman would define herself as a "Housewife," a "Working Women" or a "Career Women." For present purposes we will define "Working" Woman as a self-perceived occupational status to describe an occupation that serves as "just a job" and which was acquired primarily to generate income; whereas the "Career Woman" self-perception refers to an occupation that was acquired mainly as a means of self-fulfillment.

It is expected that these two status systems will be generally consistent in actual practice. That is, that women who hold professional-managerial occupations will view themselves as career women; women who hold clerical-sales occupations will view themselves as working women, and women who are voluntarily 'unemployed' will view themselves as housewives. This expectation is based on the premise that the status ranking system characterizing the society in which the woman lives will be internalized by her as a result of socialization processes. In essence, we come to view and value ourselves as society views and values the positions that we occupy.

Because a professional occupation is highly valued in U.S. society, a woman would be expected to view it as a means for achieving self-fulfillment and recognition. Because a clerical occupation is less-valued in our society, a woman would be expected to view it, not as self-fulfilling in and of itself, but rather as a means for obtaining other objectives - for example, the generation of additional income to purchase desired products.

To the extent that externally- and internally-ascribed status systems overlap, they may be used interchangeably in examining the consumption behavior of women. That is, we may classify women either by a set of externally-imposed categories or by the self-ascribed classification scheme that they apply to themselves. Thus the first hypothesis advanced in this research is that self-ascribed occupational status and externally-ascribed occupational status for women will be consistent.

H1: Specifically, it is hypothesized that women who hold a professional or managerial occupation will view it as a career; women who hold a clerical-sales occupation will view it as a job; and women who are voluntarily unemployed will view their occupation as that of housewife or homemaker.

It was further hypothesized that income and educational differences previously found for women based on their externally-ascribed status will be consistent with those found for their internally-ascribed status (Joyce and Gultinan 1978). That is:

H2: Career women, housewives and working women will be ordered as high, medium and low in family income, respectively.

This hypothesis is based on prior descriptive research and also on the following reasoning. First, one reason why women who voluntarily choose not to work (i.e. Homemakers) are likely to do so is because there is no immediate financial need which forces them to seek employment. Thus, they are likely to be from moderate to high income households. Second, "Working Women" have already defined themselves as working primarily "for the income;" therefore there is reason to suspect that there is inadequate household income unless they work. Third, the fact that "Career Women" are seeking self-fulfillment as a primary reason for working presupposes that there is at least adequate household income; otherwise they would be employed mainly out of necessity and not out of desire. Further, empirical reality suggests that "career" occupations are typically the most financially rewarding. Thus, the combination of a high household income (e.g. a well-employed spouse) together with her own above average earnings should place the Career Woman in the highest income rank.

H3: Career women, housewives and working women will be ordered as high, medium, and low in educational attainment, respectively.

In addition to the earlier findings that gave rise to this hypothesis (Joyce and Gultinan 1978), there are some logical inferences upon which it may be based, as well. First, educational attainment tends to be linked to income and occupational status. Therefore, it is straightforward to reason that women who work of financial necessity (i.e. Working Women) are less educated. Women who can be Homemakers are likely to be married to well-educated men who hold financially-rewarding jobs. Women who have professional occupations are likely to be highly educated. A high level of education is generally a prerequisite for holding most professional-managerial jobs and, thus, is likely a prerequisite to perceiving oneself as a Career Woman.

Retail Patronage

The second objective of the research was to extend the examination of self-ascribed status rankings to an inves-

tigation of interpersonal differences in retail patronage. The patronage area of specific concern was that of apparel purchasing. Since apparel is one of the primary means used by the individual to create a self-identity and to communicate that identity to others (e.g. Belk 1978, Holman 1980), it stands to reason that there should be apparel purchasing differences among women based on their self-ascribed status. The specific hypotheses advanced in this portion of the research were:

H4: Women will display a different pattern of apparel shopping at various store types depending upon their self-ascribed occupational status.

H5: Women will display a different pattern of apparel shopping during various times of the day and days of the week depending upon their self-ascribed occupational status.

H6: Women will display a different dimensional structure for store attributes relevant to apparel shopping depending upon their self-ascribed occupational status.

H7: Women will display a different ranking of importance for information sources regarding apparel shopping depending upon their self-ascribed occupational status.

More specific rationales for these hypothesized differences are given in subsequent discussion.

Methodology

Sample

Data for the research were generated by mailing self-administered questionnaires to stratified groups of women. The aim was to obtain proportionately equivalent representation of self-identified Housewives, Working Women and Career Women. To accomplish this, mailing lists were obtained for New York Area PTA groups, clerical and sales worker associations, and professional and business associations. Names of women were selected at random from these lists and 75 questionnaires were mailed to each type of source (e.g. PTA, clerical associations, business associations). After three weeks, 48 questionnaires had been returned on which the respondent identified herself as a Housewife, 42 questionnaires were returned on which the respondent identified herself as a Working Woman and 35 questionnaires on which the respondent identified herself as a Career Woman. Follow-up mailings (n = 50) were sent to women on the three mailing lists; and enough questionnaires were received from this second mailing to bring the total self-identified Housewives to 51, self-identified Working Women to 52 and self-identified Career Women to 50.

Because of the stratified nature of the sampling design, the approximately equal proportions of the three types of women used in the research do not, of course, mean that these groups occur with equal frequency in the general population. The purpose here was not to define parameters for the Housewife, Working Woman and Career Woman status segments, but rather to examine intergroup differences. Hence group proportional equivalence rather than population representativeness was the dominant criterion.

The women composing each group in the sample are believed to be representative of the group, within the parameters imposed by use of the mailing lists. These lists restricted the sample frame to women residing near New York City and to women belonging to certain organizations (e.g. a PTA); hence the findings from the study must be generalized from cautiously both on a demographic and

sociological basis.

Criterion

The primary criterion employed in the research was a multi-part question that requested the respondent to identify herself as belonging to one of three possible occupational groups. The classificatory statement read:

Do you see yourself primarily as a:

- (1) Homemaker, that is, your principal occupation is taking care of your home and family.
- (2) Working Woman, that is, your principal occupation is a job whose major purpose is to earn income.
- (3) Career Woman, that is, your principal occupation is a career whose major purpose is self-fulfillment.

Based on their responses to this question, women were classified as Homemakers (n = 51), Working Women (n = 52) or Career Women (n = 50).

Results

Hypothesis 1. The first hypothesis expressed the expectation that a high degree of congruence would be found between a woman's self-ascribed status and the status externally ascribed to her occupation. To test this, a contingency table was constructed comparing the three self-ascribed groupings (Homemaker, Working Woman, Career Woman) with three levels of externally-ascribed occupational status (professional-managerial, clerical-sales, housewife/unemployed). These latter status groupings were formed on the basis of actual occupation as reported by the individual.

As can be seen from the data displayed in Table 1, there is a very high degree of correspondence between self-ascribed status and actual occupation for the women in this research. Approximately three fourths of women who considered themselves "Career Women" held professional-managerial occupations; and roughly the same percentage of women who classified themselves as "Working Women" held clerical-sales positions. There was complete correspondence between women who viewed themselves as "Homemakers" and those who listed their occupation as housewife. This correspondence between occupational self-perceptions and actual occupations was highly significant statistically ($p \leq .00001$).

TABLE 1

SELF-ASCRIBED VERSUS EXTERNALLY-ASCRIBED STATUS COMPARISON

	Professional Managerial	Clerical Sales	Housewife
Homemaker	0 0.0	0 0.0	51 100.0
Working Woman	14 27.0	38 73.0	0 0.0
Career Woman	39 78.0	11 22.0	0 0.0

$\chi^2 = 191.6$

$p \leq .00001$

$n = 153$

Hypothesis 2. The second hypothesis stated the expectation that Career Women, Homemakers, and Working Women would be ordered high, medium, and low, respectively, with regard to household income. This expectation is partially confirmed by the data in Table 2. As can be seen, the Homemakers in the sample have the highest average household income, Career Women an intermediate level and Working Women the lowest. The one-way ANOVA of group mean differences indicated that both the Homemakers and Career Women had incomes significantly higher than Working Women ($p \leq .05$ and $\leq .10$, respectively) based on the Scheffe' test for simultaneous comparisons. There was not a significant difference between the income of Homemakers and Career Women, although the general direction was opposite that anticipated.

Hypothesis 3. The third hypothesis expressed the notion that Career Women would be highest, Homemakers intermediate, and Working Women lowest in educational attainment. This set of expectations was partially borne out by the results shown in Table 3. The one-way ANOVA of differences in population means indicates that Career Women are highest in educational attainment, Homemakers are intermediate, and Working Women are lowest, as was anticipated. The Scheffe' test for simultaneous comparisons was significant at the $p \leq .05$ level for the difference between Career Women and Working Women. However, there was not a statistically significant difference between the educational attainment of Career Women and Homemakers or Homemakers and Working Women as had been anticipated, although the general direction was as expected.

Since testing the first three hypotheses generally confirmed the overlap between self-ascribed and externally-assigned occupational status for women, analysis was extended to an examination of retail patronage. Specifically, the relationships between self-ascribed occupational status and several variables concerning apparel shopping were examined.

Hypothesis 4. This hypothesis expressed the belief that patronage of various types of retail stores for apparel purchases would be related to self-ascribed occupational status. Five types of stores were examined: discount stores, chain stores, department stores, specialty stores and boutiques. Respondents were asked:

"In what type of store do you usually shop for your own daywear clothing?"

The responses given by Career Women, Working Women and Homemakers were analyzed using contingency tables and are shown in Table 4. As can be seen, for only one type of store - the specialty store (e.g. Saks, Lord & Taylor) - were differences found at a statistically significant level among status groups. For specialty stores, Career Women were found to shop for daywear clothing at a proportionately greater level than either Homemakers or Working Women ($p \leq .03$). Further, it is important to note that a majority of all women respondents reported shopping most often at department stores for their daytime apparel. Women's apparel has traditionally been the forte' of department stores, and it appears that this relationship still holds among the present respondents.

Hypothesis 5. It was hypothesized that differences in day and time of shopping for apparel would be present among women having different self-ascribed occupational status. More specifically, it was believed that Career Women and Working Women would shop more on weekends and during the evenings; whereas Homemakers would utilize the "traditional" weekday shopping hours (i.e. mornings and afternoons). This was tested by asking respondents:

"When (days and hours) do you usually shop for your own clothing?"

Response categories were:

Days: Weekdays, Weekends, or Both

Hours: Mornings, Mid-day, Afternoon, Early Evening, Late Evening

TABLE 2

INCOME DIFFERENTIALS BETWEEN CAREER WOMEN,
WORKING WOMEN AND HOMEMAKERS

Career Women	Working Women	Homemakers
$\bar{x} = 22.7$	$\bar{x} = 20.0$	$\bar{x} = 24.4$
Scheffé test of simultaneous comparison		
$p \leq .10$ C/W, $p \leq .05$ H/W, $p \geq .10$ C/H		
n = 153		

TABLE 3

EDUCATIONAL ATTAINMENT DIFFERENTIALS BETWEEN
CAREER WOMEN, WORKING WOMEN AND HOMEMAKERS

Career Women	Working Women	Homemakers
$\bar{x} = 17$	$\bar{x} = 13$	$\bar{x} = 15$
Scheffé test of simultaneous comparison		
$p \leq .05$ C/W, $p \geq .10$ W/H, $p \geq .10$ C/H		
n = 153		

Responses to these questions were arranged in contingency tables and tested using the Chi-square statistic. Data are given in Table 5. As can be seen from the figures, there are some significant differences among alternative occupational status groups in shopping time. As was expected, Homemakers did most of their apparel shopping during the weekdays (84%), whereas the majority of both Career Women and Working Women did their apparel shopping on the Weekends or during a combination Weekends and Weekdays. This difference in shopping pattern was highly significant ($p \leq .0001$).

As anticipated Career Women and Working Women reported shopping in the early evening to a significantly greater extent than did Homemakers ($p \leq .0001$); whereas Homemakers were significantly more inclined to shop during the morning and afternoon hours ($p \leq .02, \leq .007$). Thus, occupational status appears to directly relate to a woman's retail patronage habits, at least as far as apparel shopping is concerned.

These findings suggest that retailers may be attracting different occupational segments of women depending upon the time of day and the day of the week. Thus, shopping time may be a central dimension for segmenting women customers. Salespersonnel, for example, could be alerted to the differential probabilities of encountering women of each occupational status group at various time periods, and sales appeals and merchandise displays could perhaps be adjusted accordingly.

TABLE 4

TYPE OF STORE USUALLY SHOPPED FOR DAYTIME APPAREL

	DISCOUNT STORE		CHAIN STORE		DEPARTMENT STORE		SPECIALTY STORE		BOUTIQUE	
	DON'T SHOP	DO SHOP	DON'T SHOP	DO SHOP	DON'T SHOP	DO SHOP	DON'T SHOP	DO SHOP	DON'T SHOP	DO SHOP
Homemaker	49	2	45	6	20	31	44	7	37	14
	96.0	4.0	88.0	12.0	39.0	61.0	86.0	14.0	72.5	27.5
Working Woman	48	4	40	2	25	27	49	3	32	20
	92.3	7.7	77.0	23.0	48.0	52.0	94.2	5.8	61.5	38.5
Career Woman	49	1	45	5	26	24	38	12	30	20
	98.0	2.0	90.0	10.0	52.0	48.0	76.0	24.0	60.0	40.0

n=153

$\chi^2 = .35$

$\chi^2 = 1.74$

$\chi^2 = 2.24$

$\chi^2 = 7.31$

$\chi^2 = 2.82$

p = .84

p = .42

p = .33

p = .03

p = .24

TABLE 5
SHOPPING TIME BY OCCUPATIONAL STATUS

	Weekday	Weekend	Both
Homemaker	43 84.0	5 10.0	3 6.0
Working Woman	24 46.0	22 42.0	6 12.0
Career Woman	18 36.0	28 56.0	4 8.0

n=153

$$\chi^2 = 29.34, p \leq .0001$$

	Mornings		Midday		Afternoon		Early Evening		Late Evening	
	no	yes	no	yes	no	yes	no	yes	no	yes
Homemaker	36	15	35	16	31	20	49	2	43	8
Working Woman	46	6	36	16	42	10	30	22	43	9
Career Woman	45	5	40	10	45	5	30	20	40	10
n=153	$\chi^2=12.17$		$\chi^2=4.0$		$\chi^2=10.02$		$\chi^2=19.7$		$\chi^2=1.16$	
	p<.02		p<.41		p<.007		p<.0001		p<.74	

Hypothesis 6. This put forward the expectation that different dimensional structures of retail-relevant attributes would characterize women having different self-ascribed occupational statuses. The specific question used to obtain data to test this hypothesis was as follows:

"How important are the following characteristics to you in deciding where to shop for your own daywear clothing?"

- (1) Broad selection of merchandise
- (2) Good service; helpful salespeople
- (3) Convenient locations and hours
- (4) Reasonable prices
- (5) Fashionable merchandise

Each of the attributes examined has been determined salient in prior research on fashion apparel (King and Ring 1980). Responses were recorded on a four-point scale ranging from 'Very Important' (4) to 'Not Important' (1). To discern if different dimensional structures existed for these attributes among the three women's occupational status groups, scores for each group were input to factor analysis (principal components with varimax rotation).¹ The results given in Table 6.

As can be seen from the data in Table 6, there are substantial differences among the three groups in terms of attribute factor structure. For Homemakers two factors were obtained; the first factor is correlated highly with selection and fashion, while the second factor is related to service, convenience and prices. In contrast, Working Women displayed a simpler, one-factor structure in their attribute perceptions. All attributes tended to be evaluated similarly by Working Women, resulting in one, overall factor upon which each attribute loaded highly.

¹ Only factors having eigenvalues greater than 1.0 are discussed.

Career Women exhibited the most complex factor structure; three factors were extracted to represent their attribute perceptions. The first factor correlated highly with convenience and prices; the second with service, and the third with fashion. Thus it appears that for the respondents in this research, there is a substantial difference in the dimensional structure of salient attributes characterizing the three occupational status groups. This finding suggests that images of varied complexity may be perceived for apparel retailers depending upon whether a woman views herself as a Homemaker, a Working Woman or a Career Woman.

Hypothesis 7. This hypothesis expressed the notion that different rankings of importance for various apparel-relevant information sources would result depending upon the self-ascribed occupational status of women. This hypothesis was tested by asking respondents the following question:

"To get ideas about what to wear during the day would you:"

- (1) Shop in stores
- (2) Read fashion magazines
- (3) Read newspaper advertisements
- (4) Watch television and movies
- (5) Wear what your friends are wearing
- (6) Wear what your work associates are wearing
- (7) Ask family members
- (8) Wear what your superiors at work are wearing

Responses were arrayed on a four-point rating scale ranging from 'Very Important' (4) to 'Not Important' (1). Based on mean values computed for each occupational status group, these scores were transformed to rankings, which are given in Table 7. The Kendall rank order correlation coefficient, tau, was used to compute the correspondence between information source rankings for pairs of occupational status groups.

TABLE 6
 FACTOR STRUCTURES FOR RETAIL-SALIENT
 ATTRIBUTES BY STATUS GROUPS

	<u>HOMEMAKERS</u>		<u>WORKING WOMEN</u>	<u>CAREER WOMEN</u>		
Selection	.82	.37	.80	.51	-.18	.46
Service	.44	.70	.75	-.21	.92	.07
Convenience	.24	.60	.79	.79	.43	.12
Prices	-.11	.87	.50	.84	-.21	.19
Fashion	.93	-.02	.79	.17	-.07	.95
Explained Variance	71%		54%	75.2%		

n=153

TABLE 7
 COMPARISONS OF RANK FOR INFORMATION
 SOURCE IMPORTANCE BETWEEN
 OCCUPATIONAL STATUS GROUPS

<u>Information Source</u>	<u>Homemakers</u>	<u>Working Women</u>	<u>Career Women</u>
Shop in Stores	1	1	1
Read Fashion Magazines	2	2	2
Read Newspaper Ads	3	3	3
Watch TV and Movies	4	6	7
Wear What Friends Wear	5	5	6
Wear What Coworkers Wear	7	4	5
Ask Family Members	6	4	6
Wear What Superiors Wear	7	6	4

Kendall's tau: Homemakers/Working Women = .59; $p \leq .01$

Homemakers/Career Women = .37; $p \leq .01$

Working Women/Career Women = .70; $p \leq .01$

n=153

The calculated values of Kendall's tau ranged from .37 to .70, all of which were significant at the $p < .01$ level. Thus, contrary to expectations, there is significant correspondence between the rank-order importance of various apparel information sources among occupational status groups. This finding is readily apparent from visual inspection of the ranked information sources provided in Table 7. The highest correlation was between Working Women and Career Women (.70); the next highest was between Working Women and Homemakers (.59); and the lowest was between Career Women and Homemakers (.37).

Rankings of the three most important apparel information sources - shopping in stores, reading fashion magazines and reading newspaper advertisements - is invariant across status groups. There were some inter-group differences, however. Wearing what coworkers or superiors wear are somewhat more important information sources for Career Women. Homemakers placed relatively more importance on watching television and movies as sources of apparel information, than did either Working Women or Career Women, perhaps because Homemakers have more time available for this type of information-seeking. However, despite these dissimilarities, the overriding pattern is one of congruence among the women's occupational status groups as regards apparel information source importance.

Further, it is of strategic importance for retailers to note that the dominant source of apparel information for all three groups of women was shopping in retail stores. This, of course, is the "source" which is most directly under the control of the retailer. The fact that all three types of women studied are turning to the store as their most important source regarding apparel information provides a substantial degree of strategic maneuverability to retailers who take advantage of this role.

Summary

This research explored the congruence between externally and internally ascribed occupational status for women, and the relationship that internally ascribed status has on various aspects of women's retail patronage. It was found that, for the respondents studied, there was very great consistency between internal and external occupational status ascriptions.

With regard to the relationship between internal status ascription and retail patronage, the following conclusions were reached: First, Career Women shop at specialty stores more than do Working Women or Homemakers. Second, Career Women and Working Women tended to shop more in the evenings and on weekends than did Homemakers. Third, Career Women, Working Women and Homemakers displayed substantially different factor structures for retail relevant attributes. Finally, rankings of importance for apparel information sources were found to be largely congruent among the three occupational status groups.

References

- Anderson, Beverly (1972), "Working vs. Non-Working Women: A Comparison of Shopping Behavior", Proceedings of the AMA Fall Conference, Houston.
- Bartos, Rena (1977), "The Moving Target: The Impact of Women's Employment on Consumer Behavior", Journal of Marketing, 41:3, 31-37.
- Belk, Russell W. (1978), "Assessing the Effects of Visible Consumption on Impression Formation", in H. K. Hunt (Ed.), Advances in Consumer Research, Vol. 5, Association for Consumer Research.
- Douglas, Susan P. (1976), "Cross-National Comparisons: A Case Study of Working and Non-Working Wives in the U.S. and France", Journal of Consumer Research, 3:1, 12-20.

Greenberg, N., Straight, B., Hassenger, W., and Raska, W., "Personality and Attitudinal Differences Between Employed and Unemployed Married Women", Journal of Social Psychology, 53, 87-96.

Harstrom, Jeanne L. and Dunsing, Marilyn M., (1978), "Socioeconomic and Social Psychological Influences on Reasons Wives Work", Journal of Consumer Research, 169-175.

Holman, Rebecca, (1980), "Clothing as Communication: An Empirical Investigation", in Jerry C. Olson (Ed.) Advances in Consumer Research, 7, Association for Consumer Research, 372-377.

Joyce, Mary and Gultinan, Joseph (1978), "The Professional Woman: A Potential Market Segment for Retailers", Journal of Retailing, 54:2, 59-70.

King, Charles W. and King, Lawrence J. (1980), "Market Positioning Across Retail Fashion Institutions", Journal of Retailing, 56:1, 37-55.

Lazer, William and Smallwood, John E. (1977) "The Changing Demographics of Women", Journal of Marketing, 41:3, 14-22.

McCall, Suzanne, (1977), "Meet the Workwife", Journal of Marketing, 41:3, 14-22.

Miller, Joanne, Schooler, Carme, Kohn, Melvin L. and Miller, Karen A. (1979), "Women and Work: The Psychological Effects of Occupational Conditions", American Journal of Sociology, 85:1, 66-94.

Nock, Steven L. and Rossi, Peter H. (1978), "Ascription versus Achievement in the Attribution of Family Social Status", American Journal of Sociology, 84:3, 565-590.

Strober, Myra H., and Weinberg, Charles B. (1977), "Working Wives and Major Family Expenditures", Journal of Consumer Research, 141-147.

Tumin, Melvin M. (1967), Social Stratification: The Forms and Functions of Inequality, Englewood Cliffs: Prentice-Hall, Inc.

CONSUMER SOCIALIZATION FACTORS IN A PATRONAGE MODEL OF CONSUMER BEHAVIOR

William R. Darden, University of Arkansas
Donna K. Darden, University of Arkansas
Roy Howell, University of Illinois
Shirley J. Miller, University of Arkansas

Abstract

The paper proposes that patronage behavior is more strongly influenced by consumer socialization than is brand choice behavior. The paper draws upon theory from sociology and business to make this case. In conclusion it strongly suggests that patronage theory has much to offer strategic planners and is therefore a research area with great potential.

Introduction

Thomas Kuhn, in The Structure of Scientific Revolutions (1970), described the pre-paradigmatic state of a science:

"In the absence of a paradigm or some candidate for paradigm, all of the facts that could possibly pertain to the development of a given science are likely to seem equally relevant. As a result, early fact-gathering is a far more nearly random activity than the one that subsequent scientific development makes familiar. Furthermore, in the absence of a reason for seeking some particular form of more recondite information, early fact-gathering is usually restricted to the wealth of data that lie ready to hand. The resulting pool of facts contains those accessible to casual observation and experiment together with some of the more esoteric data retrievable from established crafts like medicine...." (Kuhn 1970, p. 15).

It is rather obvious that Kuhn has described the history of marketing science, just as his description applies to physics, biology, etc. Early marketing science was conducted largely by practitioners of a pragmatic bent. Their variables were the ones which seemed most obvious to them, and each seemed as important as the others.

In 1963 Howard presented what was to become the Howard-Sheth Model, the most extensive and well-known system model of consumer behavior (1969). In 1966 Nicosia presented his flow diagram-simulation of consumer behavior and suggested several areas of behavioral research for marketers. The "school phase" of marketing history was established with the presentation of the "multimediation" model of Engel, Kollat and Blackwell, the point at which we now find ourselves (1973).

"Under these circumstances, the dialogue of the resulting books was often directed as much to the members of other schools as it was to nature. That pattern is not unfamiliar in a number of creative fields today, nor is it incompatible with significant discovery and invention" (Kuhn 1970, p. 13).

Kuhn continues by showing that science does not proceed by accretion, the pattern once ascribed to science by historians, but by the production of revolutions in normal science which result in paradigms. "...it remains an open question what parts of social science have yet acquired such paradigms at all."

In the absence of such a paradigm, marketing can hardly afford to ignore the schools which have developed in the closely related social sciences, since their volume of received knowledge is at least larger than our own.

Paper Objectives

Socialization is not, of course, a new variable in the study of consumer behavior. In terms of "received knowledge," socialization receives attention in almost every consumer behavior text either explicitly or implicitly. Engel, et al., (1973) includes such variables as beliefs, attitudes, motives, personality, life style, normative compliance, cultural norms, reference group and family influences, etc., in their model, all of which are the result, in the individual, of the socialization process (Berkman and Gilson 1978).

Most large system models, however, focus on brand choice behavior. For example, Engel et al., use the socialization process to explain brand choice behavior. Yet little has been written on the influence of the socialization process on patronage behavior. In addition, the consumer behavior literature rarely addresses the influence of "adult socialization" on the consumer. There is evidence that adult socialization has great impact on patronage choice behavior under a number of life cycle conditions. Furthermore, there is increasing evidence that patronage choice behavior is a major factor in explaining consumer behavior. Thus the socialization process as it applies to adults may be relatively more important to decision makers than is understanding the socialization process as it applies to adolescents (Ward 1974 and Moschis and Churchill 1978).

The objective of this paper is to present some arguments for research in the areas of adult socialization and patronage choice behavior. In particular, the paper proposes that adult socialization impinges on shopping orientations and store choice criteria, and that this relationship changes as the consumer proceeds through the stages of the family life cycle.

Figure 1 depicts the Patronage Model of Consumer Behavior, which is to be used in this paper to help illustrate the relationships between adult socialization and patronage behavior (Darden 1979).

Theoretical Background

There are numerous approaches to the study of socialization. The major concepts of the social learning theorists include imitation, shaping, reward and punishment. The developmentalists include the Freudians who are concerned with social repression of psychosexual instincts and the Piaget school with the interactive maturational unfolding of innate mechanisms. The interactionists and phenomenologists are more interested in the interaction between the situation and the role or knowledge which the individual brings to the situation in order to act, rather than react within it.

All schools of thinking, however, are agreed that, no matter what the process of socialization involves, it is an ongoing process. While it is more dramatic to watch a human infant learning its culture, the adult, too, is continually interacting with its culture in a fluid process.

Perhaps one of the first to take note of the importance of adult socialization was Erik Erikson (1950), with his "Seven Stages of Man," in which he called attention to the

"It is doubtful that one comes on a role in later life without any fragments at all of relevant socialization; the inexperienced mother may seem to know little, but she knows something, and, even more, she has response elements for the role performance that are not manifest at the conscious level."

Both socialization and resocialization are important in most adults' lives. As Babbie (1980) points out:

"Every time you interact with someone, you are being further socialized by that person, and you are socializing him or her. When you buy something in a store, you learn a little more about being a customer, and the salesperson learns a little more about being a salesperson. You may not have noticed it, but your understanding of the role of customer probably changed a little in the process."

But Brim (1966) points out there are some important significant others which affect the adult at various points in the life cycle. Brim discusses four major sources of adult resocialization (our term, not his): (1) Mobility, geographical and social; (2) Variance in customs in society (social change); (3) Changes in the content of socialization, arising from passage into different stages of the life cycle and different institutions; and (4) Where childhood socialization was incomplete.

Geographical mobility is a highly relevant factor in American society today, although there are indications that people are moving less than they did a few years ago. With the great amount of homogenization which has occurred as a result of national chains and mass media (Where can you go in America and not be understood when you ask for a "Big Mac"?), this factor may become even less important, within the society. However, it is still important, as a look at such publications as the Sterns' Roadfood (1978) or Rice and Wolf's Where to Eat in America (1979) reveals.

Social mobility has perhaps been overstressed in American society, since statistics indicate that much less vertical mobility actually occurs than the American Dream would have us believe (Robertson 1977). However, as there are strong indications that we are becoming a nation of white collar workers (Robertson 1977), and as the number of people with some college education keeps rising (Robertson 1977), social mobility is obviously an important source of adult resocialization. Much resocialization demanded by upward mobility is perhaps accomplished through anticipatory socialization (Merton 1968), and some through the mass media (*Playboy*, for example, and *Mad Magazine*; see Winick 1962; and Darden 1976).

Variance in the customs in society, or social change, affects people much more evenly, since most members of a society are subject to these pressures at relatively the same time periods. Babbie (1980:147ff) gives as examples of such factors war and social movements. These kinds of changes are perhaps more dramatic and more obvious to the participants, although some are much more obviously accepted than others.

Changes in the content of socialization which arise from different stages in the life cycle affect every normal adult in American society, although not continuously. It is expected, for example, that every adult get married, entailing a transformation from one status to another, but this transformation is expected to be performed within a definite period of time (Ritzer 1977). Moreover, it is usually expected that one's peers are undergoing such transformations themselves at relatively similar times, so that they help and give support during the resocialization process.

Brim's fourth category of resocialization factors, those which occur where childhood socialization has been incomplete, are not normally expected of a member of society, and in most cases result in institutionalization.

Normal adult socialization, as Babbie (1980) points out, is a feature of every interaction with everyone. He points out further that the main sources of such socialization are such agents as formal schools (colleges, vocational and professional schools), on the job training, and the military. One's peers are, as in childhood, the most important agents of socialization for adults, as they are for adult resocialization.

Socialization: Patronage Theory vs. Brand Choice

Socialization involves learning the norms of one's society. All human behavior is said to be normative, socialized behavior. Socialized behavior must be meaningful to the individual, and it must be learned in interaction with the individual's environment. The environment includes other actors as well as props, or material aspects of the culture. Most learning is at least as abstract as the concept "many," according to linguists.

Thus, as the behavior becomes more meaningful, the symbolic and cultural aspects become more important in its determination. Marketers have demonstrated, for example, the greater importance of group influence on more visible product choices than on those less visible (Bourne 1956) (people see your choice in cigarette brands and in part define you by your use of filters or nonfilters, etc., whereas few people see your choice of underwear and shape your identity accordingly). It is our contention that store choice, i.e., patronage behavior, is more highly visible behavior than is brand choice, for many kinds of purchases. One does not simply choose the closest retail outlet, but rather chooses that outlet which fits his image of himself, a store where he is apt to meet other people who are "like him," who are apt to be his friends (Tauber 1972), and where he can find the kinds of products which people "like him" find acceptable. Massey (1973), for example, describes the behavior of many patrons of pornography shops as being furtive and secretive, because they do not want people "like them" to see them there, wishing probably to disavow this aspect of themselves as parts of their announced identities.

As related to Brim's three important sources of adult resocialization, patronage behavior would seem to be very important.

In terms of geographical behavior, it has been found (Andreasen 1968) that people who move to another area establish their patronage habits rapidly.

Social mobility obviously involves a great deal of change in patronage behavior, as people "on the move" try to determine where people "like them" or "like they want to be" do their shopping. Much of this behavior may be exploratory, at least initially.

Variance of customs in society as a result of social change need not have any marketing implications, but most likely does. Certainly, a war can cause disruption in patronage behavior, caused by the absence of products from the marketplace, etc. Social movements probably affect most adults' patronage habits very little, although those convention cities in states which have not adopted the E.R.A. may find a difference of some degree.

Probably the largest amount of patronage behavior to be explained by socialization and resocialization occurs as a result of status changes caused by entrance into different stages of the life cycle, and by different institutions.

Socialization and Patronage Theory

Although consumer research to date has virtually equated socialization with young people (as young people are the most common referent for socialization research), it is clear that consumer socialization can and does continue

throughout the entire adult life cycle (Brim 1966; Rose 1979; Riley, et al., 1969). We will be concerned here with both adolescent and adult consumer socialization, as it is perceived that socialization to the shopping task and the retail environment is a recurring phenomenon throughout the life cycle. As discussed by Sigel and Hoskin (1977), the relative importance of early socialization as compared to adult socialization is unclear. The persistence-beyond-childhood model of socialization would accord primary importance to the socialization acquired in childhood, while at the other extreme the constant change model, "...while not excluding childhood socialization, argues that adulthood brings the organism into contact with new experiences (new settings, novel events, new responsibilities, changes in biological and social status) which have a powerful socializing impact on the individual" (Sigel and Hoskin 1977, p. 262). The true position is assumed to be somewhere between these two extreme theories.

Values

Human values, operationalized in the patronage model through the Rokeach (1973) paradigm of terminal and instrumental values (reflecting the individual's perception of some end-states of existence and modes of conduct as personally and socially preferable to alternative end-states of existence and modes of conduct) are given a prominent role in the patronage model, reflecting their centrality as constructs in consumer behavior. Values, in the Rokeach (1973) sense, are highly central and quite general, conceptualized as forming the basis for judgments in more narrowly defined situations or with specific referents. Since, in the broad sense, socialization is defined as the transmittal of culture, and since the basis of culture is the norms and bounds of acceptable variations in value systems (Clawson and Vinson 1977), it follows that the development and maintenance of a value system is, for the individual, at least in part a socialization process (Simpson 1977). This socialization may be imparted through agents such as the family, school system, peer group, or mass media in early life, while adult value socialization may be the result of the work group (occupation), geographic or social mobility, or the mass media, in addition to other factors. Rokeach (1973) has found that values differ by age group, social class, gender, and occupation. Darden, et al., (1979) have similarly shown that values are related to shopping orientations, as are the above four variables. While direction of causation is difficult to ascertain for the relationships between social class and values and the relationship between occupation and values (as will be discussed below), it can be seen that values, and thus the socialization process, play an important role in determining shopping orientations, and thus patronage behavior, both in their direct effects and as intervening variables.

Sex Role

It is generally accepted that the acquisition of a sex-role orientation is a socialization process. While the socialization agent is typically the same-sex parent, both one's peers and the mass media play a part in the sex role socialization of the individual through the identificatory process. It is felt that, in this case, one is socialized to a sex-role early in life, although it is possible for an adult to be socialized to some degree by agents reflecting the current cultural sex-role norms (Bandura 1969). While we will strictly avoid the estrogen/androgen vs. socialization argument here, suffice it to say that at least some of the sex-role variation found between males and females is due to the internalization of social norms of behavior.

The sex-role orientation of the individual with respect to shopping is a strict and important subset of that individual's general sex role orientation; the implications of

the sex-role socialization process, given the traditional sex roles in U.S. culture, on patronage behavior are numerous. To a large extent, shopping is considered the woman's domain: men are neither exposed to a shopping role model nor encouraged by their peers to be a "good" shopper. Thus, shopping tends to be a less well internalized role for the male, resulting in major differences in patronage behavior from females, who have traditionally had a strong role model in the shopping context, and thus strong anticipatory socialization to the role.

In addition to placing value on the female's family role of being a "good" shopper, the traditional female sex-role stereotype tends to sanction shopping as a legitimate social/recreational activity in which wives/mothers can engage with their peers while at the same time being "productive." Similarly, the traditional wife/mother role for which females are prepared through anticipatory socialization embodies fairly narrow life space. Shopping may thus be viewed as a sanctioned vehicle for otherwise deprived social interaction. While both of these conditions suggest different shopping orientations for males and females socialized to traditional sex-roles, and thus different modes of patronage behavior, they also suggest that the female will be more susceptible to adult socialization in the shopping task through peer influence. While this discussion is not exhaustive of the potential impacts of sex-role socialization on patronage behavior, it does indicate the importance of socialization processes in this area.

Occupation

One key aspect of social learning is "anticipatory socialization" of children and adolescents for work roles in later life (Turner 1971). Similarly, adults tend to be socialized to the work environment by their peers and the organization of the work situation with this close interaction extending the realm of socialization in some cases to roles and orientations beyond the workplace (Lorence and Mortimer 1979).

As Turner (1971) notes, the father's occupation plays a large role in determining the roles to which the adolescent will be exposed, and the nature of the "anticipatory socialization" which will occur, noting the particular impact of having a father who is in an entrepreneurial occupation.

Of particular interest here is the impact of having a parent employed in retailing. Darden, Howell and Powell (1980) have documented the extent of retail work experience in the U.S., indicating that this is not an extremely limited phenomenon, with over 35% of each of three separate random samples reporting having worked in retailing. The effect of having a parent working in retailing on the socialization process would seem to be significant, as would the effect of this unique socialization experience on subsequent patronage behavior.

Even more significant is the socialization experience acquired from working in retailing. Darden, Howell and Powell (1980) produce evidence that retail work experience, viewed as a socialization factor, is related to shopping orientations, attitudes toward retailers and retailing, and retail patronage behavior.

Mobility

Geographic mobility is one of the primary forces bringing about adult consumer socialization. A move to a new community requires a high degree of resocialization to different customs, subcultures, shopping norms and retail structure. It means, perhaps, a new job with new peers and a new organizational structure. It may mean a new set of roles for which the individual may or may not be prepared. It means the search for new role models. The farther the move (especially culturally), the greater the degree of

socialization which must take place (assuming that the individual does not engage in "expedient" socialization, wherein one learns new specific norms and behaviors but fails to internalize them, maintaining primary orientation to former norms) (Bandura 1969).

As noted by Darden, French and Howell (1978), mobile consumers have different general and shopping specific life stages, perhaps as an adaptation to the recurring resocialization process.

Age

When age is employed as a variable in consumer research in general or patronage research in particular, it is recognized that it is in fact a proxy for a number of factors, some of which are reflective of social learning. The multiple influences subsumed under age may affect patronage behavior in different ways; it is thus useful to identify them such that they can, perhaps, be measured explicitly.

In one sense, age serves as a general indicator of accumulated life experience, reflecting the extent, if not the specific nature, of the socialization process.

Age is also a component of the family life cycle, carrying with it unequal probabilities of being married, having children in the home, and the other factors which comprise the FLC, as well as altering the probability of the geographic mobility of the individual.

Age also brings with it differing levels of physical prowess and agility, thus affecting the physical cost of shopping, as well as altering the information processing style of the individual.

Finally, age reflects a cohort with similar socialization experiences based on macro-environmental events and evolution. The age cohort is also likely to reflect similar patterns of socialization based on the role-models, educational philosophy, and media available at various stages of their socialization process. Thus, when one finds that a patronage variable increases or decreases with age, one cannot partial out the cohort effects from the other effects, making inference about what the "younger" people will do when as they grow older impossible from cross-sectional data (Outler 1977), unless socialization life cycle, individual, and cohort effects are measured separately.

Conclusions and Implications

The purpose of this paper has been to use the Patronage Model of Consumer Behavior to demonstrate the impact of consumer socialization on patronage behavior. Some conclusions are listed below:

1. Consumer socialization affects patronage behavior differentially at each stage in the family life cycle.
2. While age is important in the family life cycle, it has additional implications for consumer socialization and its relation to patronage behavior.
3. Modeling, reinforcement and social interaction have significant impact on the Patronage Model of Consumer Behavior shown in Figure 1. For example, Darden, et al., (1980) suggest that retail work experience provides consumer socialization of a technical nature that significantly effects select shopping orientations. For example, those with high levels of retail work experience are likely to be quality shoppers.

4. Darden, et al., (1978) show that instrumental and terminal values are related to shopping orientations. It seems likely that consumer socialization provides the raison d'être for much of this change in values. Thus if consumer socialization changes values, it may indirectly--as well as directly--affect patronage behavior.

The introduction to this paper presented Kuhn's views on the genesis of research direction in a discipline. He stresses that scientific disciplines develop using the data at hand, rather than addressing necessarily the most critical issues. From our discussion it seems likely that patronage behavior is a critical area in marketing that deserves more attention than it has received in the past. Darden (1979) presents several reasons for the neglect of patronage research, all of them supporting Kuhn's theory.

As suggested by Kuhn a discipline can benefit from placing research emphasis in critical areas. The current emphasis on "strategic planning" in marketing could benefit from a comprehensive knowledge of patronage behavior. Strategic planning requires an understanding of the process by which consumers choose stores. Manufacturers and large chain retailers can no longer afford the inefficiency of concentrating solely on brand choice behavior in planning their distribution and promotional strategies.

References

- Andreasen, Alan R., and Durkson, Peter G. (1968), "Market Learning of New Arrivals," Journal of Marketing Research, 5, 166-176.
- Atkin, Charles K. (1976), "Children's Social Learning from Television Advertising: Research Evidence on Observational Modeling of Product Consumption," in Advances in Consumer Research, Vol. 3, ed. Beverlee B. Anderson, Cincinnati, Ohio: Association for Consumer Research, 513-519.
- Babbie, Earl R. (1980), Sociology, An Introduction. Belmont, California: Wadsworth Publishing Company, Inc.
- Bandura, Albert (1969), "Social Learning Theory of Identificatory Processes," in Handbook of Socialization Theory and Research, ed. D. A. Goslin., Chicago: Rand McNally.
- Becker, Howard S. (1964), "Personal Change in Adult Life," Sociometry, 27, 40-53.
- Becker, Howard S., and Geer, Blanche (1960), "Latent Culture: A Note on the Theory of Latent Social Roles," Administrative Science Quarterly, 5, 304-313.
- Bettleheim, Bruno (1943), "Individual and Mass Behavior in Extreme Situations," Journal of Abnormal and Social Psychology, 38, 417-452.
- Bourne, Francis S. (1956), "Group Influence in Marketing," Group Influence in Marketing and Public Relations. Foundation for Research on Human Behavior.
- Brim, Orville G., Jr. (1966), "Socialization Through the Life Cycle," in Socialization after Childhood, eds. O. Brim and S. Wheeler. New York: John Wiley & Sons, Inc.
- Calder, Bobby J., Robertson, Thomas, and Rossiter, John (1976), "Cognitive Response to Advertising: The Relation of Child to Adult Models," in Advances in Consumer Research, Vol. 3, ed. Beverlee B. Anderson, Cincinnati, Ohio: Association for Consumer Research, 536-537.
- Chaffee, Steven H. (1977), "Mass Communication in Political Socialization," in Handbook of Political Socialization, ed. S. A. Renshon. New York: Free Press, 223-258.

- Churchill, Gilbert A., Jr., and Moschis, George P. (1979), "Television and Interpersonal Influences on Adolescent Consumer Learning," Journal of Consumer Research, 6, 1, 23-35.
- Clauson, C. Joseph, and Vinson, Donald E. (1977), "Human Values: A Historical and Interdisciplinary Perspective," in Advances in Consumer Research, ed. Keith Hunt.
- Coolley, Charles Horton (1902), Human Nature and the Social Order. New York: Schocken Books.
- Cutler, Neal E. (1977), "Political Socialization Research as Generational Analysis," in Handbook of Political Socialization, ed. S. A. Renshon. New York: Free Press, 294-388.
- Cumming, Elaine, and Henry, William E. (1961), Growing Old: The Process of Disengagement. New York: Basic Books.
- Darden, Donna K. (1976), "The Media as Reference Group Relations," paper presented to the Southern Sociological Society.
- Darden, William R. (1979), "A Patronage Model of Consumer Behavior," in Competitive Structure in Retail Markets: The Department Store Perspective, eds. Ron A. Stampfl and Elizabeth Hirschman, Chicago, Illinois: American Marketing Association.
- Darden, W. R., Darden, D. K., Erdem, O., and Howell, R. (1978), "Values and Shopping Orientations," Proceedings Southwestern Marketing Association, San Antonio, Texas.
- _____, French, W. A., and Howell, R. (1978), "Mapping Market Mobility: Psychographic Profile and Media Exposure," Journal of Business Research, 7(1979), 51-74.
- _____, Howell, R., and Powell, T. E. (1980), "The Effect of Retail Work Experience on Shopping Orientations," working paper, Center for Marketing Research and Modeling, University of Arkansas, Fayetteville, Arkansas.
- Engel, J. F., Blackwell, R. D., and Kollat, D. T. (1978), Consumer Behavior, 3rd. ed. Hinsdale, Ill.: Dryden Press.
- Erikson, E. H. (1950), Childhood and Society. N.Y.: Norton.
- Goffman, E. (1961), Asylums. Chicago: Aldine.
- Gross, E. and Stone, G. P. (1964), "Embarrassment and the Analysis of Role Requirements," American Journal of Sociology, 70, 1-15.
- Howard, J. A. (1963), Marketing Management: Analysis and Planning, (rev. ed.). Homewood, Ill.: Irwin.
- _____, and Sheth, J. N. (1969), The Theory of Buyer Behavior. N. Y.: John Wiley and Sons.
- Kubler-Ross, E. (1975), Death: The Final Stage of Growth. Englewood Cliffs, N. J.: Prentice-Hall, Inc.
- Kuhn, T. S. (1970), The Structure of Scientific Revolutions. Chicago: University of Chicago Press.
- Lindesmith, A. R. (1947), Opiate Addiction. Bloomington, Ind.: Principia Press.
- _____, Strauss, A. L., and Denzin, N. K. (1977), Social Psychology, 5th. ed. N. Y.: Holt, Rinehart and Winston.
- Lorence, J. and Mortimer, J. (1979), "Work Experience and Political Orientation: A Panel Study," Social Forces, 58, 2, 651-675.
- Massey, M. (1973), "The Distribution of Pornography in the United States." Paper presented to Midwestern Sociological Society.
- Merton, R. K. (1968), Social Theory and Social Structure, 2nd. ed. N. Y.: Free Press.
- Miller, G. (1978), Odd Jobs: The World of Deviant Work. Englewood Cliffs: Prentice-Hall, Inc.
- Moore, R. L., and Stephens, L. F. (1975), "Some Communication and Demographic Determinants of Adolescent Consumer Learning," Journal of Consumer Research, 2, 80-92.
- Moore, W. E. (1969), "Occupational Socialization," in Handbook of Socialization Theory and Research, ed. D. A. Goslin. Chicago: Rand McNally, 951-979.
- Moschis, G. P., and Churchill, G. A., Jr. (1979), "Consumer Socialization: A Theoretical and Empirical Analysis," Journal of Marketing Research, 15, 599-609.
- _____, and Moore, R. L. (1979), "Decision Making Among the Young: A Socialization Perspective," The Journal of Consumer Research, 6, 101-112.
- Murphy, P. E., and Staples, W. A. (1979), "A Modernized Family Life Cycle," Journal of Consumer Research, 6, 1, 12-22.
- Nicosia, F. M. (1966), Consumer Decisions. Englewood Cliffs, N. J.: Prentice-Hall, Inc.
- Parsons, T. (1949), The Social System. N. Y.: Free Press.
- Rice, W. and Wolf, B. (1979), Where to Eat in America, 2nd. ed. N. Y.: Random House.
- Riley, M. W., Foner, A., Hess, B., and Toby, M. (1969), "Socialization for the Middle and Later Years," in Handbook of Socialization Theory and Research, ed. D. A. Goslin. Chicago: Rand McNally, 951-979.
- Ritzer, G. (1977), Working, 2nd. ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Roedder, D. L., Didow, N. M., and Calder, B. J. (1975), "A Review of Formal Theories of Consumer Socialization," in Advances in Consumer Research, Vol. 5, ed. H. Keith Hunt. Chicago: Association for Consumer Research, 528-534.
- Rokeach, M. (1973), The Nature of Human Values. N.Y.: Free Press.
- Rose, P. I., ed. (1979), Socialization and the Life Cycle. N. Y.: St. Martin Press.
- Salamon, G. (1974), Community and Occupation. N. Y.: Cambridge University Press.
- Siegel, R. A. and Hoskins, M. B. (1977), "Perspectives on Adult Political Socialization--Areas of Research," in Handbook of Political Socialization, ed. S. A. Renshon. N. Y.: Free Press, 259-293.
- Simpson, E. L. (1977), "Preference and Politics: Values in Political Psychology and Political Learning," in Handbook of Political Socialization, ed. S. A. Renshon. N. Y.: Free Press, 362-388.
- Stern, J. and Stern, M. (1978), Roadfood. N. Y.: Random House.
- Sudnow, D. (1967), Passing On: The Social Organization of Dying. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Tauber, E. M. (1972), "Why do People Shop?" Journal of Marketing, 36, 46-59.

Turner, J. H. (1971), "Entrepreneurial Environments and the Emergence of Achievement Motivation in Adolescent Males," in Socialization, E. Z. Dager, ed. Chicago: Markham Publishing Company, 110-131.

Wheeler, Stanton (1966), "The Structure of Formally Organized Social Settings," in Socialization After Childhood, eds. O. Brim and S. Wheeler. N. Y.: John Wiley and Sons, Inc.

Winick, C. (1962), "Teenagers, Satire and Mad," Merrill-Palmer Quarterly of Behavior and Development, 8, 183-203.

AN EXPLORATORY INVESTIGATION OF MEDIATING FACTORS IN RETAIL STORE IMAGE RESPONSES

Robert A. Peterson, University of Texas at Austin

Abstract

This paper proposes a general paradigm for investigating retail store images which takes into account the context of the research. An example illustrating several of the concepts set forth is provided.

Introduction

A considerable amount of research in the area of retail patronage has been concerned with the concept of store image. For example, research has been conducted regarding the role of store image in determining store loyalty and choice (e.g., Anderson and Scott 1970; Lessig 1973; Schiffman, Dash and Dillon 1977). Likewise, research has been conducted to investigate possible relationships between consumer self-images and retail store images (e.g., Bellenger, Steinberg and Stanton 1976; Dornoff and Tatham 1972), and differences between consumers' images and retailers' images (e.g., Pathak, Crissy and Sweitzer 1974-75). Indeed, an entire issue of the Journal of Retailing was devoted to store image research.¹ Finally, the importance of store image research as a managerial strategy was recently documented by King, Tigert and Ring (1979).

Moreover, because of its importance, a substantial amount of research has been dedicated to improving the measurement of retail store image. Empirical studies focusing on both validity and reliability issues have been reported in the literature (e.g., Hawkins, Best and Albaum 1975-76; McDougall and Fry 1974-75; Menezes and Elbert 1979), and several contributions made. In addition, progress has recently been made in determining the manner in which store images are formed (Reich, Ferguson and Weinberger 1977).

Purpose

The purpose of this paper is not to review and/or summarize the (ponderous) research on retail store image. Nor is it to offer new findings on store image determinants or methodological approaches to measuring store image. Rather, the purpose is more modest. It is to propose and illustrate a paradigm for more fully evaluating and interpreting retail store image data. While this paradigm can be applied in many research situations, it is particularly appropriate in store image research since there is a paucity of both physical and conceptual criteria against which to evaluate research conclusions.

Because of the desire to stimulate discussion as well as thought, no attempt is made to present an exhaustive documentation of appropriate literature. That task will be left to a later, more systematic exposition of the insights offered here. It is assumed that researchers interested in store images will already be thoroughly familiar with the appropriate literature.

A Paradigm

Investigating consumer images of retail stores is a relatively straightforward and standardized research activity. In the typical store image study a sample of consumers is selected, questionnaires administered, resulting data analyzed, and inferences drawn about image similarities and

differences among the stores investigated. Few researchers, however, are naive enough to believe that image differences observed to exist among stores---as reflected by one or more scale or question responses---can be interpreted at face value. Instead, differences must be considered within the context in which the data were collected.

In particular, image differences observed between retail stores may not arise solely as a result of perceived differences between the stores (the stimulus objects). Although the majority of the variance in consumer image responses may be due to differences between stores, other, frequently confounding, factors may account for some of the variance as well. Conceptually, responses to an image question can be thought of as reflecting a plethora of factors and influences, only one of which is stimulus object-related. Symbolically

$$R = f(SO, SC, MI, M, EN, E) \quad (1)$$

where

R represents an image response (in general or in particular);

SO represents stimulus object characteristics;

SC represents subject (consumer) characteristics;

MI represents measurement instrument characteristics;

M represents mode of data collection;

EN represents the data collection environment; and

E represents extraneous or error factors.

Subject characteristics consist of a wide variety of factors and influences. Many of them are relatively permanent and general in nature, such as demographic or personality characteristics. Others, like mood, are more transitory, while still others are specific to the research situation (e.g., response syndromes like haloing or yeasaying).

Three categories of factors are methodological in nature. Measurement instrument characteristics include type of question (open or closed end), nature of questionnaire, and so forth. Data collection mode refers to whether data are collected by means of telephone, mail or personal interviews, and whether questionnaires are self- or other-administered. Environmental factors include physical, social, and temporal (time of day, etc.) influences. The list of factors is nearly endless.

Factors within each category as well as the entire categories themselves interact and collectively result in an image response. Many of the nonstimulus object factors can be considered mediating factors in that they intervene between a stimulus and a response. Such factors can confound, bias, attenuate, or otherwise adversely influence an image response. In the extreme case an image response may actually be more an artifact of the mediating factors than be reflective of stimulus object characteristics.

In general, the above paradigm can be viewed as merely an application of concepts derived from SOR learning theory and environmental and gestalt psychology. Thus, underlying assumptions and axioms are well described elsewhere and need not be reiterated here. The remainder of this paper consists of an example illustrating, in a nonabstract manner, some of the notions expressed in the paradigm. Although the research is exploratory, it is hoped that insights will be obtained into, and an appreciation gained of, factors in store image research.

1

Journal of Retailing, 50, 4 (Winter 1974-75).

Illustration

Method

To collect requisite study data a 2X2X2 experimental design was employed. Two classes of undergraduate marketing students, 347 individuals in total, served as subjects.

Subjects were requested to complete a three-page questionnaire consisting of retail store image questions, shopping behavior questions, and several self-perception and demographic questions. The dependent variables investigated were 10 common retail store image dimensions. These dimensions were presented in the form of 7-point rating scales and addressed such characteristics as prices, merchandise quality, sales personnel, and physical layout. The specific scale content is irrelevant for the present discussion. However, approximately half of the subjects were administered the scales in a bipolar (semantic differential) format while the other half were administered the scales in a unipolar (stapel scale) format. This treatment, labeled V1, was designed to serve as a measurement instrument (MI) characteristic. In addition, half of the subjects evaluated Foley's Department Store, a large component of the Federated group, while the other half evaluated Penney's. This treatment was labeled V2; it served as a measure of stimulus object characteristics.

The third treatment was environment. About two-thirds of the subjects completed their questionnaires in a classroom with controlled conditions at one point in time. The remaining third were allowed to complete their questionnaires at their leisure, outside of a classroom environment, and return them at their convenience.² This treatment was labeled V3.

Four other variables in the subject characteristic category were also investigated. These variables were, respectively,
 ---sex (V4);
 ---shopping frequency at ___ department store (V5);
 ---general interest in department stores (V6); and
 ---mood at time of questionnaire completion (V7).
 Shopping frequency and interest were measured using 4-point rating scales. Mood was measured by means of a 6-item Likert-type summated rating scale. Scale scores ranged from 5 (good mood) to 30 (bad mood), with a mean of 16.2, and standard deviation of 3.5. Because of its temporary nature mood was deemed a "nuisance" factor.

Analysis

Initially relationships between the set of 10 dependent variables and the set of 7 independent variables were investigated through canonical analysis. This resulted in 3 canonical roots (functions) being statistically significant at the .07 level or greater. Canonical loadings for the 7 independent variables are given in Table 1. The first root was by far the most significant, with $p < .0001$; it accounted for 71 percent of the explainable variance. Together, the next two roots only accounted for 19 percent of the explainable variance.

Function I is clearly dominated by V2; 76 percent of its communality is accounted for by stimulus object characteristics. Function II is essentially an interest-frequency function since these variables load most heavily. The third function can be interpreted as reflecting the influence of mediating factors since all independent variables but stimulus object characteristics possessed

2

The low return rate of the latter group accounted for the size differences in the treatment levels.

significant canonical loadings.³ With regard to the variance explained by each independent variable, stimulus object characteristics accounted for the largest percentage of explained variance in the dependent variable set while mood accounted for the smallest percentage of explained variance.

TABLE 1
Independent Variable Canonical Loadings

Variable	Canonical Function		
	I	II	III
V1 (polarity)	.04	.13	.32
V2 (store)	.93	-.28	.08
V3 (environment)	-.02	.06	-.55
V4 (sex)	.29	.11	-.62
V5 (frequency)	.26	.52	.55
V6 (interest)	.34	.74	.39
V7 (mood)	-.02	.16	.40

To obtain greater insights into specific variable interactions an analysis of variance was conducted for each of the 10 dependent variables. However, to make each analysis more manageable, only four independent variables were retained for investigation:

---V2 (store)
 ---V4 (sex)
 ---V5 (frequency)
 ---V6 (interest).

Using a significance criterion of .05, only 27 of 150 possible effects were statistically significant. Of these 27 effects, 22 were main effects, broken down as follows:

Variable	Number of Main Effect Significant
V2 (store)	7
V4 (sex)	8
V5 (frequency)	6
V6 (interest)	1

There was no consistent pattern to the 5 interaction effects that were statistically significant. Moreover, the substantive significance of any effect was marginal at best.

Conclusion

This paper has suggested a paradigm for empirically investigating retail store images. By means of this paradigm a researcher can evaluate store image differences in a holistic sense, simultaneously taking into account perceived differences between stimulus objects and mediating (intervening) factors. This would allow, in turn, quantitative evaluation of the various factors influencing image perceptions apart from store perceptions per se.

While the results of the example are not extremely significant, they are sufficiently strong to both illustrate the concepts espoused and warrant further research. Other possible influencing factors need to be empirically explored and categorized as to their importance. Ultimately, a typology of influencing factors needs to be constructed so that store image data can be better evaluated with regard to the manner in which they are collected. The benefits of so doing are both widespread and obvious.

3

Terming a factor mediating or determinant is essentially a function of analytical purpose.

References

Andersen, Clifton R. and Scott, Richard A. (1970), "Supermarkets: Are They Really Alike?" Journal of Retailing, 46, 16-24.

Bellenger, Danny N., Steinberg, Earle and Stanton, Wilbur W. (1976), "The Congruence of Store Image and Self-Image," Journal of Retailing, 52, 17-32.

Dornoff, Ronald J. and Tatham, Ronald L. (1972), "Congruence Between Personal Image and Store Image," Journal of the Market Research Society, 14, 45-52.

Hawkins, Del I., Best, Roger and Albaum, Gerald (1975-76), "Reliability of Retail Store Images as Measured by the Stapel Scale," Journal of Retailing, 52, 31-38.

King, Charles W., Tigert, Douglas J. and Ring, Lawrence J. (1979), "Pragmatic Applications of Consumer Research in Retailing," in William Wilkie (ed.), Advances in Consumer Research, Vol. VI, Ann Arbor: Association for Consumer Research, 20-26.

Lessig, V. Parker (1973), "Consumer Store Images and Store Loyalties," Journal of Marketing, 37, 73-74.

McDougall, G. H. G. and Fry, J. N. (1974-75), "Combining Two Methods of Image Measurement," Journal of Retailing, 50, 53-61.

Menezes, Dennis and Elbert, Norbert F. (1979), "Alternative Semantic Scaling Formats for Measuring Store Image: An Evaluation," Journal of Marketing Research, 16, 80-87.

Pathak, Dev S., Crissy, William, J. E. and Sweitzer, Robert W. (1974-75), "Customer Image Versus the Retailer's Anticipated Image," Journal of Retailing, 50, 21-28, 116.

Reich, John W., Ferguson, Jeffery M. and Weinberger, Marc G. (1977), "An Information Integration Analysis of Retail Store Image," Journal of Applied Psychology, 62, 609-614.

Schiffman, Leon G., Dash, Joseph F. and Dillon, William R. (1977), "The Contribution of Store-Image Characteristics to Store-Type Choice," Journal of Retailing, 53, 3-14.

CONDITIONAL LOGIT VERSUS MDA IN THE PREDICTION OF STORE CHOICE

Stephen J. Arnold, Queens University
 Victor Roth, University of Toronto
 Douglas J. Tigert, University of Toronto

Abstract

When a research problem focusses on determinancy of choice rather than on classification or identification of a retailer's strengths and weaknesses, conditional logit analysis is clearly superior to multiple discriminant analysis. Too often, MDA is misused as a methodology for prediction of choice.

Introduction

The research described here was inspired by the recent paper of Gensch and Recker (1979) and by an earlier paper by Westin and Watson (1975). The Gensch and Recker paper provided a comparative analysis of the predictive power and diagnostic quality of conditional logit and regression analysis as applied to the problem of consumer choice of food stores. The paper raised a number of critical issues about the adequacy of a number of covariance choice models when the research question focusses on cross-sectional data rather than individual analyses and when the key research issue is determinancy of choice.

The Westin and Watson paper focussed on the issue of whether or not the population should first be segregated on the basis of "importance" characteristics before predictive models of choice are applied. In addition, the authors also raised the issue of the appropriate model for a choice framework rather than a classification framework. The research objectives for the study reported here were threefold and they build upon the key issues arising from the two papers cited above:

- i) to shed some light on the differences in the underlying assumptions of both the theoretical underpinnings and the quantitative structure of conditional logit and multiple discriminant analysis.
- ii) to test the key assumption of conditional logit related to the concept of a "common behavioural rule" (homogeneity of choice) expressed by McFadden (1974) and challenged by Westin and Watson (1975).
- iii) to examine the diagnostic qualities and predictive powers of conditional logit and discriminant analysis when applied to the same body of data.

The Multinomial, Multiattribute Logit Choice Model

Among the wide variety of competing choice models in the consumer behaviour literature, the logit model belongs to the general class of models described as covariance models. This class includes regression, MDA, logit and probit and is distinguished from other major classes such as Fishbein's (1972) learning-based model, lexicographic models, hierarchical models, multidimensional scaling and linear programming, etc. Some models incorporate individual beliefs about attribute importance while others attempt to derive such weights. The covariance models attempt to derive the weights and then use those weights in the prediction equation(s).

The logit model, which has its foundations within the psychological literature, is consistent with the theory

of sampling from a population of utility maximizing decision makers, and uses the attribute ratings of both chosen and unchosen alternatives in the choice set in order to reveal the determinant attributes. Briefly, the conditional logit model posits that when the n th individual in the population ($n = 1, \dots, N$) has a vector of measured attributes S_n and faces J alternatives which are described by vectors of attributes x_{jn} , then the individual has a utility function

$$U(S_n, x_{jn}) = V(S_n, x_{jn}) + \xi(S_n, x_{jn}) \quad (1)$$

where V is a nonstochastic function reflecting "representative tastes and ξ is stochastic and reflects the idiosyncracies of this individual in tastes for the alternative with attributes x_{jn} . If it is assumed that

$$V(S_n, x_{jn}) = Z_{jn} \theta \quad (2)$$

where $Z_{jn} = (z_{jn}^1, \dots, z_{jn}^k)$,

$$z_{jn}^k = v^k(S_n, x_{jn}), \quad k = 1, \dots, K,$$

$$\theta = (\theta_1, \dots, \theta_k),$$

a vector of unknown parameters, and that ξ is a function that varies randomly in the population with the property that in each possible alternative set $\{x_1, \dots, x_{jn}\}$, the values $\xi(S_n, x_{jn})$ are independently distributed with the Weibull (double exponential) distribution, then McFadden (1974) has shown that the probability P_{in} of the n th individual selecting the i th alternative is

$$P_{in} = \frac{e^{z_{in} \theta}}{\sum_{j=1}^J e^{z_{jn} \theta}} \quad (3)$$

Estimation of the unknown parameters θ is based on a maximum likelihood procedure i.e., finding the vector θ which maximizes the joint probability of observing the particular choice outcomes of each consumer. Mathematically, this procedure is equivalent to maximizing, by any of a variety of programs for unconstrained nonlinear optimization, the adjusted Log likelihood function

$$L = C - \sum_{n=1}^N \sum_{i=1}^J S_{in} \log \sum_{j=1}^J e^{(Z_{jn} - Z_{in}) \theta} \quad (4)$$

where Z_{in} is the vector of "ratings" by the n th consumer of the chosen i th alternative, and Z_{jn} is the vector of "ratings" by the n th consumer of the j th alternative. S_{in} is the value of the dependent variable.

The advantages of the conditional logit model over other methods of identifying determinant attributes are several-fold. The result of the maximum likelihood procedure is that the elements of θ are asymptotically efficient

and normally distributed under very general conditions and that an approximation is good even in small samples ($N > 50$). Also provided are standard errors of each of the elements of $\hat{\theta}$ thus permitting hypothesis testing and construction of confidence intervals. A determinant attribute can therefore be defined as an attribute where the null hypothesis that its coefficient is zero is rejected at the α level of significance.

The logit model can also derive the affect of a change in the choice probability as a result of a change in any of the attributes of any of the alternatives. However, the analyses of these elasticities and cross-elasticities is beyond the scope of this paper. Essentially, the programme requires that a reasonable model is one in which individuals compare all pairs of choice alternatives on the basis of their perceived differences in satisfaction on those attributes possessed by all alternatives. The programme allows for differential consideration sets across respondents so long as there is at least one pair of alternatives. Thus, the research instrument must generate the evoked set for each respondent and provide for a rating system for each alternative on all attributes. S_{in} is the value of the chosen/preferred store in each pair. It could be a dummy variable with the value "1" if the Z_{ij} pair includes the preferred store in which case all pairs of stores which do not include the preferred store are eliminated from the analysis. Alternatively, S_{in} could be the proportion of trips taken to a particular store in proportion of total dollar spending at a particular store, etc. Similarly, the $(Z_{in} - Z_{in})$ net difference scores could be based on semantic differential scales or simpler measurements such as dummy variables which express the extent to which a particular store either did nor did not possess a specific characteristic.

How Does Conditional Logit Differ From MDA

Multiple Discriminant Analysis is a fundamentally different type of model in that it is a classification rather than a choice model. The other techniques, including conditional logit, basically assume one population making choices based on their evaluation of the independent variables. MDA assumes several distinct populations (e.g., shoppers of the different chains), each having different patterns of scores on the independent variables. Discriminant analysis begins by pre-classifying respondents into market segments and then searching for a set of coefficients in n -dimensional space that will maximally separate the groups. Essentially, the discriminant programme examines across versus within group variances on the independent variables and develops a set of discriminant functions that will partition the groups in an n -dimensional space in the most efficient manner. MDA utilizes information only about the chosen alternative for each respondent and looks for patterns of differential scores on the attributes across the groups. There is no underlying relationship between attribute differentials and attribute importance in determining store choice. In conditional logit, it can be hypothesized that where large differences occur in attribute ratings between chosen and unchosen alternatives within the consideration set, then those differences should reflect determinancy of choice.

In addition, the preference probability function in conditional logit analysis overcomes several deficiencies in both discriminant and regression analyses. The computed probabilities are constrained to fall in the range of zero to one. More important, the shape of the logistic curve is theoretically consistent with the consumer choice literature. It assumes that the probability of choosing a given alternative is based on both the current position of that alternative relative to other alternatives

as well as on changes in that position based on improvements on determinant attributes. The logistic curve also recognizes the concept of "threshold" levels. Over certain portions of the curve, small improvements in market position might lead to only small improvements in the probability of choosing a particular alternative whereas at other positions on the curve, small improvements in market position might lead to large increases in the probability of choosing a particular alternative. Neither linear regression nor discriminant analysis incorporate those theoretical concepts. The following empirical section clearly demonstrates the diagnostic and predictive superiority of the multinomial logit model over MDA.

Methodology

The data analyzed were collected via telephone from a central WATS facility in Chicago, by a professional interviewing company. The market survey was conducted in the SMSA of Tampa/St. Petersburg in January, 1978 with all calling in the evening hours to insure the proper representation from working women. The sample of 903 respondents was drawn on the basis of a systematic random sampling procedure from the appropriate telephone directories covering the market area.

Questionnaire Design

The major section of the questionnaire asked respondents to name a single chain that was felt to best answer each of a series of 19 questions representing a fairly exhaustive list of grocery store attributes. Such attributes included dimensions as "lowest prices," "easiest to get to from home," "largest assortment/selection of food products," etc. The procedure, referred to as the associative technique, is simplistic in design and easy to administer by telephone. Its weakness lies in the fact that the attribute ratings on each chain by each respondent are all dummy variables. Either the chain is the best, the lowest, the cleanest, etc., or it is not. However, the diagnostic qualities of the associative technique in pinpointing strengths and weaknesses of the competitive chains in a market are well known (Tigert and Ma 1978).

In a separate section of the questionnaire, respondents were asked to specify both the most important and the second most important reason for choosing the store where they shopped most often. These "attribute importance" questions, generated by the direct questioning method, were not used in either the logit or MDA analysis but are used later in this analysis to help confirm or reject the logit and MDA results. In fact, one might argue that the logit coefficients and/or the MDA standardized discriminant function coefficients provide a measure of the validity of the direct questioning technique in uncovering determinant attributes.

The total set of outlet attributes was fairly exhaustive and included all the dimensions covered by the Lindquist (1974-75) review. They were based, however, on over 100 grocery shopping studies in Canada, the U.S., the Netherlands and the U.K. that have been completed by the authors. A summary of the development of that attribute list is reported in Arnold, Ma and Tigert (1977).

The dependent variable for the analysis was also a dummy variable represented by the specific chain at which the respondent shopped most often. For the MDA analysis, respondents were grouped into chain sets for each of the six largest chains in the Tampa/St. Petersburg market. Of the total sample of 1,000 respondents, 903 were shoppers of one of these six chains and this group forms the sample for the following analyses.

Definition of the Consideration Set

The logit analysis allows for differential consideration sets for each individual. For purposes of this analysis, a specific chain was included in an individual's evoked set if the chain received at least one mention on the 19 store attribute rating questions. Specifically, the chain had to be best on something. While it is hypothetically possible for a consumer to choose an outlet that is not perceived to be best on anything, no respondent was eliminated from the analysis on the basis of the screening criteria. As shown in the first column of Table 1, the average respondent had about 4 chains in the consideration set and therefore about 6 pairs of stores on which the net difference scores were computed. These results are consistent with the findings of Gensch and Recker (1979).

A Test of the Homogeneity of Choice Assumption

The logit analysis was first completed over the total sample of 903 respondents. Subsequently, three sub-samples of respondents were selected based on their responses to the direct questioning procedure on attribute importance. One sub-sample reported that "lowest prices" was the single most important reason for choosing the store where they shopped most often. A second group mentioned "assortment/variety of food products" and a third group rated "easiest to get to from home" as the most important attribute.

The logit model was estimated using for the choice object, the ratings of the store "shopped most often" rather than the store "shopped last" or a number of other shopping ratings included in the questionnaire. Previous work by Tigert and Arnold (1980) suggested that better predictions and higher logit coefficients result when using store shopped most often.

The strength and diagnostic quality of the logit model versus MDA was examined by comparing the hit ratios in the prediction results as well as the size and signs of the coefficients relative to the results of the direct questioning procedure on attribute importance.

Results

Table 1 reports on the results of the logit analysis and presents a series of logit coefficients for the total sample and for the three sub-samples described earlier. The coefficients are reported only for the most significant attributes from two different computer runs. In the top half of the table, only five attributes are reported. These five dimensions, covering location, price, assortment and service (2), have consistently appeared in the retailing literature as key determinants of store choice and they generated an 80 percent correct prediction in the logit model.

To investigate the improvement that could be achieved by expanding the attribute set, a second run was made with the same five attributes plus the next best three attributes. The logit coefficients for the original five attributes changed only marginally and the predictive power of the logit model improved only slightly to 81 percent correct prediction.

By far the most powerful store attribute was location/convenience ("easiest to get to from home") with a logit coefficient of 1.48 for the total sample. In the MDA analysis, this dimension was not significant at all in discriminating between the six shopper groups. The rationale for this major difference in results is directly traceable to the use of net difference scores versus absolute ratings on the chains. Table 2 sheds some light on this key concept.

TABLE 1

TAMPA/ST. PETERSBURG SUPERMARKET DATA: LOGIT COEFFICIENTS FOR MULTINOMINAL MULTIATTRIBUTE LOGIT CHOICE MODEL

STORE ATTRIBUTE	COEFFICIENTS FOR:			
	TOTAL SAMPLE	SUBSAMPLES WHO SAID EACH ATTRIBUTE WAS MOST IMPORTANT IN CHOOSING PREFERRED STORE		
		Low Price	Assortment/ Variety	Location
Location/ convenience....	1.48**	0.75	1.24	2.69**
Lowest Prices..	0.71	1.87	0.65	0.20*
Fast checkout service.....	0.51	0.35*	0.76	0.58
Friendly, court- eous service...	1.02	0.85	0.59	0.82
Assortment/ variety.....	0.71	0.09*	1.59	0.32*

Percent Correct Prediction.....	79.7%	85.6%	72.0%	90.0%
Sample Size (n).	(903)	(200)	(104)	(311)
Number of choice sets.....	(3688)	(749)	(387)	(1142)

Impact of Increasing Variable Set

Location/ convenience....	1.46	0.75	1.24	2.69
Lowest prices..	0.69	1.90	0.67	0.15*
Best quality meat ^a).....	0.61	0.17*	0.57	0.28*
Fast checkout service.....	0.48	0.34*	0.70	0.49
Friendly, court- eous staff....	0.89	0.78	0.51	0.77
Best weekly specials ^a)....	0.49	0.11*	0.37*	0.63
Best private label ^a).....	0.29	0.62	0.04*	0.29*
Assortment/ variety.....	0.61	0.55	1.54	0.36*

Percent Correct Prediction.....	81.4%	85.6%	75.2%	91.0%

* Not significant at .05 level

** Read: For the total sample of 903 respondents, the highest logit coefficient occurred for the "location/convenience" variable at 1.48. The coefficient rises to 2.69 for the sub-sample that indicated location was the most important variable in choosing their preferred store.

a) New variables

If we assume that convenience to home is a necessary condition to move a chain into the consideration set, then it would seem reasonable to hypothesize that many consumers would shop most often at the chain that was the "easiest one to get to from home". Thus, we would expect to find that in the associative technique, a large proportion of consumers would report that the chain at which they shopped most often was also the chain that was easiest to get to. The first row of Table 2 confirms that hypothesis. The average percentage across the row (weighted by chain market size) is 71 percent and there is very little variance across the chains. Thus, in terms of the analytical procedure used by MDA, the mean

TABLE 2

SUMMARY: SHARE OF MENTIONS GOING TO EACH CHAIN AMONG THOSE WHO SHOP MOST OFTEN AT THAT SAME CHAIN

Store Characteristics	Store Shopped Most Often					
	Albertson's	Kash 'N' Karry	Pantry Pride	Publix	Winn-Dixie	U-Save
1. Easiest to get to from home.....	54%*	70%	74%	68%	76%	80%
2. Lowest prices.....	46*	77	65	25	51	85
3. Best overall assortment of food products.....	87	51	34	82	33	41
4. Best overall assortment of non-food products.....	98	47	21	35	31	28
5. Best at being in-stock.....	71	65	49	87	50	63
6. Cleanest stores.....	81	46	40	91	31	27
7. Best overall customer service.....	71	65	56	90	52	59
8. Fastest Checkout counters.....	82	64	43	73	54	51
9. Most friendly, courteous staff.....	70	75	55	87	63	70
10. Best quality fresh meat.....	37	44	33	70	58	44
11. Best variety selection of fresh meat.....	56	42	30	70	52	32
12. Best quality fresh produce.....	64	33	42	78	20	33
13. Most pleasant shopping environment.	88	54	42	88	44	44
14. Best overall advertising.....	45	34	19	77	54	15
15. Best weekly specials.....	48	39	41	71	70	25
16. Best quality private label brands..	46	47	82	66	74	61
17. Best specialty baked goods dept....	71	16	16	82	27	13
18. Best delicatessen dept.....	48	5	22	83	39	10
Average weighted by attribute importance scores for each chain's shoppers for Q 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 16.	(73%)	(63%)	(55%)	(79%)	(52%)	(72%)

* READ: Among those who shop most often at ALBERTSON'S, 54% also said Albertson's was easiest to get to, 46% said Albertson's had the lowest prices, etc.

scores on this dimension across the six shopper groups would show little variance and would not therefore, provide much discriminatory power. As a consequence, in spite of the fact that location is a critical determinant of store choice, the results of the MDA would be negative in terms of the classification objectives. Attributes which do exhibit a high variance across the columns include: i) lowest prices; ii) specialty baked goods; iii) best delicatessen; iv) cleanest stores, etc. In fact, these were the dimensions that were the most significant in the MDA analysis. In short, what MDA does is provide strong diagnostics on how the chains are perceived by their own customers to be different on the various attributes but not which attributes were the most critical in determining store choice.

The Logit analysis focusses directly on store choice. It is because the chains all score fairly high on easiest to get to from home among their own customers that the net difference scores are so high. Unchosen stores received very few mentions as rating best on location.

This basic difference in the research objectives of logit and MDA lead to very different results in terms of predictive power in classifying respondents as shoppers of their preferred outlet. Table 3 shows the MDA analysis generated only 55.4 percent correct predictions compared to 81.4 percent for the logit analysis. The improvement in pred-

ictive power of the logit over the MDA was 26/55 or about 47 percent.

More important, the logit classification was very close to the actual market share positions of the chains while the MDA results were off by more than 100 percent for some chains. Yet, MDA continues to be used extensively in marketing research for purposes of classifying consumers and predicting choice. In fairness to MDA, we should report that when MDA mis-predicted, it tended to misclassify respondents into the shopper set closest to the correct chain in the perceptual map (not shown). For example, the misclassified Kash N Karry shoppers were almost all classified as either U-Save shoppers or Pantry Pride shoppers.

Testing Homogeneity of Choice

The underlying assumption of logit analysis is that all consumers have a homogeneous and "representative" set of attributes by which they maximize their utility of store choice. To test this assumption in terms of the relative strength of the logit coefficients, we re-ran the logit analysis for the three sub-samples described earlier. For the sub-sample that indicated earlier in the questionnaire that "lowest prices" were the most important store attribute in choosing the store where they shopped most often, the size of the logit coefficient on "lowest

TABLE 3

ACTUAL NUMBER OF CASES IN EACH SHOPPER GROUP AND PREDICTED NUMBER OF CASES IN EACH GROUP BY DISCRIMINANT AND LOGIT MODELS: TAMPA/ST. PETERSBURG GROCERY DATA

CHAIN	ACTUAL SHOPPERS	PREDICTED SHOPPERS BY:	
		LOGIT MODEL	DISCRIMINANT MODEL
ALBERTSON'S..	53*	59*	120*
KASH N KARRY.	211	197	111
PANTRY PRIDE.	38	36	92
PUBLIX.....	335	344	301
WINN-DIXIE...	196	193	154
U-SAVE.....	70	74	125
TOTALS....	903	903	903
PERCENT CORRECT PREDICTIONS....		81.4%	55.4%
Chance Prediction using equal group size.....			17%
Chance Prediction using largest group (Publix)			37%

* Read: In the Tampa sample, 53 shoppers indicated they shopped most often at Albertson's. The logit model classified 59 respondents as Albertson's shoppers while the discriminant model classified 120 respondents as Albertson's shoppers.

prices", rose dramatically, from 0.71 for the total sample, to 1.87 for this sub-sample. Simultaneously, the coefficient on assortment of food dropped precipitously to a non-significant level and the coefficient for location dropped by half. The conclusion about this segmentation analysis is transparent. Consumers who seek out primarily low prices are prepared to trade away both location and assortment for price. The results have major strategic significance for grocery retailing and suggest that both warehouse discount stores and limited assortment box stores in good locations have a bright future.

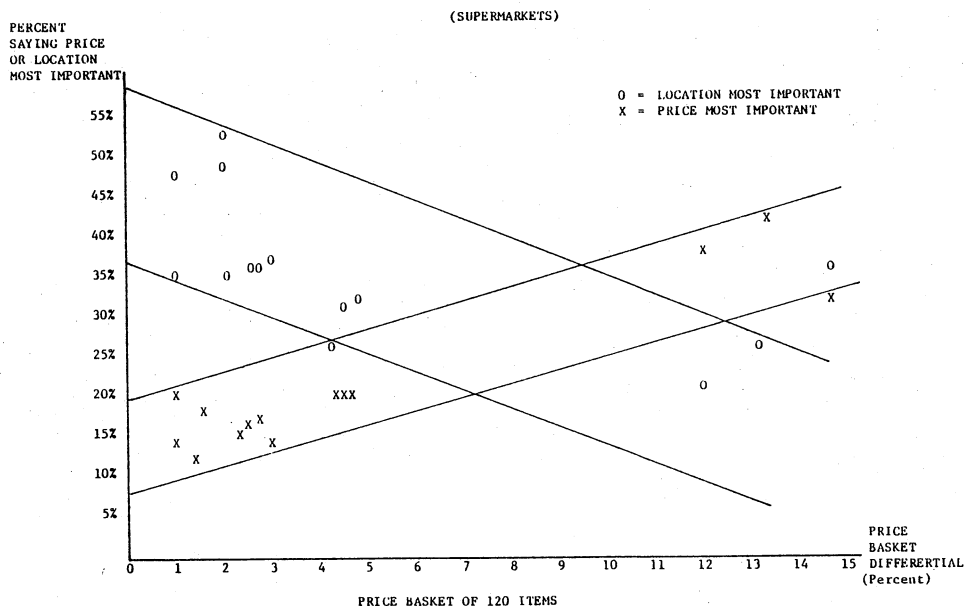
In the column labelled location in Table 1, the logit coefficient for location/convenience is so overpowering that a number of other coefficients, including the price coefficient, became non-significant. Consumers appear to be operating in a tradeoff mode between price and location. The significance of that tradeoff can be seen more clearly in Exhibit II. Across 14 major markets, in the U.S., Canada, The Netherlands, and the U.K., the authors have asked the same direct questions about attribute importance. At the same time, chains have been monitored in terms of their relative price positions by a large grocery basket of 120 items. Exhibit II reports on the proportion of respondents who said either low prices or location/convenience were the most important reasons for choosing their preferred grocery outlet. When price differentials across chains are small, upwards of 50-55 percent of consumers say they choose their preferred store on the basis of location/convenience. When price differentials rise, the consumer is sensitized to these differentials through advertising, shopping experiences (learning) and word-of-mouth behaviour. The higher the price differential, the larger is the consumer segment that trades away location for price.

Finally, note in Table 1 that the sub-sample labelled assortment/variety yielded the highest logit coefficient for the assortment/variety attribute, the highest coefficient for the "fast checkout" attribute and a moderately strong coefficient for location. This market segment could best be described by the concept of "one-stop shopping....for the hassled consumer". One can visualize a customer who says..."find me a convenient supermarket with everything I need and get me out fast". The super combo, best exemplified by Albertson's in Tampa/St. Petersburg, would seem to be ideally suited to serve this consumer segment.

Logit Coefficients Versus Attribute Importance By Direct Questioning

Table 4 reports on the results of our direct questions to respondents about the most and second most important

EXHIBIT II A FOURTEEN-CITY COMPARISON OF THE RELATIONSHIP BETWEEN PRICE DIFFERENTIALS AND STORE CHOICE CHARACTERISTICS



Source: Tigert (1980)

reasons for choosing their preferred grocery outlet. The four attributes receiving the most mentions, i.e., location, price, assortment and service, are identical to the four attributes with the highest logit coefficients in Table 1. Attributes such as best delicatessen and best specialty baked goods, that were most significant in the MDA analysis, do not even appear on the list in Table 4. More important, the relative size of the scores in the average column on the right hand side of Table 4 are closely paralleled by the relative size of the logit coefficients in Table 1. Location is clearly the dominant attribute in both tables with price, assortment and service fairly close in second, third and fourth positions.

The logit coefficients, therefore, provide strong support for the direct questioning technique in measuring attribute importance. They also provide support for the early work with gravity models which utilized only two outlet characteristics: i) location (distance from home) and ii) store size (a proxy for assortment). Both the gravity models and conditional logit attempt to incorporate information about all alternatives in the consideration set. MDA does not.

TABLE 4

TAMPA/ST. PETERSBURG MOST IMPORTANT REASONS FOR SHOPPING WHERE SHOP MOST OFTEN: ATTRIBUTE IMPORTANCE

	Most Important	Second Most Important	Total 1st & 2nd Mentions	Average
Location/Convenience....	31%*	15%	46%	23%
Low Prices/Prices.....	20	10	30	15
Assortment/Selection/Variety.	10	15	25	13
Service/Friendly Courteous/Fast				
Checkout.....	8	12	20	10
Cleanliness....	8	10	18	9
Meat Quality/Variety.....	7	7	14	7
Overall Quality.....	3	5	8	4
Weekly Specials/Coupons.....	3	3	6	3
Produce Quality/Variety.....	3	5	8	4
Nice Store/Shopping Environment....	2	3	5	2
Others.....	5	10	15	7
None/No Other..	0	5	5	3
Totals.....	100%	100%	200%	100%

* Read: 31 percent of respondents said location/convenience was the most important reason why they chose the store where they shop most often.

Conclusions

Logit analysis is a powerful tool for understanding consumer choice of food stores, particularly in diagnosing determinant attributes. While MDA is adequate for purposes of perceptual mapping and understanding weaknesses and strengths of alternative chains, its performance is weak in predicting choice and in validating determinant attributes.

A number of research steps should follow the results achieved here. First, there is a need to examine the improvement in results that could be achieved by measuring both attribute importance and store ratings through alternative scales such as semantic differentials or perhaps forced choice. Given the wealth of models now available that allow for differential evoked sets, further improvements could be made in alternative ways of defining evoked sets for individual respondents. Finally, logit analysis is only one of a variety of new models now available for analyzing consumer choice. More comparative analysis across competing models would be helpful in uncovering the most powerful alternatives.

References

- Arnold, Stephen J., Ma, Sylvia and Tigert, Douglas J. (1978), "A Comparative Analysis of Determinant Attributes in Retail Store Selection," in ed. Advances in Consumer Research, Vol VI, Ann Arbor: Association for Consumer Research.
- Arnold, Stephen J. and Tigert, Douglas J., Determinant Attributes in Consumer Choice, working paper, University of Toronto, 1980.
- Fishbein, Martin (1972), "The Search for Attitudinal-Behavioural Consistency," in Joel B. Cohen (ed.) Behavioural Science Foundation of Consumer Behaviour, New York: Free Press.
- Gensch, Dennis H. and Recker, Wilfred W. (1979), "The Multinomial, Multiattribute Choice Model," Journal of Marketing Research, XVI, 124-132.
- Lindquist, Jay D. (1974-75), "Meaning of Image," Journal of Retailing, 50, 29-38, 116.
- McFadden, Daniel (1974), "Conditional Logit Analysis of Qualitative Choice Behaviour," in ed. Paul Zarembka, Frontiers in Econometrics, New York: Academic Press, pp 105-42.
- Tigert, Douglas J. and Ma, Sylvia (1979), "In Search of A Supermarket Strategy: Albertson's Drives on Tampa/St. Petersburg," Proceedings of the 1978 Attitude Research Conference, American Marketing Association, Tarpon Springs, Florida.
- Westin, Richard B. and Watson, Peter L. (1975), "Reported and Revealed Preferences As Determinants of Mode Choice Behaviour," Journal of Marketing Research, XII, 282-289.

RESEARCH INTO SHOPPING MALL CHOICE BEHAVIOR

Roy D. Howell, University of Illinois-Urbana
Jerry D. Rogers, Southwest Missouri State University

Abstract

This paper explores some of the issues relevant to research into shopping mall choice behavior, including the measurement of patronage, situational specificity, and the level of aggregation in parameter estimation. Results of an exploratory study addressing these issues are presented. The findings suggest that shopping situation should be specified, that multiple indicators of patronage should be employed, and that parameters in a model of mall-choice behavior should initially be estimated separately for each shopping area under study.

Introduction

Consumer researchers have in recent years begun to focus more of their theory-building and empirical research efforts in the area of consumer patronage behavior (Darden 1980). Most of the research in this area has dealt with the store choice decision (Granbois 1977), as is proper and to be expected. Few retail outlets, however, exist as isolated entities. The synergistic effects of multiple retail establishments located in proximity to one another, zoning ordinances which restrict locations in which retail businesses may operate, and the limited availability of "good" free-standing retail sites tend to encourage the clustering of retail trade into relatively compact areas. This clustering may be planned and formalized as a shopping center or shopping mall, may be unplanned, as in traditional downtown shopping areas, or may be formalized *ex post facto*, as in "revitalized" downtown malls. As such, the shopping area has become a legitimate object of research both in its own right and in its effect on the stores of which it is comprised.

As Bucklin (1967) has noted, the intra-urban shopping area, as a retail entity, lies somewhere between the individual store and the urban-entity itself on a scale of disaggregation. Research on shopping center preference and patronage has reflected this positioning, with theory and methodology being drawn from trading area theory, emphasizing the mass-distance relationship (Huff 1964), and from the store choice literature, which emphasizes store "image" and its components (Lindquist 1974-75). Stanley and Sewall (1977) demonstrate how these two approaches can be combined in store-choice research, as do Jain and Mahajan (1979). Recently, Nevin and Houston (1980) have applied a model containing both gravitational and image components to the area of shopping mall choice behavior.

Research into shopping center or shopping mall choice behavior cannot, however, be simply a straightforward extension of the trade area and store choice research streams. While some of the problems present in researching shopping mall preference and patronage are similar to those faced by the researcher investigating at the level of the retail store, others are unique. In particular, three issues confronting researchers investigating shopping mall choice behavior are addressed in this paper. The first of these issues concerns the choice and measurement of image components relative to shopping malls. The second concerns the measurement of patronage behavior. Third, the level of aggregation

appropriate for parameter estimation in models of mall choice behavior is discussed.

These issues are discussed in the context of a study conducted in a medium-size midwestern SMSA. Personal interviews were conducted with two-hundred sixty middle to upper class females chosen from a two-stage area sample. Fifty-one respondents had not patronized any of the three shopping areas in the city, or had otherwise unuseable questionnaires and were thus eliminated from further analysis. The respondents were asked to rate the adequacy of two enclosed shopping malls ("B" and "P") and the "revitalized" downtown area (D) on eighteen attributes (1=very unsatisfactory, 7=very satisfactory), as implied by the adequacy importance model (Mazis, Ahtola, and Klippel 1975). Also measured was the shoppers' distance from each of the three centers, and preference and patronage measures which are discussed below.

Both malls have approximately 270,000 sq. ft., with the downtown area relatively comparable in terms of number and types of stores. All have seven to nine stores carrying women's fashion items (the shopping situation specified in this research), and represent the only major shopping areas in or around the community.

Attribute Measurement

Nevin and Houston are essentially correct in stating that, "...little work has been done on the dimensions of shopping area image" (1980, p. 84). Most applications in mall-choice research have relied heavily on the store image/attitude dimensions described by Lindquist (1974-75). While little work has been done on those dimensions specific to mall choice, some agreement seems to be emerging in the literature on the store image dimensions applicable to the shopping mall.

Aggregating observations across shopping centers studied, Hauser and Koppelman (1979) found, using factor analysis, four dimensions from sixteen attributes, labeled variety, quality and satisfaction, value, and parking. Assessing the structure of sixteen attributes separately for each of the shopping centers in their study, Nevin and Houston (1980) found three dimensions (labeled assortment, facilities, and market posture) which are similar to the dimensions found by Hauser and Koppelman (1979). What is not clear, however, is the appropriateness of items drawn from the store image literature in mall choice research. Clearly, some store image items such as price/value, credit availability, and sales personnel are applicable to the mall choice process only to the degree that the center as an entity has been successful in establishing a cohesive and consistent overall image. Similarly, the mall has attributes which are not among those composing store image, such as special events/exhibits, recreational value, and commons-area atmosphere. Additionally, the contribution of the various store images to the image of the mall, and vice versa, is a question which begs further research.

Also not firmly established is whether the dimensionality of the image/attitude items employed is consistent across centers, or whether consumers' perceptual space differs for each center studied. Hauser and

TABLE 1
Hypothesized Factor Structure of
Shopping Mall Attributes

Attribute	Factor (A)				
	1	2	3	4	5
General Attractiveness	λ_1	0	0	0	0
Quality of Merchandise	λ_2	0	0	0	0
Appeal of Atmosphere	λ_3	0	0	0	0
Cleanliness of Stores	λ_4	0	0	0	0
Attractiveness of buildings and Landscaping	λ_5	0	0	0	0
Courtesy of Sales Personnel	0	λ_6	0	0	0
Helpfulness of Personnel	0	λ_7	0	0	0
Knowledgeability "	0	λ_8	0	0	0
Availability of Latest Fashion	0	0	λ_9	0	0
Variety of Styles	0	0	λ_{10}	0	0
Comparison Shopping	0	0	λ_{11}	0	0
Amount of Advertising	0	0	0	λ_{12}	0
Quality of Advertising	0	0	0	λ_{13}	0
Information in Advertising	0	0	0	λ_{14}	0
Convenience in Center	0	0	0	0	λ_{15}
Convenience of Store Hours	0	0	0	0	λ_{16}
Parking Availability	0	0	0	0	λ_{17}
Traffic Congestion	0	0	0	0	λ_{18}

Koppelman (1979) assume a "basic structure of perceptions" and perform a factor analysis across centers. Nevin and Houston (1980) perform separate factory analyses for each center and, using a test for factor congruency, indicate "a consistent factor structure for all five shopping areas.

In this study, confirmatory factor analysis (Joreskog 1971) was employed to assess both the hypothesized factor structure and the equality of factor structures for each of the centers. The hypothesized factor structure is presented in Table 1. When non-zero covariances are allowed between the factors (Ψ not diagonal), the structure in Table 1 provides an acceptable fit for each of the three centers (average $\chi^2 = 137$ with (171-51=120) d.f., $p < .14$) and for the pooled data. While the coefficients in A and specific variances are substantially similar for each group, the hypothesis that $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma$ (equal covariance matrices for each of the centers), based on the F-statistics derived from Box's M, must be rejected. While the hierarchical hypothesis testing procedures described by Joreskog (1971) for assessing the equality of factor structures of several groups is not strictly applicable in this situation (where we are dealing with one group measured on different sets of variables), we find that the hypothesis of invariant factor pattern cannot be rejected, while the hypothesis of equal specific variances is rejected at the .05 level.

It can be inferred from this that the five factor solution with a general factor pattern such as that in Table 1 is applicable to the attributes of each of the centers, but that the magnitude of the λ in Table 1, and thus the variance of the variables not accounted for by the common factors, differs for the three centers. This further implies that the variance-covariance matrices of the five factors are not equal for each of the three centers.

In order to provide comparability of the coefficients in tests of the overall model, unit weights were employed such that summated scales were formed for each dimension of each center. Table 2 contains the reliability analysis of these and other measures used in this study. The alpha coefficients reported are consistent with the findings of the confirmatory factor analysis in that the attribute structures for B and P (the two enclosed malls) are more similar to each other than either is to the structure of Downtown. The reliabilities reported here compare favorably with those reported by Nevin and Houston (1980).

TABLE 2
Scale Analysis

Scale	Center								
	B			P			D		
	r^1	α_1^2	α	r^1	α_1^2	α	r^1	α_1^2	α
X_1 Atmosphere			<u>.86</u>			<u>.85</u>			<u>.82</u>
attractiveness & decor	.71	.81		.69	.80		.72	.74	
quality of merchandise	.61	.83		.56	.84		.53	.81	
appeal of atmosphere	.81	.78		.79	.78		.71	.75	
cleanliness of stores	.56	.85		.54	.84		.52	.80	
attractiveness	.65	.83		.67	.81		.65	.77	
X_2 Personnel			<u>.87</u>			<u>.86</u>			<u>.86</u>
courtesy	.77	.79		.73	.80		.75	.79	
helpfulness	.82	.75		.78	.75		.81	.72	
knowledgeability	.69	.87		.70	.83		.67	.85	
X_3 Fashion Shopping			<u>.80</u>			<u>.73</u>			<u>.71</u>
availability	.67	.70		.64	.54		.58	.56	
variety	.74	.61		.64	.53		.60	.51	
comparison shopping	.58	.80		.50	.72		.51	.70	
X_4 Advertising			<u>.83</u>			<u>.82</u>			<u>.85</u>
amount	.62	.83		.60	.84		.63	.83	
quality	.80	.72		.74	.71		.78	.71	
information	.76	.77		.73	.72		.72	.77	
X_5 Convenience			<u>.67</u>			<u>.75</u>			<u>.71</u>
in-center	.48	.53		.52	.67		.53	.68	
store hours	.46	.56		.50	.69		.56	.68	
parking	.52	.52		.60	.62		.59	.56	
traffic	.43	.62		.53	.68		.54	.60	
X_6 Proximity ³			<u>.69</u>			<u>.85</u>			<u>.71</u>
distance									
time									
X_7 Accessibility			<u>.78</u>			<u>.86</u>			<u>.77</u>
accessibility ⁴	.53	.71		.60	.83		.54	.76	
congestion	.63	.62		.76	.68		.71	.70	
difficulty	.62	.60		.72	.73		.67	.63	
Y_2 Patronage ³			<u>.83</u>			<u>.77</u>			<u>.81</u>
#purchases	.72	.75		.67	.66		.74	.69	
time since last purchase ⁴	.59	.81		.60	.70		.70	.71	
frequency of visit	.63	.80		.51	.77		.51	.80	
\$purchases	.70	.77		.59	.71		.56	.79	

¹ corrected item-total correlation

² alpha if item deleted

³ standardized-item scale

⁴ item reverse scored

Distance

The variety of different measures of proximity or distance have been applied in patronage research (Granbois 1977). While it is generally accepted that perceived as well as actual distance is important (Brunner and Mason 1968), many studies have relied on map distance. Others have used ratings of proximity (Gentry and Burns 1977-78) or drive time (Nevin and Houston 1980). In this study, the multiple indicator approach is employed in the measurement of distance in order to account for both perceptual variations and objective distances. One scale, labeled proximity, is composed of map distance measured in blocks and a subjective time estimate measured in minutes. The standardized item scale composed of these two measures exhibits acceptable reliability. To capture more fully the "convenience" aspect of the distance construct, respondents were asked to rate each of the centers with regard to its general availability, the degree of traffic congestion usually encountered in driving to the center, and the general degree of difficulty they encountered in making a shopping trip to the center. These items form the scale labeled accessibility.

Situational Specificity

In this study respondents were asked to rate the attributes of the centers and to report their patronage of each of the centers only in the context of shopping for women's fashion clothing items for themselves or members of their family. In this respect the study is different from most reported research in this area, wherein shopping situation has not been specified.

In store choice research, the stores included as competitors are often highly similar in the distribution of types of merchandise carried by the stores (e.g., department stores are usually analyzed vis à vis other department stores, supermarkets vis à vis other supermarkets, etc.). In such research, global measures of patronage such as number of visits over a given time period, average frequency of visit, last store visited, and other measures of this type are acceptable to the extent that the type of merchandise carried by the stores under consideration is similar. This is not an unacceptable assumption. The distribution of types of stores and thus merchandise line availability in shopping centers is subject to great variability, however, and this variability in merchandise distribution may substantially affect the number of dollars spent, frequency of visit, and other possible measures of patronage, as well as influencing the relationships between image/attitude variables, mass/distance variables, and patronage.

For example, one of the shopping areas in this study (B) is anchored by a large supermarket in addition to a national chain department store. Similarly, the presence of a large discount-drug store distinguishes center P. Observation tends to suggest that such diversity of store types among centers is the rule rather than the exception. In the case where the researcher does not specify the shopping situation, a center such as B, anchored by a supermarket, will show a higher average frequency of visit, etc., than will a center without a supermarket, with this difference quite independent of overall preference or image. Similarly, the relationship between distance and patronage is likely to be larger for a center with a large area devoted to convenience goods and frequently purchased merchandise.

This phenomenon may account for the findings of such studies as Gentry and Burns (1977-78), where proximity is found to out-perform image/attitude variables in explaining frequency of visits.

Specifically, as discussed in the brand choice area by Warsaw (1980), the probability that an individual shops at area j , $P(s_j)$, is in fact dependent on the probability that the individual encounters and shops under condition i , denoted $P(S_i)$. The probability of shopping at area j is thus

$$P(s_j) = \sum_i P(S_i)P(s_j|S_i) \quad (1)$$

Since $P(S_i)$ is not equal for all i (e.g., shopping for food is likely to occur more frequently than shopping for fashion clothing), nor for all individuals (e.g., some individuals shop for food more often than others), and $P(s_j|S_i)$ is not equal for all j given the distribution of merchandise available at the j centers in addition to perceptual factors, attempts to explain the frequency of individuals' shopping trips to j shopping areas, as in Nevin and Houston (1980) and Hauser and Koppelman (1979) are likely to meet with limited success.

By specifying a particular shopping situation, we are dealing with only one S_i . While individual differences in $P(S_i)$ remain, the analysis is not contaminated by merchandise distribution causes of variation in $P(s_j|S_i)$. It is suggested that, in research in this area, the researcher either (1) specify the shopping situation(s), or (2) include variables in the model which represent the number of stores (or square footage) devoted to various types of merchandise in each center.

Much of the research in both store and mall-choice behavior has utilized measures of affect, preference, or behavioral intentions as dependent variables. While this is useful, predicting and explaining behavior is a more rigorous and preferred task. The question is however, just what behavior or behaviors constitute patronage? A variety of measures have been utilized as operational measures of patronage (see Cranbois 1977). The most popular seems to be frequency of visit. Nevin and Houston (1980), for example, utilize shopping frequency measured on a six-point scale ranging from "less than once a year" to "once a week". Hauser and Koppelman (1980) utilize self reported frequency of visits, while Gentry and Burns (1977-78) also utilize frequency of use.

It seems that "patronage" implies more than simply frequently visiting a shopping center. Is the individual who shops frequently, but seldom buys, a "patron" to the same degree as one who not only shops frequently but also buys a large quantity of merchandise at the center? What about the individual who shops infrequently but makes large volume purchases? These considerations, along with the increased measurement reliability obtainable through the use of multiple measures, suggest the use of several indicators in the measurement of patronage.

As can be noted in Table 2, four measures of patronage were combined as a (standardized item) summated scale in this study, exhibiting reasonably good reliabilities. Purchases, frequency of visits, and dollars spent (all over the three months prior to the study) and number of weeks since last purchase were reported by respondents for each of the three centers under study. As there were no strong reasons, *a priori*, that any of the measures were more important than the others, equal weight was given to each. Alternatively, managerial or theoretical considerations may indicate unequal weighting, or the first principle component could be utilized. (Since the items are measured in different units and thus have unequal variances, each was standardized before summation).

Although neither the summated scale approach nor the weighting schemes mentioned above may be optimum in an empirical sense, the measurement of patronage is not an empirical question. In applied, managerially oriented research each of several measures of patronage may be utilized separately as a dependent variable, with the impact of image, distance and other variables estimated for each. For the development of theory in the area of patronage behavior, whether store or shopping mall patronage is concerned, a consistent operationalization of the patronage construct should be defined, and this operationalization should be theoretically richer and operationally more meaningful and reliable than simple frequency of visit.

Estimation

Yet another issue to be addressed in mall-choice research is the specification of and estimation of a model. Inherent in the model estimation issue are decisions concerning the form of the dependent variable(s), the level of aggregation at which the parameters are to be estimated, and the statistical method to be employed. Considerations concerning each of these areas and the approach employed in this study are discussed in turn below.

Dependent Variables

Patronage, however defined or measured, is the compelling choice as an ultimate dependent variable in either store

or mall centered research. However, as discussed earlier, individual differences exist in $P(S_i)$, the probability of encountering and shopping under various situations. The literature on shopping and prepurchase information search clearly documents one rather obvious conclusion: That some people do much more of it than others. Darden's patronage model (1980) suggests numerous factors which may influence the absolute amount of shopping engaged in by an individual. Demographic, life style, life cycle, and socioeconomic variables affect the amount of shopping the consumer will engage in by affecting the length and composition of the individual's "shopping list", or "need queue", and thus the probability and frequency of encountering various shopping situations. Similarly Darden (1980) and Howell (1979) have shown that individuals differ in their shopping specific life styles, or shopping orientations. Shopping orientations such as enjoyment of shopping, recreational shopping, and shopping proneness affect the predisposition to shop and thus the amount of shopping done by consumers encountering a given shopping situation.

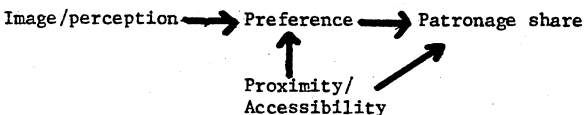
The point of this discussion is to indicate that unless one has a very rich data set containing measures such as those described above, attempts to model patronage, as such, will generally be less than satisfactory. In the absence of variables which explore the amount of shopping done by the individual, it would seem that the appropriate ultimate dependent variable should be patronage share rather than any absolute measure of patronage.

Also, most applications in the mall patronage (and store patronage) literature measure either affect, preference intentions, or behavior, in addition to cognitive attitudinal (image) variables and external facilitators (distance). The somewhat lower ability of image measures to explain patronage when included with distance measures can be to some degree accounted for by the failure to model the linkage of these through preference, as is suggested in the behavioral and brand choice literature (see Reibstein 1978), although Nevin and Houston (1980) were actually more successful in explaining frequency of visit than in explaining affect or intentions.

Preference (Y_{ij}) is measured in this study on a 1-6 scale ranging from least preferred to most preferred.

Similarly, it is not clear whether distance/proximity, however measured, is related to behavior through preference or whether it is a facilitator which moderates the affect of preference on behavior. Hauser and Koppelman (1980) address this issue directly by holding availability/accessibility constant. Nevin and Houston's (1980) results seem to indicate that distance is relatively more powerful in explaining intentions and behavior than in explaining affect.

These considerations lead to the testing of a model of the general (simplified) form:



wherein perceptions of each center affect preference for that center, which in turn affects patronage share. Proximity and accessibility are hypothesized as affecting both preference and patronage share.

Model Specification

Given the concept of determinant attributes, and Luce's choice axiom, it does not necessarily follow that having

a high (low) perception of the adequacy of a center on a particular attribute will lead to higher (lower) preference or patronage share for that center, since all of the competing centers could be perceived as being adequate (inadequate) (Bearden 1977).

Thus the adequacy of each center relative to the perceived adequacy of the other competing centers should prove a better predictor of patronage share than absolute level of adequacy. Similarly absolute proximity or accessibility should not be as effective as relative proximity or accessibility in explaining preference or patronage share.

It follows directly from the above discussion that a preferred model in this situation would be a multiplicative competitive interaction model, which has been often employed in retailing research (Stanley and Sewall 1976; Jain and Mahajan 1979; Hansen and Weinberg 1979). To estimate the MCI model through least squares techniques, however requires, in this case, that each center have a non-zero patronage share for each individual. When this is not the case, either (1) individuals who have not patronized all three centers can be deleted, or as is more often the case, (2) individuals can be aggregated by area (Stanley and Sewall 1976; Jain and Mahajan 1979). The first solution may, in many cases, result in an unacceptable and potentially biasing loss of data (Jain and Mahajan 1979), while the second approach presupposes a homogeneity of perceptions within each area. In the absence of substantial empirical evidence to the contrary, this would seem to be a somewhat unrealistic assumption.

Level of Aggregation

Also, the MCI model assumes that the parameters which relate each of the predictors in the model to the dependent variable are the same for all objects (centers) in the market. It seems that this should be treated as an empirical question rather than as an assumption.

The MCI is not unique in its assumption of the applicability of pooling across stores or centers in the parameter estimation process. The use of discriminant function analysis has similar implications. Other techniques, such as least squares regression, can be estimated either across centers or for each center individually. Both approaches are evident in the literature, although the issue is seldom explicitly addressed (Hauser and Koppelman 1980; Nevin and Houston 1980).

An examination of the regression coefficients estimated separately for each center under consideration by Nevin and Houston (1980) seems to indicate that an assumption of equal coefficients would be unwarranted, although they do not test for this explicitly. In particular, the coefficients they obtain for the downtown area in their study seem to be substantially different from those they obtain for the four mall-type shopping centers.

If the parameters relating the predictors to preference and patronage share are (statistically) equal for each of the centers, aggregation is preferred, as the increase in the number of observations used to estimate each parameter will yield estimates with lower variance. However, inappropriate aggregation can lead to biased parameter of estimates. The approach taken in this study is to estimate at the lowest level of aggregation possible; those centers with equal coefficients can be subsequently pooled and reestimated.

Based on the above considerations, the model estimated in this research is

$$Y_{1j} = \sum_{k=1}^7 \beta_{jk1} X_{jk}^* + \epsilon_{1j} \quad (2)$$

$$Y_{2j}^* = Y_{1j} Y_{1j} + \beta_{j62} X_{j6}^* + \beta_{j72} X_{j7}^* + \epsilon_{2j} \quad (3)$$

where:

Y_{1j} - (n x 1) vector of preferences for center j

$$Y_{2j}^* = \frac{Y_{2j}}{3} = \text{n-length of patronage share for center j}$$

$$\sum_{j=1} Y_{2j}$$

$$X_{jk}^* = \frac{X_{jk}}{3} = \frac{X_{jk}}{\sum_{j=1} X_{jk}} = \text{relative perceptions of center j on variable k.}$$

Y_{1j} = influence of preference for center j on patronage share of center j.

As such, this model, though cross-sectional, bears some resemblance to the market share model proposed by Houston and Weiss (1974). As a recursive model, each equation can be estimated using ordinary least-squares methods. However, since equations (2) and (3) are estimated separately for each center (j), important information contained in each error term (ϵ_{1j} and ϵ_{2j}) can be brought to bear on the estimation process.

Consider the three estimates of equation (2) estimating preference for each of the three centers. To the extent that the model fails to account for an individual's preference for shopping center "B", it probably fails similarly to account for that individual's preference for centers "P" and "D" due to factors influencing that individual's preference not included in the model. Similarly, to the extent that the error term for an individual in one of the three preference equations is small, error terms for the other two equations are likely to be small, since the model has captured to a large extent the variables salient to that individual's preference function. Thus, positive covariance is expected between the j error vectors of equation (2).

Since the three error vectors of equation (3) represent mis-estimation of dependent variables whose sum is constrained equal to one, overestimation of patronage share for one center may lead to underestimation of patronage share in the other equations (although \hat{Y}_{2j}^* is not constrained to equal 1.0). As a result, negative covariance is expected among the three error vectors estimated through equation 2. In situations where there is non-zero covariance among the error terms and the variables in the equations are not all equal, the technique developed by Zellner, Seemingly Unrelated Regressions (SUR), can produce estimates with smaller variance than OLS estimates (Johnston 1972), and is thus employed in estimating equations (2) and (3).

Results

Table 3 contains the standardized coefficients corresponding to the parameters to be estimated in equations (2) and (3). Tests for the equality of regression coefficients for the three centers reject the hypothesis of equality at or below the .05 level (Johnston 1972); the estimation of an aggregate model would be inappropriate.

TABLE 3
Regression Results (Standardized)

Predictor	B		P		D	
	Preference	Patronage Share	Preference	Patronage Share	Preference	Patronage Share
Atmosphere	.27 ^a		.29 ^a		.15 ^b	
Personnel	.20 ^a		.15 ^a		.09	
Fashion	.21 ^a		.29 ^a		.24 ^a	
Advertising	.12 ^b		-.11 ^c		.09	
Convenience	.02		.07		.03	
Proximity	.05	.06	.00	-.01	-.01	-.09 ^c
Accessibility	.23 ^a	.20 ^a	.09	.17 ^a	.25 ^a	.19 ^a
Preference		.38 ^a		.34 ^a		.42 ^a
R ² d	.46	.38	.33	.36	.34	.51

^aasymptotic t-values for SUR coefficients. None are contradictory to OLS results.

^bp < .01

^cp < .05

^dp < .10

^eR² from OLS estimates

TABLE 4
Correlation Among Error Vectors

	Patronage Equations Above Main Diagonal			Preference Equations Below Main Diagonal		
	B	P	D	B	P	D
B						
P	.19					
D	-.27	.02				

Generally, the model performs acceptably. In particular, it can be noted that accessibility is a more important predictor of preference for the downtown center than for the other three centers. It can also be noted that proximity is a good predictor of neither preference nor patronage when included in a model including perception of accessibility. Also, accessibility generally affects both preference and patronage, indicating that it acts both as a component of preference and as a facilitator in addition to its affect on preference.

As expected, there is a negative covariance of errors across the three models of patronage share. Unexpectedly, however, the contemporaneous error covariance matrix, Σ , for the equations predicting preference contains both negative and positive elements. In particular, as the model overestimates an individual's preference for center B, it underestimates preference for center D (Downtown). Thus individuals in this sample who prefer "B" do not prefer Downtown, and vice versa, while the errors for "B" and "P" positively covary. This seems to indicate a "preference map" dimension with mall type centers at one extreme and downtown at the other, which is independent of the predictors included in this model.

Conclusions

It is anticipated that this paper has raised more questions than it has answered conclusively. This is intended however, because the issues of attribute and patronage measurement, situational specificity, and model specification and estimation must be specifically and empirically addressed if the investigation of mall-choice behavior is to prove fruitful.

Other issues remain which are not addressed in this paper. One such issue is the relationship between the retail store and the shopping area. Surely this relationship is reciprocal, with store image and preference affecting the image of and preference for the shopping area and center in which it is located, while the image,

location, and preference for a shopping area affects the stores located there. The "special store" variable included by Nevin and Houston (1980) is a step in this direction; more work in this area is required.

Similarly, the model tested in this paper presupposes a one-way relationship between perceptions, preference, and behavior. However perceptions and preference are measurable at a point in time, while behavior is measurable only over a given (past) time period in cross-sectional research. This, in addition to cognitive consistency theories, might indicate a reciprocal relationship between behavior and perceptions. Research is currently underway to test this hypothesis.

References

- Bearden, William G. (1977), "Determinant Attributes of Store Patronage: Downtown Versus Outlying Shopping Centers," Journal of Retailing, 53, 29-38.
- Brunner, James A. and Mason, John L. (1968), "The Influence of Driving Time Upon Shopping Center Preferences," Journal of Marketing, 32, 57-61.
- Bucklin, Louis P. (1967), "The Concept of Mass in Intra-Urban Shopping," Journal of Marketing, 34, 32-36.
- Darden, William R. (1980), "A Model of Consumer Patronage Behavior," in Ronald W. Stampfl and Elizabeth Hirschman, eds., Competitive Structure in Retail Markets: The Department Store Perspective, American Marketing Association, 43-52.
- Gentry, James W. and Burns, Alvin C. (1977-78), "How 'Important' are Evaluative Criteria in Shopping Center Patronage?," Journal of Retailing, 53, 73-85.
- Granbois, Donald H. (1977), "Shopping Behavior and Preferences," in Selected Aspects of Consumer Behavior, National Science Foundation, Washington D.C., 259-298.
- Hansen, Michael H. and Weinberg, Charles B. (1979), "Retail Market Share in a Competitive Market," Journal of Retailing, 55, 37-46.
- Hauser, John R. and Koppelman, Frank S. (1979), "Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness," Journal of Marketing Research, 16, 495-506.
- Houston, Franklin S. and Weiss, Doyle L. (1974), "An Analysis of Competitive Market Behavior," Journal of Marketing Research, 11, 151-155.
- Huff, David L. (1964), "Defining and Estimating a Trade Area," Journal of Marketing, 28, 34-38.
- Jain, Arun K. and Mahajan, Vijay (1979), "Evaluating the Competitive Environment in Retailing Using Multiplicative Competitive Interaction Model," in Jagdish Sheth, ed., Research in Marketing - Vol. II, JAI Press, 217-235.
- Johnston, J. (1972), Econometric Methods, New York: McGraw-Hill, 238-241.
- Joreskog, K. C. (1971), "Simultaneous Factor Analysis in Several Populations," Psychometrika, 30, 4, 409-426.
- Kunkel, John H. and Berry, Leonard L. (1968), "A Behavioral Conception of Retail Image," Journal of Marketing, 32, 21-27.
- Lindquist, Jay D. (1974-75), "Meaning of Image," Journal of Retailing, 50, 29-37.
- Mahajan, Vijay, Jain, Arun K., Bergier, Michel (1977), "Parameter Estimation in Marketing Models in the Presence of Multicollinearity: An Application of Ridge Regression," Journal of Marketing Research, 14, 586-591.
- Mazis, Michael B., Ahtola, Olli T., and Klippel, R. Eugene (1975), "A Comparison of Four Multi-Attribute Models in the Prediction of Consumer Attitudes," Journal of Consumer Research, 2, 1, 38-52.
- Nevin, John R. and Houston, Michael J. (1980), "Image as a Component of Attraction to Intraurban Shopping Areas," Journal of Retailing, 56, 77-93.
- Reibstein, David J. (1978), "The Prediction of Individual Probabilities of Brand Choice," Journal of Consumer Research, 5, 163-168.
- Stanley, Thomas J. and Sewall, Murphy A. (1976), "Image Inputs to a Probabilistic Model," Journal of Marketing, 40, 48-53.

RETAIL SHOPPING AREA IMAGE: STRUCTURE AND CONGRUENCY
BETWEEN DOWNTOWN AREAS AND SHOPPING CENTERS

Michael J. Houston, University of Wisconsin-Madison
John R. Nevin, University of Wisconsin-Madison

Abstract

Consumer images of competitive retail outlets are important determinants of retail patronage decisions. The structure of image as it applies to retail shopping areas, i.e., downtown areas and shopping centers, is investigated in this paper. The underlying dimensions of shopping area image are identified and their congruency between downtown areas and shopping centers is tested. Their relative strength of association with overall evaluations of shopping areas is tested for downtown areas and shopping centers. Implications for downtown revitalization efforts are suggested.

Introduction

With the advent of major shopping centers as a type of retail institution another level of retail patronage decisions by consumers has emerged, i.e., choices between different shopping clusters or areas. Shopping clusters are represented by heterogeneous groupings of individual stores, i.e., shopping centers and downtown areas. In many communities the popularity of shopping centers has resulted in a decay of the downtown area as a major retail center. Downtown areas have been unable to effectively compete as shopping clusters with major shopping centers in these communities.

The deterioration of downtown areas has not gone without notice in many communities. Efforts to revitalize downtown areas as important retail trade centers have occurred in many cities and will continue in other cities. These efforts should be based in part on improving the capability of the downtown area to compete with shopping centers. Important input into such efforts would be an understanding of the factors that influence retail patronage decisions at the shopping area level and whether these factors operate in the same way for downtown areas as for shopping centers. With such knowledge merchant associations and municipal governments can develop appropriate marketing strategies for downtown areas.

Factors Influencing Retail Patronage Decisions

The bulk of research on retail patronage decisions has focused on choice behavior among individual stores. Two major lines of inquiry characterize this research. One approach involves the so-called gravitational model (Huff 1962) which views retail patronage as a function of store size and distance from the consumer. The gravitational model assumes stores are otherwise similar. Bucklin (1971) and Huff himself (Huff and Blue 1966) recognized that the gravitational model was inadequate if consumers perceived differences between stores on other dimensions.

The second line of inquiry, which recognizes store differences beyond size and distance, has focused on store image. Store image is the complex of a consumer's perceptions of a store on functional attributes (e.g., assortment of goods offered, price level, physical layout, etc.) and emotional attributes (e.g., perceived clientele, atmosphere, etc.). Research findings (e.g., Doyle and Fenwick 1974, Stanley and Sewall 1976) have shown that store image is significantly related to store choice.

Although to a much lesser extent than at the individual store level, both lines of inquiry have been extended to research on retail patronage decisions at the shopping area level. A few studies have confirmed the validity of size and/or distance as determinants of shopping center patronage (Brunner and Mason 1968, Bucklin 1967, Cox and Cooke 1970). Other studies (Bellenger, Robertson, and Greenberg 1977, Gentry and Burns 1977) have confirmed the importance of image-like variables in shopping center patronage. The focus in all of these studies was primarily on shopping center patronage decisions. They did not incorporate downtown areas into the design. A study that did incorporate both downtown areas and shopping centers (Bearden 1977) examined decisions at the individual store levels rather than at the shopping area level. It did confirm the importance of image in discriminating between patrons of a downtown department store and those of shopping center department stores.

Problem Definition

The research on retail patronage decisions at the shopping area level suggests that the gravity variables of size and distance and the affective variable of image can explain consumer shopping behavior with respect to downtown areas and shopping centers. This set of variables provides a framework then for determining the focus of efforts to revitalize downtown areas as major retail centers within communities. Given the nature of the gravity variables versus the image variable, it would seem that the latter offers the most practical focus for downtown revitalization efforts. Size and, especially, distance are minimally, if at all, controllable. Image, on the other hand, is largely a result of marketing efforts and the physical nature of a shopping area within given geographic boundaries. The proper manipulation of the image of downtown retail areas can enhance the competitiveness of these areas as major shopping areas. Therefore, the purposes of this paper are:

1. To investigate the nature of consumer images of shopping areas by determining the key underlying dimensions of image structure;
2. To determine if the structure employed by consumers to form shopping area image is consistent between downtown areas and shopping centers.

The second purpose is particularly important. If the structure of consumer image for downtown areas is unique then the basis for the development of image-oriented marketing strategies must differ from that of shopping centers. On the other hand, if they are similar, then downtown areas will have a common structure from which to develop image-based competitive strategies.

The possibility that downtown areas differ from shopping centers in terms of image structure would seem significant. Shopping centers are developed as integrated, self-contained shopping areas. They are typically managed by a single firm who is responsible for the promotion of an overall center. Downtown areas are much older. An evolutionary process characterizes the manner in which many of them have reached their present state.

They are a conglomeration rather than an integration of individual stores.

TABLE 1

Shopping Center Image Items

Methodology

To address the issues relating to consumer images of shopping areas a study was conducted in the Madison, Wisconsin SMSA which contains approximately 300,000 people. Five major intraurban shopping areas exist in the SMSA, four regional shopping centers and a downtown area.

Sampling

A probability sample of 2,000 households was chosen systematically from the telephone directory. Identical five-page self-administered questionnaires were mailed to sample households. A total of 827 questionnaires were returned for a response rate of 41.4%. While no direct assessment of nonresponse error was made, a comparison of the demographic profile of sample households to that of the total population suggested that older, lower-income households, a segment of the market not typically pursued by major shopping areas, were under-represented in the sample.

Development of an Image Measure

Retail shopping area image is a composite of beliefs held by a consumer on a number of dimensions. Several authors have investigated the dimensions used by consumers to form images at the retail level. For example, Lindquist (1974) in his summary of the store image literature, synthesized the frameworks of 19 studies into a set of nine categories that he called image/attitude attributes: merchandise; service; clientele; physical facilities; convenience; promotion; store atmosphere; institutional and post-transaction satisfaction. Bearden (1977) identified seven specific salient attributes of stores: price level; quality of merchandise; selection; atmosphere; location; parking facilities, and friendliness of sales people. Based on these studies and discussions with shopping center managers, a total of 16 items were generated to represent the domain of shopping area image.

The sixteen items were incorporated into a measure of consumer perceptions of each shopping area. Each item was measured on a 5-point modified semantic differential rating scale. The semantic differential format was used because it is easy to self-administer, assumes minimum verbal skills on the part of the respondents, is relatively reliable, and has been common in past research on image. See Table 1 for a listing of the 16 items and descriptions of the anchor points for each scale. Using this format respondents were asked to evaluate each of the five shopping areas with which they were familiar.

Data Analysis

In order to determine the underlying dimensions of the structure of shopping area image responses to the image measure were factor analyzed using principal components analysis and varimax rotation for each shopping area. In order to assess the stability of image structure across downtown areas and shopping centers a factor congruency test (Harman 1967, pp. 271-2) was performed for each pair of shopping areas. The result of a factor congruency test is a congruency coefficient which relates each factor of one factor matrix to each factor of another matrix. Congruency coefficients have properties similar to those of correlation coefficients. Corresponding factors are those with coefficients approaching one, while noncorresponding factors are indicated with small coefficients. Thus, stability across two factor matrices is indicated by a matrix of congruency coefficients with

<u>Item</u>	<u>Anchor Descriptors</u>
Quality of Stores	high-low
Variety of Stores	excellent-poor
Merchandise Quality	excellent-poor
Product Selection	excellent-poor
General Price Level	fair-unfair
Special Sales/Promotions	attractive-unattractive
Layout of Area	convenient-inconvenient
Parking Facilities	adequate-inadequate
Availability of Lunch/Refreshments	adequate-inadequate
Comfort Areas	adequate-inadequate
Special Events/Exhibits	attractive-unattractive
Atmosphere	friendly-unfriendly
Store Personnel	helpful-not helpful
Easy to take Children	very easy-very hard
Great Place to Spend a Few Hours	agree-disagree
A Conservative Center	agree-disagree

the diagonal values close to one and off-diagonal values that ideally approach zero.

Further insights into the nature of the image of downtown shopping centers were obtained by forming component scores of image based on the factor analysis results. Scores were computed by summing the values of items loading on the relevant factor. As a test of criterion-related validity and a basis for assessing the relative impact of each image component on global evaluations of shopping areas, these scores were regressed in stepwise fashion on a measure of general evaluative feeling (5-point scale ranging from "poor" to "excellent") for each shopping area. Also, comparisons of means for each image component yielded relative performance data for the five shopping areas.

Results

Structure of Image

For each shopping area factors were extracted until eigenvalues less than one were obtained. Factor matrices for each of the five areas are presented in Table 2. Three factors were consistently extracted for each shopping area. Explained variance ranged from 52.0% to 56.7% across the five areas. Factor loadings reveal that each factor contains a relatively unique set of items, allowing for relatively easy interpretation of the underlying dimensions of image structure.

Factor I consists primarily of six items--quality of stores, variety of stores, merchandise quality, product selection, special sales/promotions, and great place to spend a few hours. This factor appears to reflect an overall dimension associated with the perceived assortment of benefits offered to consumers by a shopping area. Factor II is indicated by parking facilities, availability of lunch/refreshments, comfort areas, easy to take children, and, to some extent, layout of area. This factor reflects a dimension related to the facilitative nature of a shopping area, i.e., the features it offers to ease the shopping effort. Factor III consists pre-

TABLE 2

Factor Matrix for Each Shopping Area

Item	Downtown			Center A			Center B			Center C			Center D		
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
Quality of Stores	<u>.73</u>	.15	.10	<u>.76</u>	.29	.13	<u>.81</u>	.14	.06	<u>.78</u>	.05	.18	<u>.82</u>	.14	.06
Variety of Stores	<u>.76</u>	.19	-.18	<u>.81</u>	.26	-.04	<u>.77</u>	.16	.02	<u>.77</u>	.20	.06	<u>.79</u>	.16	.06
Merchandise Quality	<u>.72</u>	-.04	.36	<u>.82</u>	.18	.14	<u>.82</u>	.08	.12	<u>.74</u>	.09	.25	<u>.82</u>	.11	.11
Product Selection	<u>.76</u>	.17	.08	<u>.83</u>	.21	.06	<u>.78</u>	.20	.08	<u>.74</u>	.21	.24	<u>.83</u>	.39	.10
General Price Level	.22	.16	.70	.47	.10	.42	.39	.34	.32	.26	.16	<u>.52</u>	.38	.07	<u>.54</u>
Special Sales/Promotion	<u>.53</u>	.20	.35	<u>.60</u>	.26	.33	<u>.56</u>	.38	.05	<u>.55</u>	.14	.36	<u>.63</u>	.15	.31
Layout of Area	.30	<u>.69</u>	.02	.24	<u>.56</u>	.25	.47	.18	.43	.32	.36	.34	.36	.35	.30
Parking Facilities	.09	<u>.66</u>	.19	.09	<u>.72</u>	-.13	.12	.30	.45	-.06	<u>.79</u>	-.01	.00	<u>.65</u>	.02
Availability of Lunch Refreshments	.51	.32	-.14	.26	<u>.61</u>	.09	.09	<u>.71</u>	-.06	.34	<u>.58</u>	.20	.15	<u>.63</u>	.03
Comfort Areas	.09	<u>.65</u>	.24	.24	<u>.69</u>	.14	.06	<u>.74</u>	.05	.20	<u>.65</u>	.26	.21	<u>.71</u>	.08
Special Events/Exhibits	.39	<u>.50</u>	-.14	<u>.53</u>	.37	.24	.43	.47	.13	.42	.09	<u>.54</u>	.49	.34	.26
Atmosphere	<u>.55</u>	<u>.50</u>	.03	.46	.35	<u>.57</u>	.42	<u>.55</u>	.29	.39	.13	<u>.69</u>	.45	.32	<u>.53</u>
Store Personnel	<u>.51</u>	.28	.18	.33	.38	<u>.56</u>	.38	.48	.30	.31	.10	.68	.34	.31	<u>.56</u>
Easy to Take Children	.58	<u>.75</u>	.07	.30	<u>.59</u>	.29	.22	.47	.47	.13	.38	.61	.29	<u>.52</u>	.41
Great Place to Spend a Few Hours	.56	.43	-.18	<u>.59</u>	.29	.34	.41	.44	.19	.45	.15	.49	<u>.55</u>	.24	.38
Conservative	-.11	.07	<u>.72</u>	-.04	-.04	<u>.70</u>	-.10	-.17	<u>.80</u>	-.31	.02	.48	-.14	-.09	<u>.70</u>
Eigenvalue	5.6	1.5	1.4	6.8	1.2	1.1	5.8	1.4	1.1	5.9	1.4	1.1	6.4	1.4	1.1
Cumulative % Variance	34.9	44.2	52.7	42.7	50.0	56.7	36.4	45.1	52.0	37.0	45.7	52.4	38.4	46.9	53.6

dominantly of four items--general price level, atmosphere store personnel, and conservative. This factor is less consistent across areas but appears to be associated with the market posture or positioning of the area as an integrated complex of stores.

Stability of Structure Across Areas

While visual examination of the factor matrices across the five shopping areas suggests a reasonable level of stability, more meaningful evidence is provided by the factor congruency test. The results of the congruency test for each pair of shopping areas is provided in Table 3.

The benefit-assortment factor and the facilitative factor are both highly congruent across the five shopping areas. All but one of the diagonal coefficients involving Factors I and II are above .80 and the majority are .90 or greater. Thus, it appears the first two dimensions of shopping area image are consistent between downtown and shopping centers.

Turning to Factor III--market posture/position--the stability between downtown and shopping centers disappears. Congruency coefficients for Factor III between downtown and each of the shopping centers are low, all of them being under .50. Furthermore, the congruency of Factor III between the four shopping centers, while not as strong

as for the first two factors, is relatively high. Thus market posture/position does seem to be a significant dimension of image structure in the perceptions of shopping centers but does not distinctly emerge in the case of the downtown area.

This finding is reinforced when the regression analyses of the impact of the three components of image on overall evaluations of each shopping area are considered. In all five cases the measures of the three components accounted for a significant ($p < .001$) amount of explained variance. R^2 values ranged from 45% to 50% with a mean of 47%. When the relative contributions of each component are considered, a distinct pattern emerges. For all five areas the assortment component entered the equation first and was most strongly associated with the general evaluation of the area. For the four shopping centers the market posture/position factor was second in impact, entering the equation second in all four cases. However, for the downtown area it was the weakest of the three components, entering the equation last and exhibiting the smallest standardized beta coefficient. Thus, the market posture/position component plays a more significant role in forming evaluations of shopping centers than downtown areas.

In Table 4 mean values for each of the image components and overall evaluations for each of the five areas are provided. It is evident that of the five areas the

TABLE 3

Factor Congruency Coefficients

		Downtown			Center A			Center B			Center C		
		I	II	III	I	II	III	I	II	III	I	II	III
Downtown	I												
	II												
	III												
Center A	I	.97	.64	.66									
	II	.60	.94	.65									
	III	.40	.28	.46									
Center B	I	.97	.68	.49	.98	.61	.61						
	II	.57	.86	.84	.67	.90	.77						
	III	.41	.33	.46	.51	.70	.76						
Center C	I	.98	.50	.70	.99	.50	.71	.98	.68	.45			
	II	.55	.86	.77	.59	.93	.74	.43	.80	.66			
	III	.33	.32	.42	.61	.38	.96	.62	.78	.85			
Center D	I	.88	.53	.71	.90	.51	.72	.90	.45	.71	.88	.46	.68
	II	.55	.90	.69	.63	.98	.57	.52	.73	.70	.44	.94	.43
	III	.35	.23	.48	.64	.54	.90	.62	.72	.90	.73	.69	.86

downtown area consistently performs at the worst level for the three components and in an overall sense. The greatest discrepancy occurs with respect to the facilitative component of image.

TABLE 4

Mean Values for Each Shopping Area

Variable	Downtown	Center A	Center B	Center C	Center D
Benefit Assortment	13.01	17.73	14.36	15.03	17.44
Facilitative Component	12.48	18.47	17.51	17.33	18.19
Market Posture/Position	10.93	11.75	12.00	12.01	11.49
Overall Evaluation	1.75	2.95	2.59	2.67	3.03

Implications

At first thought the implications of the above findings would be that the dimensions on which shopping area image is formed for a downtown area are different than those for shopping centers. Therefore, the marketing strategies for a downtown area would be based on a different structure. However, such a conclusion is probably

erroneous because of the nature and causes of image.

The structure of image, as indicated in the above findings, is based on measures of existing beliefs about the shopping areas. These beliefs have several sources. Beliefs from which the first two factors, benefit assortment and facilitative nature of the center, are formed result in part from behavioral or word-of-mouth experiences with a shopping area. Exposure to or information about the assortment of stores and products and the physical facilities of an area provide for the development of these underlying dimensions. The first two dimensions can be formed in the absence of promotional strategies by a shopping area, although promotion can certainly contribute to their development.

Such is not the case with the third factor. The positioning of a shopping area vis-a-vis other areas is largely the result of promotional efforts. The incongruency between downtown and each center and the congruency across all centers on this factor should not suggest that the positioning dimension is not relevant to downtown image. Rather it suggests that incomplete information from which to evaluate downtown on this dimension is available. Since such information derives largely from promotional efforts, it in turn suggests the absence of a promotional strategy for the downtown area as an integrated retail trade area. Therefore, the successful revitalization of downtown areas depends not only on the right stores carrying the right goods and physical renovations that facilitate shopping efforts but also on a promotional strategy that positions downtown

as an integrated shopping area. It appears that in the case of the community within which this study was conducted the downtown area has failed to do so.

Summary

A considerable amount of previous research suggests that image, as well as size and distance, is an important variable in explaining retail patronage decisions of consumers. Its role appears to be important at both the individual store level and the shopping area level. Its importance at the shopping area level suggests that image should be considered in efforts to revitalize downtown areas as major retail centers.

A study which examines the nature of consumer images of major shopping areas has been reported. Measures of consumer perceptions of a downtown area and four shopping centers on 16 image items were factor analyzed to reveal the underlying structure of consumer images of shopping areas. Three major dimensions emerged: the assortment of benefits offered by an area, the facilitative nature of an area, and the market posture assumed by an area.

Factor congruency tests were performed to determine the stability of image structure between downtown and shopping centers. Results suggested that the first two dimensions were quite stable across all five areas. The third dimension exhibited stability across the four centers but was not stable in comparisons between downtown and each center. Regression analyses of consumer perceptions of the three dimensions on global evaluations of the five areas provided further confirmation by showing that the third dimension was the weakest of the three for the downtown area but second strongest in its association with general feelings about each of the four shopping centers.

These results, while limited to one community, suggest that downtown areas may be deficient in promoting themselves as integrated shopping units. Successful revitalization of downtown areas depend upon promotional strategies that clearly position them within the retail environment. A failure to do so will hinder consumer abilities to form impressions of downtown areas on a major dimension of image, a key variable in explaining consumer retail patronage decisions. Therefore, downtown merchant associations and/or municipal governments, as unifying agents, should develop promotional efforts for the downtown area beyond those of individual stores.

References

Bearden, William O. (1977), "Determinant Attributes of Store Patronage: Downtown Versus Outlying Shopping Areas," Journal of Retailing, 53, 15-22.

Bellenger, Danny N., Robertson, Dan H., and Barnett A. Greenberg (1977), "Shopping Center Patronage Motives," Journal of Retailing, 53, 29-38.

Brunner, James A. and Mason, John L. (1968) "The Influence of Driving Time Upon Shopping Center Performance," Journal of Marketing, 32, 57-61.

Bucklin, Louis P. (1967) "The Concept of Mass in Intra-Urban Shopping," Journal of Marketing, 31, 37-42.

_____ (1971) "Retail Gravity Models and Consumer Choice: A Theoretical and Empirical Critique," Economic Geography, 47, 489-97.

Cox, William E. and Cooke, Ernest F. (1974), "Other Dimensions Involved in Shopping Center Performance," Journal of Marketing, 34, 12-17.

Doyle, Peter and Fenwick, Ian (1974) "How Store Image Affects Shopping Habits in Grocery Chains," Journal of Retailing, 50, 39-52.

Gentry, James W. and Burns, Alvin C. (1977) "How Important Are Evaluative Criteria in Shopping Center Patronage," Journal of Retailing, 53, 73-86.

Harman, Harry H. (1967), Modern Factor Analysis, 2nd ed., Chicago: University of Chicago Press.

Huff, David L. (1962), Determination of Intra-Urban Retail Trade Areas, Los Angeles: University of California Real Estate Research Program.

_____, and Blue, Larry (1966), A Programmed Solution for Estimating Retail Sales Potential, Lawrence: University of Kansas Center for Regional Studies.

Lindquist, Jay D. (1974) "Meaning of Image: A Survey of Empirical and Hypothetical Evidence," Journal of Retailing 50, 29-38.

Stanley, Thomas J. and Sewall, Murphy (1976) "Image Inputs to a Probabilistic Model: Predicting Retail Potential," Journal of Marketing, 40, 48-53.

DEVIANCE AND DISSATISFACTION: AN EXPLORATORY STUDY

Michael K. Mills, University of Southern California

Abstract

This paper reports the results of an exploratory study which tested the notion that the power of a retail store would directly influence both the level of consumer dissatisfaction and the incidence of deviant consumer acts occurring in that store. The results support the hypothesized relationships.

Introduction

The recent wave of consumerism and the available literature indicate that consumer dissatisfaction is quite widespread. Much of this dissatisfaction occurs at the retail level (cf. Valle and Wallendorf 1977, Miller 1976, Kraft 1977, Krishnan and Mills 1978, Bearden and Mason 1979), and may subsequently be directed at retail stores in the form of various "deviant consumer behaviors" occurring in the retail context (cf. Mills 1979 for a review of the deviant consumer behavior problem).

The research presented in this paper utilized the notions of power theory as these may apply to explaining the link between consumer dissatisfaction and the incidence of deviant consumer behavior in retail stores. The basic notion that was tested in this study was that the power of a retail store (as perceived by the customer) would directly influence both the level of consumer dissatisfaction with that store and the incidence of deviant consumer behavior occurring therein.

Evidence from previous empirical work would tend to support the contention that the power context (image) of a store influences the degree of customers' deviant acts. In several studies (Kraut 1976, Robin 1963) highly ranked reasons given for shoplifting included the fact that the person wanted to get back at the store, had political motives, or felt shoplifting was acceptable.

Power is implicated further by a second factor that shows that most shoplifting and other deviant consumer behavior is done by juveniles (Cohen and Stark 1974), the "weaklings" of adult society. Further, rebellion, alienation, and powerlessness are important reasons given by apprehended juveniles for their deviant behavior (Pedrini 1972, Thall 1973).

The effects of store image and attitudes held toward a store as they affect the level of consumer satisfaction have also been explored, though not with respect to either deviance or power preceptions (cf. Swan 1977), but the extrapolation is a fairly intuitive one.

Thus, based on the literature and a seemingly sensible theoretical extrapolation, this study was an attempt to answer the following hypotheses.

H₁: The greater the customer's perceptions of power differences between store and customer, the greater will be the incidence of deviant consumer behavior as a counterpower move by consumers.

H₂: The greater the consumer's perceptions of power differences between store and consumer, the lower will be the level of consumer satisfaction.

H₃: There will be a significant relationship between the

level of consumer satisfaction with regard to retail power contexts and to the incidences of deviant consumer acts occurring in those stores.

Methodology

A usable sample of 153 retail shoppers, gathered through mall and similar interventions, responded to a written survey instrument which featured seven scenarios of retail stores representing one or another type of retail outlet, randomly arranged, and which corresponded to retail power types as classified by the Bonoma (1976) power theory. (See Mills and Bonoma 1979 for pretests of the methodology.)

The Bonoma (1976) theory identifies three prototypical power systems that occur in everyday socio-cultural contexts, and which form a kind of "power continuum." The unilateral power system is characterized by a strong versus weak relationship among the interactive partners, by its focus on target compliance, and by the existence of conflict over scarce resources or mutually exclusive behavioral preferences. The mixed or bargaining power system falls between the unilateral and the group welfare (bilateral) power systems. Bargaining as an influence mode is used to determine outcomes, because force or rewards are not generally or effectively used between equals. The bilateral or group welfare system is characterized by a kind of social altruism as the determining mechanism of exchange.

Thus a "power continuum" ranging from high (unilateral) to low (group welfare) is formed by these three social systems, which seem to have their equivalents in retail settings--as operationalized in the scenarios. For example, the scenario for store A--a unilateral department store, read as follows:

Store A is a large, "prestige" department store in the downtown area. It occupies 200,000 square feet and nine floors of its own building. Store A aims its marketing efforts primarily at the upper-middle income ranges. Its pricing structure is rather high; it is "known" for its clothes and furniture. Significant sales and discounts are only infrequently offered. Store A has its own credit cards and maintains a strict credit policy. Parking is difficult; a parking rebate is offered to buying customers.

The scenario for store A, like all scenarios, was accompanied by a 5 x 7 inch black and white photograph of the store interior.

For each scenario, respondents were asked to anonymously self-report both the extent and types of their own deviant consumer behavior within these various retail power contexts, and were asked as well to report how likely they felt it was that relevant others would engage in these acts.¹ Respondents were also asked to indicate their

¹Several previous researchers have submitted convincing evidence for the use of the self-report methodology in deviance studies (cf. Hirschi 1972, p. 57-64, Short and Nye 1957, Wilkes 1978). Projective "third person techniques" (Tull and Hawkins 1976) have a long history of use in marketing studies (cf. Haire 1950, Hill 1968).

general level of satisfaction/dissatisfaction at the retail level for each of the store "power types" and completed a full demographic profile.

Measures of Interest and Operationalization

Measurement of the dependent variables of perceived store power and the incidence of deviance was essentially identical to that used in earlier pilot tests (Mills and Bonoma 1979). Each scenario was followed by a list of fifteen questions, broken up into three types. The store-power manipulation check section asked five questions about the perceived power of the store depicted in the scenario as viewed by the subject. Following from the power theory base, the five questions asked whether the store made too much money, whether the store was very powerful relative to the respondent, whether there was a lot of conflict between the store depicted in the scenario and the respondent, whether the respondent could influence what was stocked in this type of store, and if the respondent had been treated wrongly in that store, whether he or she could effectively communicate with someone about it. The coding scheme for the manipulation check was a Likert scale varying from one to five in the theoretically dictated direction, except for the questions pertaining to whether the subject could influence what was stocked in the store, or could communicate with someone if they were treated wrongly in the store context, which were reverse scored. The five manipulation check measures were added to get a summary index for each respondent on each scenario.

The remaining questions were concerned with the incidence of five deviant consumer acts, including destroying or damaging merchandise, fraudulent merchandise returns, shoplifting, vandalism, and fraudulent complaints. The questions were asked in two different ways. One set of questions asked whether the subject had engaged in these actions. The other set of questions asked how likely subjects thought it would be that "someone" would engage in the behaviors so described. Both sets of "deviance" questions used a five-point Likert scale which ranged from (1) very unlikely (never) to (5) very likely (always).

These "self" and "other" questions were included for an express purpose. It was felt that subjects' self-reporting of deviance would likely be understated. On the other hand, it was regarded as quite likely that subjects would endorse that a relevant other might engage in such behaviors. The "self" questions described above were combined to form a summary index variable which could range from five (all "never") to 25 (all "always"). The "someone" questions were also summed to form a composite variable varying from five (all "very unlikely") to 25 (all "very likely").

The measure of consumer satisfaction was the subject's responses to questions inquiring about the level of satisfaction from shopping at each of the store types, and was measured on a five point Likert scale ranging from five (always satisfied) to one (never satisfied) for each store.

Results

Frequency of Deviant Acts

Table 1 shows, for each of the five deviant acts investigated in the study, the average percentage of sample respondents who self-reported that they always, very often, sometimes, frequently, or never commit these acts. As is evident from Table 1, the relative percentages of individuals in each category are quite consistent across acts. Examination of the table shows further that the percentage of individuals who always or very often engage in these deviant acts is quite small, although it should

be noted that the average percentage figures shown here are quite consistent with previous sampling estimates of the percentages of individuals engaging in each activity.

Table 2 shows the percentage breakdowns for dissatisfied consumers. As is evident from Table 2, the average percentage of dissatisfied consumers who engage in each of the deviant acts is higher than that for the entire sample population.

TABLE 1
Percent of Sample Population Engaging
in Specific Deviant Consumer Acts

I. Destroy-Damage Merchandise					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
1.7	1	4.8	3.3	89	
II. Fraudulent Returns					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
1	1.5	5.5	4.5	87.5	
III. Shoplifting					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
1.2	1.2	3.6	3.3	90.6	
IV. Vandalism					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
1	.5	3.5	2.2	92.6	
V. Fraudulent Complaints					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
1.1	2	7	4.4	85.3	

TABLE 2
Percent of Dissatisfied Respondents Engaging
in Specific Deviant Consumer Acts

I. Destroy-Damage Merchandise					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
3.8	3.2	10.1	4.8	78.1	
II. Fraudulent Returns					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
4.2	6.3	6.3	3.2	80	
III: Shoplifting					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
4.3	3.3	16.1	2.8	73.5	
IV. Vandalism					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
1.8	1.0	3.5	2.8	90.9	
V. Fraudulent Complaints					
<u>Always</u>	<u>Very Often</u>	<u>Sometimes</u>	<u>Seldom</u>	<u>Never</u>	
4.5	6.3	13.6	4.3	71.3	

Manipulation Check

Table 3 shows the results for the manipulation check (power variable). A repeated measures analysis of variance run, with Scheffe (1953) comparisons, was done on the data. As shown in Table 3, the manipulation check was

successful. The overall F-test computed on the summary of the five manipulation check questions yielded a highly significant value of 84.45, reliable at the .001 level. Individual Scheffe (1953) comparisons--made considering all store types, department stores only, and across different store types--were all significant at the .01 level. (The "Big Chicken" test, also successful at the .01 level, pitted two equally large but "power different" department stores against each other to test an alternative notion, suggested by a colleague that store "size" and not power was the important dimension.) Thus the manipulations were all significant in the direction predicted by the theory. That is, consumers' perceptions of unilateral, bargaining, and bilateral store scenarios was such that they represented a "power continuum" ranging from high (unilateral) to low (bilateral).

TABLE 3
Results on Manipulation

Planned Comparisons. F (6,912) = 84.45 (p<.0001)

- Overall:

Unilateral = Scenario $\overline{A + G}$	= 16.803	.01
Bargaining = $\overline{C + E}$	= 14.513	.01
Bilateral = $\overline{D + F}$	= 11.156	.01
- Within Department Stores only:

Unilateral = Scenario $\overline{A + G}$	= 16.803	.01
Bargaining = Scenario E	= 14.091	.01
Bilateral = Scenario F	= 11.477	.01
- Across Retailers:

Unilateral = Scenario $\overline{A + G}$	= 16.803	.01
Bargaining = Scenario C	= 14.934	.01
Bilateral = Scenario D	= 10.836	.01
- "Big Chicken" Theory:

Unilateral = Scenario $\overline{A + G}$	= 16.803	.01
Mixed Big = Scenario B	= 14.542	

Hypothesis One

Tables 4 and 5 show the results for the levels of deviance reported that formed the basis for testing Hypothesis One. To save space, only the "someone" (Table 4) results will be discussed here (the "self" results, while not as strong as the "someone" results, generally confirm the hypothesis in question). As shown in Table 4 the overall F-test yielded a highly significant value of 55.19, reliable at the .0001 level. Planned comparisons made for all stores considered, within department stores only, and across retail outlets were all significant in the direction predicted by the theory at the .01 level, with the exception of the across retailers-bargaining versus bilateral comparisons. The "Big Chicken" comparison, while in the direction predicted by the theory, was nonsignificant.

Thus, Hypothesis One was confirmed. The data showed that more powerful stores, as measured by the composite store power index, were felt likely to have a greater incidence of deviant consumer acts committed in them than less powerful stores. We turn now to the test of Hypothesis Two.

TABLE 4
"Someone" Results

Planned Comparisons. F (6,912) = 55.19 (p<.0001)

- Overall:

Unilateral = Scenario $\overline{A + G}$	= 16.709	.01
Bargaining = $\overline{C + E}$	= 13.751	.01
Bilateral = $\overline{D + F}$	= 11.362	.01
- Within Department Stores only:

Unilateral = Scenario $\overline{A + G}$	= 16.709	.01
Bargaining = Scenario E	= 15.261	.01
Bilateral = Scenario F	= 10.954	.01
- Across Retailers:

Unilateral = Scenario $\overline{A + G}$	= 16.709	.01
Bargaining = Scenario C	= 12.241	.01
Bilateral = Scenario D	= 11.771	ns
- "Big Chicken" Theory:

Unilateral = Scenario $\overline{A + G}$	= 16.709	ns
Mixed Big = Scenario B	= 16.130	

TABLE 5
"Self" Results

Planned Comparisons. F (6,912) = 11.82 (p<.0001)

- Overall:

Unilateral = Scenario $\overline{A + G}$	= 6.869	.05
Bargaining = $\overline{C + E}$	= 5.905	.01
Bilateral = $\overline{D + F}$	= 5.493	ns
- Within Department Stores only:

Unilateral = Scenario $\overline{A + G}$	= 6.751	ns
Bargaining = Scenario E	= 6.104	.01
Bilateral = Scenario F	= 5.457	ns
- Across Retailers:

Unilateral = Scenario $\overline{A + G}$	= 6.751	.01
Bargaining = Scenario C	= 5.705	.01
Bilateral = Scenario D	= 5.529	ns
- "Big Chicken" Theory:

Unilateral = Scenario $\overline{A + G}$	= 6.751	ns
Mixed Big = Scenario B	= 6.392	

Hypothesis Two

Table 6 shows the results for the test of Hypothesis Two. Once again a repeated measures analysis of variance was performed on the data, with the resulting overall F-test

yielding a highly significant value of 9.86, reliable at the .001 level. Planned comparisons performed on the means yielded somewhat mixed results, however.

TABLE 6
Results on Satisfaction

Planned Comparisons. $F(6,912) = 9.86$ ($p < .0001$)

1. Overall:		
Unilateral = Scenario $\overline{A + G}$ = 2.63	.05	
Bargaining = $\overline{C + E}$ = 2.379		ns
Bilateral = $\overline{D + F}$ = 2.562		ns
2. Within Department Stores only:		
Unilateral = Scenario $\overline{A + G}$ = 2.63	.05	
Bargaining = Scenario E = 2.882		.01
Bilateral = Scenario F = 3.47	.01	
3. Across Retailers:		
Unilateral = Scenario $\overline{A + G}$ = 2.63	.01	
Bargaining = Scenario C = 1.876		.01
Bilateral = Scenario D = 1.654		ns
4. "Big Chicken" Theory:		
Unilateral = Scenario $\overline{A + G}$ = 2.63	.01	
Mixed Big = Scenario B = 3.046		

As shown in Table 6, for the overall comparisons one significant relationship was noted, but it is a direction contrary to the hypothesized relationship. However, significant results were seen when consumers' mean levels of satisfaction were considered with respect to department stores only. Here, the direction of all comparisons was in harmony with the hypothesized relationship and was statistically significant as well. Across retailers, however, two significant, but directionally not predicted, results were seen. Finally, the "Big Chicken" test was significant at the .01 level in the direction predicted by the theory. Additionally, it is quite apparent that, across all store types, customers' mean levels of satisfaction with these stores was quite low.

Thus, the test of Hypothesis Two yielded somewhat mixed results. A possible explanation for these results may be rooted in the nature of the comparisons made, as well as in the nature of the store types themselves. That is, significant findings in the direction of the hypothesized relationships were noted in both the within department stores only and the "Big Chicken" tests. These comparisons were of similar types of retail outlets (i.e., department stores), rather than of differing types, and hence these comparisons may be more indicative of the phenomena under investigation. It remains for additional research to resolve this issue.

Hypothesis Three

Table 7 shows the results for the correlational analysis that was used as the statistical test of Hypothesis Three. As shown in Table 7, there is the expected negative relationship between the level of shopping satisfaction in each of the stores and the incidence of deviant acts committed in those stores. However, the size of the correlation coefficients is quite small, albeit statisti-

cally significant in three instances, and hence some caution is required in interpreting these results.

TABLE 7
Correlations Between Shopping Satisfaction
in Store and Deviance in Store

Store						
A	B	C	D	E	F	G
-.1441	-.2031	.0257	-.0719	-.1437	-.1259	-.0701
(.038)	(.006)	(.376)	(.189)	(.038)	(.061)	(.195)

Conclusion

This study was an attempt to examine the possible relationship between perceived power of a retail store, consumer dissatisfaction, and deviant consumer behavior in retail stores. The results seem to indicate that stores perceived as more powerful by the consumer have higher levels of deviant consumer behavior occurring in them and, at least in some instances, have lower reported levels of consumer satisfaction than less powerful stores. Further, it appears that there is a relationship between the level of consumer satisfaction with a store and the incidence of deviant consumer behavior committed against that store.

While some caution is necessary in interpreting these results given the exploratory nature of the study (and because of the small sizes of the correlation coefficients) the results are, nonetheless, interesting and point the way towards additional research in this area. This is particularly true with regard to further tests of the importance of power perceptions on consumer dissatisfaction. While some attempt has been made by Valle and Wallendorf (1977) and others to bring theory to the question of consumer dissatisfaction and complaining behavior, it appears likely that power theory may shed some new light in this area (as well as on the question of consumer deviance). It is the author's hope that this study will serve as an impetus to more research in these and related areas.

References

- Bearden, W. O. and Mason, J. B. (1979), "Elderly Use of In-Store Information Sources and Dimensions of Product Satisfaction/Dissatisfaction," *Journal of Retailing*, 55(1), 79-91.
- Bonoma, T. V. (1976), "Conflict, Cooperation and Trust in Three Power Systems," *Behavioral Science*, 21, 499-514.
- Cohen, L. and Stark, R. (1974), "Discriminatory Labeling and the Five Finger Discount," *Journal of Research in Crime and Delinquency*, 2(1), 25-39.
- Haire, M. (1950), "Projective Techniques in Marketing Research," *Journal of Marketing*, 14, 649-656.
- Hill, C. R. (1968), "Haire's Classic Instant Coffee Study --18 Years Later," *Journalism Quarterly*, 45, 466-472.
- Hirschi, T. (1972), *Causes of Delinquency*, Berkeley, CA: University of California Press, 57-64.
- Kraft, F. B. (1977) "Characteristics of Consumer Complainers and Complaint and Repatronage Behavior," in *Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, ed. Ralph L. Day, Bloomington: Indiana University, 79-84.
- Krishnan, S. and Mills, M. K. (1978), "Dissatisfaction with Retail Stores and Repatronage Behavior." Paper presented at the Third Annual Conference on Consumer Satisfaction, Dissatisfaction and Complaining Behavior.

Kraut, R. E. (1976), "Deterrent and Definitional Influences on Shoplifting," Social Problems, 23(3), 356-358.

Miller, J. A. (1976) "Store Satisfaction and Aspiration Theory," Journal of Retailing, 52(3), 65-84.

Mills, M. K. (1979), A Power Contextual Analysis of Deviant Consumer Behavior in Retail Stores, Pittsburgh: University of Pittsburgh.

Mills, M. K. and Bonoma, T. V. (1979), "Deviant Consumer Behavior: New Challenge for Marketing Research,"

Proceedings, American Marketing Association, Annual Marketing Educators' Conference.

Pedrini, D. T. (1972), "Psychology and the Problems of Shoplifting." Paper presented at the Midwest Section of the American Business Association Conference.

Robin, G. (1963), "Patterns of Department Store Shoplifting," Crime and Delinquency, 9, 163-172.

Scheffe, H. (1953), "A Method for Judging All Contrasts in the Analysis of Variance," Biometrika, 40, 87-104.

Short, J. F. and Nye, F. I. (1957), "Reported Behavior as a Criterion of Deviant Behavior," Social Problems, 5, 207-213.

Swan, J. E. (1977), "Consumer Satisfaction with a Retail Store Related to the Fulfillment of Expectations on an Initial Shopping Trip," in Consumer Satisfaction, Dissatisfaction and Complaining Behavior, ed. Ralph L. Day, Bloomington: Indiana University, 10-17.

Thall, M. M. (1973), Behavioral Components of Adolescent Shoplifting, Ann Arbor: University of Michigan.

Tull, D. S. and Hawkins, D. I. (1976), Marketing Research: Meaning, Measurement and Method, New York: MacMillan.

Valle, V. and Wallendorf, M. (1977), "Consumer Attributions of the Cause of Their Product Satisfaction and Dissatisfaction," in Consumer Satisfaction, Dissatisfaction and Complaining Behavior, ed. Ralph L. Day, Bloomington: Indiana University, 26-30.

Wilkes, R. (1978), "Fraudulent Behavior by Consumers," Journal of Marketing, 42(4), 67-75.

COMPREHENSIVE IDENTIFICATION OF CONSUMERS' MARKETPLACE
PROBLEMS AND WHAT THEY DO ABOUT THEM

F. Kelly Shuptrine, University of South Carolina
Gerhard Wenglorz, University of South Carolina

Abstract

A representative sample of households in a medium-sized southeastern city were surveyed by personal interviews concerning the last product of any type that the respondent could remember having any type of problem. In contrast to other studies reporting one-fourth to one-third of the respondents having a problem, this survey found 76.3% of the respondents experiencing some recent problem of any type and this percentage increased to 83% when those who said they didn't have any problems were asked a "how sure" question. Significant comparisons of those who had problems/no problems, made a complaint/no complaint, and were successful/unsuccessful complainers are also provided.

Introduction

Numerous studies about consumers and their complaints emerged with the rise of the "consumerism movement" during the last two decades. Such attention helped to result in such acts as the Consumer Credit Protection Act (1968), Poison Prevention Packaging Act (1970), and the Consumer Product Safety Act (1972). Consumer Protection Agencies, Consumer Affairs Offices, Consumer Organizations, and related consumer associations have evolved in this period.

However, our knowledge in the area of consumer problems in the marketplace and their resolution are fragmented. Some studies have dealt with consumer's attitudes and opinions toward the marketplace and others have recorded behavioral information.

Reviewing some recent studies, Barksdale and Darden (1972) note that younger, more liberal subjects were more critical of marketing practices and more impressed with the accomplishments of the consumer movement. In a summary of 912 women consumers (Hustad and Pessemier 1973), it was found that those who felt negative toward business tended to be younger women who are intellectually, socially, and economically upscale.

In their study of consumer dissatisfaction with household appliances Mason and Himes (1974) found 44% of this sample to have experienced dissatisfaction severe enough to register a complaint. The initial retailer was the most effective at satisfying the consumer's discontent resolving 55% of the problems with repair centers following behind at 16%. As many as 18% of the problems reported were still unresolved at the time of the study.

Using Lundstrom's (1976) Consumer Discontent Index, Lundstrom and Kerin (1976) attempted to identify psychological and demographic correlates of consumer discontent. They reported consumer discontent to be statistically related in a positive direction to alienation, especially the powerlessness and social isolation components, anomic, being male, age, income level and occupation level. Consumer discontent existed predominately in older males from the middle to upper middle class and not in the lower socio-economic groups.

A study on elderly reactions to unsatisfactory purchases indicates that propensity-to-complain declines with age (Zaltman, Srivastava, Deshpande 1978). Kraft (1978) in a

paper on characteristics of consumer complainers notes that more complaints are made on large-ticket items, by whites than blacks, and by younger than older respondents.

Fornell and Westbrook (1979) report that those who are more assertive and aggressive are more likely to exhibit complaining behavior. A comprehensive study (Shuptrine, Thomas, Sharma 1979) on the consumer complaint process involving durables found that of those who complained and reported successful resolution, no significant demographic differences between successful/unsuccessful complainers were found. A study by Day and Landon (1976) suggests a more comprehensive model of complaining behavior is needed and provided major input with a national study for the FTC on a large number of product categories cross-classified by specific consumer complaints or complaint categories.

Another group of studies has investigated who complains about the marketplace and the products and services they complain about. In these studies the dependent variable is behavioral, action or inaction, whereas in the first series of studies the dependent variable was an attitude. In another part of their study on a sample of female's dissatisfaction with appliances Mason and Himes (1973) found those who took action tended to be from larger households, had higher incomes, were primarily middle aged and tended to own their own home. Miller (1974) reports that the number of complaints or suggestions for improvements a customer expresses for a store tends to come from younger, more mobile and better educated women. Lielfield, *et al.*, (1975) found that Canadian consumers who had written complaint letters were middle aged, better educated, earned higher incomes, and had professional heads of household. According to Hair, *et al.*, (1976) chronic complainers have negative attitudes toward business and marketing practices. They attend church less frequently, belong to fewer organizations, younger, less educated and have lower incomes.¹

By including both an attitudinal and behavioral dimension in their study Warland, *et al.*, (1975) found that two types of dissatisfied consumers emerge--dissatisfied consumer activists (Upset-Action Group) and dissatisfied passive consumers (Upset-No Action Group). The Upset-Action both reads and participates in consumer and political activities. They are younger, have higher social status, income and education. The Upset-No Action Group have lower incomes, are less educated and do not engage in consumer and political actions. They exhibit low social involvement and are the most politically alienated. The authors feel that the Upset-No Action Group have potential to become an important force in the consumer movement if given leadership, direction and education. They are not a segment business can afford to ignore. Wall, Dickey, Talarzyk (1977) also found that complainers are more upscale in socio-economic status.

¹Many additional studies concerning consumer complaints are available from the four annual conferences on consumer satisfaction, dissatisfaction, and complaining behavior (Day and Hunt 1976-1979). Other research concerning complaining theories, models and correlates are available in previous ACR proceedings. Much of the research findings available are quite similar to the research reviewed in this paper.

On a more comprehensive study of consumer complaints (Andreasen, Best 1977), it was found that one in five purchases of products and services resulted in consumers' dissatisfaction with something other than price, less than half of these problems resulted in a consumer complaint, and about two out of three complainers ended up with a satisfactory resolution of their problem. A study by Shuptrine, Thomas, and Sharma (1979) reported earlier on durables found about 70% of those complaining receiving satisfaction.

In another study (Advertising Age, June 21, 1976) consumer organizations report that one-fourth of all households have some sort of marketplace problem and of these only one in three are actually reported. Where complaints are registered, however, the study finds about 56% of the consumers get satisfaction. Low-income consumers are least likely to complain, while the likelihood that the consumer will complain increases if it is an expensive item, if the problem is clear-cut rather than a matter of judgement, and if it is a big-tag credit transaction.

And finally, a study by the U.S. Office of Consumer Affairs reported in the Marketing News (February 22, 1980, p. 8) on a nationwide sample of 2500 households says:

Most private industries and government agencies are inept when it comes to handling consumer complaints and do an inadequate job of informing the public on how to file and follow through on their protests.

The study found that about one-third of the sample had one or more consumer problems in the last year and about 75% of all complaints were about the poor quality of products and services. Most complaints, consistent with previous research, were directed at the auto industry followed by appliances. A majority of the consumers said their first contact with reporting their complaint was with the retailer. About 75% of the consumers will complain if the value of the product is between \$6 and \$10; whereas, 95% will complain about a product costing \$500 to \$1000.

This brief review of the literature indicates that there are many inconsistencies when comparing studies. Some research only looks at those who complain (who are often different from those who don't), others examine attitudes and behavior of those with problems (though not much persistence is seen in determining whether people have had any type of problem), and others have examined briefly the consumer complaint process and comparisons of successful and unsuccessful complainers. However, when summarizing, inconsistent profiles develop; with some studies saying discontent and complaining are related to high income and education and others report they are related to low income and education and so on.

Because of the fragmented approaches that have been largely used so far, the intent of this study was to provide a more systematic approach to the area of consumer marketplace experiences. Specifically, the study is designed to:

- Probe strongly to identify any type of problem(s) consumers have had in the marketplace.
- See if they had a problem, whether they would attempt complaint resolution.
- Examine the process complainers would go through in attempting resolution.
- Compare profiles of successful and unsuccessful complainers.
- To obtain opinions of non-complainers as to why they didn't attempt resolution.

The results of the study will help to serve as one of the building blocks needed for a comprehensive theory of complaint behavior.

Method

A questionnaire was designed along similar lines as a previous questionnaire used in a consumer durable complaint study. It was pretested and revised twice on small samples (25) of households reflective of the population of interest. Major emphasis was placed on the first question as to whether the respondent had a recent problem (of any size) with products or marketplace activities. To really be confident, a later question for those who said no was a how sure are you question to see if the first question was supported.

A random sample of 275 households was selected from a cross-section of census tracts in a middle-sized south-eastern city. A criss-cross telephone directory was utilized to randomly select residential addresses (business addresses were not included). There were 224 usable questionnaires from the sample of 275 (81.5%) which were obtained from a personal interview with an adult household member at the chosen address.

Results

Overview

Summary responses of the total sample (224 respondents) on some questions and demographics are presented in the Appendix. The most common problems respondents had (consistent with prior research) were automotive related (25.7%) and appliances -- small (10.5%) and large (17.5%). The main types of problems consumers had were: product not functioning as claimed (25.4%), was poorly constructed (39.1%), some problem with the actual functioning of the product (33.7%) and problems with repairs/service (19.5%). The product was under warranty 37.7% of the time when problems arose and 32.9% of the products were never covered with warranties. The dollar value of the products that people had problems with ranged from less than \$5 (17.7%) to over \$500 (28%).

The demographic composition of the sample was widely dispersed by age, profession, education, and income. Twenty-one percent were single and 67.4% were married. There were 37.4% males and 62.2% females and 22.5% blacks and 77.5% whites. From prior research in this metropolitan area, the demographic profile was fairly representative of the entire population.

Complaint Resolution Efforts

Table 1 presents efforts by those who had a problem and complained. This effort could be classified as a first stage consumer complaint process. Most complaints were made to the retailer selling the product (82.3%) and was made by a personal visit (62.8%) or telephone call (38.9%). Most complaints were answered (86.7%) either immediately or within 1-7 days (28.1%). There were 68.4% of the complaints adjusted to the complainant's satisfaction--refund (11.4%), exchange (30%), and repaired with charge (48.6%). A summary of resolution for this first stage complaining was success for 61% of the complainers, complaint not answered for 13.6%, and complaint was answered but not resolved for 25.4%.

Consumers That Have Problems

Several previous studies cited have stated that 1/4 to 1/3 of the sampled populations had a problem. This questionnaire was set up to really check these earlier findings and thus a major emphasis was put on seeing how many of our

Table 1
Efforts by Complainers to Resolve Their Problems (N=113)*

Complaint to Whom	No.	%
Retailer	93	82.3
Manufacturer	11	9.7
Better Business Bureau	3	2.7
Consumer Affairs	3	2.7
Local Repair Office	3	2.7
Other	7	6.2
Total	120	106.3
How Communicated		
Personal Visit	71	62.8
Phone	44	38.9
Letter	10	8.8
Other	1	.9
Total	126	111.4
Answered		
Yes	98	86.7
No	15	13.3
Total	113	100.0
Reply Time		
Immediately	49	51.0
1-7 Days	27	28.1
3-15 Days	4	4.2
16-30 Days	6	6.3
1-12 Months	9	9.4
1 Year and more	1	1.0
Total	96	100.0
Adjusted to Satisfaction		
Yes	67	68.4
No	28	28.6
No Answer	3	3.0
Total	98	100.0
Action Taken		
Refund	8	11.4
Exchange	21	30.0
Repaired with Charge	34	48.6
Repaired without Charge	4	5.7
No Answer	3	4.3
Total	70	100.0
Summary of Effort		
Results		
Complaint, was not answered	15	13.6
Complaint, was not resolved	28	25.4
Success	57	51.0
Total	110	100.0

*Sample and total percentage might change due to multiple responses.

sample have had any type of problem in the recent past. There were 171 out of 224 or 76.3% who stated they had a problem of some type. With a follow-up question, another 15 or an additional 6.7% admitted to some problem. So, in total, our study identified 83% of the respondents having experienced a recent problem (of any magnitude) in the marketplace.

A comparison was made of those who had problems and those who didn't by respondent's demographic characteristics. The significant findings are presented in Table 2. Younger respondents were more likely to express having problems than older ones. The self-employed (93%), white (88%) and blue collar (91%), students (100%), and housewives (98%) were more likely to have problems than the retired (71%) and the unemployed (50%). The higher one's educational background, the more likely they were to have had problems (college graduates 93%); whereas those with less education were less likely to have problems. Whites (85%) were significantly more likely to have problems than Blacks (70%).

Table 2
Comparisons of Those Who Had Problems With Those Who Did Not

Age Group	Distribution According to Age		No.
	Had Problems %	No Problems %	
0-19	100.0	0.0	7
20-24	37.0	13.0	23
25-34	93.2	6.8	44
35-44	85.7	14.3	28
45-54	81.1	18.9	37
55-64	70.0	30.0	30
65+	74.2	25.8	31
			Total 200

χ^2 calculated = 10.5, significant at the .10 level.

	Distribution According to Profession		No.
	Had Problems %	No Problems %	
Self-Employed	92.9	7.1	14
White Collar	87.7	12.3	65
Blue Collar	90.5	9.5	21
Student	100.0	0.0	16
Housewife	97.5	2.5	40
Retired	70.8	29.2	24
Unemployed	50.0	50.0	6
			Total 186

χ^2 calculated = 14.5, significant at less than the .05 level.

	Distribution According to Education		No.
	Had Problems %	No Problems %	
Grade School	60.0	40.0	5
Some High School	69.6	30.4	23
High School Graduate	73.1	26.9	52
Some College	91.4	8.6	35
College Graduate	93.3	6.7	60
Post-Graduate	61.9	38.1	21
Technical College (2 year)	85.7	14.3	7
			Total 203

χ^2 calculated = 16.1, significant at less than the .02 level.

	Distribution According to Race		No.
	Had Problems %	No Problems %	
Black	69.3	30.2	43
White	85.4	14.6	164
			Total 207

χ^2 calculated = 4.6, significant at less than the .05 level.

Consumers That Complain

Comparisons were made between those who have a problem and complain and those who don't on product characteristics and respondent demographics. The significant findings are presented in Table 3. According to product group, clothes and miscellaneous items (small ticket items) were less likely to be complained about, 44% and 50% respectively. Automotive related (82%), furnishings (83%), and house related problems (83%) were product groups most likely to have a complaint made. In line with other studies, the more expensive the product the more likely one was to make a complaint. By profession, the self-employed (77%), white collar employees (77%), and housewives (77%) were more likely to complain. Blue collar workers (37%) and the retired (53%) were much less likely to make a complaint.

Successful Complainers

Significant comparisons between successful and unsuccessful complainers are presented in Table 4. Only two factors were found to be significantly related-profession and educational level of the respondent. The self-employed (80%) and the retired (78%) were more likely to be successful. Blue collar workers were least likely to achieve resolution

Table 3
Comparisons Between Those Who
Complained and Did Not Complain

Product	Distribution According to Product Group		No.
	Problems, Complained %	Problems, Did Not Complain %	
Automotive Related	81.8	18.2	44
Small Appliances	72.2	27.8	18
Major Appliances	76.7	23.3	30
Furnishings	83.3	16.7	6
Food/Household			
Cleaning Items	75.0	25.0	21
Clothes	43.8	56.2	16
House Related	83.3	16.7	6
Miscellaneous	50.0	50.0	30
			Total 171

χ^2 calculated = 20.3, significant at less than the .01 level.

Distribution According to the Value of the Product

	Problems, Complained %	Problems, Did Not Complain %	No.
< \$5	37.9	62.1	29
5 < 10	66.7	33.3	6
11 < 50	61.4	38.6	44
51 < 100	70.0	30.0	10
101 < 300	66.7	33.3	18
301 < 500	100.0	0.0	10
> 500	71.7	28.3	46
			Total 163

χ^2 calculated = 23.4, significant at less than the .001 level.

Distribution According to Profession

	Problems, Complained %	Problems, Did Not Complain %	No.
Self-Employed	76.9	23.1	13
White Collar	77.2	22.8	57
Blue Collar	36.8	63.2	19
Student	56.3	43.7	16
Housewife	76.9	23.1	39
Retired	52.9	47.1	17
Unemployed	66.7	33.3	3
			Total 164

χ^2 calculated = 15.3, significant at less than the .02 level.

(only 14%). Those with some college (68%) and college graduates (76%) were most likely to be successful. Those with some high school training and high school graduates were less successful, 50% and 50% respectively.

Followup on Noncomplainers

An additional question was asked of those who had a problem but didn't complain (see Table 5). About 86% of the non-complainers gave reasons for not complaining of: too much trouble (21%), price too low, not worth the time (30%), won't do any good (26%), and waste of time (9%). All of these seem to indicate a feeling of separation from the seller and the buyer's feelings that they will not get any satisfaction if they complain. When we contrast this view with the results in Table 1 where 61% of the complainers got satisfaction on their first complaint, we feel that the consumer may be giving up too quickly. The business firms may be much more responsive than what many people perceive them to be.

Discussion

A representative sample of households in a medium-sized southeastern city was contacted and answered a detailed questionnaire concerning any problems they have had in the marketplace and what they did or did not do about them.

Table 4
Comparisons Between Successful
and Unsuccessful Complainers

	Distribution According to Profession		No.
	Successful %	Unsuccessful %	
Self-Employed	80.0	20.0	10
White Collar	61.4	38.6	44
Blue Collar	14.3	85.7	7
Student	55.6	44.4	9
Housewife	66.7	33.3	30
Retired	77.8	22.2	9
Unemployed	100.0	0.0	2
			Total 111

χ^2 calculated = 10.8, significant at less than the .10 level.

Distribution According to Education

	Successful %	Unsuccessful %	No.
Grade School	0.0	100.0	1
Some High School	50.0	50.0	6
High School Graduate	50.0	50.0	24
Some College	68.2	31.8	22
College Graduate	76.2	23.8	42
Post-Graduate	50.0	50.0	14
Technical College (2 year)	25.0	75.0	4
			Total 113

χ^2 calculated = 10.6, significant at the .10 level.

Table 5
Reasons Given by Those Who Had a Problem but
Did Not Make A Complaint (N = 53)

Reason for Not Complaining	Number	Percent ³
Too much trouble	11	20.8
Price too low, not worth the time	16	30.2
Won't do any good	14	26.4
Waste of time	5	9.4
Warranty is no good, given the run around	4	7.5
Tried, but business moved	2	3.8
Too much time involved	1	1.9
Miscellaneous reasons	3	5.7
No sales slip anymore	1	1.9
No one feels responsible	1	1.9
TOTAL	53	100.3%

³Percentage greater than 100% due to multiple responses.

The study was systematic in that comparisons were made of those who had problems/no problems, complainers/noncomplainers, and successful complainers/unsuccessful. The complaint process was also examined for those who made a complaint.

Though one cannot generalize these findings to the entire U. S. population, the results are representative of the southeastern area from which they were taken and provide useful points of comparison with other related studies. It is unique in that it is much more comprehensive in that it doesn't examine only those people who complain, or those who are successful complainers, or those with problems, but includes all of these groups.

A major result of this study was the identification of such a large number of people who have had problems in the marketplace. Previous research (Warland, et al. 1975, Advertising Age 1976, Andreasen, Best 1977, Marketing News 1980) have noted between 20 to 35% of their samples with some consumer problem within the last year or less. Our

survey found 76.3% of our respondents experiencing some recent problem of any type and this increased to 83% when those who said they didn't have any problems were asked a "how sure" question. These included very large items such as houses and automobiles down to small items such as cottage cheese, records and bread. Now it is debatable as to whether all studies are comparable, especially when some look only at certain marketplace problems, or ask if you were good and mad, and so on, but it appears to us that there are many more problems than most studies lead us to believe. (For corrective or educational actions, reports that are broken down into more detailed problem/service categories such as Andreasen and Best (1977) would likely be more useful for private or public change agents).

The nationwide study done for the U. S. Office of Consumer Affairs (Marketing News, February 22, 1980) only found about one-third of those questioned to have had one or more consumer problems within the last year. We feel that our sample is definitely not that much different from the U. S. populace. In fact, it is in a rather conservative, pro-business area and one would expect even fewer consumer problems with goods and services.

In examining the complaint resolution process, it was noted that most complaints were directed to the retailer, communicated by personal visit or phone, responded to immediately or within less than 7 days, was answered, and when adjusted to the customer's satisfaction, was with a refund, exchange, or repaired with charges. Most notable was the fact that 61% of those making an attempt to resolve a problem obtained satisfaction. We did not follow-up to see if those who were rebuffed on their first attempt persisted and obtained satisfaction. So, the final group receiving satisfaction is undoubtedly greater than 61%.

The following statements are descriptive of comparisons of those who had problems/no problems, made a complaint/no complaint, and were successful/unsuccessful complainers.

Problems/No Problems

- Those with problems tend to be younger while older respondents have fewer problems.
- The retired and unemployed tend to report fewer problems than other professions.
- College graduates are much more likely to report consumer problems than high school graduates.
- Whites report a greater number of problems than blacks.

Complaint/No Complaint

- Those with automotive related, house related, and house furnishings problems are more likely to complain. Those with clothing problems and small miscellaneous problems are much less likely to complain.
- The more expensive the item in question, the more likely one is to complain.
- The self-employed, white collar workers, and housewives are more likely to be complainers. Blue collar workers and the retired are more likely to be non-complainers.

Successful/Unsuccessful Complainers

- The self-employed and retired are more likely to report satisfaction. Blue collar workers are more likely to be unsuccessful.
- Those with some college and college graduates

are more likely to be successful. Those with some high school and high school graduates tend to be more unsuccessful.

These findings are similar to other studies and are more systematic in that all of these issues are examined in the same study. They do point out areas where substantive educational efforts could be beneficial in improving the consumer's welfare.

And finally, of those who had a problem but did not complain, we note from Table 5 that about 86% of the reasons given for not complaining relate to their perceptions that it was too much trouble, price too low, not worth the time, won't do any good, and waste of time. Now, it very well may not be in the consumer's most economic use of time and travel to worry about a small ticket item. But, for other items that these people feel they won't get any satisfaction on, it appears from this study and other reported studies that they would likely experience a 60-70% chance of satisfaction if they would attempt to resolve their problem.

In conclusion, there appears that a serious review is needed of the existing attitude among many businesses that a relatively small number of complaints received is a measure of general consumer satisfaction. Our findings show that a very large number of problems are encountered in the marketplace by consumers with many unwilling to make a complaint about their problems. Obviously, one way of improving the situation is to better educate consumers and have them become more responsible. Profiles developed in this and previous studies will be useful inputs for designing educational programs. Another way is to use government or consumer organizations as intervenors. A better way might be for business firms to recognize the extended product concept which implies satisfaction after purchase. So, instead of leaving the burden of the complaint process entirely with the consumer the business firm needs to enter the picture and provide adequate mechanisms for communicating and resolving problems.

Appendix Summary Responses of Total Sample

	Number	Percent
I. <u>Distribution By Problem Type</u>		
1. Automobile related	44	25.7
2. Appliances: small	18	10.5
major	30	17.5
3. Home Furnishings	6	3.5
4. Food/Household cleaning items	21	12.5
5. Clothes	16	9.4
6. House related	6	3.5
7. Miscellaneous	30	17.6
	171	100.0
II. <u>Type of Problem with the Product</u> (N = 169)		
	Number ^a	Percent
1. Product did not function (was not sold) as claimed in promotion	43	25.4
2. Product was poorly constructed	66	39.1
3. Some matters involving the actual functioning of the product	57	33.7
4. Product did not fit the user	3	1.3
5. Problem in getting satisfactory servicing or repair on the product when malfunction occurred	33	19.5
6. Outdated	9	5.3
7. Price of service too high	1	.6
8. Miscellaneous	15	8.9
	327	134.3

^aMultiple responses were used.

References

III. Location of Purchase	Number	Percent
1. Furniture Appliance Store	17	9.9
2. Department Store	32	18.7
3. Discount Store	17	9.9
4. Building Supply Store	1	.6
5. Auto Dealers (new or used)	39	22.3
6. Supermarket/Grocery Store	23	13.5
7. Tire Retailers	2	1.2
8. Specialty Clothing Store	4	2.3
9. Other Specialty Store/Firm	36	21.1
	171	100.00

IV. Product Under Warranty	Number	Percent
1. Yes	64	37.7
2. No	40	23.5
3. Don't know	10	5.9
4. Not covered	56	32.9
	170	100.0

V. Dollar-value of the Purchase	Number	Percent
1. Less than \$5	29	17.7
2. Between \$5 and \$10	6	3.7
3. Between \$11 and \$50	44	26.3
4. Between \$51 and \$100	10	6.1
5. Between \$101 and \$500	18	11.0
6. Between \$501 and \$500	10	6.1
7. Value over \$500	46	28.0
8. Don't know	1	.7
	164	100.0

Demographic Composition (Population, N=224)

Age:	Number	Percent
Blank	9	4.0
0 - 19	7	3.1
20 - 24	24	10.7
25 - 34	46	20.5
35 - 44	30	13.4
45 - 54	40	17.9
55 - 64	32	14.3
65 over	36	16.1
	224	100.0

Profession:	Number	Percent
Blank	14	6.3
Self employed	16	7.1
Manager	4	1.8
White Collar	63	28.1
Blue Collar	23	10.3
Student	16	7.1
Housewife	54	24.1
Retired	28	12.5
Unemployed	6	2.7
	224	100.0

Education:	Number	Percent
Blank	7	3.1
Grade School	6	2.6
Some High School	25	11.2
Graduate, High School	55	24.6
Some College	38	17.0
Graduate, College	62	27.7
Post Baccalaureate Education	24	10.7
Technical College	7	3.1
	224	100.0

Income:	Number	Percent
0 - 4,999	47	21.0
5,000 - 9,999	13	5.8
10,000 - 14,999	33	14.5
15,000 - 19,999	25	11.2
20,000 - 24,999	23	10.3
25,000 over	46	20.5
	224	100.0

Marital Status:	Number	Percent
Blank	6	2.7
Single	47	21.0
Married	151	67.4
Widowed	20	8.9
	224	100.0

Sex:	Number	Percent
Male	83	37.4
Female	139	62.6
	222	100.0

Race:	Number	Percent
Blank	50	22.5
White	172	77.5
	222	100.0

Andreasen, Alan R. and Best, Arthur (1977), "Consumers Complain--Does Business Respond?" Harvard Business Review, 4, 55, 93-101.

Barksdale, Hiram C. and Darden, William R. (1976), "Consumer Attitudes Toward Marketing and Consumerism," Journal of Marketing, 36, 28-35.

Day, Ralph L. and Landon, E. Laird (1976), "Collecting Comprehensive Complaint Data by Survey Research," in Advances in Consumer Research, Vol. 3, B. B. Anderson, Ed., Cincinnati: Association for Consumer Research, 263-268.

Fornell, Claes and Westbrook, R. A. (1977), "An Exploratory Study of Assertiveness, Aggressiveness, and Consumer Complaining Behavior," in Advances in Consumer Research, Vol. 6, W. L. Wilkie, ed., Miami: Association for Consumer Research, 105-110.

Hustad, Thomas P. and Pessemier, Edgar A. (1973), "Will the Real Consumer Activist Please Stand Up: An Examination of Consumer's Opinions About Marketing Practices," Journal of Marketing Research, 10, 319-324.

Kraft, F. B. (1978), "Characteristics of Consumer Complainers and Complaint and Reparatronage Behavior," paper presented at the Symposium on Consumer Satisfaction/Dissatisfaction and Complaining Behavior, Chicago.

Liefeld, J. P., Edgecombe, H. C. and Wolfe, Linda (1975), "Demographic Characteristics of Canadian Consumer Complainers," Journal of Consumer Affairs, 9, 73-80.

Lundstrom, William and Kerin, Roger (1976), "Psychological and Demographic Correlates of Consumer Discontent," in Proceedings, Howard C. Schneider, editor, American Institute for Decision Sciences, 488.

Mason, Joseph Barry and Himes, Jr., Samuel H. (1973), "An Explanatory Behavioral and Socio-Economic Profile of Consumer Action About Dissatisfaction with Selected Household Appliances," Journal of Consumer Affairs, 7, 121-127.

Miller, John A. (1970), "Who is the Discontented Consumer?," in Thomas V. Greer, Combined Proceedings, Chicago: American Marketing Association, 486-490.

"Most Consumer Complaints Fall on Deaf Ears: Survey," Marketing News, 13 (February 21, 1980), 8.

"Most Unhappy Shoppers Don't Voice Complaints," Advertising Age, 47 (June 21, 1976), 27.

Shuptrine, F. K., Thomas, W. R., and Sharma, A. (1979), "The Plight of the Discontented Consumer," Journal of Consumer Studies and Home Economics, 3, 161-171.

Wall, M., Dickey, L. E., and Talarzyk, W. Wayne (1977), "Predicting and Profiling Consumer Satisfaction and Propensity to Complain," paper presented at the Research Symposium on Consumer Satisfaction/Dissatisfaction and Complaining Behavior, University of Indiana.

Warland, Rex H., Herrman, Robert O. and Willits, Jane, (1975), "Dissatisfied Consumers: Who Gets Upset and Who Takes Action," Journal of Consumer Affairs, 9, 148-163.

Zaltman, G., Srivastava, R. K., and Deshpande, R. C. (1978), "Perceptions of Unfair Marketing Practices: Consumerism Implications," in Advances in Consumer Research, Vol. 5, H. Keith Hunt, ed., Chicago: Association for Consumer Research, 247-253.

AN INTEGRATED VIEW OF THE STORE CHOICE/PATRONAGE PROCESS

Donald Granbois, Indiana University

Abstract

Three papers, each touching on one or more aspects of store choice/patronage behavior, are reviewed here with the goal of relating each to a simple phase model. Contributions, methodological issues and ideas for future research growing out of this review are presented.

Introduction

Beyond their common concern with consumers' retail patronage behavior, the three papers presented vary considerably in method, the underlying bodies of theory and research drawn on, and the degree of innovativeness the authors have shown in conceptualizing the behaviors studied. Each paper reports an investigation of one or more phases of a sequence of mental processes and overt behaviors consumers exhibit in their store choice/patronage behavior. Beyond the usual compliments, criticisms and suggestions triggered by the individual papers under review which are expected in a discussion paper, each paper's relation to a simple phase model of store choice/patronage will be shown. Suggestions for the authors' future research growing out of this attempt at integration will be made.

The Store Choice/Patronage Process

Elements in the store choice/patronage process can be categorized as four linked phases, including:

Image Formation → Patronage →
Post-patronage Evaluation → Further Behavior

Image formation includes those mental processes whereby information and experience are processed and evaluated resulting in predispositions which generally guide patronage. Patronage refers both to customers' actual shopping trip behavior -- number and sequence of store visits -- and their in-store behavior, as goods are acquired, interactions with sales personnel and other customers are entered into, etc. Post-patronage evaluation has been thought of both as a comparison of actual and expected behavior and as an attribution process through which consumers attempt to assess the causes of the outcomes of their patronage behavior. Further behavior includes remedy-seeking among unhappy consumers, word-of-mouth (both positive and negative) reflecting the outcome of the post-patronage evaluation process, and even punitive behavior directed at retailers.

Retail Shopping Area Image

Consumers probably have at least rudimentary knowledge about available stores and shopping clusters prior to most store visits, although the extent and character of this knowledge is seldom if ever the subject of research. As in the literature on brand and product choice, there has been a tendency for researchers to assume that an evaluation process based on two or more attributes underlies an overall affective orientation towards individual stores. Houston and Nevin, like earlier researchers, refer to the output of this evaluation process as the store's "image," the "complex of a consumer's perceptions of a store on attributes." It seems possible, however, that an affective orientation towards a store sufficient to trigger a store visit can develop with little or no cognitive evaluation,

a position consistent with a considerable body of evidence in both psychology (Zajonc 1980) and consumer behavior (Olshavsky and Granbois 1979). Respondents' seeming ability to assign ratings to listed alternatives in terms of criteria specified on a questionnaire does not support the conclusion that we carry such neatly-structured images around in our brains! There is ample precedent for Houston and Nevin's assumption, however, although studies asking if and how such images are formed would be quite interesting in the context of retail shopping clusters.

Houston and Nevin's research on Madison shopping clusters assumed a multi-attribute structure comprised of 16 dimensions underlies shoppers' images of the five clusters studied. Respondents provided up to 80 judgments (depending upon their familiarity with the clusters). The 16 cluster attributes were represented by short, very abstract phrases of 2 or 3 words. Again, there is much precedent for this approach, but one wonders what meanings respondents associate with such terse statements.

To the extent that consumers do not have complete knowledge about stores or shopping clusters, "images" may best be thought of as expectations or predictions. As such, these may be biased by information processing errors and deficiencies such as the incorrect application of "surrogate indicators." There is, in fact, already evidence in the literature that such biases do indeed occur in judgments about individual stores' price levels and locations (Brown 1969, Thompson 1963, Olshavsky, MacKay and Sentell 1975).

Finally, the 16 shopping cluster dimensions were taken from the store image literature, with some modifications apparently based on the judgment of the researchers after consulting with shopping center managers. Surprisingly, two attributes of great potential importance -- location (distance and/or driving time from home or other trip origin) and depth of assortment -- were excluded (assortment depth may be captured somewhat by the dimension "Product Selection," although other meanings may well have been interpreted by respondents). Gravitational studies cited by the authors have found these two variables to effectively predict aggregate movement of fashion goods shoppers even when measured with crude proxy variables such as map distance and shopping goods stores' square footage.

Houston and Nevin's factor analysis findings provide what seems to me to be rather indirect evidence of an unsurprising conclusion; that is, downtown Madison as a shopping cluster is perceived differently from the four planned shopping centers in the area. The nature of these differences is only hinted at in Table 4, where downtown is shown somewhat lower in respondents' evaluations than all four shopping centers.

Since even summary measures of the Madison shoppers patronage behavior are not reported, it is not possible to assess the predictive validity of the study's results, so we are not able to evaluate the importance, or the nature of the less favorable evaluation given downtown. For all we know, Madison patrons could be shopping downtown despite lower evaluations on the study's 16 attributes. The paper's conclusion, then, about the need for repositioning downtown through promotion seems premature, since it is by no means clear that the image dimensions studied are

salient ones, driving actual patronage behavior. Beyond this, mere knowledge that downtown is rated lower on "General Price Level," "Friendly Atmosphere," "Helpful Store Personnel" and "Conservative" seems insufficient guidance for creating an effective promotional campaign to stimulate patronage. Indeed, price and personnel policy changes might well prove to be more appropriate responses by downtown merchants.

Consumers' Marketplace Problems

Shuptrine and Wenglorz, in another city, report a study fitting our store choice/patronage model somewhat less directly. Like many earlier studies, their findings reveal some consumers having some problems seeking remedy from sellers, which overwhelmingly were the retailers selling the product (82.3% of all complaints). This behavior, an important type of "Further Behavior" identified in our phase model, was found to be considerably more prevalent than in earlier studies, perhaps because questions were purposely included to probe and encourage reports of problems. The study's primary contribution is its conclusion that consumer problems may be more widespread than has been shown in earlier research, thus focusing even greater concern on the retailer as an important factor in remedying consumer dissatisfaction. Beyond this, the study is totally descriptive in character, its many cross-classifications revealing product and consumer-type variations in both problem incidence and tendency to complain quite similar to many earlier studies.

While the research serves to heighten our concern for the level of consumer problems reported (even though over 2/3 were reportedly adjusted satisfactorily) it provides no insights into the evaluation process whereby the consumer concludes a problem exists. Post-patronage evaluation processes are hardly touched on in consumer research, yet this seems to be a necessary next step if we are to begin to understand and explain the phenomenon of consumer dissatisfaction and subsequent behaviors. Such understanding and explanation need to precede serious recommendations for policy action. For example, possible consumer education efforts are difficult to conceive of without far greater understanding of consumer expectations formation; how product experience is interpreted and evaluated; and the nature of the attribution process as consumers attempt to decide whether they themselves, the manufacturer, the retailer, or some of the party should be blamed for less-than-satisfactory experience.

Deviance and Dissatisfaction

Mills' research fits our phase model too in that the five deviant behaviors studied occur either during the shopping process itself (part of the Patronage phase) or as part of Further Behavior. Since respondents predicted both their own and others' likely behavior in response to "scenarios" comprised of store descriptions and photographs, the perceptions evoked as the settings in which the deviant behaviors of interest were to be imagined by respondents seem to correspond to the concept of store image as it is normally defined. An interesting theoretical twist distinguishes the study in that the retail unit's power relative to that of customers is added as a significant dimension of retail image. Self-reported and projected deviant behavior was hypothesized to vary with the three power "types" represented by the various scenarios, as was consumer satisfaction.

Relating the perceptions evoked by Mills' scenarios to the concept of store image focuses attention on the interesting issue of how consumers form impressions of retail power. Apparently the image elements included in the written scenarios and photographs were successful in evoking the "proper" perception of power among respondents; from the written sample scenario presented, these elements seemed

to include building size and type, market segment served, price policy, credit policy, and parking. Since we are shown just one scenario and don't know the nature of the process used in developing the written and photographic stimuli, we can only speculate about why the scenarios contained the elements that they did. Given the results of the study, which tended to confirm the importance of perceived power as an influence on self-reported and projected deviant behavior, the issue raised here has much practical significance. Other things equal, manipulating those elements perceived as indicators of retail power or changing consumers' beliefs about the relationships which exist between store attributes or policies and store power could conceivably both increase consumer satisfaction and reduce deviant behavior directed at stores.

Some Methodological Concerns

The surveys by Shuptrine and Wenglorz and by Houston and Nevin both utilized good-sized samples selected randomly from city populations, a method still not common in academic research. The cost and effort in such relatively ambitious surveys is substantial, placing a burden on the researchers to be sure the research question not only requires but is worthy of such a costly setting. Judging by the conclusions and recommendations offered by these two papers, neither study seems highly cost efficient, but it is possible that additional findings not reported here were obtained in each case.

Shuptrine and Wenglorz's research largely replicated the research question already investigated in a number of earlier studies. The justification appeared to be their desire to provide a more valid estimate of the true incidence of consumer problems growing out of their realization that earlier research may have understated the importance of consumer dissatisfaction. Assuming this objective to be the purpose of the research, one questions their estimate of the incidence of various consumer problems, since each respondent apparently reported on a single problem. An alternative method would have identified product categories first, then established the incidence of recent purchases in each, followed by probing questions getting at types and frequencies of problem experiences.

Houston and Nevin appeared to miss an opportunity to get at true differences in consumers' perceptions of downtown versus shopping center attributes by imposing the same set of image dimensions on responses for both kinds of shopping cluster. Unless questions not reported here covered both familiarity and actual patronage of the shopping clusters investigated, some very important validating data were excluded from the study.

While factor analysis simplifies interpretation by reducing large numbers of variables into a much smaller number of dimensions, the nature of the question pursued here seems to call for more rather than less detailed analysis. Each of the 16 initial image dimensions represents a somewhat complex set of variables (how many aspects of "atmosphere" characterize downtown Madison?) each worthy of investigation if workable proposals for stimulating increased downtown patronage are to be derived. At the least, examination of the profiles of each shopping cluster on all 16 attributes should be made.

Mills' study recruited a group of shoppers in malls and elsewhere with no apparent attempt to sample a defined population. Details about the scenarios used and their method of development are not known. In Mills' design, hypothesis testing was based on totally hypothetical situations presented on the basis of quite artificial stimuli. There are real questions of external validity in this

approach. As a preliminary (and perhaps not very expensive) test, the study is of value. Judgments of Mills' results, though, should probably await verification in research relying less on artificial scenarios. (Admittedly, given the sensitive nature of the issues investigated, more realistic designs don't come easily to mind). Since the behaviors in question are deviant acts, the validity problem is still more serious. It would be useful to know if actual levels of deviant acts by consumers experienced by retail outlets in each of the three power types vary as would be predicted by Mills' hypothesis. Measures of satisfaction with actual stores, classified similarly into the three types, would provide a stronger test of the hypothesis relating satisfaction to perceived power.

Future Research

The linkage between consumers' perceptions of shopping clusters' attributes and actual patronage behavior has seldom been investigated, and Houston and Nevin's study represents an important start in what should be a fruitful stream of research. More detailed study of perceptions and misperceptions of shopping cluster characteristics (including location/convenience factors!) seems called for, followed by validating studies in which relationships between shoppers' evaluations of shopping clusters are compared with their actual patronage behavior.

The next step in Shuptrine and Wenglorz's dissatisfaction and complaining behavior research might well follow up their finding that probing questions appear to stimulate recall of dissatisfying experiences. Perhaps different levels of probing questions could be devised and the level of incidence of dissatisfaction indicated by each level could be experimentally determined. Further evidence that dissatisfaction is in part an artifact of the persistence of the interviewer would certainly force some reinterpretation of earlier research, and provide further incentive for researchers in this increasingly-popular area to devise more comprehensive (and perhaps less overt) measures of dissatisfaction. A second direction worthy of much additional effort is the exploration of evaluation processes giving rise to varying levels of satisfaction and dissatisfaction.

Deviant consumer behavior is clearly important to study but the sensitivity of the issues involved serve to magnify our usual problems in devising valid self-report measures. Nonetheless, Mills' hypotheses are of sufficient interest to warrant his continued efforts on the topic. More thorough exploration of the retail firm's attributes underlying consumers' perceptions of the firm's power type might give insights into aspects of store design, advertising media and copy, salespersons' behaviors and other controllable dimensions of the retail "mix" that could be manipulated to influence consumers' perceptions.

References

- Brown, F. E. (1969), "Price Image versus Price Reality," Journal of Marketing Research, 6, 185-91.
- Olshavsky, R. W. and Granbois, D. H. (1979), "Consumer Decision Making -- Fact or Fiction?", Journal of Consumer Research, 6, 93-100.
- Olshavsky, R. W., MacKay, D., and Sentell, G. (1975), "Perceptual Maps of Supermarket Locations," Journal of Applied Psychology, 60, 80-86.
- Thompson, D. L. (1963), "New Concept: 'Subjective Distance' or Store Impressions Affect Estimates of Travel Time," Journal of Retailing, 39, 1-6.
- Zajonc, R. B. (1980), "Feeling and Thinking: Preferences Need No Inferences," American Psychologist, 35, 151-175.

PRODUCT AND BRAND USER STEREOTYPES AMONG SOCIAL CLASSES

J. Michael Munson, University of Santa Clara
W. Austin Spivey, University of Santa Clara

Abstract

This research investigated the relation of social class to product and brand-user stereotyping. The data base included over 200 females evaluating 48 brands in 8 product categories. As hypothesized, significant differences between upper, middle and lower social classes were observed for both product and brand-user stereotypes.

Introduction

Marketers have given considerable attention to social class and opinion leadership. Both constructs are felt to play an important part in decisions about market segmentation and the selection of mass media aimed to reach specific target markets. Interestingly, a review of the marketing and communications research highlights the fact that these constructs have been studied independently. There is little, if any, explicit recognition of theoretic linkage. In addition, a review implies that the perceived usefulness of the constructs may vary depending upon specific product class, situation, and target market. Indeed, some seem to feel that one construct is "better" than the other.

Some researchers decry the difficulties associated with operationalizing opinion leadership. Rogers (1972) as well as King and Summers (1970) bemoan the problems in identifying and using opinion leaders directly. Robertson (1970) comments on the difficulty in simulating opinion leadership via testimonials, slice-of-life vignettes, and man-on-the-street recommendations. Loudon and Della Britta (1979) point out how hard it is to stimulate leaders to talk about a product and exert personal influence. There are conflicting findings about whether the influence of opinion leaders tends to be product specific (e.g., Robertson and Myers 1969; Silk 1966) or generalizable to overlapping product categories (e.g., King and Summers 1970).

Other researchers question the usefulness of social class. The "dated" nature of much social class research--most of the studies are 15 to 20 years old--leads some to question the relationship between social class and today's consumption patterns. Some believe that social classes are disappearing or becoming more homogeneous. Bieda and Kassarian (1969) suggest that differences in product preferences among social classes may have disappeared because of common exposure to mass media. Their suggestion implicitly argues that mass media has had a leveling influence on the values and life style aspirations of all individuals. If this is true, the stereotypes of specific brand or product users should be relatively homogeneous across social strata.

In light of the weaknesses associated with both social class and opinion leadership, it seems incorrect to view either construct as better than the other. Rather, it would seem advertisers should concentrate more effort on "integrating" the potential inherent in each. It may well be that many of the effects of opinion leadership are contingent upon social class considerations. Indeed, Klapper (1970) views mass media communications as operating through mediating factors--group membership, interpersonal contacts, selective exposure and defense mechanisms--so that it typically is a contributing agent but not the sole cause of audience reactions.

The desire to integrate the potential inherent in the two

constructs provided the impetus for this research. Two key questions are addressed based upon Bieda and Kassarian's (1969) suggestion that differences in product perception among social classes have disappeared:

- (i) do perceptions of brand and product user stereotypes differ across social classes?
- (ii) if so, for what types of product classes?

The Study

If social class is to become a resurgent key to implementing opinion leadership, advertisers must answer a key question: to what extent do differential stereotypes exist across social classes at both the micro brand level and the macro level of product category. If common exposure to the mass media has acted as a leveling influence across all individuals, as Bieda and Kassarian suggest (1969), then differences in product perception among social classes may have disappeared.

Given that prior marketing research on social class has focused on one of three areas--shopping, media and product class usage behavior (as opposed to brand usage)--little is known about how perceptions of specific brands may differ across social class. Moreover, much of the previous research is based on samples of university students which limits the generalizability of results.

Study Hypotheses

The need to fill in these gaps dictated that two working hypotheses be formulated to guide the study. Stated in alternative form, they are:

- H1: Stereotypes of people who own or prefer a specific brand will differ across social classes.
- H2: Stereotypes of people who own or prefer a specific category of product will differ across social classes.

Two corollaries (C1 and C2) were formulated from the working hypotheses. The first related to the magnitude of any observed differences among social classes:

- C1: Any differences in stereotypes will be largest between lower and upper social classes.

The second corollary related to observed differences for a product category:

- C2: Differences in user stereotypes will exist only for value-expressive product categories.

Specific support for H1, H2 and C1 comes from many sources. Warner (1960) suggested that the further apart two people are in class level, the further apart will be their usage and understanding of a product. Martineau (1957) noted social class helps to define the hierarchy of things on which people wish to spend money. The further apart two people are in social class, the greater the difference will be in perceived likelihood of purchase. Levy (1966) states "...As one looks at the social class ladder, the further he goes down it, the more difference he sees in life styles and values, and thus, the difference in

consumption patterns becomes greater..."

C2 is based upon research by Locander and Spivey (1978). They discussed the basic differences between utilitarian and value-expressive products. Value-expressive products help an individual to give positive expression to their values and self concept. For these types of products, the key concern is the stereotype of the person who owns or prefers the product. On the other hand, for utilitarian products, user stereotyping is relatively unimportant compared to specific product attributes. Another point is noteworthy. Value-expression and utilitarianism are not opposite ends of a continuum; rather, they are independent dimensions (Spivey 1977). Therefore, a product quite possibly can be considered utilitarian and still provide an opportunity for self-expression. Generally, however, one would expect to observe few differences across social classes regarding the perceptions of utilitarian product users.

Method

Sample Design

In order to test the hypotheses and the corollaries, 203 women were interviewed in their homes. They were chosen by using a stratified, multi-stage probability sample with quotas. Approximately equal sample sizes were obtained for each social class: 64, upper; 87, middle; and 52 lower.

The women were chosen from a city frequently used for test marketing. The sampling universe was the total number of city households. Initial stratification was done on the basis of income. Census tracts were used as the primary sampling units; each tract was designated as representative of either upper, middle, or lower social class. Interviewers were sent to random starting points and instructed to proceed in a clockwise direction around each block until a quota of six interviews per block had been completed. Block quotas were filled subject to two restrictions: (1) at least one and preferably two of each six respondents were employed; (2) no women under 18 years or no more than two women 65 years old or older be included so as to make the characteristics of the sample as similar as possible to the general population. Some call-backs were required to fulfill the quota within each block.

Product Typology

Across the entire sample of 203, a wide range of product categories and brand names were incorporated into the survey. Two primary criteria were used to select the products which were included in the study. First, the products had to represent categories which could be designated as convenience, shopping and specialty goods. Second, the products had to reflect a wide range with respect to their ability to communicate symbolically. Twelve judges familiar with these criteria were used to categorize 23 product categories. Of the 23, a smaller subset of 8 product categories which reflected the highest levels of inter-judge agreement were selected for inclusion in the final survey.

Two product categories high on the value-expressive dimension (VED) were selected: automobiles and magazines. Three categories moderately high on VED were chosen: washing machines, vacuum cleaners and gasoline. Three product classes low on VED were selected: brassieres, deodorants, and laundry detergents. Available research supports these categorizations. For example, Lessig and Park (1978) found automobiles to have the highest score, 3.65, on the value-expressive dimension among the 20 products studied; magazines ranked eight with a score of 3.08; laundry detergent ranked 17 with a score of only 2.42. Present re-

search also suggests that for at least 6 of the 8 product categories, wives may be the dominant spouse in any decision processes involving their purchase (e.g., Fabian 1965, Davis and Riguax 1974). Within each product category six brands were selected for possible evaluation on the basis of greatest familiarity among a demographically representative pilot sample. Hence, the hypotheses were tested over a comprehensive product typology which included 8 product categories comprised of 48 individual brands.

Operational Definitions

Brand or product user stereotypes were obtained by having each respondent evaluate how she viewed "a person who owns or prefers" a particular brand over 16 semantic differential scales. These scales (e.g., successful-unsuccessful, informed-uninformed, etc.) have been shown in previous research to be well suited to assessing product and brand user stereotypes for various product categories (e.g., Ross 1971; Munson 1973; Munson and Spivey 1979). The scales represent major dimensions of how the individual views either himself and/or others and "...represent clear examples of self-concept definition" (Bills, et.al. 1951). These scale items have demonstrated adequate test-retest reliability over a three week interval, with reliability coefficients (r) ranging from $r = .72$ ($p < .01$) to $r = .68$ ($p < .01$). Additionally, Ross (1967) has undertaken specific converging operations in demonstrating the construct validity of the measure while others (Munson 1973 and Ross 1967) have established their freedom from social-desirability effects. Social class was determined for each respondent by using the Socioeconomic Index for Occupations (Duncan 1961). This index provides social class values for each of the detailed occupational classifications of the census based on a combination of occupational prestige ratings, educational status, income level, and age distribution.

Analytical Technique

The two hypotheses and the corollaries were tested by using one-way, univariate analysis of variance (ANOVA) coupled with a form of the chi-square statistic. For each ANOVA calculated the response variable was the sum of an individual's responses to all 16 semantic differential scales which indexed her stereotype of a specific brand user. Social class was the "treatment" effect of interest. H1, about differences in perceptions for specific brands, was tested within the ANOVA framework by using the standard F-ratio computed for each brand (Hicks 1973). To test H2, the probability of differences in perception at the macro level of product category, a chi-square statistic was computed based upon the ANOVA results associated with each brand within a product category (Winer 1962). The appropriate chi-square statistics were computed based upon contrasts calculated at the micro level of brand. The reader interested in a detail explanation of the relationship between ANOVA and the chi-square statistic used to test H2 should see Appendix 1 where the latter statistic for testing the difference in product category perceptions between lower and upper class is derived.

C1, about differences among social classes regarding perceptions of both brand and product class, was tested by using the principle of orthogonal contrasts (Hicks 1973). A linear and a quadratic contrast were considered simultaneously: the linear contrast considered the difference in mean perception between upper and lower social classes; the quadratic contrast considered the difference in mean perception between the middle class and the average of the upper and lower classes. C2, about differences among social classes existing only for perceptions about value-expressive products, was tested by noting the pattern of results. No specific test statistic was necessary since the hypothesis could be rejected simply by contradiction (Daniel and Terrell 1979).

Results

A summary of the ANOVA and chi-square results is presented for all product classes and brands in Table 1. Note that only the linear contrasts and chi-square are shown; quadratic contrasts are omitted in the interest of clarity. In addition, note that the results are ordered: highly value-expressive products are mentioned first; more utilitarian products are mentioned last.

Value-expressive Brands and Products

Significant differences in brand-user stereotypes were found between upper and lower social classes for 9 of the brands tested (Table 1): Chrysler, Ford, Buick, Standard, Texaco, General Electric, Vogue, Ladies Home Journal and Playboy. These significant differences in brand user stereotypes were most pronounced between lower and upper class--as indicated by the significant linear F-ratios. In only one product category, washing machines, were significant quadratic effects observed. A significant difference in perception for Sears Kenmore washer was found between middle class compared to both lower and upper class. This indicates that the perceptions of lower and upper classes sometimes merge for this brand. Similar results were observed regarding differences among social classes about product categories. As shown by the chi-square statistics derived from the linear F-ratios, three of the eight categories--automobiles, gasolines, and magazines--were perceived differently by the lower social class compared to the upper class. In addition, middle class perceptions were different than the other classes with respect to washing machines as indicated by a quadratic contrast.

Utilitarian Brands and Products

Results for the utilitarian product categories--brassieres, deodorants, and laundry detergents--show relatively few significant differences in stereotypes at either the brand level or macro product category level. Only Right Guard and Salvo were perceived differently.

Conclusions About the Hypotheses

Generally, H1, H2, and C1 were supported; C2 was not supported.

The results for the individual product categories clearly support previous research. The findings for magazines are consistent with Warner and Lunt's (1941) earlier Yankee-City research, while those for automobiles are consistent with Coleman (1960), Peters (1970), and Akers (1970). Peters (1970), for example, found occupational class income (defined as a combination of family income and occupation) to be differentially related to automobile buying behavior for foreign economy, intermediate-sized, and compact cars.

Findings were as expected for automobiles and magazines because other research has shown that among women, these two product classes are perceived as more conspicuous and important than many others (Cohen and Barban 1970). Both may be perceived as having considerable potential to facilitate symbolic communication. The results for gasoline may, in part, be attributed to a type of complimentary relation they have with automobiles. These two products may represent logically associated categories within the consumer's cognitive structure. The findings were not quite as expected for washing machines. Given the similarity of this product to refrigerators, differences between social classes were expected based upon the results of Lessig and Park (1978). But the differences were expected between upper and lower social classes and not between middle class and both extremes.

Implications

Social Classes Have Different Perceptions

Overall, these results refute the contention of Bieda and Kassarjian (1969) that differences in social class have disappeared because of common exposure to mass media. True, as indicated by these results, there is unanimity of perception, but only for some products. Further, it is clear that differences in brand perception across social class are related to the generic product's ability to communicate symbolically. The highly value-expressive product categories had many brands with significant differences in user stereotypes while the utilitarian categories had few. Moreover, for highly value-expressive products, the product categories themselves were perceived as different by upper and lower social class.

The pattern of significant differences in user stereotype across social classes observed for value-expressive but not for utilitarian products has implications for researchers studying opinion leadership. To better understand the processes by which opinion leaders exert their influence, greater knowledge about their social class is essential. The results of this study imply that marketing efforts which attempt to employ either social class or opinion leadership in isolation will be less effective than efforts which consider both simultaneously. Focusing exclusively on one is an inadequate representation of a theoretic system in which both are interrelated. This is true for two reasons. First, the construct "opinion leadership" only assumes meaning within the context of some relevant social system in which the opinion leader is implicated. From a pragmatic viewpoint of identifying market segments and selecting media to reach target segments, social class is the most easily identifiable "social system" construct within which an opinion leader can operate. Second, where instances of opinion leadership occur, it would seem that complex social system and communication effects can intervene between both the opinion leaders' reception of the message from the mass media and his relaying of the message to others.

Appendix 1

Table 1A shows the derivation of the chi-square statistic derived from the linear contrasts associated with the various brands (makes) of automobiles. Brands within the product category can be regarded as a series of k experiments designed to test the same hypothesis (e.g., H2). The probabilities of the observed outcomes with respect to the common hypothesis tested can be designated P(1), P(2),... P(k). If the experimenter desires to make an overall probability statement, the following statistic may be used (Winer 1962):

$$\chi^2 = 2 \sum u_i \text{ where } u_i = -\ln P(i) \quad (1)$$

Under the hypothesis that the observed probabilities are a random sample from a population of probabilities having a mean of .50, the chi-square statistic in equation (1) has a sampling distribution which is approximated by the chi-square distribution having 2k degrees of freedom.

For example, the chi-square based on linear contrasts for automobiles in Table 1A may be determined as follows. The probabilities of obtaining the linear F ratio or one more extreme is given in the column headed Probability. The natural log of the probability with the sign changed is given in the next column. The numerical value of the chi-square statistic is equal to twice the sum of the negative natural log of each probability. The number of experiments (k) or brands equals six; the degrees of freedom is 2(k). An arbitrary alpha level of .05 was selected against which to test the null hypothesis. Hence, the probability of obtaining a chi-square statistic of 28.30 or larger is less than .005 when (chi-square .995 = 28.30) the mean

TABLE I
DIFFERENCES IN BRAND AND PRODUCT USER STEREOTYPES
ACROSS UPPER, MIDDLE AND LOWER SOCIAL CLASSES

Product Category	Linear F-Ratio (Brand level)	Chi-Square based on linear F-ratio (Product Level)
<u>Automobiles</u>		
Cadillac	.24	34.29***
Chrysler	12.76***	
Ford	9.55***	
Volkswagen	.17	
Chevrolet	.06	
Buick	3.25*	
<u>Magazines</u>		
Vogue	4.38**	27.45***
Life	.00	
Ladies Home Journal	3.92**	
Readers Digest	.91	
T.V. Guide	.01	
Playboy	7.01***	
<u>Washing Machines</u>		
General Electric	3.32*	16.04 ^b
Maytag	.09	
Sears Kenmore	1.96 ^a	
R.C.A.	.64	
Wards	.86	
Frigidaire	.02	
<u>Vacuum Cleaners</u>		
Hoover	1.79	13.09
General Electric	1.16	
Sunbeam	.10	
Electrolux	1.33	
Sears Kenmore	.47	
Eureka	.70	
<u>Gasolines</u>		
Standard	4.36**	22.42**
Shell	1.16	
Texaco	3.99*	
Clark	.89	
Liberty	2.25	
Arco	3.02	
<u>Detergents</u>		
Tide	.06	17.50
All	1.07	
Cheer	.61	
Pollution Free	2.36	
Cold Power	2.02	
Salvo	3.63*	
<u>Deodorants</u>		
Arid Extra Dry	.19	15.22
Secret	.24	
Right Guard	3.33*	
Five Day	2.68	
Ban	.03	
Mum	.02	
<u>Brassieres</u>		
Maldenform	.16	8.32
Sears	.13	
Playtex	.03	
Lovable	.01	
Bali	.07	
Penneys	.70	

Note: All F-ratios are based on at least $df = 2, 50$. All Chi-square values have $df = 2K$.

^a showed a significant quadratic contrast

^b showed significant quadratic chi-square

* $p \leq .10$. ** $p \leq .05$. *** $p \leq .01$.

Appendix 1 Continued

probability in the population is .50.

Using the chi-square statistic to reject the null hypothesis is, in effect, accepting the series of individual brand-user stereotype evaluations as combined evidence indicating that the common hypothesis (H2) in each of the experiments can be rejected if inferences are to be made with respect to the combined series of experiments.

TABLE A1
DIFFERENCES IN BRAND USER STEREOTYPES OF AUTOMOBILES
BY SOCIAL CLASS AND CHI-SQUARE STATISTICS BASED
ON LINEAR F RATIOS

Makes of Automobiles	Linear F-ratio Observed	Probability	-ln (Prob)
Cadillac	.248	.500	0.693
Chrysler	12.767	.001***	6.908
Ford	9.547	.003***	5.809
Volkswagen	0.174	.500	0.693
Chevrolet	0.006	.500	0.693
Buick	3.249	.095*	2.353
			17.149

Chi-square based on Linear F-ratios = $\chi^2 = 2(17.149) = 34.298$

$\chi^2_{99.5} = 28.3$

* $p \leq .10$

*** $p \leq .01$

References

Akers, F.C., (1968), "Negro and White-Automobile-Buying Behavior: New Evidence," Journal of Marketing Research, 5, 283.

Bieda, J.C., and Kassarjian, H.H., (1969), "An Overview of Market Segmentation," in B.A. Morin ed., Marketing in a Changing World, Chicago: American Marketing Association, 249-253.

Bills, R.E., Vance, E.L., and McLean, O.S., (1951), "An Index of Adjustment and Values," Journal of Consulting Psychology, 15, 257-261.

Cohen, J.B., and Barban, A., (1970). "An Interactive Consumer Product Typological System, A Progress Report and Partial Evaluation," Working Series in Marketing Research, No. 12, College of Business Administration, Pennsylvania State University.

Coleman, R.P., (1970), "The Significance of Social Stratification in Selling," Proceedings of the American Marketing Association, ed., M.L. Bell, 171-184.

Daniel, W.W., and Terrell, J.C., (1979), "Business Statistics: Basic Concepts and Methodology," Boston: Houghton Mifflin Company

Davis, H.L., and Rigaux, B.P., (1974), "Perception of Marital Roles in Decision Processes," Journal of Consumer Research, 1.

Duncan, O.C., (1961), "Socioeconomic Index for All Occupations," in A.J. Reiss, Jr., Occupations for Social Status. New York: Free Press.

Fabian, L., (1965), "Expenditure Patterns of the American Family," National Industrial Conference Board, sponsored by Life Magazine.

Hicks, C.R., (1973), Fundamental Concepts in the Design of Experiments. New York: Holt, Rinehart and Winston.

King, C.W., and Summers, J.O., (1979), "Overlap of Opinion Leadership Across Consumer Product Categories," Journal of Marketing Research, 7, 43-50.

Klapper, J., (1970), The Effects of Mass Communication. Glencoe, Ill: The Free Press, 508.

Lessig, V.P., and Park, C.W., (1978), "Promotional Perspectives of Reference Group Influence: Advertising Implication," Journal of Advertising 41-47.

Levy, S.J., (1959), "Symbols for Sale," Harvard Business Review, 37, 117-124.

Locander, W.B., and Spivey, W.A., (1978), "A Functional Approach to Attitude Measurement," Journal of Marketing Research, 15, 576-587.

Loudon, D.L., and Della Britta, A.J., (1979), Consumer Behavior: Concepts and Applications. New York: McGraw Hill Book Co. 189-214, 263-288.

Martineau, P., (1957), Motivation in Advertising. New York: McGraw-Hill Book Company.

Munson, J.M., (1973), "An Investigation of Self Concept Congruity and Brand Preferences: Toward a Predictive Model," unpublished Ph.D. dissertation, University of Illinois, Champaign-Urbana.

Munson, J.M., and Spivey, W.A., (1979). "Assessing Self Concept," Proceedings Association for Consumer Research, Vol. VII, J.C. Olson, (ed.), 598-603.

Peters, R.A., (1970), "Perceived-Price Quality Relationship: Experimental Evidence," Journal of Marketing Research, 7, 525-528.

Robertson, R.S., (1970), Consumer Behavior. Glenview: Scott Foresman, 81-97.

Robertson, T.S., and Myers, J.H., (1969), "Personality Correlates of Opinion Leadership and Innovative Buying Behavior," Journal of Marketing Research, 6, 164-168.

Rogers, E.M., (1972), Diffusion of Innovations. New York: The Free Press, 1-42.

Ross, I., (1967), "The Relationship between Self Concept and Judgements Concerning Selected Consumer Products," unpublished doctoral dissertation, Ann Arbor, Michigan: University Microfilms.

Silk, A.H., (1966), "Overlap Among Self-Designated Opinion Leaders: A Study of Selected Dental Products and Services," Journal of Marketing Research, 3, 255-259.

Spivey, W.A., (1977), "An Experimental Evaluation of Attitude Change for Attitude Functions Serving Combinations of the Utilitarian and Value-Expressing Attitude Functions," unpublished Ph.D. dissertation, University of Houston.

Warner, W.L., and Lunt, P.S., (1941), "The Social Life of a Modern Community," Yankee City Series 1.

Warner, W.L., (1960), Social Class in America. New York: Harper and Row.

Winer, B.J., (1962), Statistical Principles in Experimental Design. New York: McGraw Hill Book Company.

EXTENSIONS OF THE BASIC SOCIAL CLASS
MODEL EMPLOYED IN CONSUMER RESEARCH

Terence A. Shimp, University of South Carolina
J. Thomas Yokum, University of South Carolina

Abstract

Social class research in consumer behavior has a long history, but the contributions have indeed been minimal. This is partly because the model that has dominated thinking and research is both inconsistent with major trends in American society and incompatible with the socialization perspective that represents a major orientation in studying consumer behavior. This paper proposes extensions to the standard procedure for operationalizing social class and presents data that support the rationale for these extensions.

Introduction

An historical examination of a scientific discipline entails an exhaustive review of that discipline's literature. Analysis of contemporary journal articles and conference proceedings provides insight into those scientific units (i.e., theories, concepts, and methods) that are the subject of current interest, while a review of textbooks over a period of years reveals those enduring scientific units that have transcended the discipline's occasional excursions into faddish areas.

Such an analysis of the consumer behavior literature would likely lead to a major conclusion that many of the cherished theories and concepts which have remained virtually intact for years in textbooks have done so in the absence of any significant advances in insight or explanatory ability. The social class concept is a notable example. Consumer behavior texts invariably include a chapter on social stratification, and the material coverage is, as invariably, highly similar. This literature presumably reflects current knowledge and thinking, yet the scientific value and practical usefulness of social class for understanding and predicting consumption phenomena is dubious.

This paper presents a critical analysis of the manner in which social class has been treated in consumer behavior and offers suggestions for improvement. These suggestions are predicated on the belief that the conceptual model of social class prevalent in consumer behavior is both inconsistent with major trends in American society and incompatible with the socialization perspective that represents a dominant orientation in studying consumer behavior.

Social Class in Consumer Behavior

Consumer behavior interest in social class has waxed and waned over the past two decades. Martineau's (1958) seminal article provided the momentum for other solid works which appeared in the 1960s. Most notable were the writings of Coleman (1960), Carman (1965), and Levy (1966). Where Levy demonstrated that significant differences do exist among social classes with regard to various general behaviors, Coleman cautioned that the social class concept has been misunderstood and over-simplified. Carman, though critical of the overly simplistic manner in which researchers had applied the social class concept, presented a strong case supporting its potential usefulness.

The reflection and conceptualization personified by these works was followed by a period of empirical activity. Rich and Jain (1968) examined the relationship between social class and shopping behavior and concluded that social class has dubious usefulness in understanding this behavior.

Their skeptical perspective was countered by Wasson (1969), who argued, with a modicum of supportive evidence, that social class is superior to income as a predictor of consumer behavior. The social class vs. income debate was initiated, and a series of studies appeared (Hisrich and Peters 1974; Mathews and Slocum 1969; Myers, Stanton, and Haug 1971; Myers and Mount 1973; Slocum and Mathews 1970).

Empirical activity has been sporadic since then, exceptions are the store patronage research by Foxall (1975) and Prasad (1975) and Jain's (1975) novel application of conjoint analysis for delineating individuals' implicit social class concepts.

Interest in social class has not vanished, however. This is reflected in Nicosia and Mayer's (1976) writing on the sociology of consumption and in the provocative piece by Zaltman and Wallendorf (1977). These latter writers argue convincingly that a sociological perspective is needed to supplement the prevailing psychological orientation if a comprehensive understanding of consumer behavior is to be achieved. Their sentiments are echoed by Sheth (1977) who claims that demographic and SES variables, including social class, should not be discarded prematurely; instead, more sophisticated measurement and application are needed. Carman (1978) arrives at a similar conclusion while asserting that social class research has ignored his earlier criticisms (Carman 1965).

Critical Analysis

A variety of reasons account for why social class has provided minimal explanation of consumption phenomena. In fact, there are nearly as many reasons as there are empirical studies (cf. Dominquez and Page 1978; Zaltman and Wallendorf 1979). This critical analysis will provide a useful framework for suggesting improvements in the application of social class to consumer behavior. First, however, it will be useful to examine the exact nature of social class and also explore the reason why social class should provide a useful explanation of consumption phenomena.

The Nature of Social Class

A fundamental difficulty in working with the social class concept is arriving at a clear understanding of its meaning. From a strict sociological perspective, social class is just one dimension of a more general social stratification construct. Class is a power-based concept, including both economic and political elements. But stratification also includes a prestige dimension, or what sociologists term "status."

Carman (1965) has argued that a third dimension, cultural class, is one of primary relevance to consumer behavior. Cultural classes are self-perpetuating sub-cultures which differ with respect to the value placed on education, solidarity of the family, religious involvement, media exposure, recreational activities, etc. Cultural class can be operationalized with real variables (e.g., values, attitudes, behaviors) or with such proxy variables as occupation and education (Carman 1965).

Viewed from the perspective that a particular social class is a group of individuals who share a common culture and manifest similar life styles, the concept of social class

should be useful for understanding and predicting various consumption phenomena: Since human behavior is determined in large part by the particular elements of culture which are learned and transmitted from generation-to-generation, different social classes should exhibit differences in values, motives, and other precursors to consumption behavior.

Why has it "Failed?"

Applications of social class in consumer behavior have not been extensive. The only sustained research has involved the social class vs. income controversy, and this research cannot be regarded as definitive since non-equivalent tests were employed, including different products, diverse data collection procedures, and different operationalizations of social class. The failure of social class rests more on the prevailing orientation that has guided the research.

The basic research approach has been to: administer a standard social class index (e.g., Warner's or Hollingshead's); use the household head's (typically husband's) score as the basis for determining the family's social class; and to statistically test whether different social classes differ with regard to product-, store-, or brand-choice behavior.

Three major assumptions are implicit in this "basic model." First, it is assumed that social class is capable, by itself, of explaining choice behavior, oftentimes without regard to whether or not the consumption behavior in question is subject to cultural influence. A second assumption is the tacit belief that a husband's social class by itself is the sole determinant of a household's class position. An additional assumption is that current consumption behavior can be explained by present social class, without regard to one's class standing during that period of primary socialization influence. These assumptions are now discussed in detail as the "choice behavior," "husband-only," and "present social class" fallacies.

The Choice Behavior Fallacy. The only substantive conclusion that can be drawn from empirical applications of social class is that neither social class nor income is an adequate explanan of consumer choice behavior. It is simplistic to expect a single variable to account for behavior that is caused by multiple determinants. Kassarian (1971) has made this point cogently in his review of personality research, and it is equally applicable here.

A second aspect of the choice behavior fallacy stems from the fact that although social class is an indicator of life style, different products can be used to fulfill the same life style, and the same product can be consumed by different people, in different ways, for different reasons (cf., Kernan 1977). It is thus unreasonable to expect social class to explain and predict choice behavior. Its proper role should probably be to account for the values and motives that underlie this behavior rather than attempting to explain the behavior per se, especially when that behavior may not be subject to cultural influence (e.g., brand choice in a low-unit-value product category).

The Husband-Only Fallacy. Measurement of social class based exclusively on the class position of the household head, who is typically the husband in conventional American families, assumes that the husband's social class is the sole determinant of a household's class standing. In years past this assumption may have been tenable (in paternalistic societies the assumption still holds), but in contemporary America it is not. Indeed, nearly a third of all U.S. households include a wife who is working outside the home at least part-time (Pralle 1980). In addition to augmenting the family's purchasing power, this employment status, together with educational achievements, must have some effect on the family's class standing.

The social class categorization of households based exclusively on the husband's class indicators would be less un-

acceptable if there were a high degree of similarity (homogamy) in husbands' and wives' backgrounds and educations. Research performed by Blau and Duncan (1967) indeed reflects this. "For men 45 to 54 years of age, the correlation between husband's and wife's education is .580, and for men 55 to 64 years old it is no less than .632" (p. 190). These correlations, though apparently invalidating the above argument, actually suggest two reasons why a husband-only measure of social class is inadequate: (1) although husbands and wives share common backgrounds and educations in the majority of families-of-procreation, a large number of pairings are not homogamous; (2) assuming the lower correlation for the younger age cohort reflects a trend (Blau and Duncan did not suggest this), then American families are becoming even less homogamous.

The Present Social Class Fallacy. Operationalizing social class based on the husband's present class position assumes implicitly that every household that is classified into the same social class will exhibit the same consumption behavior because they share this one commonality. The assumption would not be so limiting if American society were free of vertical mobility, but this is not the case. The father-son correlation for occupational status is of the order of .4 (Blau and Duncan 1967). Evidently, a very considerable amount of status modification does occur.

The consumer socialization literature (Moschis and Churchill 1978; Ward 1974; Ward, Wackman, and Wartella 1977) affords an explanation of why this extensive status modification invalidates the assumption that current choice behavior is determined by one's present social class position alone. Basic consumption skills are acquired through a lifetime of information receipt and value inculcation. One's present values, beliefs, and attitudes are partially attributed to recent acculturation but, in addition, are in large part the result of the childhood enculturation process.

Since behavior is partly a function of deep-seated values, beliefs, etc., it follows that consumption behavior is largely determined by past social position in addition to present position. In short, people arrive at the same class position in different ways; just because they occupy the same class position at a particular point in time (the time at which a cross-sectional study is performed) is no reason why their consumption behavior should be the same.

This point takes on added significance when the nature of vertical mobility is examined. The greatest mobility occurs among the "sons" of skilled and semi-skilled workers, i.e., children from upper-lower and lower-middle class families are moving both up and down the class hierarchy. Since the mass market for most businesses (and researchers also) consists of the upper-lower, lower-middle, and upper-middle classes, the respondents in a social class study are the ones who are most likely to occupy a particular social class position that is different from the one in which they received their initial consumption socialization.

Extended Operationalizations of Social Class

Revision and extension of the basic model that has dominated social class research in consumer behavior is long overdue. Previous extension efforts have achieved some success. Coleman's (1960) consumer privilege concept and Peter's (1970) relative occupational class income concept have both enhanced the ability of social class to predict select consumption phenomena. Neither extension goes far enough, however,

A number of extensions are possible. However, the present effort is not intended to provide an exhaustive treatment. The aim, instead, is merely to illustrate two extended concepts and to suggest the consumption conditions when each may augment the explanatory ability of the standard operationalization of social class. The logic of discovery

rather than that of justification provides the philosophy of science rationale for these efforts (Hunt 1976).

A Husband-and-Wife Combined Measure of Social Class

The identification of a household's social class based exclusively on the husband's occupation, education, or other social class indicators has been criticized as a fundamental weakness of social class research in consumer behavior (Zaltman and Wallendorf 1979). Sociologists (e.g., DeJong et al. 1971) have also called for a re-examination of the use of husband-only indicators as the sole determinants of a family's social position. However, how husband's and wife's individual social class indicators should be combined to form an overall family social class score is an open issue.

One method would be to simply average husband's and wife's separate social classes (Haug 1973). This method, though computationally convenient, would be wrong conceptually. The United States is a "leveling-up" society: the marriage of an individual from an upper social class with one from a lower class would not produce a middle-class couple. An alternative would be to assign the higher social class of the husband or wife to the family as a whole (Haug 1973; Haug and Sussman 1971). Conceptual problems also plague this approach. In fact, it would yield results different from the husband-only approach only in those situations where a wife's social class exceeds her husband's.

Both methods ignore the joint influence of husband and wife in consumption-oriented decision making. They would structure a process wherein all consumption decisions would be predicted using the same set of spouse weightings. What is needed is a weighting scheme that assigns weights based on each spouse's relative influence in a specific decision process.

We propose a procedure that places weights on husbands' and wives' social classes proportionate to their relative influence in the consumption situation at issue. The procedure is based on two premises. First, it is assumed that much consumption behavior involves joint decision making. The foundation for this premise is solid (e.g., Davis 1976). A second premise is that the particular social class concept and corresponding operationalization used in consumer research should be situation-specific. To make an analogy, just as a single measure of IQ is inadequate across all cultures, a single social class measure is inadequate across all consumption situations.

For example, if the objective is to predict differences in product use for a product which involves a husband-dominant decision and whose use is not related to deep-seated enculturated values (say the purchase of a lawnmower), then the standard, husband-only concept may serve quite well. If, however, the objective is to account for the frequency of family vacation behavior, which is known to involve joint decision making (Davis and Rigaux 1974) and which likely is influenced greatly by childhood experience, then it is doubtful that the standard concept will account for much variance in vacation frequency.

Implementation of the proposed husband-and-wife-combined measure of social class necessitates two types of measurements. One would entail measurement of the relative influence of husband and wife in the consumption activity at issue; the other measure would involve indicators of husband's and wife's social class. Each is examined separately, and then we turn to the issue of indexing them into an overall family social class score.

Measurement of relative influence of husband and wife would require precise specification of the consumption situation, since the roles of husbands and wives are variable by decision process stage. Once this is done, it becomes a

relatively simple matter of assigning influence to each partner. A common procedure would be to have husband and wife rate themselves on rating scales ranging from "husband decided" to "wife decided" (Davis 1976). Another method would be to have respondents divide a number of points between husband and wife, using a constant sum scale, to indicate each partner's influence (Haley and Overholser 1975).

The separate social classes of husbands and wives could be measured using traditional indices: Warner's Index of Status Characteristics, Hollingshead's Index of Social Position, and Carman's Index of Cultural Classes. Although the validity of these has been challenged in terms of both the choice and weighting of indicators (Dominguez and Page 1978; Haug 1977; Haug and Sussman 1971; Jackson and Curtis 1968), we will assume for simplification sake that the indices are valid so that we can turn to the more fundamental concern of the paper.

The procedure for combining the separate social classes of husbands and wives is illustrated below. The discussion assumes that separate measures of social class have been obtained for both husband and wife, using for illustration Hollingshead's two-factor index. This procedure places each respondent into one of seven occupational and educational categories. The occupational score is weighted by seven, the educational score by four, and the indexed score is then assigned to one of five social classes (cf., Hawkins, Coney, and Best 1980). Using this procedure, the derived family social class would be a weighted sum of each spouse's social class:

$$F_{SC} = \sum_{i=1}^2 w_i SC_i \quad (1)$$

where F_{SC} is the family's overall social class; w_i is the weight for each spouse, derived from the measure of relative role for the decision process at issue; and SC_i is each spouse's social class, ranging from class 1 (the highest) to class 5 (the lowest), following Hollingshead's convention.

A simple numerical example will illustrate the procedure. Consider a decision making situation that varies from household-to-household in terms of relative influence of husband and wife. Assume for illustration that the decision involves the choice of vacation site. Data for four hypothetical families are presented.

	Decision Process Weight: Wife	Wife's SC	Decision Process Weight: Husband	Husband's SC	Family SC
Family 1	0.8	3	0.2	2	2.8 (2)*
Family 2	0.0	3	1.0	2	2.0 (2)
Family 3	0.3	2	0.7	3	2.7 (3)
Family 4	1.0	2	0.0	3	2.0 (3)

*Values in parentheses represent each family's social class based on the husband-only procedure

These illustrations show how social class assignment for a family is weighted in the direction of the spouse who has a greater decision influence. Where the husband-only measure would have categorized Family 1 as class 2, the husband-and-wife-combined measure yields an indexed score that is closer to the wife's class position, since she plays the dominant role in the particular decision. The combined procedure has the opposite effect for Family 3, where the weighting emphasizes the husband's dominant influence. The illustration for Family 4 reflects a situation where the family would be assigned to the wife's social class, because she has sole influence for this particular decision.

Family 2 reflects the only case where the husband-only index provides an equivalent result to our combined procedure. This occurs because the husband is the sole decision maker. Such a situation is the only occasion where it is theoretically appropriate and empirically prudent to designate a family's social class based exclusively on the husband's class position.

A Conjoint, Past-and-Present Measure of Social Class

The focus now turns to the issue of devising a social class index that extends beyond a family-of-procreation's present social class by incorporating the social classes of the families-of-orientation in which husband and wife received their early consumption socialization. In other words, what we are proposing is a social class index which combines "past" and "present" social classes. The considerable amount of intergenerational mobility or status modification provides the theoretical justification for such an index. It is unrealistic to expect to explain a household's current consumption behavior when this behavior in many instances is due to habits, choice rules, and other choice-determining factors that were acquired when family members occupied different (typically lower) class positions during childhood.

The rationale for combining past and present social classes is conceptually sound, but procedures for accomplishing this are not so obvious. For present purposes, we will simply conceptualize what the determining factors might be. Again, the premise is invoked that the amount of weight or influence that is assigned to the to-be-indexed components, past and present social classes, must vary by consumption situation, since there is also variability in the determinants of the particular behavior that the social class measure is attempting to explain. Vacationing behavior illustrates this point. While the decision to take a summer vacation may, for a particular family, be influenced greatly by what husband and wife were accustomed to doing as children, the decision concerning the particular vacation destination may be entirely independent of their past customs or experiences.

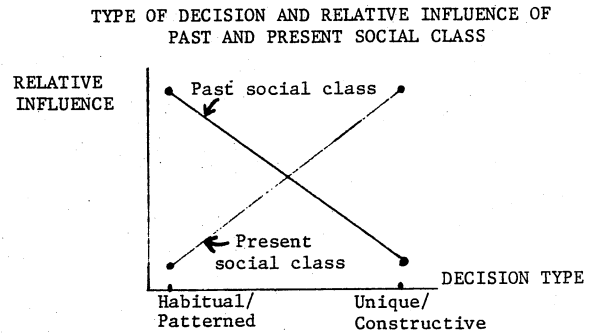
In a more general sense, we propose that the relative roles of past and present social position are a function of how "unique" the particular decision is. Where many decisions are highly habitual and based on past patterns (e.g., consumption decision relating to child-rearing practices), other decisions are unique and require the construction or creation of choice procedures (e.g., the choice of whether to try an innovative birth control method).

This logic is similar to Bettman and Zin's (1977) distinction between stored rules and constructive mechanisms, where the latter is likely employed when a choice is made for the first time or when a changed situation is encountered. We hypothesize that the more unique or constructive the decision, the less the role of past social class as a determinant of current consumption activity. In such a situation, the indexing procedure should assign less weight to past social class. The proposition is portrayed in the Figure.

The empirical upshot is the need for a procedure to identify the relative influence of past and present social classes in context of the consumption decision at issue. Though obviously not a simple task, it would be possible to devise a series of questions to get at the extent to which a family-of-procreation's decision is based on a constructive mechanism, or, instead, results from stored rules acquired during husband's and wife's respective families-of-orientation.

We close this section by suggesting that the empirical difficulties are not as insurmountable as might appear. This is predicated on our view that social class should in the first place be employed as an explanan only in those situations where the consumption activity is truly influenced by cultural class antecedents (Carman 1965). Provided that research is restricted to such situations, then adults in families-of-procreation should be capable of providing reasonably accurate responses concerning the types of behavior performed by their parents during the period of initial consumption socialization in their families-of-orientation.

FIGURE 1



Data Supporting the Need for Extended Conceptualizations of Social Class

Implicit in the foregoing arguments are two crucial assumptions. One is that husbands' and wives' social classes are indeed inconsistent in a relatively large number of households: If the amount of discrepancy were minimal, the combination of their separate classes would be moot, since the husband-only method would capture all the relevant information. A second assumption is that intergenerational mobility is sufficiently extensive to warrant the effort to devise a procedure for combining past and present classes.

A study was performed to test these assumptions. A statewide sample of approximately 800 South Carolina households received a questionnaire in 1979; measures of social class were included among many other questions. The initial mailing and a follow-up generated an 80% response rate. Respondents were somewhat upscale, mostly white, and predominantly from urban locales.

They were asked to describe as specifically as possible the occupation and education of: the husband, the wife, the husband's father, and wife's father. They were instructed to identify their fathers' occupations and highest educational achievement at that period when they (respondents) were "growing up," i.e., "when you were in elementary school." The intent was to capture fathers' social classes at the time of respondents primary consumption socialization.

Hollingshead's two-factor index was used to assign each household's responses to four possible social classes associated with husband, husband's father, wife, and wife's father. Incomplete data were extensive due to such reasons as inability or unwillingness of respondents to designate their father's occupation or education; respondents were retired or had student status, and thus no current occupation could be assigned, a spouse was not employed outside the home; respondent was unmarried; etc.

One person coded all the questionnaires on two separate occasions, with a minimum separation of one month between recoding the same questionnaire. The agreement was substantial. When coding differences could not be reconciled, the questionnaire was removed from further analysis.

Congruity of Husbands' and Wives' Social Classes

The relationship between husbands' and wives' social classes is presented in Table 1. These data represent only those households where both husband and wife were employed outside the home. A total of 311 cases, or almost exactly one half of the total responses, provided this data.

The main diagonal in Table 1 reflects instances of status congruency--a total of 46.9% of the 311 cases. Entries above the diagonal indicate instances where the husband's

TABLE 1

CONGRUENCY BETWEEN HUSBANDS' AND WIVES' SOCIAL CLASSES

Husband's social class	Wife's social class given husband's class					Totals
	1	2	3	4	5	
1 ^a	0	5	6	0	0	11
2	0	38	75	0	0	113
3	0	24	94	5	0	123
4	0	13	34	14	0	61
5	0	0	2	1	0	3
Totals	0	80	211	20	0	311

^aIn the Hollingshead two-factor procedure, lower numbers represent higher social classes.

social class exceeds his wife's. There were 29.3% such cases. Though not surprising, it is noteworthy that the higher the husband's social class, the greater the probability that the wife's class was lower than his, with the obverse holding as well.

The most revealing result concerns the prevalence of cases where the wife's social class exceeds her husband's. There were a total of 23.8% entries, or nearly one-fourth of all cases, below the diagonal. Considering the upscale sample involved in this analysis, it is likely that much greater frequency of wife's-class-greater-than-husband's would be found in a more representative sample. The implication is clear: The traditional, husband-only measure of social class probably under-represents the true social class and purchasing power of a large percentage of families.

Intergenerational Mobility

Table 2 presents data indicating the relationship between the social class of husbands and wives in families-of-procreation with their respective father's social class. There were a total of 457 cases where data were available on husband's and husband's-father's social classes, and a total of 325 cases of wife's and wife's-father's social classes.

Entries on the main diagonal reflect the absence of intergenerational mobility, i.e., children in families-of-procreation possess the same social class as their fathers. The proportion of such cases is essentially equivalent for husbands and wives, 26.3% for the former and 24.9% for the latter. In other words, approximately three-fourths of all families-of-procreation have experienced status modification.

Upward mobility is indicated in the entries below the diagonal, which reflect fathers' social classes as lower than children's. In fact, 65.2% of all the husbands experienced upward mobility, while a slightly larger 67.1% of the wives experienced this mobility.

Entries above the diagonal reveal relatively few instances of downward mobility; there were only 8.5% cases of this nature for husbands, and 8% such cases for wives. These data are harmonious with previous results (e.g., Blau and Duncan 1967) indicating the extensive upward mobility in American society.

Discussion

These data have obvious limitations; foremost, perhaps, is the non-representative sample. However, the data are presented for illustration purposes only, and are not intended as definitive. Limitations aside, the data clearly manifest extensive intergenerational mobility as well as

TABLE 2

RELATIONSHIP BETWEEN FAMILY-OF-PROCREATION AND FAMILY-OF-ORIENTATION SOCIAL CLASS

Father's social class	Husband's and wife's social class given their father's class										Totals	
	1		2		3		4		5		H	W
	H	W	H	W	H	W	H	W	H	W	H	W
1	HF	1	14		4		0		0			19
	WF		0		4		7		0		0	11
2	HF	11	36		9		2		0			58
	WF		0		11		14		1		0	26
3	HF	7	59		46		6		1			119
	WF		2		27		60		0		0	89
4	HF	6	59		92		37		3			197
	WF		0		40		105		10		0	155
5	HF	1	19		21		23		0			64
	WF		0		6		30		8		0	44
Totals	HF	26	187		172		68		4			457
	WF		2		88		216		19		0	325

Note: H=Husband's social class; W=Wife's class; HF=Husband's-father's class; WF=Wife's-father's class.

significant discrepancies between the social classes of husbands and wives in those households where both partners have employment outside the home.

The upshot is that the traditional approach to social class research in consumer behavior is conceptually and empirically flawed. Measuring only the husband's social class and then attempting to predict choice behavior based on this information alone assumes that the husband's class is the sole indicator of a family's class standing. This assumption is untenable in view of the fact that women from approximately one-third of all households are in the work force, and many possess indicators of higher class than their husbands'. We have proposed a method which takes into consideration the social class of both husband and wife, and combines the two by weighting each by the degree of influence each partner has in the particular decision process that is subject of inquiry.

It is also unreasonable to assume that current behavior is solely a function of a household's present class standing. A present-only operationalization disregards past socialization influence. The empirical problem resulting from this is that a sample of households assigned to the same social class by a present-only measure may be quite heterogeneous with regard to values, motives, and other real determinants of consumption behavior.

It is one thing to recommend the indexing of past-and-present social classes into an overall conjoint scale. Doing so is another matter. We have conceptualized a general procedure for determining the relative weights of past and present classes, but specific procedures await future research.

The procedure for combining the separate social classes of husbands and wives is feasible, however; it is merely a matter of doing the research to ascertain whether a combined husband-and-wife index provides superior predictions to the simple, husband-only procedure that has dominated past research.

REFERENCES

- Bettman, James R., and Zins, Michel A. (1977), "Constructive Processes in Consumer Choice," Journal of Consumer Research, 4, 75-85.
- Blau, Peter M., and Duncan, Otis D. (1967), The American Occupational Structure, New York: John Wiley & Sons.
- Carman, James M. (1965), The Application of Social Class in Market Segmentation, Berkeley: Institute of Business and Economic Research.
- _____ (1978), "Values and Consumption Patterns, A Closed Loop," in Advances in Consumer Research, ed., H. Keith Hunt, Ann Arbor: Association for Consumer Research.
- Coleman, Richard P. (1960), "The Significance of Social Stratification in Selling," in Marketing: A Maturing Discipline, ed., Martin L. Bell, Chicago: A.M.A.
- Davis, Harry L. (1976), "Decision Making within the Household," Journal of Consumer Research, 2, 241-260.
- _____, and Rigaux, Benny P. (1974), "Perception of Marital Roles in Decision Processes," Journal of Consumer Research, 1, 51-61.
- DeJong, P. Y., Brewer, M. J., and Robin, S. S. (1971), "Patterns of Female Intergenerational Occupational Mobility," American Sociological Review, 36, 1033-42.
- Dominquez, Luis V., and Page, Albert L. (1978), "Stratification in Consumer Behavior Research: A Re-examination," Case Western Reserve Univ. Working Paper.
- Foxall, Gordon R. (1975), "Communication Note: Social Factors in Consumer Choice: Replication and Extension," Journal of Consumer Research, 2, 60-64.
- Haley, Russel, Overholser, C. C., and Associates (1975), "Purchase Influence: Measures of Husband/Wife Influence on Buying Decisions," New Canaan, CT: Haley, Overholser, and Associates, Inc.
- Haug, Marie R. (1973), "Social Class Measurement and Women's Occupational Roles," Social Forces, 52, 86-98.
- _____ (1977), "Measurement in Social Stratification," Annual Review of Sociology, 3, 51-77.
- _____, and Sussman, Marvin B. (1971), "The Indiscriminate State of Social Class Measurement," Social Forces, 48, 549-63.
- Hawkins, Del I., Coney, Kenneth A., and Best, Roger J. (1980), Consumer Behavior: Implications for Marketing Strategy, Dallas: Business Publications, Inc.
- Hisrich, Robert D., and Peters, Michael P. (1974), "Selecting the Superior Segmentation Correlate," Journal of Marketing, 38, 60-63.
- Hunt, Shelby D. (1976), Marketing Theory: Conceptual Foundations of Research in Marketing, Columbus, OH: Grid.
- Jackson, E. F., and Curtis, R. F. (1968), "Conceptualization and Measurements in the Study of Social Stratification," in Methodology in Social Research, eds., H. M. Blalock and A. B. Blalock, New York: McGraw-Hill.
- Jain, Arun K. (1975), "A Method for Investigating and Representing an Implicit Theory of Social Class," Journal of Consumer Research, 2, 53-59.
- Kassarjian, Harold H. (1971), "Personality and Consumer Behavior: A Review," Journal of Marketing Research, 8, 409-19.
- Kernan, Jerome B. (1977), "Retrospective Comment on Martin-eau's 'Social Classes and Spending Behavior,'" in Classics in Consumer Behavior, ed. Louis E. Boone, Tulsa, OK: The Petroleum Publishing Co.
- Levy, Sidney J. (1966), "Social Class and Consumer Behavior," in On Knowing the Consumer, ed. Joseph W. Newman, New York: John Wiley & Sons.
- Martineau, Peirre (1958), "Social Classes and Spending Behavior," Journal of Marketing, 23, 121-30.
- Mathews, Herbert L., and Slocum, John W. (1969), "Social Class and Commercial Bank Credit Card Usage," Journal of Marketing, 33, 71-78.
- Moschis, George P., and Churchill, Gilbert A. (1978), "Consumer Socialization: A Theoretical and Empirical Analysis," Journal of Marketing Research, 15, 599-609.
- Myers, James H., and Mount, John F. (1973), "More on Social Classes Vs. Income As Correlates of Buying Behavior," Journal of Marketing, 37, 71-73.
- _____, Stanton, Roger P., and Haug, Arne F. (1971), "Correlates of Buying Behavior: Social Class Vs. Income," Journal of Marketing, 35, 8-15.
- Nicosia, Francesco M., and Mayer, Robert N. (1976), "Toward a Sociology of Consumption," Journal of Consumer Research, 3, 65-75.
- Peters, William H. (1970), "Relative Occupational Class Income: A Significant Variable in the Marketing of Automobiles," Journal of Marketing, 34, 74-77.
- Pralle, Michael E. (1980), "Survey Dispels Myth About Growing Dual-Earner Market," Marketing News, May 16, 17-18.
- Prasad, V. Kanti (1975), "Socio-economic Product Risk and Patronage Preferences of Retail Shoppers," Journal of Marketing, 39, 42-47.
- Rich, Stuart U., and Jain, Subhash C. (1968), "Social Class as Predictor of Shopping Behavior," Journal of Marketing Research, 5, 41-49.
- Sheth, Jagdish N. (1977), "Demographics in Consumer Behavior," Journal of Business Research, 5, 129-38.
- Slocum, John W., and Mathews, H. Lee (1970), "Social Class and Income as Indicators of Consumer Credit Behavior," Journal of Marketing, 34, 69-73.
- Ward, Scott (1974), "Consumer Socialization," Journal of Consumer Research, 2, 1-16.
- _____, Wackman, Daniel B. and Wartella, Ellen (1977), How Children Learn to Buy, Beverly Hills: Sage Publications.
- Wasson, Chester R. (1969), "Is It Time to Quit Thinking of Income Classes," Journal of Marketing, 33, 54-57.
- Zaltman, Gerald, and Wallendorf, Melanie (1977), "Sociology: The Missing Chunk or How We've Missed the Boat," A.M.A. Educators' Proceedings, Series #41, 235-38.
- _____, (1979), Consumer Behavior: Basic Findings and Management Implications, New York: John Wiley & Sons.

AN ANALYSIS OF ALCOHOL ADVERTISING USING
FRENCH AND RAVEN'S THEORY OF SOCIAL INFLUENCE¹

Scott B. MacKenzie, University of California at Los Angeles,²
Judy L. Zaichkowsky, University of California at Los Angeles

Abstract

Previous research on the content of advertising has found limited use of information based appeals and has asserted that advertising is typically non-informative in nature. The present study investigated this assertion using a theoretical classification system based on French and Raven's Theory of Social Influence to classify alcohol advertisements. Results of the study showed that within the product category studied, information based appeals were used far more frequently than previously thought. In particular, wine ads were found to concentrate primarily on informational and expert based appeals, while liquor and beer ads concentrated more on referent, reward and coercive appeals.

Introduction

Perhaps one of the most often asked questions in advertising research is "What does advertising do?" Economists have contended that the value of advertising is in the information it provides the consumer. However, it is not clear exactly how much information is actually conveyed in most advertisements. Similarly, in recent years, public policy makers have expressed their concern with how advertisements influence the consumers' choice of products and if advertisements supply consumers with enough information to make "informed choices". In response to these concerns research into the information content of advertising has been carried out by various researchers (Resnik and Stern 1977, Marquez 1977, Shimp 1979 and Pollay et al 1980). Results of these studies show only limited support for the contention that advertising provides information to consumers upon which they can then make informed choices. But if advertisements are not making information based appeals then what types of appeals are being made? It is the purpose of this study to examine the content of advertising using a theoretical framework that will provide insight not only into the information content of advertising but also into the types of non-informational appeals being used.

Review of the Literature

Research dealing with content analysis of television advertising (Resnik and Stern 1977 and Pollay et al 1980) has indicated that only about 50% of all advertising supplies some information cues to the consumer. Resnik and Stern (1977) analyzed the content of 378 videotaped television ads randomly drawn from those aired on the major networks during April 1975, while Pollay et al (1980), using exactly the same methodology analyzed the content of 884 television ads from 1971, 1973 and 1977.

These studies were very generous with their definition of informational cues and looked for these cues along fourteen and twenty dimensions respectively. All specific claims were accepted as valid and classified as information. Assertions in the ad copy were coded so that ads could be

classified as providing information on any or all dimensions, (rather than coding just the dominant theme or 'copy premise'). For example, an informational cue of convenience would be an assertion by the ad that the product is easy to use and/or easier than some alternative. Some of the other cues classified were price, safety, availability, taste, guarantees, etc.. The Resnik and Stern (1977) findings indicated only about 1% of the ads contained three or more informational cues, some 16% provided two cues, and slightly more than half (50.8%) provided no informational cues at all. The Pollay et al (1980) findings also concluded that the advertising studied presented little in the way of informational cues to the consumer. One of the problems with studying the content of advertising in this way is that no information is gained about the types of non-informational appeals being made. If in approximately half of the ads studied no informational cues were given, then what are the advertisers saying in their commercials?

A study by Marquez (1977)³ using print advertisements attempted to address this issue. Using a variety of nationally distributed magazines, he classified 600 advertisements on the dimensions of persuasion, information and intimidation. An advertisement was classified as basic persuasion if it did not contain specific facts about the product and if the claims made could not be verified. Basic information is the converse of basic persuasion and was defined as the specific, relevant and verifiable facts about an advertised product. Intimidation was defined in this study as an act of creating a desire for the product through the use of fear appeals.

Marquez reported that 43.3% of advertisements were basic persuasion, 22.8% of the advertisements were basic information, 22.4% of the ads were high persuasion/low information, 10.2% were high information/low persuasion ads and that 1.2% were primarily intimidation based. Persuasive appeals were thus used twice as often as information based appeals. In addition, he found that the different types of appeals were used more frequently for some product categories than others. Wine and liquor advertisements contained more persuasive appeals than any other product category, while institutional advertising relied heavily on information based appeals.

Marquez's study was a step in the right direction in that he attempted to classify, albeit broadly, both the informational and non-informational types of appeals being made in the advertisements. However, his operational definitions completely lacked any theoretical basis. The definition of persuasion was derived from the dictionary as "an act of causing (someone) to do or believe something, especially by reasoning, urging, etc.; convince". Similarly the definition of information was also based on the dictionary definition of the word, "something told or facts learned; news or knowledge". These definitions are not mutually exclusive. The "reasoning" mentioned in the definition of persuasion could possibly be expected to include "facts, news or knowledge".

What is needed is a classification system that is more soundly based on theory. It is the purpose of this study

¹The authors would like to thank Carol Scott and Jim Bettman for their comments on an earlier draft of this paper.

²Scott MacKenzie and Judy Zaichkowsky are doctoral students, Department of Marketing, Graduate School of Management, UCLA.

³This study made no mention of who exactly judged the ads or of any reliability measure usually associated with content analysis.

to propose such a classification system and use it to investigate the types of appeals being made in advertising. The theoretical framework proposed is French and Raven's theory of social influence, and the advertisements classified will be printed alcoholic beverage ads. This medium and product category were chosen in order to make the ads in this study as similar as possible to those used by Marquez, so that any differences in the results obtained could be attributed to the different classification systems used.

Marquez found wine and liquor ads during 1973-1976 to be highly persuasive in nature and very low in basic information. The present study samples print alcohol advertising from the same time period. Ideally in order to make this study comparable to the previous studies a random sample of all types of ads should have been used. However, since this research is somewhat exploratory in nature, it was deemed sufficient to select one category of advertisements previously judged as non-informational and compare the results using a classification system that is conceptually more precise.

French and Raven's Theory of Social Influence

French and Raven (1959) define influence as a change in a person's cognition, attitude, or behavior, which has its origin in another person or group. Their theory is appropriate for the study of advertising because advertising can really be defined as an attempt on the part of one person or group (the sponsoring firm) to influence the purchase behavior of another person or group (the consumer).

In extending the theory, Raven (1965) identified two types of influence; socially dependent and socially independent influence. Independent influence depends entirely on the "content of the communication" between the influencing agent and the actor, whereas dependent influence depends not only on "what is said" but on "who said it". Raven (1965) theorized that independent influence is the result of a basic change in cognitive elements having its basis in information communicated by the influencing agent. Here the content of the communication is of primary importance, not the nature of the influencing agent.

Five different types of dependent influence were identified: reward,⁴ coercive, expert, referent and legitimate. Reward power stems from the ability of the influencing agent to mediate rewards for the influencee; coercive power results from the ability of the agent to mediate punishment. (It should be noted that rewards can either be physical or symbolic). Expert power (or influence) stems from the attribution of superior knowledge or ability to the influencing agent. The influencing agent then functions as a guide for the other person indicating the path which will most likely lead him to his goals. It is necessary both for the person to think that the influencing agent knows the truth, and for the person to trust that the influencing agent is telling the truth, for expert influence to occur. Referent power results from the person's identification with the influencing agent and his desire to maintain similarity with him. By identification they mean a feeling of oneness between the influencing agent and the person or a desire for such an identity. Legitimate power stems from internalized values in the person which dictate that the influencing agent has a legitimate right to influence them. In other words, the person feels obligated to accept this influence. In all cases the concept of legitimacy involves some sort of code or standard accepted by the individual, by virtue of which the external agent

4

Following Raven (1965), the terms power and influence are used somewhat interchangeably. Raven defines power as potential influence and influence as kinetic power.

can assert his power. (Only a brief description is possible here, see French and Raven 1959 for more details).

In this study a group of judges were asked to determine the extent to which each of the above mentioned types of appeals were presented in a set of advertisements. The judges responded on seven point scales ranging from "definitely not implied" to "definitely implied". The influence bases were operationalized as follows:⁵

1) Information: The ad explains why this is a good product, citing available evidence such as quality of ingredients, caloric content, price, taste, etc.. An example of an informational appeal would be,

"Lite beer has one third less calories than your regular beer."

2) Expert: The ad implies that someone who is an expert, who knows about such products, thinks this product is the best and recommends it. For example, having a wine master stating that, in his opinion, a certain product is the best, would be an expert appeal.

3) Referent: The ad implies that other people, whom you admire or want to identify with, use the product. The following is an example of a referent appeal for Ambrosia Liqueur,

"Are you an Ambrosiac? Have you ever seen the sun rise? Did you ever swim in the nude? Do you enjoy walking on autumn leaves? Have you ever seen the same movie three times?"

4) Reward: The ad implies that if you use the product you will experience some positive consequences from others; i.e., social approval, acceptance, sexual success. The following appeal implies some increase in social approval or sexual success if the product is used,

"Gene Robins recently introduced his girlfriend to Ballantine's Scotch. She, in turn, introduced five friends to Ballantine's. Now Gene Robins has six girl friends. Moral: It pays to be loyal."

5) Coercive: The ad implies that if you do not use this product you will suffer some negative consequences from others; i.e., loss of social approval or prestige, rejection. The following appeal implies some loss of social prestige if the proper wine is not selected,

"Will your chilled wine get a chilly reception? Will the white be wrong? The Chablis shabby?"

These operationalizations of information, expert and referent power are directly derived from French and Raven's definitions. However, the definitions of reward and coercive power used in this study were not exactly the same as French and Raven's. The definitions of reward and coercive had to be broadened to include rewards and punishments mediated by third parties because there obviously was no direct contact between the sponsors of the advertisements and their target audiences. Appeals of this type are really a mixture of information and reward/coercion. The ad gives the person information, but the information is not about the product or its benefits, but concerns what other people will do if the product is used. Therefore, in this study a reward or coercion based appeal is one that gives information concerning what other people (who presumably have direct contact with the target audience) will do if the person does or does not do as the advertisement says.

Legitimate based appeals were excluded from the analysis

⁵The authors would like to thank Bert Raven for his help in operationalizing these concepts.

because it is difficult to imagine any truly "legitimate" advertising themes for alcoholic beverages. Such an appeal would have to persuade you to buy the product because you feel you 'ought' to buy it. Even if such appeals have been used on occasion in the past it is felt that their occurrence would be so rare that their exclusion would not significantly effect the analysis.

Methodology

The Sample

The sample of beer, wine and liquor ads chosen for the study was drawn from a large group of nationally distributed magazines. This medium is representative of alcohol advertising because all liquor and most wine advertising appears in print media. While much of the advertising for beer appears on television, industry experts and advertising agency executives indicated that the same themes and campaigns were used on television and in print advertising. Thus, even though most beer advertising does not appear in print, the print ads are still representative of beer advertising in general.

For each of the eight years between 1971 and 1978, fifteen ads were selected; five liquor ads, five wine ads and five beer ads. One ad for each beverage was randomly selected from each of the five magazines carrying the most advertising pages for the product categories during the year. These five top magazines accounted for 30% to 55% of the total alcoholic beverage advertising pages appearing during the year.

Using the sampling procedure described above, 120 ads were selected. A 35mm color slide was made of each ad. The ads were then randomly assigned to one of the three coding sessions.

Subjects

The judges were introductory psychology students who participated as a part of a course requirement. Ninety coders participated, 30 in each coding session. However, not all of the judges completed the task, thus reducing the number of coders in each session to approximately 24.

Procedure

At the beginning of each coding session the experimenter carefully explained the various power bases to the judges giving them numerous examples, and then answering any questions that arose. All of the judges indicated that they fully understood each of the types of appeals they were to rate. The judges were instructed to consider only the advertisements they were currently viewing when making their judgements and to ignore any images or concepts they had developed from prior experience. (Obviously in some instances it is impossible to ignore past experiences, however in this situation, given the relatively straight-forward judgment task, it seemed reasonable to assume this was possible).

Each slide was projected onto a screen and rated by a group of approximately 24 judges. The judges were asked to: 1) Rate the extent to which each of the five social influence appeals was used in each advertisement and 2) Indicate which of the five appeals they thought was the main point of the ad.

Results and Discussion

The purpose of this exploratory study was to examine the types of appeals made in alcohol advertisements and to assess the usefulness of French and Raven's Theory of Social Influence in classifying them. Consequently, the analyses focused on determining: 1) the extent to which each of the theoretical bases of social influence was used

in the ads studied, and 2) how reliable the appeals could be classified.

The extent to which each of the five types of appeals was present in a given ad was determined by calculating the mean rating for each type of appeal averaging across all judges who rated the ad. For a given product (i.e. beer, wine or liquor), the strength of each type of appeal was assessed by averaging the ratings across all judges and all the ads in that category. The product class mean ratings for each type of appeal were the result of averaging across all judges and all ads. Table 1 contains the mean ratings for each type of appeal for the three products and over all products.

TABLE 1

MEAN RATINGS OF ADVERTISING CONTENT

	<u>REFERENT</u>	<u>INFO</u>	<u>COERCIVE</u>	<u>EXPERT</u>	<u>REWARD</u>
Liquor	3.09	3.49	2.14	2.19	3.83
Wine	2.38	4.57	1.84	3.20	3.20
Beer	2.83	3.23	1.91	2.09	3.74
All Ads	2.75	3.78	1.95	2.51	3.58

The strength of each type of appeal or the extent to which each type of appeal was present is indicated by the magnitude of the mean ratings. From the bottom line of Table 1 it can be seen that information had the highest mean rating (3.78) followed by reward (3.58), referent (2.75), expert (2.51), and coercive (1.95). These means are significantly different at the .001 level, thus indicating that, overall, information based appeals are the strongest type of appeal made in alcohol advertising.

Table 2, which contains the percentage of ads in which each type of appeal was judged to be the main point, also supports this ordering of effects. Thirty-nine percent of the ads were judged to have information based appeals as their main point. Again, reward, referent, expert, and coercive, followed in precisely the same order.

TABLE 2

PERCENT OF ADVERTISING HAVING EACH TYPE OF APPEAL AS THE MAIN POINT

	<u>REFERENT</u>	<u>INFO</u>	<u>COERCIVE</u>	<u>EXPERT</u>	<u>REWARD</u>	<u>TOTAL</u>
Liquor	22%	34%	5%	7%	32%	100%
Wine	10%	51%	4%	17%	17%	100%
Beer	20%	32%	4%	6%	38%	100%
Total	17%	39%	4%	10%	29%	100%

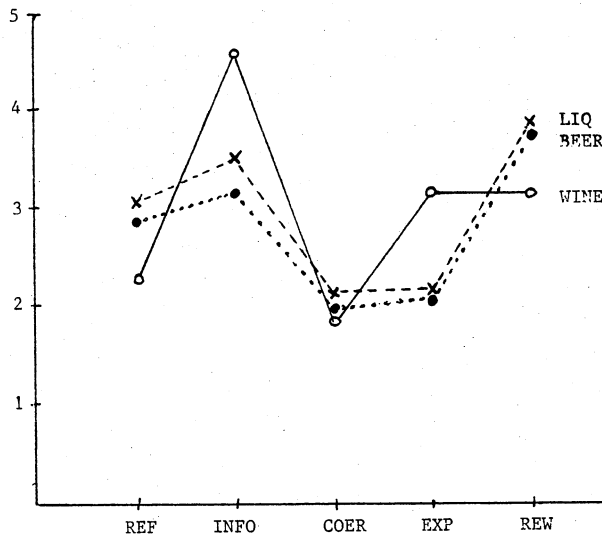
The strength of the information based appeals observed in this study are in direct opposition to Marquez's findings. Both studies sampled ads from virtually the same product category published during the same years. (Marquez sampled wine and liquor ads published between 1973 and 1976, this study sampled wine, liquor and beer ads published between 1971 and 1978). Yet Marquez observed that the ratio of 'persuasive' to 'informational' ads in this product category was 37 to 1. In this study, if all the source dependent appeals (reward, coercion, expert, and referent) are combined into a 'persuasive' category similar to Marquez's, the ratio of persuasive appeals to information based appeals is only 1.5 to 1 (see Table 2). A difference of this mag-

nitude indicates how sensitive the results are to the type of classification system chosen. The choice of the classification system thus becomes an extremely important decision, one that should be as soundly based on existing theory as possible.

The above discussion has been concerned with the overall nature of alcohol advertising as a product class. In looking at the results for beer, wine, and liquor separately, some interesting differences in advertising strategy become apparent. A multivariate analysis of variance, using the judges ratings of referent, information, coercive, expert and reward appeals as the dependent measures and the type of product (beer, wine or liquor) as the grouping variable, was performed to investigate the possibility that the ads for the three types of products reflected different advertising strategies. An overall F of 52.1, which was significant at the .001 level, was obtained. Group level differences across the three product types were tested for by summing each judge's ratings on the five types of appeals and then using this summed variable as the dependent measure in a univariate analysis of variance where product type was again the grouping variable ($F = 22.15, p < .001$). It was also important to determine whether there was an interaction between the type of product advertised and the kind of appeal being used. To test this the following four quantities were computed for each judge on every ad: referent minus information; information minus coercion; coercion minus expert; and expert minus reward. These quantities then became the dependent measures in a multivariate analysis of variance designed to test for product-appeal interactions. An F of 57.27 ($p < .001$) was obtained thus indicating that the profiles for the three types of products are not parallel. In sum, as shown in Figure 1, the three types of alcoholic beverages use significantly different advertising strategies.

FIGURE 1

ADVERTISING STRATEGIES BY PRODUCT TYPE



From the figure it appears that liquor and beer ads are basically the same but that they differ in content from the wine ads. This is in fact the case for referent, information, expert and reward mean ratings. The wine mean ratings for these variables are all significantly different ($p < .001$) from either the liquor or beer means, while the liquor and beer means never differ significantly. (The fact that the coercion mean ratings did not follow this pattern

is not too disturbing since coercion was the weakest of the five types of appeals and was the main point of only 4% of the ads studied).

Furthermore, it can be seen that wine ads use significantly ($p < .001$) stronger information and expert based appeals, while liquor and beer ads use significantly stronger referent, reward and coercive appeals. (Table 2 confirms this pattern of results). It appears that liquor and beer are sold primarily on the basis of "social" appeals, while wine is sold on the basis of "rational" appeals. Referent, reward and coercive appeals are all based on the social consequences of using or not using the product. Information and expert based appeals are in a sense more rational because the decision to purchase the product is based on either information about the product or on expert testimony concerning the product. We might say the person purchases the product because they believe it is better than the competition, rather than because they are concerned with what other people will do or think.

As is typically the case in content analyses, the reliability of the judges ratings was assessed by determining the percentage of times their judgments agreed. In order to do this it was necessary to transform the judges ratings, which were continuous in our study, into dichotomous variables. This was done by recoding the judges ratings so that a rating of 1, 2, or 3, was scored as a zero on the new variable and a 4, 5, 6, or 7 was scored as a one. The newly created variable crudely represents the absence (0) or presence (1) of a given type of appeal. The percentage of agreement between the judges was then calculated for each appeal on each ad. These percentages were then averaged across the appropriate groups of ads to produce the percentage agreements shown in Table 3.

TABLE 3

PERCENT AGREEMENT BETWEEN JUDGES

	REFERENT	INFO	COERCIVE	EXPERT	REWARD	AVERAGE
Liquor	75%	75%	81%	80%	70%	76%
Wine	80%	78%	90%	72%	66%	77%
Beer	75%	79%	86%	82%	67%	78%
Average	77%	77%	86%	78%	68%	77%

Across all types of ads and all types of appeals the judges agreed an average of 77% of the time (93% of the ads had greater than 70% inter-judge agreement). In order to determine how sensitive these percentages were to the way in which the judge's ratings were recoded, an alternate recoding was tried, (1, 2 versus 3, 4, 5, 6, 7). The overall percentage of agreement was 75%. On the basis of these figures in Table 3, we can therefore conclude that the judges' ratings of the types of appeals used in the ads were quite reliable.

Summary and Conclusions

One of the most interesting findings of this exploratory study was the extent to which the advertisements were classified as having informational based appeals. This finding is in contradiction to Marquez's (1977) study and could be due to two points; 1) the fact that the present study only looked at alcohol advertisements, so there was no room for comparison across product categories, and 2) the different classification schemes used in the two studies had a significant impact on the conclusions reached. It could be the case that the use of French and Raven's framework results in a higher percentage of the appeals being classified as informational, but that this increase occurs for all product

categories, thus leaving the relative amount of information unchanged.

Perhaps the main contribution of this paper is the theoretical framework proposed to study and classify the advertisements. French and Raven's Theory of Social Influence not only allows one to distinguish between informational appeals and non-informational appeals but also distinguishes between the various types of non-informational appeals. This information would be very helpful in assessing the value of advertising from both an economic and public policy point of view. For example, public policy makers might become less concerned with controlling advertising for product categories that concentrate on 'rational' appeals and perhaps more concerned with product categories that rely on ads using 'social' or 'emotional' appeals.

Although the present study used print media, the classification scheme also seems adaptable to television advertising. The next step in developing the application and suitability of the French and Raven framework, would be to classify a cross section of print advertisements to see if the results are comparable. It would also be interesting to reclassify the Resnik and Stern (1977) or the Pollay et al (1980) television advertisements to determine what kinds of appeals are being made in that 50% non-informational category.

Bibliography

French, J.R.P. and Raven, Bertram H., (1959), "The Bases of Social Power", in D. Cartwright (ed.), Studies in Social Power. Ann Arbor: University of Michigan Press.

Marquez, Francis (1977), "Advertising Content: Persuasion, Information, or Intimidation", Journalism Quarterly, 3, 482-491.

Pollay, Richard W., Zaichkowsky, Judy L., and Fryer, Christina (1980), "Regulation Hasn't Changed Advertising Much", Journalism Quarterly, 3

Raven, Bertram H., (1965), "Social Influences and Power", in I.D. Steiner and M. Fishbein (eds.), Current Studies in Social Psychology. New York: Holt, Rinehart and Winston; pp. 371-382.

Resnik, Allan and Stern, Bruce L., (1977), "An Analysis of Information Content in Television Advertising", Journal of Marketing, 1, 50-53.

Shimp, Terrance A. (1979), "Social Psychological (Mis)Representations in Television Advertising", Journal of Consumer Affairs, 2, 28-40.

THE INFLUENCE OF GEOGRAPHIC SUBCULTURES IN THE UNITED STATES

Del I. Hawkins, University of Oregon
 Don Roupe, University of Oregon
 Kenneth A. Coney, Arizona State University

Abstract

Geographic subcultures can be an important variable for detecting differences in consumptive and nonconsumptive behaviors. Physical region and urban-rural-suburban categories provide the basis for specifically identifying regional subcultures, while the nature of geographical influence is defined as being physical or psychological. There appears to be reason to believe that geographic subcultures are fairly stable and that significant behavioral differences of interest to marketers do exist. An explanatory model of geographic subcultural influence is presented, as well as a format for analyzing that influence in the development of marketing strategy.

Introduction

Are there differences in consumption behaviors between individuals that can be predicted or explained by their membership in differing geographic subcultures? If so, what is the nature of the differences? Are these differences likely to increase, decrease, or remain stable during the 1980's? These important questions have largely been ignored in the marketing literature. (For an exception, see Revzan 1978). Yet, as this paper will demonstrate, geographic subcultures exert strong influences on the consumption patterns of their members. These influences have important implications for both marketing practice and the development of marketing theory.

The purpose of this paper is: (1) to describe the nature of geographic subcultures; (2) to point out some of the consumption and nonconsumption behaviors influenced by geographic subcultures; and (3) to present a process for developing marketing strategies based on geographic subcultures.

Nature of Geographic Subcultures

A subculture has been defined as a segment of a culture which shares distinguishing patterns of behavior (Robertson 1970). It is important to remember that subculture members are also members of the dominant culture. Thus, subculture members generally have many behaviors in common with members of the broader culture. However, for a group to constitute a subculture, its members must also share certain behavior patterns that are not shared by most members of the dominant culture. When these unique behaviors affect the consumption process, they become of interest to marketers.

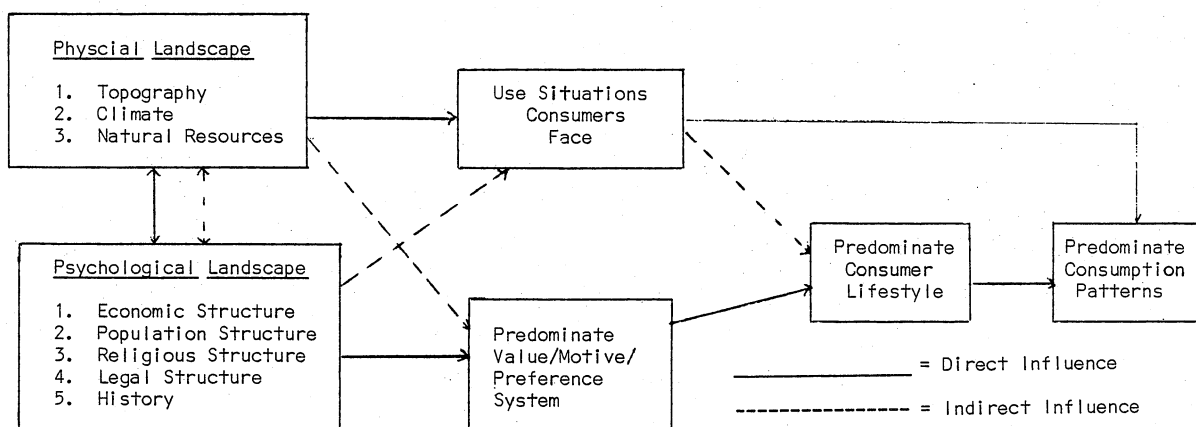
Based on the above definition of a general subculture, a geographic subculture can be defined as a part of the country whose residents share patterns of behavior that are distinct from those of the remainder of the country. There are two fundamental schemes for isolating and defining geographic subcultures: region of the country categories and urban-suburban-rural categories. These schemes may be used either singularly or in combination. Secondary data sources such as the Census Bureau, Target Group Index, and others generally provide underlying data using one or both of these schemes. The factors producing and maintaining the unique behavior patterns associated with a subculture are of the same type for either classification system.

A Model of Geographic Subculture Influences

Figure 1 presents a model of the factors that compose a geographic subculture and the relationship of those factors to the consumption process. The first and most obvious factor is physically bound and influence consumer

FIGURE 1

MODEL OF GEOGRAPHIC SUBCULTURE INFLUENCE ON CONSUMPTION



use situations, and hence predominate consumption patterns, directly. For example, individuals living in very cold climates face particular consumption use situations because of the climate. Warm clothing, antifreeze, insulation and so forth are examples of product purchases caused by physical features of a geographic subculture. A change in physical features (eg. moving to a warm climate) will generally change the use situations consumers face and hence the predominate consumption patterns related to those situations. Physical geographical factors also can affect the predominate value/motive/preference system that is the basis of consumer lifestyle. This influence is more indirect, however.

The second type of influence proposed in the model is psychologically bound. These influences affect relatively permanent characteristics of the individual and have a more enduring nature. Should an individual leave a physical geographical region, the attitudes and behaviors formed as a result of psychological bounds are likely to be carried along. Likewise, an individual moving into a new geographic location will not generally adopt the geographically bound attitudes and behaviors at once, but will have to undergo a period of acculturation before adopting some or all of the new psychologically bound behaviors. Psychological influences directly affect a geographical area's predominate value/motive/preference system and hence the predominate consumption lifestyle of the region. This lifestyle in turn is directly related to consumption patterns since a major reason for many purchase decisions is the desire to maintain or enhance a desired lifestyle.

A brief examination of the major elements of the model will clarify further the nature of geographic influence.

Physical Landscape

The physical landscape is composed of the area's topography, climate, and resources. The topography and climate have a major direct influence on geographically bound behaviors. In combination with the area's resources, they also have a direct influence on the human landscape. That is, the physical landscape determined, in part, the type of people who settled in the region and how those individuals evolved over time. Likewise, the individuals who settled in the area affected the natural resources through either development or depletion activities. The physical environment also affects the value/motive/preference system. For example, relatively "hostile" climates are likely to produce groups that value survival-related skills and activities.

Human Landscape

The human landscape is composed of the economic, population, religious, historic, and legal structures of the area. Obviously, these are not independent variables. Furthermore, each variable may exert direct influences on the value/motive/preference system or the consumption process independent of the geographic region. For example, the Mormon religion influences its followers independently of the area of the country that they are in. However, it becomes a factor in a geographic subculture when it influences the predominate value/motive/preference system of the area, or when it directly influences the behaviors of most members of the region even if they are not Mormons.

Predominate Value/Motive/Preference System

As an area develops a human history, it also develops a value/motive/preference system that is shared in varying degrees by most of the members of the area. New members of the area tend to acquire this value/motive/preference system over time (acculturation). A value can be defined as a widely held belief that affirms what is desirable and has some impact on activities (Nicosia and Mayer 1976). For example, the West Coast and the Northeast regions of the

United States differ rather sharply on the tradition-change value. Learned motives, such as the need for achievement, also vary across geographic regions (McClelland 1961). Preference refers to acquired tastes in such areas as food and architecture which obviously vary from region to region.

Predominate Lifestyle Preferences

Lifestyle refers to how an individual or family chooses to live. It is how one uses discretionary time, the things one owns, how one uses those things, and the meanings those things have to the individuals involved. One hears of suburban lifestyles, informal lifestyles, traditional lifestyles, and so forth. Lifestyles are heavily based on the individual's value/motive/preference system. Therefore, geographic areas often have a generalized lifestyle that is desired by a majority of the members of the area.

Stability of Geographic Subcultures

Given the high rate of geographic mobility, the popularity of national mass media, increased urbanization, increased industrialization, increased travel, and the increased standardization of the education system, it has been suggested that geographic subcultures have or will soon cease to function (see Berger 1960; Mayo 1964; Glenn and Alston 1967; and McKinnery and Bourque 1971). This is sometimes referred to as the massification theory.

However, it appears more likely that, while geographic subcultures, like the broader culture, are changing, they do not appear to be losing their distinctive characteristics (Glenn 1967; Glenn and Simmons 1967; and Peterson and Demaggio 1975). A number of reasons appear to account for this.

Geographic mobility generally involves moves within the same geographic region. Furthermore, individuals often move to a region because they admire the perceived values or lifestyle of that region. Thus, many of those moving into a region already accept many of the region's cultural values. Finally, those moving into a region generally come from a variety of other regions and thus are relatively heterogeneous. The easiest pattern of behavior is to conform to the norms of the existing group. This in time generally leads to acceptance of the regional norms.

Mass media cut through regional areas. However, viewing patterns differ between regions and numerous regional media exist. Furthermore, given the nature of most mass media content, it is more likely to reinforce widely shared cultural values than it is to challenge the unique values associated with individual regional subcultures. In addition, when Glenn and Simmons (1967) examined the effect of mass communication upon different geographic regions, they concluded that it may actually increase the variation between regions because of the differing effect it had in different localities. Similar analyses can explain the limited "massification" impact of increased industrialization, travel, and standardization in the education system.

Research Traditions on Geographical Subcultures

Historians, sociologists, anthropologists and geographers have all tried to conceptually and operationally define and apply the notion of geographical subcultures to their own respective areas of interest. Sociologists, however, are primarily responsible for the evolution of research on geographic subcultures. The founding father of sociology's geographic perspective, Howard W. Odum, characterized a geographical region as having: (1) spacial limits and bounds, (2) flexibility in those bounds, (3) structural and functional aspects that denominate a region, (4) people

"culturally conditioned" through time and spacial relationships (Odum 1942; Odum and Moore 1938). Hertzler (1944) took a social psychologists' perspective towards the notion of geographic regionalism. Hertzler stressed that the underlying concept of a region was its natural landscape. The topography, climate and natural resources determines the specific lifestyle or "social life" of the group. This, in turn, fashions the "cultural landscape" of the group. This landscape includes the economic structure, the composition of the population, the historical social processes of the area and the combination of cultures each successive population increment has contributed to the geographic region. From these variables the regional culture is developed and this makes representatives of that region easily identifiable. Interactions between individuals rest upon this regional base and, in turn, behavior and personality patterns result through the regional "conditioning".

Anthropologists have primarily relied upon a "community analysis" approach. That is, anthropologists have focused their attention upon specific communities and made inferences about the area's culture. When anthropologists analyze a community, they become familiar with the "cultural context" that extends beyond its historical and current social characteristics (Park 1972). Arensberg (1954,1955) utilized this community study approach in an attempt to develop a classification scheme and typology of American culture and geographic subcultures. Arensberg felt that communities which are properly sampled do reflect their culture and to some degree, the cultural qualities of the surrounding area. He then suggested that anthropologists could develop a model that compared communities and effectively delineated regions. This model would examine five variables (individual, space, time, function, and structure and process) and the relationships they had with one another inside a community setting. Gillin (1955) built upon Arensberg's theory and actually formulated a regional descriptive structure. Community cultural data was utilized by Gillin to develop a "list of values" dominant in the American national culture as well as various regional cultures.

As can be seen from the above discussion, a number of research traditions exist with respect to geographic subcultures. While these traditions can provide both data and methodological guides to the marketing researcher, it is likely that a product category specific approach will be required by marketers. This means that marketing researchers should avoid the mistake of "blind" borrowing from other research traditions (an error that has hampered several areas of consumer research, such as personality).

Geographic Subcultural Influences on Consumption and Nonconsumption Behaviors

A popular method used to analyze the significance of geographic subcultures is to analyze regional diversities in respect to a "common social indicator." Homicide and violence (Wolfgang and Ferracuti 1967), educational performance (Ferriss 1969; Sheldon and Moore 1968), infant mortality, fertility (Ram 1976), prejudice (Middleton 1976), authoritarianism (Williams 1966), poverty (Miller 1971), and urbanization (Grasmick 1974; Fisher 1975) are some of the primary social indicators used to understand geographical diversity.

Williams (1966) asserts that there seems to be an association between an individual's level of authoritarianism and prejudice and the geographic setting he/she originates from or lives in. In respect to urbanization, Fisher's article (1975) supports the belief that the increased concentration of population will: (1) produce increased subcultural diversity, (2) strengthen existing diversity patterns, and (3) excite the level of diffusion of subcultural diversity.

Geographic subcultures can also influence any of the steps in the consumption process. It is rather easy to find examples of variations in behavior between regions in the consumption process. For example, more antifreeze is sold in Maine than in Florida. This is due primarily to a variation in problem recognition caused by geographically bound influences (climate). The proportion of beverages sold in returnable bottles is substantially higher in Oregon than in California. This variation in disposition behavior is caused by psychologically bound influences reflected in Oregon state law.

Table 1 illustrates variations in the consumption and preparation of coffee across four broad regions of the United States. As can be seen, there are dramatic differences. Similar differences can be found for other product categories, other types of geographic areas, and other aspects of the consumption process. For example, ratings of television show types showed "adventure" shows to have the lowest ratings of six show types in the Northeast and the West Central regions and the highest rating in the South (A.C. Nielsen).

Table 1
Regional Variation in the Consumption and Preparation of Coffee

	East	Midwest	South	West
Percent consuming (on a winter day)				
Coffee	63%	60%	58%	55%
Tea	32	19	32	20
Soft Drink	47	49	57	39
Milk	44	51	50	59
Fruit Juice	51	41	40	47
How coffee is consumed at home--				
Black	15	39	30	48
Sweeteners only	6	10	14	14
Sweeteners and creaming agents	53	27	38	24
Creaming agents only	26	24	18	14
How coffee is prepared--				
Electric percolator	33	58	48	53
Nonelectric percolator	38	12	13	12
Drip pot	8	4	8	1
Automatic electric drip maker	13	21	25	21
Other	8	5	6	13
Container used for drinking coffee--				
Cup	49	59	58	37
Mug	44	34	35	55
Both	6	7	7	7

Source: Derived from Coffee Drinking in the United States: Winter 1976 (Pan-American Coffee Bureau 1976)

Even based on the limited evidence presented in Table 1, it is clear that marketing practitioners should consider geographic subcultures as a potentially useful variable when developing marketing strategies. The next section of this paper presents a simplified framework for systematically analyzing the potential influence of a geographic subculture.

A Format for Developing Marketing Strategies
Based on Geographic Subcultures

Figure 2 presents a useful format for analyzing potential geographic influences and determining whether or not those differences will vary, both in nature of and intensity of, between and within geographic regions. The intent of the format presented is not to provide marketing managers with a "formula" that will identify geographic influences automatically. Rather, the format provides a systematic procedure for looking at the decision making process of a specific target market for a specific product across differing geographic regions and boundaries. This systematic consideration of possible geographical influence is an attempt to provide marketing managers with a way to become aware of geographical influence and the ability to pinpoint where in the decision making process the influence occurs. This should then make the development of appropriate marketing strategy more complete and effective.

The geographic influences and differences in the influences by region can best be translated into marketing strategy if looked at from a consumer decision process focus. In other words, the consumer should be looked at from a problem solving/information processing point of view and strategies developed that relate to one or more relevant parts of the decision making process. The decision process utilized in Figure 2 is broken down into problem recognition, information search, evaluation and selection, outlet selection and implementation, product use, product disposition, and satisfaction evaluation.

It may be helpful in understanding the format to use a hypothetical example of the geographic influences on the purchase and use decision of home freezers. Rural subcultures experience a larger degree of problem recognition related

to home freezers than do urban and suburban residences. Furthermore, this problem recognition is of a different nature. Perhaps the primary problem recognition factor in urban and suburban areas is for the convenience offered by a freezer while rural residents recognize the economic benefits of freezing home grown foods. Information search is shown to be less in rural areas due to the limited number of retail outlets (geographically bound) as well as being different due to differing media habits (psychologically bound). Outlet selection is different for rural residents because of a greater reliance on home shopping (mail-order). The disposition of a used freezer is more difficult in an urban areas because of a generally lesser amount of storage space. It also would seem natural to expect that there would be a great deal of variation between regions, even when looking at the rural category, e.g. the Northwest versus the deep South.

Given an analysis such as the one presented, a marketing manager would be alerted to the fact that a standardized distribution or advertising campaign for home freezers across the United States may not be as effective as a geographically based one. Additionally, when one examines the lifestyle characteristics of specific geographical areas, segmentation strategies may become apparent.

Conclusions

In this paper, we tried to illustrate the usefulness if not the necessity of considering the potential influence of geographic subcultures in the development of marketing strategies. In addition, we presented a structured framework to aid in analyzing the influence of geographic subcultures. It is our position that using a single marketing mix throughout the United States without considering poten-

FIGURE 2

A FORMAT FOR ANALYZING GEOGRAPHIC INFLUENCE

	REGION 1						REGION N					
	Physical			Psychological			Physical			Psychological		
DECISION MAKING PROCESS	U	S	R	U	S	R	U	S	R	U	S	R
1. Problem Recognition												
2. Information Search												
3. Evaluation and Selection												
4. Outlet Selection and Implementation												
5. Product Use												
6. Product Disposition												
7. Evaluation												

U = urban S = suburban R = rural

tial geographic variations makes no more sense than applying the marketing mix in a variety of countries without first examining the cultural variations among those countries.

While less obvious, geographic subcultures may also influence the development of consumer behavior theory. For example, the responsiveness to various persuasive techniques (authority figures, fear appeals, and so forth) may vary across geographic regions because of differing values. Studies conducted in only one geographic region may produce results that are specific to that region. Likewise, national studies that are aggregated for analysis may hide the fact that the variable being examined has differing impacts across geographic regions.

It is hoped that this paper will alert both marketing practitioners and academicians to the potential advantages of recognizing the strong geographic influences within this culture.

References

- Arensberg, C.M. (1954), "The Community Study Method," American Journal of Sociology, Vol. 60, 2.
- _____ (1955), "American Communities," American Anthropologist, 2, Vol. 57.
- Berger, B. (1960), Working-Class Suburb, Berkeley: University of California Press.
- Ferriss, A.L. (1969), Indicators of Trends in American Education, New York: Russell Sage Foundation, Co.
- Fisher, C.S. (1975), "Toward a Subcultural Theory of Urbanism," American Journal of Sociology, 6, vol. 80, 1319.
- Gillin, S. (1955), "National and Regional Cultural Values in the United States," Social Forces, 107.
- Glenn, N.D. and Alson, J. (March 1967), "Rural-Urban Differences in Reported Attitudes and Behavior," Southwestern Social Science Quarterly.
- _____. (December 1967), "Massification vs. Diversification," Social Forces, 2, Vol. 46, 172.
- _____ and Simmons, J.L. (Summer 1967), "Are Regional Differences Diminishing," Public Opinion Quarterly, 2, Vol. 31.
- Grasmick, H.G. (Winter 1974), "Rural Culture and the Wallace Movement in the South," Sociology, 4, Vol. 39, 465.
- Hertzler, J.O. "Some Notes on the Social Psychology of Regionalism," Social Forces, 144, 331.
- Mayo, S.C. (October 1964), "Social Change, Social Movements and the Disappearing South," Social Forces, 1.
- McClelland, D.C. (1961), The Achieving Society, New York: Free Press.
- McKinney, J.C. and Bourgue, "The Changing South: National Incorporation of a Region," American Sociology Review, June 1971.
- Middleton, R. (February 1976), "Regional Differences in Prejudice," American Sociological Review, 41, 94.
- Miller, W.B. (Summer 1971), "Subculture, Social Reform and the Culture of Poverty," Human Organizations, 30.
- Nicosia, F.M. and Mayer, R.N. (September 1976), "Toward a Sociology of Consumption," Journal of Consumer Research, 67.
- Odum, H.W. (May 1942), "A Sociological Approach to the Study and Practice of American Regionalism," Social Forces, Vol. 20, 425.

CONSUMERISM IN DEVELOPING COUNTRIES - THE BRAZILIAN EXPERIENCE

John Stanton, Temple University
Rajan Chandran, Temple University
Jeffrey Lowenhar, Temple University

Abstract

Consumerism in developing countries has not been analyzed and discussed in current literature. In part this has been due to the lack of an appropriate methodological framework. This paper adopts a model developed by Kotler to the analysis of consumerism in developing countries. Consumerism in Brazil is explored as a case example. Guidelines for policy are suggested.

Consumerism in Developing Countries The Brazilian Experience

Introduction

Consumerism in the United States has received much attention in recent business literature. Most of the articles dealing with the topic have been descriptive, commenting on its importance and implications, and/or prescriptive, outlining alternative activities and policies for consumers, governments, and firms. However very little attention has been paid to consumerism abroad, especially among developed and developing countries where consumerism is in its infancy or early growth stage. In these countries, protest over consumer issues has advanced from basic causes such as lack of food, clothing and shelter; to more sophisticated issues like consumer information reforms, consumer education, and consumer protection. Since a large part of American trade is with these developed and developing countries, government policy makers, consumer behavior researchers and the business community must be aware of and concerned about the growth of consumerism in these countries and its economic and social ramifications. For example, promotions used by the baby food subsidiaries of Nestles and Bristol Myers have been attacked in many developing countries (Sethi 1978). These firms have been held responsible by its critics for the unintended deaths of infants, which resulted from the misuse of their products. There has also been an increasing incidence of government intervention and control of all aspects of business life abroad. It is essential that American business firms strategically intervene when confronted with the growth of consumerism in developing and developed countries, rather than take a benign stance. Governmental policymakers can also learn from studying consumerism abroad by noting the methods used by other governments in tackling this issue. For example, Gazda (1977) has noted that the Canadian Government has influenced the growth of consumerism by timely and active responses. Despite its importance, international consumerism has not been systematically discussed in current literature. In part, this has been due to the lack of a methodological framework for examining and comparing consumerism in developing and developed countries.

In this paper the authors will attempt to (1) discuss the global use and interpretation of the term "consumerism" by examining some current consumer literature (2) analyze consumerism in Brazil using a theoretical framework developed by Kotler, (3) highlight the social and economic implications and (4) suggest guidelines for consumer policy. The article focuses on Brazil because: (1) it is a developing country faced with problems which are typical of those encountered in developing countries; (2) it contributes significantly to the overall Latin American economy by virtue of being the biggest country in Latin America and (3) the authors have been able to gather first

hand information during recent visits.

Some Current Definitions

A determination of whether historical and current consumer movements can be adequately characterized by universal definitions requires an analysis of current definitions and examples.

Buskirk and Rothe (1970) in a oft quoted definition of consumerism characterize it as:

"The organized efforts of consumers (whatever the economic system they live under) seeking redress, restitution and remedy for dissatisfactions they have accumulated in the acquisition of a standard of living." (emphasis ours)

Hermann (1970) on the other hand explains consumerism as:

"A conglomeration of separate groups, each with its own particular concerns, which sometimes form temporary alliances on particular issues."

Kotler (1972) defines consumerism as:

"A social movement (under any system) where buyers seek to augment their rights and powers in relation to sellers." (emphasis ours)

The common feature of these definitions is the organized efforts of consumers, or identifiable consumer groups in any system to right perceived and/or actual wrongs. (Coppet 1974) The wrongs may have been committed by businesses in countries where the free enterprise system exists, or by government in countries where the government plays an active role in the economic system. Kotler's definition is broadest in that he interprets consumerism as a social movement where a group of people coalesce to accomplish a shared goal. Kotler's model provides a framework for examining and comparing consumerism in different countries. Kotler explains the social factors contributing to consumerism as a social movement. First, there must be present (1) Structural Conduciveness: an environment conducive to consumerism - an educated group of people, growth of income, certain problematic issues that require collective behavior. These factors induce greater expectations, which when unfulfilled produce (2) structural strains in the existing socioeconomic system. These structural strains contribute to the (3) generalized belief that business has not been responsive to society's expectations of performance. This general belief is then ignited by external factors that act as a catalyst or a (4) precipitating factor. Professional consumer spokespersons, politicians and the media (5) mobilize for action to solve the identified problems. Businesses usually act, but if their actions are perceived as inadequate, some form of legislation or (6) social controls result.

Hendon (1975) documents the fact that, historically, some manifestations of consumer discontentments have been recorded as social movements. Consumers have organized and sought redress at various times under different political systems. He cites the Grain riots (1775), Revellion riots (1789), and Grocery riots (1792-93) in France as examples of organized protests based on consumer issues. In Eng-

land the Grain riots (1736), Price control movement riots (1776), Garden riots (1780), Church and King riots (1790-92) and Rebecca riots (1839-42) also were based on consumer issues. In Poland the recent food riots were the result of consumers seeking redress for some basic dissatisfactions. In the United States the Boston Tea Party (1760) the Whiskey rebellion (1790), collective consumer movements of the 1900's and 1930's, and the more recent meat price agitations were likewise based on consumer dissatisfactions. Hendon notes that until a nation is economically mature, consumer issues usually revolve over basic issues like food, clothing and shelter and tend to be violent. As the country develops, problems of consumption outweigh problems of production, consumer issues become more sophisticated and less violent. In the economically mature countries, consumer issues focus upon auto safety, nutritional labelling, advertising to children, wool products labelling, boat safety, Consumer Credit Protection and the like.

Gaedeke and Udo Aka (1974) queried diplomatic representatives of 58 nations on consumer concerns in their respective countries and seemed to confirm Hendon's theories. They found that the degree and focus of consumer concerns and consumer protection programs depended upon the level of industrial development of the country. As countries develop, consumer concerns and the need for protection assume increased meaning and importance and become vital elements in business and government policy.

In summary, it is safe to conclude that the above definitions of consumerism are typical and generally accepted. The consumer movement has always existed in some form throughout history in many areas of the world under a spectrum of political systems. The United States is not, and has not been, a leader in worldwide consumerism. Consumerism is a social movement dependent upon national economic conditions. Its progression in different countries can be analyzed using the Kotlerian framework. The economic maturity of a country dictates the sophistication of the topics involved, the violence of the movement, and future.

Brazil and Consumerism

Having discussed and explained consumerism, for the purposes of the article we will attempt to assess the status of consumerism in Brazil. First, a very short introduction to the country will be provided which highlights the importance of studying consumerism in Brazil.

Brazil has an estimated population of 120 million people; it is the most populous country in Latin America and has the seventh largest population in the world. The approximate annual growth rate of the population is 2.8%. Half the population is located in urban areas; approximately 35% reside in Brazil's eight largest cities. The 1979 GNP was approximately \$200 billion.

Geographically a large country, Brazil has extremely rich mineral resources, particularly iron ore, manganese and semi precious gems. In addition, a new kind of bauxite was recently discovered in the Amazon region. Brazil's industry is quite strong, with exports of \$10.1 billion in coffee, \$2.2 billion in iron ore, and \$1 billion in soybeans. With the aid of severe restrictions on imports, Brazil has generally maintained a fairly good balance of payments. Recently oil imports have posed a problem, but they are likely to be reduced by up to 20% when a new alcohol program is operationalized. This will hopefully result in a more favorable balance of payments position.

Foreign investment represents a small but important part of Brazil's capital base. Although the United States makes a significant portion (4.5 billion) of the approximately \$10 billion of foreign investments, other countries

such as Japan and Germany are increasing their capital commitments.

The world's financial institutions expect the continued growth and development of Brazil. Brazil is currently Latin America's largest borrower from international financial institutions, with \$50 billion in commitments to commercial banks and other institutions such as the World Bank and International Monetary Fund. (Business Week 1978)

After examining its economic status, one can conclude that Brazil is one of the most important emerging countries. In addition, recent social changes make the study of consumerism in Brazil more attractive. Military control has been reduced and there is a movement towards a more democratic society. Secondly, improvements in education, and better dissemination of information through the mass media indicate that factors conducive to consumerism evolving into a broad social movement are present.

Application of the Model to Brazil

Using the framework developed by Kotler we can analyze consumerism in Brazil in greater detail. (See Figure 1)

1. Structural Conduciveness

Brazil has four structurally conducive characteristics that permit or create the potential for a consumer movement: a) advances in income and education; b) growth of technology; c) exploitation of the environment and d) disparities between the quality of local and exported products.

(a) Income and Educational Advances

Brazil, like other developing countries, has experienced a steady increase in per capita income each year, although the average income is still far below that of developed countries. One estimate of the per capita income of residents of the areas surrounding Rio de Janeiro is \$556 per month (Pesquisa de comercio 1974), an estimate for similar areas of Sao Paulo is \$350 per month. (Cunningham, Moore, Cunningham 1974) The average number of automobiles per family is estimated to be .72, although the average for families with at least one car is 1.32. (Pesquisa de Comercio 1974) A 1974 study of Brazil indicated that 41.3% of the population of Sao Paulo could be classified as middle class. (Pesquisa de Comercio 1974) Consumers visit shops frequently and spend considerable time shopping for food, clothing and appliances. The fact that should be stressed is the prominence of middle income consumers who now have the opportunity to buy goods and services. Furthermore, this growing middle class has higher levels of education. In Rio and San Paulo a majority of the residents have had at least secondary school education. This is particularly important since an educated middle class plays an important role in advancing consumerism due to its higher expectations. (Diamond, Ward and Farber 1976) When these expectations are not met, frustration usually results, setting the stage for consumerism.

(b) Growth of Marketing Technology

The sudden growth of imported technology, both in marketing and manufacturing, plays an important role in creating social contradictions and strains. The problem of marketing technology, especially unwholesome marketing technology, is exemplified by the exposure of the Brazilian consumer to subtle marketing campaigns that could cause confusion for more sophisticated consumers. For example, consider pricing tactics in Brazil. Various pricing alternatives are offered to consumers in advertisements in Sunday newspapers in several large Brazilian cities. The main alternatives are: 1. the monthly payment price often stated without specifying the actual number of required payments. 2. the full price if payment is made at time of purchase (called a vista). 3. both the a vista price and the monthly payment price (both with and without the number of payments).

This type of pricing makes it virtually impossible to shop comparatively. Yet for many product categories, Brazilian shoppers use the newspaper as the primary source of information. When information was available to compare the monthly price of a product with the a vista price, the time payment method was extremely costly, even after considering the effect of inflation. Confusing marketing strategies, shoddy products dominate the market system. Games, prizes, discounts, and other gimmicks cause additional mis-information to be communicated to the Brazilian consumer.

(c) Exploitation of Environment

The progressive exploitation of the environment is a factor that has special significance in Brazil. Many Brazilian cities are situated in areas of magnificent natural beauty, but the development of business and urban housing has been accomplished with utter disregard for the ecological balance. Copacabana in Rio de Janeiro is a classic example of this situation.

Pollution is yet another problem. Many of the largest Brazilian cities have pollution levels that are far worse than those that resulted in consumer and governmental action in comparable American cities. In a recent public opinion poll of the 11 million inhabitants of the Greater Sao Paulo area, residents indicated that their major concern was air pollution. (Brazil Herald 1977) Brazilians are proud of their cities and their beaches, and the decay of both has aroused many residents.

(d) Quality of Local versus Exported Products

A factor unique to developing countries contributes to social contradictions: the quality difference between products produced for local consumption and those produced for export. In Brazil the finest coffee and leather is exported and not available for local consumption. One Brazilian automobile (PUMA) has 50 safety options that are available only on the exported model. (Ribeiro 1974) In essence, export products are of higher quality and have better features than those produced for domestic consumption. These inconsistencies and limitations become a major impetus for change.

In summarizing this section, one can conclude that the presence of an educated population, a prominent new middle class demanding quality goods and services, and the increasing exploitation of the environment, form the basic elements for consumerism and that the potential for consumer action exists.

2. Structural Strains

The existence of the above elements lead to social contradictions resulting in conflicting attitudes and deteriorating relations between consumers, businesses and government. These conflicting attitudes and relations in turn produce structural strains among parts of the socioeconomic system causing frustration and discontentment.

Economic discontentment is very real in Brazil and other developing countries. Inflation, cited as a structural strain in the United States is a major economic problem in Brazil. Average monthly inflation is estimated to be about 6% unadjusted or about 78% in 1979. Consumers are aware that their salaries do not rise at the same rapid rate. There is pressure on consumers to spend all their income, because attempts at saving are futile due to inflation.

Political discontentment is a major problem in Brazil. Unlike other developing countries, Brazil has a stable government, but it is a military government. There is much discussion about returning the government to a true democracy, and some steps have been taken in that direction this past year. However, the interaction between the military and the population appears to be a special strain,

although the military is not without internal discontentment. The labor movement has claimed the right to strike and the middle class is increasingly discontented with the country's economic institutions and unresponsive bureaucracies.

The first two antecedent conditions for the growth of consumerism structural conduciveness and strains are evident in Brazil. If these were all that were needed for consumerism, one would expect that Brazil would soon have a very active and vocal consumer movement. Yet a consumer movement has not appeared. The last four conditions appear to be acting as restraining influences. Of course, the question of the degree of development required before the conditions can coalesce is subjective. Recent social changes and the authors' personal experience this past year suggest that the other components of the change model are emerging in importance.

3. Generalized Belief

The existence of stress leads to the proliferation of certain consumer beliefs. As a social movement consumerism cannot develop unless there is a generalized belief about problems in the marketplace. This generalized belief, however, is not well developed in Brazil. ANDEC (Associacao Nacional de Defesa do Consumidor) is trying to create this awareness and generalized belief, however it has a limited budget and staff. Several legislators have an interest in the consumer's cause, but they lack clout. Nina Ribeiro, Deputato Federal, addressed the congress on the plight of the Brazilian Consumer as early as 1971. He has a regular TV show on consumer problems in Rio. Ualo Neto, a former biology teacher is now head of the environmental protection division in the Ministry of Interior, and is active in environmental protection.

In general, it appears these efforts have made only a slight impression on Brazilian consumers. There are two major reasons for this problem. First, the press in Brazil is not absolutely free. It does little to communicate consumer beliefs. Second, there is a predominance of fatalism among Brazilians as with the members of most underdeveloped countries. (Neilson and Stanton 1972) Although Brazil is now emerging from its underdeveloped economic state, its culture, attitudes, values and other social characteristics are slower to change. Social-psychological changes take longer than economic changes and require that consumers perceive their attempts at change will be ultimately successful. If consumer confidence does not exist, a generalized belief about problems and solutions will not develop.

The consumer movement is less likely to flourish without generalized consumer beliefs. Consumerism needs sustained consumer or general population support. In Brazil although these conditions are presently not readily apparent, a basic structure is developing that would make the maturing of the remaining conditions faster than would be expected. Inflation and other economic pressures are growing, it is likely that consumerism will soon become familiar to Brazilians. Some of the impetus, of course, will hinge on the future actions of the Brazilian government.

4. Precipitating Factors

The existence of a common belief is a necessary condition for a social movement to evolve. However, a catalyst is needed to precipitate the movement. Catalysts or precipitating factors are usually specific events or major problems that spark the growth of the social movement, in this case, consumerism.

Precipitating factors can develop quickly and without warning. For example, in the United States Ralph Nader was a lawyer and unknown freelance writer in 1964; one year later he was the author of the well-known book Unsafe at any Speed and ignited the consumer movement. (McCarry 1973) A Brazilian could just as quickly come to the fore-

front of consumerism in Brazil. A major roadblock to any precipitating factor emerging is the National Security Laws of Brazil. When liberally interpreted, it prohibits all actions perceived as contrary to the interests of the government. If the government does not create an environment that is conducive to consumerism, its genesis will be slow, and probably violent.

This condition should not be viewed too pessimistically, however. For example, the existence of similar laws in Poland did not hinder consumers from rioting over high food prices. Further sharp rises in prices or any other problem in the Brazilian economic environment may become the spark necessary to precipitate a consumer movement. Another catalyst might be the natural environment. With increasing urbanization, groups are forming to address environmental issues. Although one may question whether "all" sides believe this issue is important, progress on raising the level of the population's awareness of this issue is evident.

5. Mobilization for Action

The success of a consumer movement ultimately depends upon how effectively it is organized and managed. This can be accomplished either by a single dynamic individual or groups of individuals. Working links are usually established with people in universities and other professions to acquire technical expertise and credibility. Politicians, prestige seekers and other groups coalesce and the movement is born.

In Brazil, although still small, special purpose consumer groups are beginning to form. A group in Sao Paulo organized a Walk-a-thon to focus attention on pollution within the city. Residents in Rio made an overnight trip (in a caravan of forty cars and buses) to the neighboring state of Espirito Santo to protest government attempts to convert a world-famous biological reserve into a commercial palm plantation. Other residents in Rio have blocked construction of high rise apartments around one of the city's most beautiful parks. Residents of Sao Paulo also battled the city over plans to build a paper factory at the headwaters of what is said to be the state's last unpolluted river. Although these are at most rudimentary beginnings, they indicate the genesis of a structure that could support rapid growth if encouraged. A few politicians are now campaigning on consumer issues, as consumer awareness develops, more and more politicians will use this strategy. Consumerism is a non-political social issue that lends itself well to the "bandwagon" effect and will in all likelihood draw followers in the times ahead.

Another element for mobilization is the mass media. In Brazil (as in other developing countries), there is censorship of the media. Although censorship is usually not directed at issues like consumerism, it can often be capricious, thereby restraining the media from controversial issues in general. Lack of wide media coverage has to a certain extent stifled the organization of a consumer movement in Brazil, but with the return of certain freedoms under the new government there could be better mobilization for action.

6. Social Control

The progression of the consumer movement in a country will ultimately depend upon the reaction of business and government leaders. Counterveiling forces must be considered in assessing the progress and course of consumer movements. If business or government leaders fail to respond appropriately, consumer activists are encouraged. In the United States, legislators and the business community mostly ignored the early signs of growing consumerism. This response created greater frustration and resulted in an even more organized and concerted consumer effort.

It is too early to determine how the Brazilian domestic and multinational business community will respond to the activities of consumerists. The government appears to be trying to participate in the process, but like all bureaucratic bodies it moves slowly. Still, increasing budgets, new departments, and so forth indicate that the government is not ignoring the potential consumer movement. It is also possible that the domestic business community may learn from the mistakes of developed countries and react more quickly to protests and complaints, although this appears improbable. Multinational corporations in Brazil could play a more active role in determining future consumer needs and rights rather than reject proposals for change. They can listen to "reasonable" proposals and act jointly with the government as influential change agents, and thus negate the need for overly stringent legislation. Certain industries, like the Brazilian advertising industry, are developing codes and procedures that are attempting to provide consumer protection from fraudulent and misleading advertising. This is an initiative taken by the industry because it feels it has professional expertise that policy makers lack.

Summary and Implications

When the previously discussed conditions were present in the United States in the 1960's, the consumer movement advanced to where it was recognized and accepted as a part of the American society. In developing countries like Brazil, consumerism appears to be in an infant stage. But, as can be seen in this article, many of the antecedent conditions that gave rise to the American movement are present in Brazil. The emergence of a generalized belief and certain precipitating factors could activate the dormant movement. It only took approximately 15 years for consumerism to become a household expression in America. The United States, however, was at a much higher level of economic development than Brazil at the time.

It is difficult to predict when or how quickly the movement will flourish in Brazil or other developing countries. A great deal depends upon the government's role in the process. The significance of the government's position has both positive and negative potential implications. The government of developing countries have traditionally played active roles in stimulating public sectors of the economies to grow faster than would be expected in free market situations. It appears that Brazil's government will continue to play a major role in the country's economic development. For example, Brazil's Minister of Finance, Henrique Simonsen has said he prefers a hybrid economic system to minimize the imperfections of both free markets and planned economies. He believes that it is not a question of finding out which works better under ideal conditions but which makes the fewer mistakes under normal conditions. (Simonsen 1977) The government in Brazil has not played an active role so far.

Likewise, until the government realizes that consumerism is an expected and important adjunct to economic development, it will not encourage the business sector to support the consumer movement or introduce the legislation to force compliance. For that reason, it will be important for the newly formed consumer groups to demonstrate to government officials that in the long run the effect of the consumer movement is to provide better, safer and more efficient products and thus stimulate economic growth rather than hinder it.

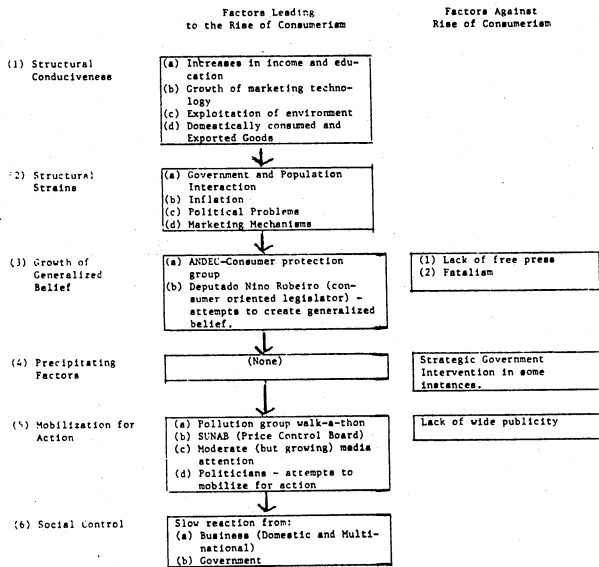
Two situations illustrate this point. Used car dealers are required by law to disclose all defects and potential problems to the buyer. The effect of this law was lower car prices, since in general prices were higher when defects could be "forgotten." It was estimated that the savings to the consumers of one state was \$40 million in one year. These savings are now available to the consu-

mers to either save, which encourages investment, or spend. (Anderson and Whitten 1977)

In Brazil new legislation, although still unclear, is being considered to force retailers to give both the a vista price and the monthly price with the number of payments. The impact on consumers might be to encourage comparative shopping. Monthly payments might still be selected due to lack of funds or credit availability but the purchase decision will be based on accurate and comprehensible information. Consumers who select the a vista price will have additional monies to spend on other items rather than on interest.

Consumerism is inevitable in developing countries, as conditions conducive to its development emerge. The direction of consumerism in Brazil is yet to be determined and its leadership is still unclear. The multinational community can and should engage in a more positive role with the governments of developing countries in future consumer issues. Rather than maintaining a passive or reactive stance, the multinational firms can utilize their experiences in developed countries to shape consumer programs in developing nations that reflect their interests. This should not be thought of as an attempt to subvert the consumer movement, only to suggest that the mutual interests of both groups may be more closely attained if they share a joint role in establishing policy.

FIGURE I
Factors Leading to the Rise of Consumerism in Brazil*



*Adapted from Kotler, Philip, "What Consumerism means to Marketers", Harvard Business Review, May-June, 1972, pp. 48-57.

References

Anderson, J. and Whitten, L. (November 13, 1977), "Shoddy Used Car Business," Sundicated Column Brazil Herald, 7.

Brazil Herald (November 29, 1977) "World Ecology Revolution Finally Makes it to Brazil," 12.

Business Week (December 5, 1978), "The Brazilian Gamble."

Buskirk, R.H., and Rothe, J.T., "Consumerism - An Interpretation," Journal of Marketing, 34, 61-65.

Coppet, J.I., "Consumerism from a Behavioral Perspective," Akron Business and Economic Review, 5, 23-28.

Cunningham, W., Moore, R., Cunningham, I., "Urban Markets in Industrializing Countries: The Sao Paulo Experience," Journal of Marketing, 39, 13-19.

Diamond, S., Ward, S., and Farber, R., "Consumer Problems and Consumerism: Analysis of Calls to a Consumer Hot Line," Journal of Marketing, 40, 58-63.

Gazda, G.H., "Descriptive Theory of Consumerism," Journal of the Academy of Marketing Science, Proceedings of the International Marketing Conference, 5, 33-34.

Gaedeke, R.M., Aka, U.U., "Toward the Internationalization of Consumerism," California Management Review, 17, 86-92.

Hendon, D.W., "Toward a Theory of Consumerism," Business Horizons, 18, 16-24.

Hermann, R.D., "Consumerism: In Goals, Organization and Future," Journal of Marketing, 34, 55-60.

Kotler, P., "What Consumerism Means for Marketers," Harvard Business Review, 50, 48-57.

McCarry, C., "Citizen Nader," Signet Books, New York, 1973.

Neilson, R., and Stanton, J., "Information Sensitivity, Spending Behavior and Fatalism," Proceedings of the 1972 American Marketing Association, 205-208.

Pesquisa da Comercio, Relatoria e cenalise, unpublished report Guonabara, Brazil, 1974.

Ribiero, N., (1974), Em Defesa do Consumidor, Centro de Documentacao e informacao Coordinacao de Publicacoes.

Sethi, S.P., and Post, J.E., Infant Formula Marketing in Less Developed Countries, An analysis of Secondary effects in Proceedings of the 1978 Educators Conference, ed. Jain, S.C., AMA, 43, 271-275.

Simonsen, Henrique, Unpublished report National Business Convention (CONCLAP), Rio de Janeiro, November 6, 1977.

TESTING THE STABILITY OF MARKET
SEGMENTATION ANALYSIS¹

Tony Schellinck, Dalhousie University²
Ian Fenwick, Northeastern University²

Abstract

The validity of segments formed by search procedures is of major concern to marketers. Currently split-half analysis is the most recommended method, but this requires large samples. This paper presents an alternative validation method which does not require large samples. The procedure is illustrated using a segmentation by AID analysis.

Introduction

Market segmentation is well entrenched as an integral part of marketing theory, but market segmentation analysis still faces many problems. As Wind (1978) points out: The validity of segmentation research is by far the most crucial question facing management. Do the segments discovered in a segmentation study exist in the population? Is the estimated segment size accurate? And how accurate are the estimated segment responses to the firm's marketing actions?

These questions are particularly important when the researcher uses a clustering based segmentation method. Powerful search techniques scan the data and will almost certainly identify a structure. The authors have yet to find a real data set that did not yield distinct, interpretable clusters from at least one clustering algorithm. Unfortunately the identification of such "segments" in the data is no guarantee of their existence on the ground. The susceptibility of search procedures to sampling error is well known (Doyle and Fenwick 1975) and there is a real danger that the segmentations produced may be highly unstable. Typically, consumer researchers have few a priori hypotheses about their segments and consequently can rationalize almost any segmentation, spurious or not.

Validating segment structures is recognized as a particularly difficult task (Anderberg 1973; Frank and Massey 1975; Hartigan 1975; Sherman and Sheth 1975). Replication is the most frequently recommended validation method, usually in the form of a split-half or hold-out analysis (Frank and Massey 1975; Mehrotra and Wells 1978). These methods analyze one half of the data and use the other half to test the analysis' predictions. In the case of segmentations, groupings derived from one half of the sample may be compared with those derived from the other.

Split-half analysis has two major problems. First, for a viable split-half test, half the data set has to be sufficient for the analysis technique being used. Consequently the researcher has to collect a sample twice the normal size. In practical terms this often translates into either no attempt at validation or else two analyses both based on inadequate samples.

Second, split-half testing will identify discrepancies between the analyses applied to the different halves, but will offer no clue as to which is the better representation of reality. If the split-halves fail to agree in some respect (e.g. factor structures differ; segmentations change; or predictive power is low) it is usually impossible to tell whether the differences are due to idiosyncracies of the particular split used or genuine indications of an unstable analysis. If the sample had been split in a different way, different variations in the analyses might have been observed and different results labelled unstable. Indeed researcher's reaction to split-half disagreement is varied. While most applications simply drop all parts of the analysis that are not identical in both split-halves, some actually retain all results that are of importance in either split-half (Roberts and Wortzel 1979).

It would be of much greater practical use to have a validation method that was based on more than two sample variations. This could test stability and offer a majority opinion of what the true structure was like. The method presented, here called "N + 1 Sample Analysis", allows just such an estimation of stability without requiring additional data collection.

N + 1 Sample Analysis

N + 1 Sample Analysis is based on the jackknife procedure (Tukey 1958). The jackknife requires that the available sample be split, at random, into N equal-size subsets. The analysis to be jackknifed is carried out on the whole sample, and then repeated omitting each subset of the sample in turn. Each of these analyses on the sample with a subset omitted yields estimates \hat{p}_i , $i=1, N$. The jackknifed estimator, $J(\hat{p})$, is a weighted difference between the mean of these estimates and the results for the whole sample analysis:

$$J(\hat{p}) = N \hat{p}_{\text{all}} - (n-1) \bar{\hat{p}}, \quad (1)$$

where \hat{p}_{all} is the estimate derived from the whole sample analysis and $\bar{\hat{p}}$ the mean of the \hat{p}_i values.

Jackknifing will often reduce bias and enables significance tests to be performed (for a further discussion of the jackknife and its use in marketing, see Fenwick 1979).

As it stands the jackknife procedure cannot be applied to testing segment stability. It is designed for a parametric rather than grouping methods. However the principle of N + 1 separate analyses - one on the whole sample and N on samples with a subset omitted - is distinctly relevant. The N analyses of samples with a subset omitted can be used to validate the grouping formed in the whole sample analysis. In the example presented here N + 1 Sample Analysis is used to validate a segmentation produced using AID.

AID Analysis as a Segmentation Method

AID (Automatic Interaction Detection) developed by Morgan and Sonquist (1963) is frequently used as a segmentation

¹Part of this research was supported by a grant from the Dalhousie School of Business Administration and the assistance of Marie Kalbfleisch.

²Tony Schellinck is Assistant Professor of Marketing at Dalhousie University; Ian Fenwick is Visiting Associate Professor of Marketing at Northeastern University.

method (Assael and Roscoe 1976; Sheth and Roscoe 1972; Gensch 1978; Assael 1970). AID employs a sequential dichotomization technique to partition the sample into progressively smaller groups so as to maximize between group differences. This analysis forms a tree diagram the end groups of which are used to define market segments.

AID is notoriously unstable. Its developers cautioned against its use on samples smaller than 1000 and validation was strongly recommended. Validation has in fact rarely been attempted, and where it has AID has not held up well (Doyle and Fenwick 1975). Nonetheless AID is very appealing as a segmentation tool. It offers the capability to handle ratio, interval, ordinal or nominal data. It produces an analysis that is easily explained to management and therefore has a good chance of implementation.

N + 1 Sample Analysis allows AID to be validated without using a large sample. Moreover the method enables individual segments to be assessed, so stable groupings may be retained and unstable ones discarded. Similarly, outliers, sample members that do not fit within the segmentation, may be identified and either discarded or perhaps set aside for more detailed analysis.

An Application

The example presented concerns a segmentation analysis of data collected from telephone interviews with 200 households. Using normal methods this sample size would probably prevent any split-half validation and certainly prohibit the use of AID. Yet samples of this size are by no means unusual, particularly when the interview is detailed and the incidence of qualified respondents low. The

dependent variable measures the adoption of a service. The predictors used here are 20 general AIO items, found in pretesting to be related to the service in question.

The N + 1 Sample Analysis involves three stages. First, AID is applied to the whole sample. Using the "standard" stopping criteria (split eligibility of .01, split reducibility .01, minimum group size for splitting 40) in our case produced eight splits, forming 9 end-groups (segments). The AID tree from this analysis appears in Figure 1.

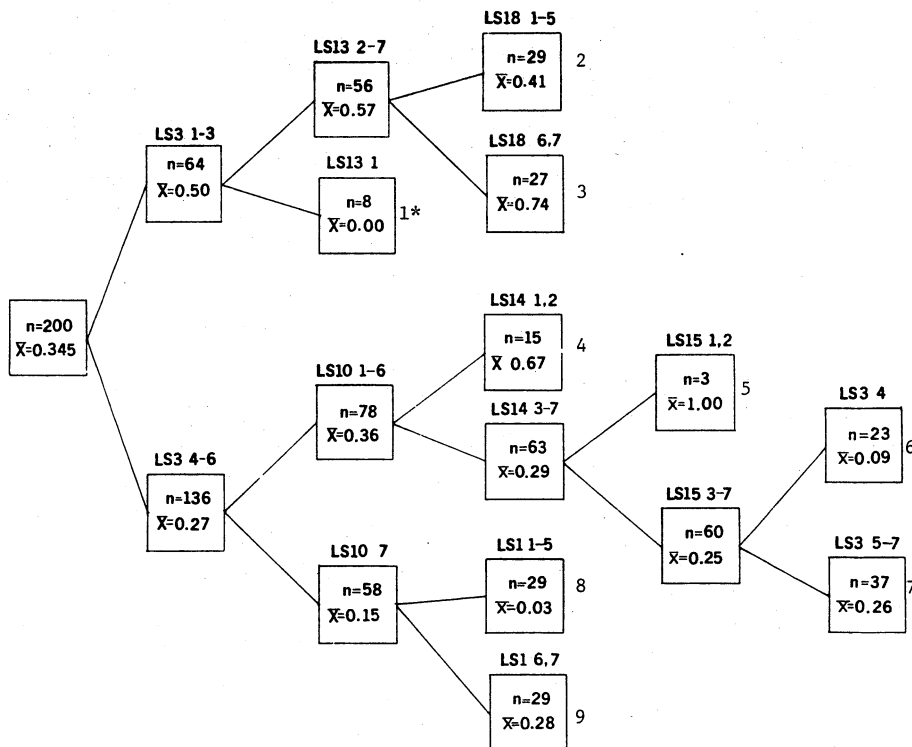
Second, the sample is split, at random, into equal-sized subsets and the AID analysis repeated, omitting each subset in turn. Although using identical stopping criteria these analyses are likely to produce rather dissimilar results. In the example presented here, 10 subsets of 20 cases each were formed, and so 10 further AID analyses were performed, each analysis omitting 20 cases. Both the configurations of the AID trees and the predictors entering the analyses varied considerably over the different runs.

The AID analysis appears to be affected by small sample variations. However, changes in the shape of the tree and changes in the predictors used are not by themselves evidence of unstable segments. If there is multicollinearity it is quite possible for the membership of the end-groups (segments) to remain unchanged although the variables defining them alter. So different AID trees could in fact be grouping together the same individuals. If we are to test the stability of segments we must look at segments, not the variables defining them.

The final stage of the N + 1 Sample Analysis is to compare

FIGURE 1

AID SEGMENTATION ANALYSIS ON WHOLE SAMPLE



*Original segment identification number.

the segment found in the whole sample analysis with those obtained from each of the runs with a data subset omitted. Before any meaningful segment comparison can be made it is necessary to align the segments - i.e., make sure we are comparing the most similar groups. Accordingly, for our example, the 9 segments derived in each of the runs with a data subset omitted were arbitrarily lettered. Ten tables can now be formed comparing each of the subset omitted analyses with the whole sample analysis. An example of one of these tables is shown in Table 1.

be the only members in a particular subset analysis segment. And the number of outliers, individual cases that are not well modelled by the segmentation.

Table 2 presents shared membership percentages for this set of data. Segment 1 is perfectly stable, 100% of its members are always classified together. Segment 2 is close to perfect stability with an average shared membership of over 90%. Segments 6 and 7 are rather unstable: on average they retain only slightly more than half of their

TABLE 1
A Comparison of the Whole Sample
Segmentation with that from the First Subset Omitted Analysis

ORIGINAL SEGMENT	A	B	C	D	NEW SEGMENTS		G	H	I	OMITTED CASES	TOTAL
					E	F					
1	6 100% 100%	-	-	-	-	-	-	-	-	2	8
2	-	24 100% 100%	-	-	-	-	-	-	-	5	29
3	-	-	23 100% 100%	-	-	-	-	-	-	4	27
4	-	-	-	6* 46.2% 42.9%	2 15.4% 6.7%	5 38.5% 17.2%	-	-	-	2	15
5	-	-	-	1 33.3% 7.1%	1 33.3% 3.3%	1 33.3% 3.3%	-	-	-	0	3
6	-	-	-	1 4.6% 7.1%	11 50% 36.7%	10 45.5% 34.5%	-	-	-	1	23
7	-	-	-	6 17.1% 42.9%	16 45.7% 53.3%	13 37.1% 44.8%	-	-	-	2	37
8	-	-	-	-	-	-	27 100% 100%	-	-	2	29
9	-	-	-	-	-	-	26 96.3% 100%	1 3.7% 100%	-	2	29

*Read as 6 of the cases classified as being in segment 4 in the whole sample analysis were placed in segment D of the first subset omitted analysis. These 6 cases represent 46.2% of segment 4 from the whole sample analysis and 42.9% of segment D of the first subset omitted analysis column. All percentages have been adjusted to exclude omitted cases.

For each of the segments formed in the whole sample analysis this table shows how that segment's membership was distributed over the segments formed in the first subset omitted analysis. For example, all the members placed in segment 1 in the whole sample analysis were in segment A of this particular subset omitted analysis. The members of segment 4 were considerably more dispersed: 46% turned up in segment D, 15% in segment E and 39% in segment E. The columns of the table are now re-ordered so as to maximize the diagonal elements, and the columns numbered. This procedure is repeated for each of the ten tables. This ensures that segments with the same number are indeed the most similar, i.e., that we are comparing like with like.

It is now possible to evaluate the segmentation. Three measures can be calculated. Shared membership, or common core, this estimates the proportion of a segment that would still be placed together were the analysis repeated. That is, if the original segment is stable then members should consistently show up in a common segment in each of the repeated analyses. Distinctiveness, an indication of the size of the shared membership relative to the segment as a whole, Members of the original segment should

original members.

Shared membership is not the only criterion however. In particular shared membership could be high as a result of comparing a small whole sample analysis segment with a large subset omitted analysis segment. We also require that the segment be distinctive or unique, i.e., the shared membership should be a major part of the whole segment. Table 3 shows the percentage of shared membership in each segment.

Notice segment 1 is unique. No other cases are ever included with those from segment 1. Segment 3 is rather less distinct. For example, in the 7th analysis, although shared membership was 92% (Table 2), this membership made up only 55% of the total segment. Segments 4,6,7, and 9 are particularly indistinct with a large proportion of their members drifting in and out.

Finally, it is possible that low shared membership may be the result of outliers. A few cases in the whole sample analysis segment may fail to be consistently segmented and so cloud the performance of the segment as a whole. Table 4 presents a frequency distribution for each segment

TABLE 2
Shared Membership Percentage of
Original Segment Classified in the
Same Group

Subset Omitted Analyses

Original Segment	I	II	III	IV	V	VI	VII	VIII	IX	X	Average
1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2	100.0	63.0	100.0	89.3	100.0	100.0	78.3	84.6	100.0	100.0	91.5
3	100.0	59.1	100.0	66.7	100.0	100.0	91.7	70.8	100.0	100.0	88.8
4	46.2	28.6	100.0	33.3	100.0	41.7	100.0	50.0	30.8	100.0	63.1
5	33.3	33.3	100.0	100.0	100.0	33.3	100.0	100.0	33.3	100.0	73.3
6	50.0	45.5	68.5	28.6	60.0	38.9	100.0	40.0	42.9	100.0	57.4
7	45.7	37.5	60.6	29.4	51.5	32.4	81.8	34.3	45.2	100.0	51.8
8	100.0	48.0	68.0	39.3	72.0	42.3	100.0	42.3	61.5	76.0	64.9
9	96.3	42.9	57.7	34.6	96.0	42.9	100.0	48.0	44.0	100.0	66.2

1 All percentages have been adjusted to exclude cases not used in the particular analysis.

TABLE 3
Segment Distinctiveness: Shared
Membership as a Percentage of Segment Size

Subset Omitted Analyses

Original Segment	I	II	III	IV	V	VI	VII	VIII	IX	X	Average
1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2	100.0	65.4	100.0	64.1	100.0	100.0	45.0	56.4	100.0	100.0	83.1
3	100.0	56.5	100.0	35.9	100.0	100.0	55.0	43.6	100.0	100.0	79.1
4	42.9	12.1	100.0	17.2	100.0	15.2	100.0	41.2	7.8	100.0	53.6
5	7.1	100.0	100.0	9.1	100.0	10.0	100.0	75.0	25.0	100.0	62.6
6	36.7	27.0	39.4	46.2	41.4	20.6	79.3	22.9	17.6	100.0	33.1
7	53.3	36.4	60.6	32.3	58.6	32.4	100.0	34.3	42.4	100.0	55.0
8	100.0	31.6	53.1	32.4	100.0	30.6	100.0	31.4	31.4	100.0	61.1
9	100.0	31.6	46.9	26.5	77.4	33.3	100.0	37.5	21.6	81.3	55.6

1 All percentage have been adjusted to exclude cases not used in the particular analysis.

showing the number of times each case remains within the common core. Thus, for segment 1 all members are consistently segmented in every run in which they appear (notice the maximum number of subset omitted analyses for any case is 9, in one run it is in the subset excluded). In contrast segment 7 has not a single case that is consistently segmented in every run, and 5 cases only appear in this segment twice. However there are no bimodal distributions: segments with poor cases lack good cases as well. So outliers do not appear to be biasing the percentages in tables 2 and 3. If some segments had displayed a concentration of outliers these could have been excluded and Tables 2 and 3 re-computed.

Conclusions

The N + 1 Sample Analysis described has enabled the stability of a segmentation to be assessed, despite small sample size. Overall, 2 segments were found to have unacceptably low shared membership and 2 more to be insufficiently distinct. For the purposes of this analysis a segment was considered unstable if it failed to retain, on average, 60% of its original members or if the membership retained made up less than 60% of the whole segment. Clearly these cut-off proportions will vary with the purpose of the analysis. Exploratory research may well

TABLE 4 -

Outliers: Frequency Distribution of Cases by
Number of Correct Classifications

Original Segments	Number of Classifications in Designated Segments									Segment Size	
	0	1	2	3	4	5	6	7	8		9
1										8*	8
2							1	6	7	5	29
3							1	4	16	6	27
4					1	6	5	3			15
5						1		2			3
6					6	8	6	2	2		23
7			5	7	8	7	5	2	4		37
8			1		4	6	9	6	2	1	29
9				2	5	3	8	6	4	1	29

*Read as 8 members of segment 1 classified together in all 9 analyses in which they appeared.

accept segmentations considerably below those used here; high investment decisions will probably wish to impose more stringent requirements.

At the end of an $N + 1$ Sample Analysis the researcher has a clear idea of the segments in the data and of their likely stability. An informed decision can be made on which segments to exploit, which to reject and which require further investigation. Although the example presented here uses the AID algorithm the method is suitable for any segmenting search procedure.

Of course nothing can make up for any unrepresentativeness in the sample. Like the jackknife itself, $N + 1$ Sample Analysis merely makes more intensive use of the available data; the relevance of its results depends on the sample. Similarly, unstable segments show that the sample used is too varied to support that segmentation, not that a larger or more representative sample would not ensure stability. $N + 1$ Sample Analysis is intended to bolster confidence in segmentation results, not replace large sample analysis. Despite all these reservations this method does offer unique advantages to researchers forced to use small samples, and provides at least a first step towards full-scale validation.

References

- Anderberg, Michael R. (1973), Cluster Analysis for Applications, New York: Academic Press.
- Assael, Henry (1979), "Segmenting Markets by Group Purchasing Behavior: An Application of the AID Technique," Journal of Marketing Research, 7, 153-8.
- Assael, Henry, and Roscoe, A.M., Jr. (1976), "Approaches to Market Segmentation Analysis," Journal of Marketing, 40, 67-76.

Doyle, P. and Fenwick, I. (1975), "Pitfalls of A.I.D. Analysis," Journal of Marketing Research, 12, 408-413.

Fenwick, Ian (1979), "Techniques in Market Measurement: The Jack-Knife," Journal of Marketing Research, 16, 410-414.

Frank, R.E. and Massey, W.F. (1975), "Noise Reduction in Segmentation Research," in J.U. Farley and J.A. Howard (eds.), Control of Error in Market Research Data.

Gensch, Dennis H. (1978), "Image-Measurement Segmentation," Journal of Marketing Research, 15, 384-94.

Hartigan, John A. (1975), Clustering Algorithms, Toronto: John Wiley & Sons.

Mehrotra, Sunil and Wells, William D. (1977), "Psychographics and Buyer Behavior": Theory and Recent Empirical Findings," in Arch Woodside, Jagdish Sheth and Peter Bennett (eds.), Consumer and Industrial Buying Behavior, Elsevier North-Holland, N.Y., 49-66.

Morgan, James N., and Sonquist, John A. (1963), "Problems in the Analysis of Survey Data, and a Proposal," Journal of the American Statistical Association, 58, 415-34.

Roberts, M.L. and Wortzel, L.H. (1979), "New Life-Style Determinants of Women's Food Shopping Behavior," Journal of Marketing, 43, 28-39.

Sherman, Lawrence and Sheth, Jagdish N. (1975), "Cluster Analysis and Its Applications in Marketing Research," Working Paper, University of Illinois.

Sheth, Jagdish N., and Roscoe, A. Marvin, Jr. (1972), "Demographic Segmentation of Long Distance Behavior: Data Analysis and Inductive Model Building," Faculty Working Paper, College of Commerce and Business Administration, University of Illinois at Urbana - Champaign.

Sonquist, J.A. and Morgan, J.N. (1964), The Detection of Interaction Effects, Ann Arbor: Institute for Social Research, University of Michigan.

Tukey, J.W. (1958), "Bias and Confidence in Not-Quite Large Samples," (abstract) Annals of Mathematical Statistics, 29, 614.

Wind, Yoram (1978), "Issues and Advances in Segmentation Research," Journal of Marketing Research, 15, 317-337.

VIVE LES DIFFERENCES

Roger M. Heeler, York University, Canada

Abstract

The three papers in this session are so different that there are no appropriate overview remarks. Each will be considered independently, vive les differences.

Geographic Subcultures - Hawkins et al

Basic marketing texts indicate many variables by which a market may be segmented. Geography is commonly one of these variables and one that according to the authors receives insufficient attention. The authors provide an interesting and succinct review of geographic differences as viewed by various disciplines, and illustrate the differences in consumption patterns that may occur with beverage consumption data. Strangely, the only reference to international marketing where geographic differences are the foundation of the field, is in a one sentence aside at the very end of the article.

In missing the international literature, the authors have missed one of the most compelling practical reasons for geographic segmentation, that is differential reachability. Segmentation requires both heterogeneity of consumers between groups (which the authors discuss) and an ability to reach these groups differentially with the marketing armoury. Geographic units are much easier to target for differential price, product, distribution, advertising etc. devices than units formed on many other segmenting variables, for example AIO measures. Practical marketers make full use of these possibilities.

There is an element of ambiguity in the authors' concept of geographic subcultures. Most of the paper is concerned with specific territorially contiguous areas such as "the West". These groupings are those most easily reached by differential targetting. The authors final example concerns "urban" versus "rural" USA. These two groups occur nationally, like "old" and "young" and are correspondingly more difficult to target.

Another concept missed through ignoring the international marketing literature is the exportability of many items simply because they are foreign. For example, "New York Bars" are popular outside of New York in part because of their foreignness elsewhere. Specialities from one area provide big opportunities elsewhere - look at Colonel Sanders!

The pre-publication version of this paper was missing half its references. If the published version is too, a supplementary listing should be provided to the audience.

Reviewers of this paper noted that (a) the lifestyle differences between regions should be examined per se (b) to what extent are differences a function of other variables such as population age in California (c) what is the importance of geography relative to other variables (d) how distinct are the regions.

Exporting Consumerism - Stanton et al

Consumerism having become big business in the USA, it is not surprising that academic entrepreneurs have turned to the task of examining its exportability. The Stanton et al paper examines the prospects for consumerism in Brazil which they regard as a paradigm for developing countries. For structural analysis they use Kotler's (1972) framework of the conditions likely to facilitate or limit the emergence of consumerism.

It is apparent that consumerism is still a difficult subject to define. In its widest sense, Kotler's (1972) definition

"A social movement (under any system) where buyers seek to augment their rights and powers in relation to sellers" could embrace everything from Marie Antoinette's advice on cake, and the subsequent spate of head chopping, government change, and rewriting of European history that emerged, to the benign concerns of middle american burghers over the display of unit prices. Where should the balance be struck? The authors do not seem sure, and perhaps avoid the issue by considering only the concerns of the minority of Brazilians who are both relatively wealthy and live in cities. The political and economic miseries of the majority are ignored, just as American winter visitors to Montego Bay can keep the horrors of Kingston from their vision. In this authors opinion the emphasis is probably appropriate for discussion of consumerism. As practiced, consumerism is a second level luxury for the very rich of the world, who are well beyond worrying about where the next meal is coming from. It is social concern with the gloves on.

This genteel limitation of consumerism makes the authors contribution relatively valid. Brazil is not a representative third world country. Most are much poorer and do not have the steady economic expansion of Brazil though many others also have a totalitarian regime of some type or another. But in major cities in many of these countries, a middle class for whom consumerism is a possibility exists. The authors should be projecting to this mini universe of Spartas around the world rather than to developing countries in general.

The analysis itself lacks the immediacy that might be added to the paper if it had been written by a Brazilian. But a useful array of secondary sources are used to indicate some of the pre-conditions and impediments to consumerism amongst the urban middle class. A couple of specific consumerism actions are quoted, but unfortunately the pervasiveness of consumerism actions is not well delineated. Towards the end the authors acknowledge that the government controls all and the press is restricted, so that consumerism might actually have a difficult time flourishing in Brazil. Perhaps they should have added an important prior corollary to Kotler's list of consumerism conditions. One must have a society that possesses and values liberal expression of competing viewpoints, and appropriate media to carry these expressions before the genteel force of consumerism can flourish. If this condition does not exist, the steady flow of consumerism may be replaced by the occasional eruptions of revolution.

Segment Stability - Schellinck and Fenwick

This paper uses Tukey's (1958) hold out procedure to test for the stability of segments found by an AID fishing trip. Although Tukey's technique has been well known to a generation of social scientists, it is less used than it deserves, so a current example is of value to consumer behaviour researchers.

The word "validity" is in many ways too strong a word to use in association with the techniques. Validity carries overtones of full testing of the proper meaning of a pre-hypothesized concept or measure (see Campbell and Fiske

(1959) or any social science research methods text). The jackknife is more of a reliability test, especially useful for showing the stability of fishing trip analysis. The authors show how this may be made to work for AID segments and acknowledge the degree of judgmental art that remains even after the science.

The analysis does not allow a small sample to stand in for a large one as the authors seem to imply. For example, the authors make much of one sub-segment that is found to be 100% stable through the hold out procedures. This segment consists of six cases drawn from a sample of 200. This sample has a confidence interval, considering random sampling error alone, of 1 case to 11 cases. If this segment is to be used for actual management purposes, clearly a more precise population proportion estimate is needed, which can only come from a larger sample.

Reviewers remarks included (a) an important topic for users of survey research (b) lots of guesswork still left in the analysis.

References

Campbell, D.T. and Fiske, D.W. (1959) Convergent and Discriminant Validation by the Multitrait - Multimethod Matrix," Psychological Bulletin, 56,81-105.

Kotler, P. (1972) "What Consumerism Means for Marketers" Harvard Business Review, 50, 48-57.

Tukey, J.W. (1958) "Bias and Confidence in Not-Quite Large Samples," (abstract) Annals of Mathematical Statistics, 29, 614.

MOTHERS' ATTITUDES AND PERCEPTIONS OF CHILDREN'S INFLUENCE
AND THEIR EFFECT ON FAMILY CONSUMPTION¹

Mary Lou Roberts, Boston University
Lawrence H. Wortzel, Boston University
Robert L. Berkeley, Boston University

Abstract

A model is proposed in which a range of mothers' attitudes affect the amount of influence they perceive their children to have and the amount of influence, in turn, affects the amount of family consumption. Findings indicate that at least three attitudinal dimensions -- economic, health-related and liberal versus conservative -- affect the degree to which mothers report children are influential. Degree of influence had a significant effect on the amount of family usage of a set of products, all of which are appropriate for consumption by a single individual. Areas in which further research is needed are discussed, as are managerial implications.

The role of children as influencers in family purchasing decisions is an accepted part of the conventional wisdom of marketing. Empirical documentation of this role, however, is scant. Most studies have focused on only one product (Berey and Pollay 1968, Caron and Ward 1975, Frideres 1973), a small set of products (Jenkins 1978, Mehrota and Torges 1976; Moschis and Moore 1979; Popper 1979, Szybillo, Sosanie and Tenebein 1979; Sybillo and Soxanie 1976, Ward, Popper and Wackman 1970) or a single consumption decision (Nelson 1978), although at least one study did investigate a relatively large array of products (Ward and Wackman 1972).

The paucity of studies is not surprising in view of the difficulties of collecting valid and reliable data on family decision processes and consumption behavior (Davis 1977). The well-documented problems in collecting data on husband-wife influence processes are exacerbated when one tries to collect data on the input of children, especially young ones, to the process (Robertson and Feldman 1975, Rossiter 1978, Szybillo, Sosanie, and Tenebein 1979, Ward 1978). Hence, the important question of the effect of children's influence on the type and magnitude of family consumption remains empirically unanswered.

The existence of a data base from a large representative sample makes it possible to approach this question in an exploratory vein. In so doing, we are suggesting a simple model in which parental attitudes, in this case the mother's, will affect the amount of influence she perceives her children to have, and that this perception will, in turn, affect the amount of that product category consumed by the family.

In one of the earliest studies in this field Berey and Pollay (1968) identified the attitude child-centeredness as important in determining whether or not the mother yielded to the child's influence attempts. Ward and Wackman (1972) later found that the age of the child was a determinant of the frequency of parental yielding. The importance of the age variable is obvious, but other parental attitudes might prove to be of equal or greater value in explaining the success, or lack thereof, of children's influence attempts. General attitudes toward family financial matters, media and society all seem to be possible

determinants of parental yielding. Specific attitudes toward food prices and nutrition would seem to have special relevance to yielding to requests for food products. The importance mothers place on consuming, or not consuming, certain types of food and on nutrition in general should affect their willingness to yield to children's requests. In terms of both the effect of attitudes on influence and of influence on consumption, it is necessary to control for the age of the child.

Looking at our basic model from the mother's viewpoint is appropriate for several reasons. The mother is clearly a primary socialization agent during the early years of a child's life, though it is equally clear that peer groups become increasingly important as the child grows older. Peer groups may achieve dominant influence during adolescence (Moschis and Moore 1979). Early in the child's life, the mother purchases most of the products consumed by the child. As they become older children obtain more discretionary spending power and adolescents may actually purchase many, if not most, of the products they consume singly. However, the preponderance of evidence indicates that, throughout the family life cycle, the mother is the chief purchasing agent for products consumed jointly by various members of the family. The provision of food, especially, is an important part of mothering throughout the family cycle. If however, other family members, in this case children, do act as influencers, their influence may affect both the type and amounts of products actually consumed by the family.

Looking at the model from the mother's viewpoint suffers from the limitations of monomethod data collected from a single family member (Szybillo, Sosanie, and Tenebein 1979). However, these data have the advantages of robustness and a degree of replication which provides a useful reliability check.

Sample and Methodology

The data were obtained from the 1979 Needham, Harper and Steers Life Style Study, which uses the Market Facts mail panel. Needham, Harper and Steers describe the data base as follows:

"The Market Facts' mail panel is balanced for geographic region, age, income, and degree of urbanization. The very poor, the very rich, the transient, and minority populations are not well represented in the panel. Needham, Harper and Steers place a further restriction on the sample by requiring all individuals to be married. This latter restriction, coupled with the general characteristics of the Market Fact's mail panel, tends to confine the representativeness of the Life Style sample to stable, middle class households... (it) has proven to be an effective barometer of middle America."

The research questions to be answered using these data are:

1. Do mother's attitudes toward a variety of family-related and social issues influence their perceptions of the amount of influence their children have on their brand choices in selected product categories?
2. Does the amount of influence children have affect the amount of family consumption in that particular product

¹The authors wish to thank Needham, Harper & Steers Advertising, Inc. for providing access to its 1979 Life Style Study data which were used in this project.

category?

Data for mother's attitudes (Table 1) was obtained from the AIO Section of the questionnaire. Since the influence questions dealt primarily with food products, although they also included some items of children's clothing, we considered it desirable to include specific attitudes towards food (Concern for Nutrition), food prices, and money matters in general (Price Conscious, Family Financial Matters), and homemaking (Involved Homemaker). We also included more general measures of attitudes, social attitudes toward media (Concerned About Sex and Violence) and toward society. Some of the social attitudes seem to be focused inwardly on personal and family concerns (Traditional, Family Togetherness, Concerned About Our Society), while others suggest a reaching out for new ideas and experiences (Cosmopolitan, Venturesome).

The attitude scales were constructed by judgmentally selecting from the approximately 200 available AIO items, measured on a six-point agree-disagree scale, a set of items which appeared to tap the attitude of interest. Each set of items was then subjected to reliability analysis. Unreliable items were removed until examination of the item-total correlations and matrix indicated that no further improvement in alpha was possible. Cronbach's alpha is presented twice for each scale, once for all women in the sample (n=1601) and again for the subsample who had children or grandchildren aged 0-19 (n=1150). All remaining data deals with the subsample only.

Children's influence on brand choice for 35 products was measured on a four point 'Almost All the Time' to 'Never' scale. Because of the lack of precision in this measure, the responses were factor analyzed (Table 2). After

TABLE 1
MOTHERS' ATTITUDES SCALES²

Scales	Reliability (Cronbach's alpha)	
	All Women (n=1601)	Only Women Reporting Children's Influence (n=1150)
<u>Concerned About Sex and Violence</u>		
There is too much violence on prime time television.		
There is too much emphasis on sex today.		
There is too much sex on prime time television.		
TV commercials place too much emphasis on sex.	.73	.73
<u>Concerned About Nutrition</u> (12 items)		
I am more concerned about nutrition than most of my friends are.		
I try to eat natural foods most of the time.		
All food advertising should reveal the calories in the product.	.78	.78
<u>Price Conscious</u>		
I never know how much to tip.		
I always check prices even on small items.		
I'm eating out less often because of rising restaurant prices.		
I like to save and redeem savings stamps.		
I shop a lot for specials.	.49	.51
<u>Family Financial Matters</u>		
Our family is too heavily in debt today.		
Investing in the stock market is too risky for most families.		
My greatest achievements are still ahead of me.		
I use credit cards because I can pay the bill off slowly.		
Five years from now our family income will probably be a lot higher than it is now.	.68	.69
<u>Involved Homemaker</u> (11 items)		
When I see a full ashtray or wastebasket, I want it emptied immediately.		
I am a homebody.		
The kind of dirt you can't see is worse than the kind of dirt you can see.	.66	.68
<u>Cosmopolitan</u>		
I am interested in the cultures of other countries.		
I like to visit places that are totally different from my home.		
I would like to take a trip around the world.		
I would like to spend a year in London or Paris	.64	.65

MOTHERS' ATTITUDES SCALES CONTINUED

Scales	Reliability (Cronbach's alpha)	
	All Women (n=1601)	Only Women Reporting Children's Influence (n=1150)
<u>Traditional</u> (20 Items)		
I have somewhat old-fashioned tastes and habits. A woman's place is in the home. Everything is changing too fast today.	.74	.74
<u>Venturesome</u> (14 items)		
I like to think I am a bit of a swinger. I like the feeling of speed. I would like the type of work which would keep me constantly on the move.	.73	.74
<u>Concerned About Our Society</u> (21 items)		
Most big companies are just out for themselves. The energy shortage is a hoax created by the govern- ment, utilities, and corporations. Unions have too much power in America today.	.73	.71
<u>Family Togetherness</u> (20 items)		
A wife's first obligation is to her husband, not her children. When children are ill, parents should drop everything to see to their comfort. Children are the most important thing in a marriage.	.69	.68

²Sample items only are shown for long scales. Complete scales are available from the authors.

TABLE 2

CHILDREN'S INFLUENCE FACTORS

Item	Factor Loading	Reliability (Cronbach's alpha)	Item	Factor Loading	Reliability (Cronbach's alpha)
<u>Children's/Pet Foods</u>			<u>Clothing/Cereal/Cookies</u>		
Packaged dry pasta	.79		Children's shoes	.83	
Dog Food	.67		Children's Clothes	.82	
Cheese	.60		Cold cereal	.50	
Peanut butter	.58		Toys	.49	
Cake Mix	.54		Packaged cookies	.40	
Hot dogs	.53		% of variance explained	.05	.77
Vitamins	.51		<u>Sweets and Snacks</u>		
Soup	.48		Candy	.64	
Jam and Jelly	.45		Cookie mixes	.62	
Cat Food	.44		Canned pudding	.61	
Brownie mix	.43		Canned fruit drink	.57	
Ice Cream	.42		Snack cakes	.54	
Potato chips	.42		Canned pasta	.51	
% of variance explained	.32	.82	Cold cereal	.50	
<u>Gum</u>			Ice cream	.49	
Sugarless bubble gum	.81		Powdered soft drink	.47	
Regular bubble gum	.77		Potato chips	.47	
Regular chewing gum	.71		Cake mixes	.47	
Sugarless chewing gum	.67		Jam/jelly	.46	
% of variance explained	.06	.80	Brownie mixes	.46	
			Restaurants	.44	
			Carbonated soft drinks	.41	
			% of variance explained	.04	.85

varimax rotation, four factors were obtained which explained over 47% of the variance, a quartimax rotation was also carried out to see if all loadings could be forced onto one factor. Quartimax rotation also produced four factors, all having high correlations with the original four (unrotated) factors. Results of the varimax rotations were therefore used in the analyses.

The four influence factors, which bear strong resemblance to those in an analysis of earlier Needham, Harper and Steers data (Mehrota and Torges 1976), were called Children's/Pet Foods, Gum, Clothing/Cereal/Cookies, and Sweets and Snacks. The analysis consists of cross-tabulations of the attitude scale scores against the influence factors followed by crosstabulations of the influence factors against reported family use, measured on a six-point 'Less than once a Month' to 'More than once a Day' scale, of all available products in the influence factors. Limiting our analysis to a relatively few scales and factors tends to preclude finding a few significant relationships by chance alone.

Results

In analyzing the data we used cross-tabulations rather than correlations because the influence categories and the use frequency categories do not, strictly speaking, constitute scales in which categories are equidistant. The predictor variables, which are the mean scores from each attitude scale, have been collapsed into three categories representing high, medium and low levels of agreement with the attitudes making up the scale. The criterion variables represent, respectively, four influence categories and three usage frequency categories.

The analysis resulted in the production of forty-four three by four tables representing attitude/influence relationships. Space limitations make it impossible to present all of these data; therefore we will present only the probability of significance ($p \leq$) Chi-Square statistic for each table. (A complete set of tables is available from the authors on request). The analysis was conducted using as the research population all mothers and grandmothers who reported children's influence.

The Relationship Between Attitudes and Perceived Influence of Children

Table 3 shows the Chi-Square relationship between all of the attitude scales and each of the four product categories. All of the statistically significant relationships are linear. The pattern of results, while reasonable, is not always easily explainable.

High concern over Family Financial Matters predicted low levels of child influence for all product factors except Gum, yet Price Consciousness predicted low influence levels only for Children's/Pet Foods. Thus, one dimension which determines the extent of children's influence is economic, but it appears that general economic status is of more importance than specific price consciousness. Concerned about Nutrition predicted children's influence in three of the four categories. The fact that it did not predict for chewing gum may simply reflect mother's recognition that gum neither adds to nor subtracts from the child's nutritional well-being and is therefore of little consequence.

Some of the social-attitude-related scales also demonstrated significant predictive power. Although Traditional Orientation predicted a low degree of influence for Clothing/Cereal/Cookies and Children's/Pet Foods, high Venturesomeness predicted higher degrees of influence for Children's/Pet Foods, Gum and Clothing/Cereal/Cookies. A higher degree of Concern about Our Society and Family Togetherness (conservative dimensions) predicted significantly lower degrees of influence for all four product categories. It is not unreasonable to characterize these scales as constituting a rough measure of a liberal versus a conservative orientation toward child rearing and thus to conclude that a conservative orientation is indicative of a smaller role for the child in purchase decision-making.

Thus, there are at least three dimensions which affect the extent to which children are described by these mothers as having influenced purchase decisions for four groups of products. One dimension is economic, the second is health-related (nutrition) and the third is a liberal versus a conservative orientation. We shall next explore, by looking at the relationships between influence factors and product use, whether this influence results in greater purchases of the products by the family.

Controlling for Children's Ages

A further analysis was performed, controlling for the presence or absence of children in four age categories. The categories used were 0-2, 3-7, 8-12 and 13-19 years of age, the first two of which approximate the stages of cognitive development. In the majority of cases where the two-way crosstabs were significant, a consistent pattern emerged in the three-way tables.

This is an example in which Concerned About Nutrition attitude scale is crosstabbed by the Children's/Pet Food influence factor and controlled by the presence and absence of children in the four age categories (only the significance level of the Chi-Square is shown). The presence

TABLE 3

SUMMARY OF RELATIONSHIPS BETWEEN MOTHER'S ATTITUDE SCALES AND CHILDREN'S INFLUENCE FACTORS

Significance Level of Chi-Square Statistic ($p \leq$)

Attitude Scale	By Children's Pet Foods	By Gum	By Clothing/ Cereal/Cookies	By Sweets and Snacks
Concerned About Sex and Violence	.18	.19	.07	.96
Concerned About Nutrition	.000	.27	.04	.01
Price Conscious	.008	.25	.31	.30
Family Financial Matters	.04	.60	.02	.005
Involved Homemaker	.51	.27	.34	.13
Cosmopolitan	.59	.93	.48	.30
Traditional	.01	.83	.03	.48
Venturesome	.20	.76	.17	.03
Concerned About Our Society	.001	.08	.06	.02
Family Togetherness	.001	.001	.03	.01

Age Categories	p ≤	Age Categories	p ≤
0 - 2	.000/.59	8 - 12	.37/.0001
3 - 7	.47/.0001	13 - 19	.01/.001

of children aged 0-2 is highly significant; the direction of the relationship is that the more concerned mother is about nutrition the less she allows her children to influence her, which makes good intuitive sense. However, it is the absence of children in the other three categories that shows significance, not their presence. In these tables the relationships also follow the direction seen in the two-way crosstabs in which more concern led to less influence. It is difficult to explain how the absence of children would have this effect, and for this reason the complete set of controlled crosstabs is not presented for either attitudes/Children's influence or influence/family product usage in which the same pattern also tended to emerge.

A further attempt was made to control for children's age by creating subsamples of families with children in specific age groups. However, only one group of families (those with children in the 0-2 age range) emerged as a separable subsample; all other families had children whose ages spanned two or more of the categories, even when age categories were re-defined as preschool (0-5), young (6-12) and teens (13-19).

Another controlling variable which separated families according to the age category into which the greatest number of children fell, or by the age of the oldest child(ren) when there were equal numbers of children in more than one category, was also used. This variable provided no further explanation of the variance in the data, and was discarded. Having children in other than the primary category appeared to create confounding effects.

The Relationships Between Children's Influence and Family Product Usage

Table 4 shows the Chi-Square relationships between influence factors and product usage. In this analysis, we will be comparing the extent to which mothers say their children influence purchases of each product factor with frequency of use of each of the available products within these factors.

This table shows that the relevant influence factor affected the household's usage of a substantial number of the products studied. In each case where there was a significant effect it was positive; a higher degree of influence was accompanied by a higher rate of product usage by the family. The common denominator among all the products, with the exception of canned dog food, is that they can easily be consumed alone by one of the family members. It is reasonable, therefore, that influence should predict household demand.

A comparison of the results in this table with those in Table 3, which presented the relationship between attitudes and influence, suggests that it may be important not only to identify whether children are influential but whether their influence is directed primarily at generic product category demand or toward brand level demand.

Implications for Research and for Marketing Action

Our findings should be considered as tentative and exploratory. In this paper we have only begun to explore data available even in the data set we used. Next analytical steps should aim towards developing multidimensional classifications of attitudes, perception of children's influence and consequential purchase behavior. Two techniques that might be employed in this analysis are multivariate analysis of variance (MANOVA) and cluster analysis. Both these techniques should contribute to our understanding of the relationship. A Manova analysis would use the set of attitude factors as predictors and the individual products within product influence factors as criterion variables. A cluster analysis would use all three sets of variables, attitude scales, influence factors and use frequencies to develop behavioral market segments.

Future research should also concentrate on obtaining data which indicates the influence each child has on the family's decision making process. In developing and testing models of influence, attention ought to be paid not only to the age of the children, but to the spacing between children. It may well be that children grouped closely together exert more influence, due not only to greater numbers but to allied interests, than those with more years between. It may also be that differences, attitudinal as well as cognitive, in individual children affect the degree

TABLE 4

SUMMARY OF RELATIONSHIPS BETWEEN CHILDREN'S INFLUENCE FACTORS AND FAMILY PRODUCT USAGE

<u>Children's/Pet Foods</u>	p ≤	<u>Clothing/Cereal/Cookies</u>	p ≤
Packaged pasta	.004	Packaged cookies	.36
Canned pasta products	.08	<u>Sweets and Snacks</u>	
Canned dog food	.13	Cookie Mixes	.84
Cheese	.61	Packaged cookies	.02
Peanut butter	.45	Snack cakes	.03
Hot dogs	.13	Canned pasta	.59
Soup mix	.01	Ice cream	.002
Jam/Jelly	.49	Powdered soft drink mix	.0004
Canned Cat food	.67	Cake Mix	.16
Cake mix	.23	Jam/Jelly	.46
Brownie Mix	.05	Brownie Mix	.14
Ice cream	.05		

to which they attempt and are able to exert influence.

The implications of our findings for marketing management are obvious. Evidently, children affect generic demand in some product categories as well as affecting brand choice. Therefore, it should be worthwhile to estimate the potential market size that can result from children's demand and to direct advertising to them.

Evidently, certain children also affect brand level demand in some product categories. Our findings suggest possible marketing strategies based on these findings. For example, children seem to be more influential when their mothers have a liberal orientation, and less influential when their mothers are conservative. One might then advertise primarily to children in the liberal Northeast and advertise the same brand primarily to mothers in the Midwest. Or, one could limit advertising to mothers to conservative-oriented vehicles. In either instance the advertisements should take into account product-specific parental attitudes and emphasize how the products contribute to children's well-being.

Conclusion

This analysis carries the work on children as influencers and consumers one important step further by expanding the scope of previous research on the effect of parental attitudes on children's influence and by demonstrating that children's influence does affect the frequency, and presumably therefore the magnitude, of household consumption of certain products. Although the difficulties inherent in extending this stream of research are great, the conceptual and practical insight to be gained from it will certainly justify the effort.

References

- Berey, Lewis A. and Pollay, Richard W. (1968), "The Influencing Role of the Child in Family Decision Making," Journal of Marketing Research, 5, 70-72.
- Caron, Andre and Ward, Scott (1957), "Gift Decisions by Kids and Parents," Journal of Advertising Research, 15, 15-20.
- Davis, Harry L. (1977), "Decision Making Within the Household," in Robert Ferber ed., Selected Aspects of Consumer Behavior, Washington, D.C., U.S. Government Printing Office.
- Frideres, James S. (1973), "Advertising, Buying Patterns and Children," Journal of Advertising Research, 13, 34-46.
- Jenkins, Roger L. (1978), "The Influence of Children in Family Decision-Making: Parents' Perceptions," in William L. Wilkie ed., Advances in Consumer Research, Vol. VI, 413-418.
- Mehrota, Sunil and Torges, Sandra (1976), "Determinants of Children's Influence on Mothers' Buying Behavior," in William D. Perreault, Jr. Ed., Advances in Consumer Research, Vol. IV, 56-60.
- Moschis, George P. and Moore, Roy L. (1979), "Decision Making Among the Young: A Socialization Perspective," Journal of Consumer Research, 6, 101-112.
- Nelson, James E. (1978), "Children As Information Sources in the Family Decision to Eat Out," in William L. Wilkie ed., Advances in Consumer Research, Vol. VI, 419-423.
- Popper, Edward T. (1979), "Mothers Mediation of Children's Purchase Requests," in 1979 American Marketing Association Combined Proceedings, 645-648.
- Robertson, Thomas S. and Feldman, Shel (1975), "Children As Consumers: The Need for Multitheoretical Perspectives," in Beverlee B. Anderson ed., Advances in Consumer Research, Vol. III, 508-512.
- Rossiter, John R. (1978), "Children's Consumer Research: A Call for Rigor," in William L. Wilkie ed., Advances in Consumer Research, Vol. VI, 424-426.
- Szybillo, George J., Sosanie, Arlene K. and Tenebein, Aaron (1979), "Family Member Influence in Household Decision Making," Journal of Consumer Research, 6, 312-316.
- _____, _____, _____ (1976), "Family Decision Making: Husband, Wife and Children," in William D. Perreault, Jr. ed., Advances in Consumer Research, Vol. IV, 469-49.
- Ward, Scott (1974), "Consumer Socialization," Journal of Consumer Research, 1, 1-17.
- _____. (1978), "Research on Marketing and Children: Upside or Downside on the Product Life Cycle?," in William L. Wilkie ed., Advances in Consumer Research, Vol. VI, 427-429.
- _____ and Wackman, Daniel B. (1972), "Children's Purchase Influence Attempts and Parental Yielding," Journal of Marketing Research, 9, 316-319.
- _____, Popper, Edward and Wackman, Daniel (1977), Parents Under Pressure: Influences on Mothers' Responses to Children's Purchase Requests, Cambridge: Marketing Science Institute.

FOUR SITUATIONS AND THEIR PERCEIVED EFFECTS ON
HUSBAND AND WIFE PURCHASE DECISION MAKING

Alvin C. Burns, Louisiana State University
Stephen P. DeVere, Louisiana State University

Abstract

This study investigates the effects of four purchase decision making settings - at home alone, at the store alone, at home with a salesman present, and at the store with a salesman present - on husbands' and wives' perceptions of three dependent variables: relative influence, discussion times, and resolution modes given disagreement. An equivalent groups design was used with five subdecisions for each of four products. Physical surroundings and social circumstances were found to affect estimated discussion times and disagreement resolutions, but no effects were evident for perceived shared influence.

Introduction

Recent findings in the area of husband and wife purchase decision making influence have begun to reveal the complexity of the topic. It is common knowledge that influence distributions vary across products, across decisions within products, across phases in the decision making process, across families, and across spouses (Davis, 1976). Although context has been largely ignored in husband and wife purchase decision making studies, there is reason to believe that variance will also be found across situations in which the interaction takes place. In terms of individual purchase decision making, ample support for this contention comes from the arguments by Belk (1975) and Russell and Mehrabian (1976).

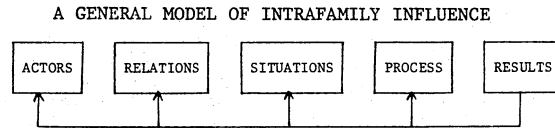
The husband and wife decision process would seem to be especially rich with regard to the opportunity for differences between situations inasmuch as the process necessarily occurs through interpersonal communication channels, most of which are overt. That is, spouses must communicate to each other their thoughts, opinions, and preferences. Furthermore, they interact to reach decisions. This requisite for joint decision making brings the process out into an observable realm. In fact, Burns and Granbois (1979) contend that analysis of tape recordings of husband and wives undertaking decisions is comparable to the protocol analysis approach applied to individuals and will greatly help in the understanding of collective purchase decision making.

One can envision situations in which the communication between spouses might be affected by the context. Obviously, individuals will differ in their sensitivities for situational factors, but when there exists the possibility of scrutiny of the conversation by a third party or parties, and the prospect of social evaluation of roles, it is conceivable that situation will influence the purchase decision making process. Consequently, this study may be regarded as an initial attempt to ascertain the effects of four different purchase contexts on three dimensions of the husband and wife purchase decision making process.

Conceptual Framework and Hypotheses

Much of the background for this study is described in a working paper by the authors (1979). This particular paper represents an initial empirical study investigating selected aspects of the comprehensive model proposed in the first paper. The model delineates attributes comprising each of the five "blocks" represented in the general schema:

FIGURE 1



The study focuses on two situational dimensions in which husband and wife purchase decisions can occur: the physical surroundings and the social circumstances. Taken as a 2x2 paradigm, the four situations of interest are: (1) at home alone; (2) in the store alone; (3) at home with a salesperson present; or (4) in the store with a salesperson present. Each of these circumstances implies different degrees of potential scrutiny by commercial agents. Similarly, the store atmosphere implies a less familiar context than does the home situation. Couched somewhat differently, one could contend that the greater the potential for social and commercial scrutiny and pressure, the more pressure for normative compliance, while the opposite would be the case in situations where influence interaction takes place in the privacy of one's home. Consequently, it seems appropriate to hypothesize differences in interaction results from situation to situation.

Three dependent variables are addressed in this study, all drawn from the "Results" block. Perceived distribution of purchase influence has been investigated as the primary dependent variable in most family purchase decision making studies (Davis 1976, Burns and Granbois 1979). The second dependent variable is the estimated time of discussion for various purchase decisions. This variable has not received attention, except in observation studies or analysis of spouses' discussions such as those reported by Kenkel (1957). The temporal dimension appears to be a logical impact point for situational influences. The third dependent variable, "disagreement outcomes," is analogous to the "recognized authority" variable studied and reported by Burns and Granbois (1977) wherein spouses indicated whose choice would be the final decision in the event of a disagreement on the resolution of a purchase decision. Again, the nature of disagreement outcomes is logically linked to discussion context. While other dependent variables are described in the model, these three were selected as representative and as ones most likely to be affected by the independent variables.

The study's hypotheses are stipulated below.

Hypothesis 1

Differences will be evident in all three dependent variables across the two situational dimensions.

Hypothesis 2

The "At Home Alone" situation will be associated with: (a) more equal sharing of influence; (b) longer purchase decision discussions; and (c) more joint resolution of disagreements than will the "At the Store with Salesman Present" situation.

The expectations of the behaviors of the dependent variables under the other two situations are not clear at this point. Consequently, only the two extremes are addressed

in this study.

Other hypotheses were also generated for the study as a result of knowledge about the husband and wife purchase decision making process. In general, the following hypotheses require support as corroboratory evidence that the current study is consistent with other studies.

Hypothesis 3

Husbands' and wives' perceptions of purchase decision influence will differ.

Hypothesis 4

Differences will be evident in purchase decision influence across products.

Hypothesis 5

Differences will be evident in purchase decision influence across product subdecisions.

Method

The Four Situations

Brief scenarios were constructed for each of the four situations, and all four descriptions are included in the Appendix. In each, the spouses were instructed to maintain a mental image of the context while responding to the questionnaire. In addition, the beginning of each dependent variable section in the questionnaire reminded respondents of the purchase situation they were to retain. Admittedly, these scenarios are weak stimuli, but the authors believe that significant findings in such a case are compelling.

Products and Decisions Studied

Four products were chosen which could "fit" into any and all of the situations. Additionally, certain constraints on the product choice were imposed by the nature of the sample of married couples (see below). The four products included were: (1) vacuum cleaner; (2) life insurance; (3) encyclopedia set; and (4) pots and pans. For each product, five subdecisions were distinguished: (1) verify the need to buy one; (2) what price to pay; (3) how to pay for it; (4) what brand to buy; and (5) what model or type to buy. These five subdecisions or close variations have been studied by others (Burns and Granbois 1979).

Dependent Variables

Relative purchase influence was operationalized as a specification out of 100 percentage points (constant sum scale) as to the amount of influence the husband and wife, respectively, would have for each of the five subdecisions for each of the four products. Jenkins (1978) has shown this approach to be somewhat more reliable than others, while Szybillo, Sosanie, and Tenenbein (1979) have shown the equivalence of a constant sum scale to 5-point scales. Discussion time was operationalized as an estimate of how long the couple would talk about each of the subdecisions for each product. Finally, disagreement outcomes were indicated by a specification of how the decision would be resolved assuming some disagreement did arise. Spouses could indicate "husband" or "wife," signifying that that person's choice would be selected, or "together," indicating that the disagreement would require a joint discussion and determination of the outcome. Test-retest correlations were determined to assess reliability. Forty-two respondents (21 husbands and 21 wives) were telephoned within two weeks of the survey and requestioned on representative items from each variable. The correlations were .88 and .77 for the influence and time estimates, respectively.

Also, 79% gave perfectly consistent responses to the disagreement outcomes item retested. Two respondents could not recall the situation presented; all others recalled it correctly (95%).

Research Design and Sample

Questionnaire length and contamination effects precluded asking each respondent to go through all four situations; consequently, a split-plot design with repeated measures on products and decisions was adopted. The sample was drawn from married students living in married student housing on the LSU campus. Only American students were used in the sample, and spouses received identical questionnaires. A field worker introduced the study to respondents and left the questionnaires with the couple. Questionnaires were color-coded; moreover, spouses were requested to sign a statement certifying that they had not compared answers and made any changes as a result. Sample selection was based on the desire to minimize variability between groups and convenience. The final sample size for the study was 148 couples with the following group sizes: At Home Alone, 38 couples; Alone in the Store, 37 couples; At Home with a Salesman, 39 couples; and In the Store with a Salesman Present, 34 couples.

Findings

Equivalence Tests

Equivalence tests verified the equality of the four groups. Several demographic variables were subjected to analysis of variance, and no statistically significant differences were determined. The demographic variables and their means were: Husband's Age (24.7 years); Wife's Age (23.2 years); Years Married (2.7 years); Number of Children (.6 children); Husband's College Education (4.6 years), Wife's College Education (3.4 years); Husband's Workweek (22.2 hours); and Wife's Workweek (27.8 hours).

Purchase Influence

The investigation of the distribution of purchase influence began with the mean percentage allocations for husbands and wives assigned by husbands and assigned by wives which are presented for inspection in Table 1. This information served as the data for an analysis of variance determination of the hypothesized differences. For this analysis, only the estimate of the husband's influence was used due to the constant sum interdependent nature of the scale. Because of the confounding efforts of the repeated measures, multiple error terms are required. The couple was used as the basic unit of analysis. The results of this analysis, and salient means are reported in Tables 2 and 3, respectively. The corroboratory hypotheses for products and decisions were supported; however, the hypothesis of differences in spouses' perceptions was not. Main effects for situations are not evident, but significant interaction effects were found, however, between physical situation and product as well as decision. At best, therefore, situation can be viewed as a moderator variable for the main effects of these two variable already known to influence perceived purchase decision making influence distributions.

Estimated Discussion Times

The analysis of the estimated time of discussion for each subdecision assumed a similar tact. The means for the husbands and wives groups are reported by situation in Table 4. Tables 5 and 6 contain analysis of variance results and relevant means, respectively. In this instance significant differences were found for the main effects of all independent variables (physical situation, social situation, spouse, product and decision); while interaction effects were determined for isolated cases.

TABLE 1
AVERAGE INFLUENCE ASSESSMENTS BY SPOUSES ACROSS SITUATIONS*

PRODUCES AND DECISIONS	SITUATIONS							
	At Home Alone		Alone At The Store		At Home With A Salesman		In The Store With A Salesman	
	Husbands Group	Wives Group	Husbands Group	Wives Group	Husbands Group	Wives Group	Husbands Group	Wives Group
Vacuum Cleaner Decisions								
The need to buy one	40%	35%	38%	43%	41%	34%	42%	35%
What price to pay	55%	51%	57%	47%	51%	50%	55%	52%
How to pay for it	62%	57%	63%	52%	55%	55%	59%	53%
What brand to buy	45%	40%	41%	39%	45%	41%	41%	40%
What model or type	40%	34%	35%	37%	41%	39%	30%	35%
Life Insurance Decisions								
The need to buy one	60%	63%	60%	57%	58%	60%	59%	61%
What price to pay	63%	65%	65%	61%	59%	59%	60%	60%
How to pay for it	64%	60%	67%	58%	55%	54%	57%	56%
What brand to buy	60%	69%	68%	62%	62%	64%	59%	62%
What model or type	63%	69%	68%	64%	62%	63%	60%	60%
Encyclopedia Decisions								
The need to buy one	50%	50%	49%	49%	55%	50%	53%	53%
What price to pay	54%	55%	60%	54%	53%	50%	55%	53%
How to pay for it	61%	56%	62%	55%	54%	55%	55%	52%
What brand to buy	52%	50%	51%	51%	53%	53%	55%	55%
What model or type	52%	50%	52%	52%	52%	53%	55%	54%
Pots and Pans Decisions								
The need to buy one	29%	23%	24%	32%	30%	30%	31%	29%
What price to pay	42%	37%	39%	38%	35%	42%	45%	45%
How to pay for it	57%	48%	51%	62%	46%	48%	56%	47%
What brand to buy	29%	24%	27%	47%	32%	31%	32%	28%
What model or type	28%	23%	25%	29%	30%	31%	32%	27%
Number of Respondents	38	38	37	37	39	39	34	34

*Interpret as Husband Influence %. Wife Influence=(100%-Husband Influence)

TABLE 2
ANALYSIS OF VARIANCE RESULTS FOR INFLUENCE ASSESSMENTS

Source	S.S.	d.f.	F Value	Significance Probability
Physical Situation	142.6	1	.07	
Social Situation	11.1	1	.07	
Physical X Special	10.0	1	.00	
Error a	315016.6	144	5.28	
Spouse	2801.5	1	1.89	
Spouse X Physical	1659.4	1	1.12	
Spouse X Social	480.2	1	.33	
Spouse X Physical X Social	418.2	1	.28	
Error b	425194.2	288	3.65	
Product	571554.5	3	414.70	.01 d
Decision	113376.7	4	61.7	.01
Product X Physical	6614.7	3	4.8	.01
Product X Social	735.9	3	.5	
Product X Decision	111798.8	12	20.3	.01
Product X Spouse	2870.2	3	2.1	.10
Decision X Physical	4460.2	4	2.4	.05
Decision X Social	1271.1	4	.7	
Decision X Spouse	2923.6	4	1.6	
Product X Decision X Physical	2359.7	12	.4	
Product X Decision X Social	1948.6	12	.4	
Product X Decision X Spouse	1435.1	12	.3	
Product X Physical X Social	218.4	3	.2	
Product X Physical X Spouse	108.7	3	.1	
Product X Social X Spouse	2661.5	3	1.9	
Product X Decision X Physical X Social	2329.8	12	.4	
X Social				
Product X Decision X Physical X Spouse	1182.8	12	.2	
Product X Decision X Social X Spouse	1835.9	12	.3	
Product X Physical X Social X Spouse	5660.9	3	4.1	.01
Decision X Physical X Social	1623.7	4	.9	
Decision X Physical X Spouse	1341.14	4	.7	
Decision X Social X Spouse	1920.6	4	1.0	
Decision X Physical X Social X Spouse	678.2	4	.4	
Product X Decision X Physical X Social X Spouse	852352.2	159	11.7	.01
Error c	2715908.7	5911	3.27	

a Couples (Situations)
b Couples (Situations x Spouse)
c Couples (Products x Decisions x Situations x Spouse)
d Only those $\leq .10$ reported

TABLE 3
MEANS FOR INFLUENCE ASSESSMENTS*

Subject	Mean Value
Products	
Vacuum Cleaner	45.1%
Life Insurance	61.4%
Encyclopedia	53.3%
Pots and Pans	35.1%
Decisions	
The need to buy one	44.4%
What price to pay	52.0%
How to pay for it	55.5%
What brand to buy	46.5%
What model or type	45.1%
Situations	
At Home Alone	49.0%
Alone At the Store	48.9%
At Home With a Salesman	48.2%
In The Store With a Salesman	48.6%
Spouse	
Husbands	49.5%
Wives	47.9%

* Husband Influence is reported; Wife Influence is (100% - Husband Influence)

Further presentation of the means is in Table 7 which compares the husbands group to the wives group. Duncan multiple range tests revealed that in relative terms, the estimates are consistent across spouses, yet differences persist in absolute levels with wives always specifying more estimated time than did husbands. Product and sub-decision effects differences are in evidence with the Pots and Pans decisions expected by both spouses to take the least time and life insurance decisions expected to take the most time. Need verification is expected to take the longest time by both spouses, while the resolution of how to pay is anticipated to be shortest in duration. The "At Home Alone" situation is perceived to be the one in which the discussion times would be the longest while the "At the Store With a Salesman Present" discussions are estimated to take about half as long, on the average. The second hypothesis is therefore supported.

TABLE 4
AVERAGE ESTIMATED DISCUSSION TIMES IN MINUTES

PRODUCTS AND DECISIONS	SITUATIONS							
	At Home Alone		Alone At The Store		At Home With A Salesman		In The Store With A Salesman	
	Husbands Group	Wives Group	Husbands Group	Wives Group	Husbands Group	Wives Group	Husbands Group	Wives Group
<u>Vacuum Cleaner Decisions</u>								
The need to buy one	23.4	23.3	13.0	17.1	12.5	14.0	9.2	7.9
What price to pay	18.3	18.2	11.8	14.1	13.1	12.7	10.0	11.3
How to pay for it	11.2	12.0	7.9	13.1	7.6	8.5	4.1	6.9
What brand to buy	18.6	20.5	13.9	14.9	13.0	11.9	8.6	9.6
What model or type	17.4	18.6	12.2	13.7	11.8	10.7	7.7	8.2
<u>Life Insurance Decisions</u>								
The need to buy one	33.8	33.6	23.2	27.1	28.3	25.7	21.2	15.9
What price to pay	25.2	27.6	19.0	23.1	17.2	22.9	17.9	17.2
How to pay for it	13.3	21.1	11.7	18.1	11.2	17.5	11.8	9.2
What brand to buy	24.4	30.6	25.6	23.5	20.3	23.5	16.6	16.9
What model or type	21.0	31.5	20.2	19.8	18.1	21.7	17.4	16.8
<u>Encyclopedia Decisions</u>								
The need to buy one	24.3	27.2	16.9	14.5	18.9	14.8	13.3	12.4
What price to pay	14.2	21.2	12.0	12.9	14.1	12.7	9.0	12.8
How to pay for it	9.7	15.7	7.0	11.5	7.8	8.8	6.6	9.0
What brand to buy	13.8	22.5	14.4	12.7	11.2	10.8	8.6	11.4
What model or type	11.4	19.6	10.9	12.0	10.0	9.9	6.2	9.2
<u>Pots and Pans Decisions</u>								
The need to buy one	11.7	17.3	7.8	10.0	9.6	11.6	7.4	9.2
What price to pay	10.3	16.3	7.7	10.1	9.0	10.9	6.4	10.0
How to pay for it	7.4	10.9	5.4	9.2	6.5	6.7	4.1	6.7
What brand to buy	7.6	15.6	9.0	10.5	7.5	8.9	6.5	7.8
What model or type	7.5	15.2	7.2	8.7	7.3	9.0	6.3	6.7

TABLE 5
ANALYSIS OF VARIANCE RESULTS FOR DISCUSSION TIME ESTIMATES

Source	S.S.	d.f.	F Value	Significance Probability
Physical Situation	27387.8	1	7.42	.01 d
Social Situation	21064.9	1	5.71	.05
Physical X Special	2620.9	1	.71	
Error a	520235.3	141	16.60	
Spouse	6713.9	1	2.29	.10
Spouse X Physical	1835.2	1	.63	
Spouse X Social	795.1	1	.27	
Spouse X Physical X Social	505.8	1	.17	
Error b	825604.4	282	17.13	
Product	111729.4	3	117.9	.01
Decision	34401.4	4	27.2	.01
Product X Physical	2451.7	3	2.6	.05
Product X Social	1026.4	3	1.1	
Product X Decision	7654.2	12	2.0	.05
Product X Spouse	705.9	3	.7	
Decision X Physical	1489.6	4	1.1	
Decision X Social	3203.5	4	2.5	.10
Decision X Spouse	1234.5	4	1.0	
Product X Decision X Physical	574.2	12	.2	
Product X Decision X Social	462.4	12	.1	
Product X Decision X Spouse	1360.4	12	.4	
Product X Physical X Social	2097.4	3	2.2	
Product X Physical X Spouse	151.9	3	.2	
Product X Social X Spouse	1333.3	3	1.4	
Product X Decision X Physical X Social	564.5	12	.1	
Product X Decision X Physical X Spouse	373.1	12	.1	
Product X Decision X Social X Spouse	637.2	12	.2	
Product X Physical X Social X Spouse	2097.4	3	2.2	
Decision X Physical X Social X Spouse	982.2	4	.8	
Decision X Physical X Spouse X Spouse	181.2	4	.1	
Decision X Social X Spouse X Spouse	797.2	4	.6	
Decision X Physical X Social X Spouse X Spouse	982.2	4	.8	
Product X Decision X Physical X Social X Spouse X Spouse	239386.7	159	.5	
Error c	1828510.4	5787	2.77	

- a Couples (Situations)
- b Couples (Situations x Spouse)
- c Couples (Products x Decisions x Situations x Spouse)
- d Only those $\leq .10$ reported

TABLE 6
MEANS FOR DISCUSSION TIME ESTIMATES

Subject	Estimated Minutes
<u>Products</u>	
Vacuum Cleaner	12.9
Life Insurance	21.2
Encyclopedia	13.2
Pots and Pans	9.2
<u>Decisions</u>	
The need to buy one	17.5
What price to pay	14.7
How to pay for it	10.0
What brand to buy	14.8
What model or type	13.4
<u>Situations</u>	
At Home Alone	18.6
Alone At the Store	13.8
At Home With a Salesman	13.2
In The Store With a Salesman	10.3
<u>Spouse</u>	
Husbands	13.4
Wives	14.8

TABLE 7
MEANS FOR DISCUSSION TIME ESTIMATES:
HUSBANDS GROUP COMPARED TO WIVES GROUP

Subject	Husbands Group	Wives Group
<u>Products</u>		
Vacuum Cleaner	12.4	13.4
Life Insurance	20.0	22.4
Encyclopedia	12.1	14.2
Pots and Pans	7.7	10.6
<u>Decisions</u>		
The need to buy one	17.3	17.7
What price to pay	13.6	15.9
How to pay for it	8.4	11.6
What brand to buy	13.9	15.8
What model or type	12.1	14.6
<u>Situations</u>		
At Home Alone	16.3	20.9
Alone At the Store	12.8	14.8
At Home With a Salesman	12.7	13.7
In The Store With a Salesman	9.9	10.7

Disagreement Outcomes

The final set of analyses concerned the recognized authority variable. Preliminary investigation began with husbands and wives groups distributions for each situation treatment. The distributions are too cumbersome for display here. Analysis of differences was performed with the use of a generalized least squares linear model approach developed by Grizzle, Starmer, and Koch (1969). The method partitions multinomial responses to test for main and interaction effects. The three corroboratory hypotheses were supported, and social situation main effects were evident but no support was found for physical situation effects. Significant interaction also emerged. Table 9 compares the two situations of interest for the second hypothesis. Only the percent of husbands and wives stipulating "joint" resolution is reported. The hypothesis is not supported. In fact, the reverse of the predicted result seems evident: more joint resolution is believed to occur in the store with salesman present than at home alone.

TABLE 8

ANALYSIS OF CATEGORICAL RESPONSES FOR DISAGREEMENT OUTCOMES

Source	d.f.	Chi-Square	Significance Probability ^a
Physical Situation	2	.5	
Social Situation	2	18.9	.0001
Product	6	825.1	.0001
Decision		224.9	.0001
Spouse	2	8.5	.01
Physical X Social	2	.2	
Physical X Product	6	13.3	.04
Physical X Decision	8	24.4	.002
Physical X Spouse	2	.7	
Social X Product	6	4.9	.06
Social X Decision	8	12.0	
Social X Spouse	2	3.1	
Product X Decision	24	124.6	.0001
Product X Spouse	6	6.1	
Decision X Spouse	8	23.4	.003
Physical X Social X Product	6	14.1	.03
Physical X Social X Decision	8	7.4	
Physical X Social X Spouse	2	11.1	.004

a Only those \leq .10 reported

Discussion

The findings of this exploratory investigation of situational influences on husband and wife purchase decision making have revealed that spouses are sensitive to context. In particular, both husbands and wives believe that discussions held at home alone will last longer than when they were held in the store with a salesman present. Also, the character of resolution of disagreements is affected by social circumstances. More joint resolution of disagreements is believed to occur in the store and salesman situation. Specification of the amount of shared influence purchase, however, appears independent of the physical or social surroundings in which it is envisioned to occur.

Precisely why this pattern has been found is a matter of speculation, and the external validity of the experiment may be criticized. Conceivably, these couples require a stronger stimulus than the one presented. Counterarguments to this contention are elicited in the significant findings for estimated times and disagreement outcomes, both of which followed the influence assessments section of the questionnaire. A more logical explanation, it seems, emerges from the sample's characteristics. Well-educated, young, newly married couples with few children and in an obviously income-constrained period of their marriages probably engage in high degrees of perceived equal decision making regardless of physical or social circumstances.

TABLE 9

COMPARISON OF DISAGREEMENT OUTCOMES*

Products and Decisions	Situations			
	At Home Alone		In the Store with a Salesman	
	Husbands	Wives	Husbands	Wives
Vacuum Cleaner				
The need to buy one	48%	48%	57%	42%
What price to pay	46%	59%	63%	80%
How to pay for it	49%	56%	54%	66%
What brand to buy	36%	38%	40%	43%
What model or type	38%	36%	40%	43%
Life Insurance Decisions				
The need to buy	56%	54%	51%	66%
What price to pay	56%	41%	46%	66%
How to pay for it	54%	51%	54%	71%
What brand to buy	44%	44%	49%	49%
What model or type	41%	41%	49%	49%
Encyclopedia Decisions				
The need to buy one	72%	64%	74%	80%
What price to pay	62%	64%	69%	80%
How to pay for it	62%	59%	63%	66%
What brand to buy	62%	62%	71%	71%
What model or type	67%	67%	74%	80%
Pots and Pans Decisions				
The need to buy	28%	15%	43%	37%
What price to pay	41%	41%	60%	51%
How to pay for it	35%	59%	66%	69%
What brand to buy	28%	15%	29%	23%
What model or type	28%	15%	20%	23%

* Only the percent of "joint" specifications is reported.

This study of situational influences has served to demonstrate the multidimensionality of the husband and wife purchase decision making process. Clearly these and other dimensions deserve more research consideration with sensitivity for context.

Appendix

Descriptions of the Four Situations

At Home Alone

For each of the following purchase decisions, please assume that you and your spouse are discussing them at home alone. That is, there are no other people in the room during your discussion. By "alone" we mean there would be no children, friends, neighbors or relatives present. It is a discussion between you and your spouse concerning various aspects of the purchase listed below.

Alone at the Store

For each of the following purchase decisions, please assume that you and your spouse are discussing them while standing in a store. Other customers are in the store, some of whom are walking by and some of whom are shopping for items at various places in the store. You are not being helped by a salesperson. Even though sales people are within view, no salesperson is helping you at present, and no salesperson is about to approach you. If you have children, they are not with you in the store, and there are no friends, neighbors, relatives with you either.

At Home with Salesman Present

For each of the following purchase decisions, please assume that you and your spouse are discussing them at home with a salesman present. That is, the salesperson for that product has come to your home and made a sales presentation to you. You are now discussing the various aspects of the purchase of that particular product. Assume that no children, if you have children, are present. Also no friends, neighbors, or relatives are with you in this room -- only you, your spouse, and the salesperson are present.

In the Store with Salesman Present

For each of the following purchase decisions, please assume that you and your spouse are in a store discussing them after a sales presentation has been made to you. That is, there are people walking through the store and shopping for various items at various places in the store, and you are discussing the various aspects of this purchase with the salesman standing beside you and listening to your conversation. If you have any children, they are not present with you at this time, and there are no friends, neighbors, or relatives with you at the time. The salesperson will be present throughout the conversation.

References

- Belk, Russell W. (1975), "Situational Variables and Consumer Behavior," Journal of Consumer Research, 2, 157-64.
- Burns, A.C. and Granbois, D. (1977), "Factors Moderating the Resolution of Preference Conflict in Family Automobile Purchasing," Journal of Marketing Research, (February), 77-86.
- _____ and _____. (forthcoming), "Advancing the Study of Family Purchase Decision Making," Advances in Consumer Research, Vol. VII.
- Davis, H.L. (1976), "Decision Making Within the Household," Journal of Consumer Research, (March), 2, 241-60.
- DeVere, Stephen P, and Burns, Alvin C. "An Expanded Model of Family Purchase Decision Making," LSU working paper.
- Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969) "Analysis of Categorical Data by Linear Models," Biometrics 25.
- Jenkins, R. "Minimizing Bias in the Measurement of Spousal Influence in Family Decision Making by Self-Report Methods," Research Frontiers in Marketing: Dialogues and Directions, 1978 Educators' Proceedings, Sabhash, C. Jan (ed.), 238-243.
- Kenkel, W.F. (1957), "Influence Differentiation in Family Decision Making," Sociology and Social Research, (September-October), 42, 18-25.
- Russell, J.A. and Mehrabian, A. (1976), "Environmental Variables in Consumer Research," Journal of Consumer Research, 3, 62-63.
- Szybillo, G.J., Sosanie, A.K. and Tenenbein, A. (1979), "Family Member Influence in Household Decision Making," Journal of Consumer Research, 6, 312-16.

THE ELDERLY CONSUMER: PAST, PRESENT, AND FUTURE

H. Lee Meadow, Virginia Polytechnic Institute and State University
Stephen C. Cosmas, Virginia Polytechnic Institute and State University
Andy Plotkin, Bridgewater State College

Abstract

This paper reviews literature relating to research on the elderly consumer. Relevant empirical findings and propositions are examined to provide a base for future theoretical development. Specifically, the findings and propositions are discussed in a manner conducive for continuing research on this increasingly important population segment. Finally, future directions are indicated to guide consumer researchers in theory development.

Introduction

In the last twenty years, a number of primary and secondary research studies have attempted to describe the elderly consumer. However, it has only been of late, due to increasing size and buying power, that the elderly consumer segment has become critically important. Given this development, it would seem judicious that an examination should be undertaken of what is currently known about the elderly consumers to aid in future consumer research efforts.

The purpose of this paper is to examine the findings of past elderly consumer research in order to provide for an overview of the nature and scope of the elderly segment. Toward this end, relevant findings will be presented to the exclusion of detailed methodological and theoretical criticisms. These findings should establish a foundation for future marketing and consumer research effort.

Literature Review

The Elderly Consumer: The Call for Attention

In the 1960's, U.S. demographers began to chart a growing population trend. The upcoming decades would be characterized by a senior citizen market, defined as those 65+, which would begin to grow at a faster pace than ever before experienced. Because of increasing life spans due to preventive medicine, the aging of the baby boom cohorts, the control of some death causing illnesses, and the increased quality of environmental living conditions, the elderly were predicted to become a major force in terms of size in the U.S. marketplace (Business Week 1960).

This recognition of the growing elderly population led to a number of calls for research to estimate the size and character of the elderly segment. Concurrently, calls were made for marketer reaction to create products and services that would meet the "special" needs of what was then perceived to be a unique \$40 billion market segment (Gidlow 1961, Goeldner and Munn 1964, Morse 1964). However, all did not agree that the elderly segment was important and/or unique enough for generating research and development. Reinecke (1964) concluded from his investigation that there was no unique elderly market, elderly consumption patterns were not homogeneous, and the elderly could not be discerned from other population groupings.

For whatever reasons, a period of three years passed in which interest in the elderly consumer waned. In 1967, Samli provided some empirical evidence that the elderly were indeed a unique demographic market segment. In addition, Goldstein (1968) used Bureau of Labor Statistics

(BLS) data from 1950 to 1960 to independently confirm the contention that the aged market indeed had unique consumption patterns in terms of goods purchased (a finding that he had earlier reported in the Journal of Gerontology, 1965).

Still, few if any attempts were made by researchers and practitioners either to indicate methods for marketing or attempting to market to the elderly (Forbes 1979, Nation's Business 1971). The lack of interest in this market appeared to be based on intuitive judgments of marketing practitioners that the aging (estimated now to be a \$60-\$200 billion market) did not possess unique wants or needs or even want to be reminded that they were growing old (Business Week 1971, Media Decisions 1973, U.S. Department of Commerce 1973). However, this position did not entirely eliminate the academic interest in the elderly consumer. Generally, two to four consumer behavior studies have appeared annually to the present time addressing the aged market.

To gain marketing reaction, Block (1974) and Waddell (1975) demonstrated that the elderly were deserving of special attention as a result of a higher level of victimization suffered relative to other population segments. Due to fixed income, inflation, lack of social interaction, and psychological insecurity among other factors, the elderly, it was asserted, are particularly susceptible to fraudulent practices. It was felt that a concerted marketing institution effort might ameliorate these factors and significantly benefit this disadvantaged group.

Louden (1976) also raised the issue of whether it was economically justifiable to ignore the aged market. Demographically, he argued, that not only did they constitute a major purchasing power, but that the elderly would increase this power in the future through rapid population growth (expected by the year 2000 to grow to 20% of population).

Most recently the banner has again been raised by business observers calling for a marshalling of effort to reach the elderly consumer (Business Week 1979). The findings indicate that reaching the "olderster" or "maturity" market (defined as 45+) could pay off (and has for a number of firms as illustrated by the examples reported). It was shown that: (1) the 45+ market accounts for 39.2 million households or 53.9% of all households, (2) the 45+ segment has an average income of \$13,388 (\$16,619 for the 45-64 year old and \$8,063 for the 65+ person), (3) the 45+ segment has 52.4% of all spendable income available, and (4) the 45+ segment accounts for 59.9% of all discretionary income.

To conclude, marketers and consumer behaviorists have begun to show interest in the aged market. However, even with the many calls for research that have been generated throughout the past 20 years, and in spite of the subsequent research efforts focusing on this group, it would appear that not enough is known about this population segment in terms of description, prediction, and explanation of behavioral phenomenon to fully develop conceptualized decision-making frameworks.

A Review of General Behavioral and Decision-Making Findings

Samli and Palubinskas (1972) conducted exploratory research in a west coast SMSA senior citizen center to explore two aspects of consumer behavior: consumption patterns and buying behavior activity. The research generated a number of initial findings among which were: (1) food accounts for a substantial proportion of the elderly's incomes, (2) the elderly are generally limited to weekend shopping periods because of restricted transportation sources, and competing activities such as hobbies, civic functions, and community activities, and (3) the elderly are heavy users of mass communication sources for consumer information.

Mason and Smith (1974) conducted a survey in an elderly public housing complex of a major SMSA. This was done in order to develop: (1) a shopping behavior profile of the low income senior citizen, (2) to determine the basic sources the elderly utilized in obtaining consumer information, (3) and to outline the travelling behavior of the elderly in making major purchases. Their findings suggested that the senior citizen: (1) relies slightly on in-home shopping services, (2) utilizes both newspapers and previous experiences as prime sources of information, and (3) shops as part of a life style, not just for particular purchases.

Reinecke (1975)--the author of the 1964 study which purported that a distinct elderly segment did not exist--examined retailing outlets in order to see the compatibility between the institutions and the needs of the senior citizen. He found that the buying needs and location of the elderly made the current array of supermarkets and shopping centers non-usable in meeting elderly consumption needs and thus, were not part of their system of buying.

Concurrently, using the younger consumer as a comparative group, Bernhardt and Kinnear (1975) conducted a random sample survey of senior citizens in a southeastern SMSA in order to develop a consumer profile of the senior consumer. This profile included such consumer phenomena as shopping behavior, credit card usage, media habits, and leisure time activities. Their findings indicated that the senior citizens were: (1) less likely to use discount stores and more likely to use traditional department stores than their younger counterparts; (2) much less likely to have credit cards than the younger group; (3) more likely to read newspapers than those under 35 but less likely than the group 35-64 years old; (4) much less likely to listen to FM radio rather than AM than the younger group; (5) much more likely to watch more daytime TV than younger groups; and (6) less likely to engage in reading, attend sporting events, go to movies, go out to eat, or attend religious services than those younger.

During that same period, Tongren (1975) conducted a study of the income and credit usage of the elderly by analyzing census data and using a random sample of elderly from the middle Atlantic states. He concluded that income of the elderly has doubled in the previous decade and that this trend will continue in the future due to: (1) higher investment income and returns on working assets than those of previous elderly generations and (2) continued expansion of private pensions in terms of dollar size and people covered. The first finding that the elderly possess more purchasing power than past elderly due to increased returns on assets and investments was confirmed in a study of retirement associations in Houston, where the following analysis resulted:

To look at the 65-and-over group as low income shoppers can be misleading in two ways. First, it means over-looking the possibility that a senior citizen with a \$500-per-month income may

be ready for a round-the-world trip. Assuming that income breaks down to \$180 for Social Security, \$320 from investments, the \$320 may be a 5 percent return on about \$77,000 (Gelb 1978, p. 45).

Tongren (1975) also found that 80% of the elderly used some form of consumer credit for periods less than 30 days. This finding appears contrary to Bernhardt and Kinnear (1975), until one examines Tongren's data carefully and observes that it implies that as age increases, the use of credit diminishes.

Martin (1976) conducted research on a convenience sample from southwest Missouri in order to examine behavior differences exhibited in varying age groups of fashion buyers. Using life cycle as his grouping classification system, he concluded from his analysis that the following major differences exist among generations: (1) the elderly emphasize alternative choices less than other generational groups, (2) the elderly have less formed predispositions prior to the shopping than other generational groupings, and (3) the elderly rely more heavily on newspaper and salespersons for marketing information than other generational groupings.

Towle and Martin (1976) began to question, as did Reinecke (1964), the assumption that the elderly age segment was homogeneous. However, unlike Reinecke (1964), their own assumption was that the elderly could be categorized into various segments, each different than younger age groupings. Using psychographic and buying study data (and demographic data which did not work) from a national panel, a cluster analysis was performed. They concluded that not only were the elderly different from their younger counterparts, but the following sub-segments were found:

- (1) Saver/Planner (buys unknown brands)
 - a. Psychographic description--frank, self-assured and confident
 - b. Possessed 25.1% of the elderly market
- (2) Brand Loyalist (Does not buy for approval of friends)
 - a. Psychographic description--brave and not stubborn
 - b. Possessed 8.4% of the elderly market
- (3) Information Seeker (persuasible)
 - a. Psychographic description--kind and sincere
 - b. Possessed 10.1% of the elderly market
- (4) Economy Shopper (not brand loyal).
 - a. Psychographic description--not brave, not demanding, not egotistical, and bland personality
 - b. Possessed 10.6% of the elderly market
- (5) Laggard (not persuasive)
 - a. Psychographic description--liberal, unreserved, and cold to others
 - b. Possessed 11.2% of the elderly market
- (6) Conspicuous Consumer (change brands and see approval from friends)
 - a. Psychographic description--dominant, egotistical, and stubborn
 - b. Possessed 34.6% of the elderly market

Mason and Beardon (1978) conducted research on a random sample of elderly in a southeastern SMSA to develop a profile of shopping behavior of elderly consumers. Their findings suggested: (1) that like other age groups, the elderly shop for a variety of reasons other than buying, e.g., exercise, leisure, and recreation, and (2) that they tend to use friends, personal experience, price comparisons, and mass media as information sources in making marketing decisions.

Finally, Deshpande and Krishnan (1979) using the coping skill conceptual frame of Botwinick (1973) examined how the elderly perceived their marketplace interactions. They investigated several decision rule strategies, via factor analysis, of a national panel of elderly consumers and then ran a discriminant analysis between consumers who claimed to have good versus bad marketplace experiences. They found that the order of prediction factors of decision rule strategies associated with bad shopping experiences (from most to least) was: flexible shopping rules, conservative shopping rules, brand conscious rules, sales pressure avoidance rules, and information avoidance rules. As they state:

The flexible shopping rules have to do with ignoring budgeted limits, going beyond shopping list items, and disregarding detailed information on contracts. This may open up elderly buyers to make unwise or hazardous purchases that they later regret. Contrast this with the Brand Conscious or Sales Pressure Avoidant Decision rules which imply more methodical and painstaking prepurchase activity. These rule sets seem intuitively to lead to better buying as evidence by less post purchase dissatisfaction. (p. 586)

In summary, the behavioral and decision-making findings generally show the elderly to have the following overall characteristics: (1) they tend to be somewhat limited in mobility; (2) they use large quantities of marketer generated information; (3) they engage in shopping for more than just buying reasons; (4) they have a similar pattern of leisure to younger age groups, just lower in quantity; (5) they have more sources of incomes than ever before because of pre-retirement investments; (6) they use long-term credit less as age progresses; (7) they are not just one homogeneous segment but many smaller segments, each differing from their younger counterparts; and (8) they depend on a wide array of shopping rules that result in varying amount of coping skill levels in dealing with marketing institutions.

The Aged Market: A Review of Information Processing Findings

Consumer information processing research has been the basis for developing several overall consumer buying models (e.g., Engel et al., 1978, Bettman 1979). The study of consumer information processing generally includes the following variables: nature of consumer's search activities, amount of information seeking by consumers, factors that influence the search process, type of information sought, sources of information, evaluating alternative factors influencing the amount of evaluation, and the results of the evaluation (Loudon and Della Bitta 1979, pp. 459-481).

Schiffman (1971; 1972; 1973) in studying the elderly consumers in a senior citizen community asked the following: (1) what was the impact that internal (research experiences) and external (marketing communications) sources of information had on new trial (product purchases) behavior, (2) what is the relationship between new trial and perceived risk, and (3) what is the relationship between product-related social interaction variables and general social interaction variables? His findings indicated that: (1) both internal and external sources of information affect the elderly's consumer decision-making, (2) the higher the perceived risk, the less likely the product trial, (3) there is a high significant relationship between new product related social interaction and general social interaction, and (4) socially active elderly are more innovative than social isolates.

Phillips and Sternthal (1977) provided the first major synthesis of related behavioral literature dealing with

age differences in information processing with the accent on the elderly consumer. Their focus was twofold: (1) to examine the differential sensitivity to processing information between the young and the old and (2) to determine when, in terms of age, these differentials are likely to occur. To address these issues, research on patterns of information exposure, learning, and influencibility were reviewed.

First, with regard to information exposure, Phillips and Sternthal (1977) in reviewing Frederick (1973), Graney and Graney (1974), and Schramm (1969) were able to conclude that the elderly have greater exposure and reliance on mass media for information relative to younger age groupings. Further, contrary to popular opinion, the elderly show similar patterns to their younger counterparts in information exposure and reliance on the extended family.

Second, with regard to learning, Phillips and Sternthal (1977) in reviewing Arenberg (1967), Botwinick (1973), Rabbitt (1965), Birren (1974), and Schonfield (1974) among others, concluded that age decrements can indeed exist in levels of learning, continuous attention, and information processing speed. These decrements tended to be especially apparent when the information presented is not self-paced and when distractions are present. When one goes outside the Phillips and Sternthal review and examines problem-solving contexts, there is confirmation of the existence of age related decrements especially with increases in problem-solving complexity (Young 1966, Arenberg 1968, Wetherneck 1966).

Finally, with regard to influencibility, Phillips and Sternthal (1977) in reviewing Whittaker and Meade (1967) and Klein (1972) among others, concluded that the empirical evidence was inconsistent as to differences in influencibility relative to younger age groupings. In most cases, influencibility was at least situation specific and could be present due to such factors as social isolation or self-perception of incompetence.

As a conclusion to their review, Phillips and Sternthal state that age related factors significantly affect the way individuals conduct information processing activities. However, they caution their readers that:

These differences do not necessarily occur at age 65; rather, they are related to the social, psychological and physical changes that accompany aging (p. 453).

Phillips and Sternthal's (1977) work on information processing provides a paradigm for explaining the consumer behavior of the aged by systematically relating a variety of studies from studies from various disciplines. Methodologically known as intersubjective certifiability (Popper 1968), Phillips and Sternthal's work is a crucial phase in developing a comprehensive consumer behavior theory for the elderly.

During the same year of the Phillips and Sternthal review, French and Crask (1977) conducted research on the relationship of persuasibility and media appeals using a random sample of elderly from a major SMSA. Using hypotheses developed from Phillips and Sternthal (1977) conceptualizations and Layton's (1975) documented findings of diminished sensory capabilities associated with age, French and Crask found the following results: (1) the elderly perceive no differences in source credibility of advertisements carried by different media, (2) advertising in general has a significant impact on senior citizen buying behavior, and (3) the young-old are more influenced by advertising than old-old--using the Neugarten (1974) elderly classification scheme.

Beardon and Mason (forthcoming) studied a random sample of elderly consumers from a southeastern SMSA and compared their findings to a national sample of grocery shoppers in order to examine "the familiarity with and use of grocery shopping aids and other consumer oriented information" provided by marketers. Additionally, product unavailability and complaint experiences were also examined. Their research was based on the Krosk and Srivastava (1977) findings that indicated there was lessened sensitivity to unfairness and less tendency to be assertive about complaining among elderly in comparison to other groups. Beardon and Mason found that: (1) differences between elderly and a national sample of representative grocery shoppers existed, in that the elderly use and are less familiar with consumer information made available in the grocery outlet and (2) that dissatisfaction with food was higher for the national sample than the elderly sample due to decreased sensitivity of the older shoppers to negative shopping occurrences.

A similar study was conducted by Lawther (1978) using national panel data, which examined social integration of the elderly and the relationship this variable had with unfairness awareness, complaint actions, and information usage. Previous research had shown that social integration affects brand preferences of consumers in general and that social integration affected the behavior of the elderly in particular (Stafford 1966, Larsen 1978). Utilizing the paradigm of social integration which is composed of social values, social roles and informed group membership, she explored the impact social integration had on elderly consumer behavior (Rosow 1974). Some of her findings indicated that: the greater the degree of social integration. The greater the awareness of unfair marketing practices (e.g., bait and switch tactics) and the greater the likelihood of complaining about a buying experience that was less than satisfying, the more likely is the use of marketer dominated sources of information, as opposed to consumer dominated sources.

In summary, the information processing research has been useful in developing an initial descriptive understanding of the elderly's consumer behavior. From the reviewed studies, the following relationships were found to be most significant: (1) social interaction/integration is highly correlated to consumer decision processes and the utilization of marketing related sources, (2) age differences do indeed exist between young and old in information processing abilities, and (3) age related insensitivity to marketing unfairness and less assertive behavior in the fact of such practices, appear to put the elderly at a comparative disadvantage relative to younger age groups.

Future Directions

Researchers in the past twenty years have made significant inroads toward an understanding of the elderly consumer. Information about behavioral patterns that heretofore went unrecognized have now begun to come together into an overall generally accepted picture of the aged segment. Advancement from this point on rests upon subjecting the empirically found relationships to causal analyses based on sound theoretical foundations. If the elderly are to completely be understood and behavioral explanations offered, researchers must pursue the root causes for the observed behavioral phenomenon. Once such research is conducted it will be legitimate to assert that not only are the elderly deserving of attention, but the foundation will have been laid for developing normative decision actions.

To provide specific direction for future research effort directed at the elderly consumer, the following issues must be addressed. One, the issue of definition is paramount for consistent and coordinated empirical work.

The labels of elderly, aged, and senior citizen all imply constructs that need precise definition for continuous research in future years. In essence what is required is a precise set of bounds delimiting the character of this population segment.

Two, a priori cause and effect relationships need to be conceptualized for elderly consumer behavior. A sound theoretical basis for all researched relationships must be established for empirically verifying root causes. For example, an important arena for future research is the area of whether or not age is the driving force or just an intervening variable spuriously related to behavior that masks cohort and/or period effects. Answers to cause and effect issues such as the age related effects could begin to provide enough answers to correctly develop necessary decision actions by public and private sectors on behalf of the elderly in meeting their specific needs.

Three, research effort should be directed at providing inputs into policy decisions geared toward meeting the needs of the elderly consumer. In a time of contracting resources and financial restraint decision makers may be forced into 'either/or' decisions at the expense of the elderly consumer. Its an issue such as this that makes it imperative for consumer researchers to generate findings that will have policy impact.

Four, there is the question of elderly consumer quality of life. In recent years policy makers, who have tried to satisfy the needs and enhance the well-being of the elderly consumer, have encountered difficulty in measuring the outcomes of their decisions. In order to overcome this difficult, a coherent conceptual framework needs to be developed that can describe, predict, and explain the system of variables, relationships, and processes that produce a given level of well-being, or quality of life. Such research could be extremely important in providing for the consumer needs of the elderly.

Finally, before research is begun, a caution should be noted. Though the preponderance of empirical evidence supports the contention that the elderly, though not totally homogenous, are significantly different from other population segments, the reader should be aware that this is not a universally held position. It has been suggested by some gerontologists that society is "moving in the direction of what might be called an age-irrelevant society; and it can be argued that age, like race or sex, is diminishing in importance as a regulator of behavior (Neugarten and Hagestad 1976, p. 52). Still, even in light of this cogent argument, it is these authors' conclusion that the differences indicated in the consumer research examined warrant the contention that the elderly comprise a recognizable if not totally homogeneous market segment. This is sufficient reason for continuing the effort by consumer researchers to seek an understanding of the behavioral phenomena surrounding this increasingly important population segment.

References

- Arenberg, D. (1968), "Concept Problem Solving in Young and Old Adults," Journal of Gerontology, 23, 279-282.
- _____ (1967), "Regression Analysis of Verbal Learning on Adult Age at Two Anticipation Intervals," Journal of Gerontology, 11, 42-27.
- Beardon, W. O., and Mason, J. B. (Forthcoming), "Elderly Use of In-Store Information Sources in Supermarket Purchase Decisions and Dimensions of Product Satisfaction/Dissatisfaction," Journal of Retailing.

- Bernhardt, K. L., and Kinnear, T. C. (1975), "Profiling the Senior Citizen Market," in Advances in Consumer Research-Vol. 3, ed. B. B. Anderson, Cincinnati: Association for Consumer Research. pp. 449-452.
- Bettman, J. R. (1979), An Information Processing Theory of Consumer Choice, Reading, MA: Addison-Wesley Publishing Company.
- Birren, J. (1974), "Translations in Gerontology from Lab to Life: Psychophysiology and Speed of Response," American Psychologist, 29, 808-815.
- Block, J. E. (1974), "The Aged Consumer and the Market Place: A Critical Review," Marquette Business Review, 18, 73-81.
- Botwinick, J. (1973), Aging and Behavior, New York: Springer Publishing Company.
- Business Week (1960), "How the Old Age Market Looks," February 13, pp. 37-38.
- _____ (1971), "The Power of the Aging in the Marketplace," November 20, pp. 52-58.
- _____ (1979), "Media and Advertising: The Over-50 Set," November 19, pp. 194-195.
- Deshpandé, R., and Krishnan, S. (1979), "The Use of Buying Decision Rules by the Elderly: Public Policy Implications of Perceived Post Purchase Dissatisfaction," in AMA Educators' Conference Proceedings, eds. N. Beckwith et al., Chicago: American Marketing Association. pp. 583-588.
- Engel, J. F., Blackwell, R. D., and Kollat, D. T. (1978), Consumer Behavior, (3rd ed.) Hinsdale, IL: The Dryden Press.
- Forbes (1969), "The Forgotten Generation," January 15, pp. 22-29.
- Frederick, K. (1973), "The Aged Consumer: Socialization and Integration of the Aged into the Servicing Social Structures," unpublished Ph.D. dissertation, University of Pennsylvania.
- French, W. A., and Crask, M. R. (1977), "The Credibility of Media Advertising for the Elderly," in AMA Educators' Conference Proceedings, eds. B. A. Greenberg and D. N. Bellinger, Chicago: American Marketing Association, pp. 74-77.
- Gelb, B. D. (1978), "Exploring the Gray Market Segment," MSU Business Topics, 26, 41-46.
- Gidlow, E. (1961), "The Senior Market," Sales Management, October, pp. 35-39, 108-111.
- Goeldner, C. R., and Munn, H. L. (1964), "The Significance of the Retirement Market," Journal of Retailing, 40, 43-52, 60.
- Goldstein, S. (1965), "Changing Income and Consumption Patterns of the Aged, 1950-1960," Journal of Gerontology, 20, 455-461.
- _____ (1968), "The Aged Segment of the Market, 1950 and 1960," Journal of Marketing, 32, 62-68.
- Graney, M., and Graney, E. (1974), "Communication Activity Substitutions in Old Age," Journal of Communications, 24, 88-96.
- Klein, R. (1972), "Age, Sex, and Task Difficulty as Predictors of Social Conformity," Journal of Gerontology, 27, 229-236.
- Krosk, N., and Srivastava, R. K. (1977), "Preliminary Examination of the Sources and Handling of Consumer Problems by the Elderly," unpublished manuscript, Elderly Consumer Project, University of Pittsburgh.
- Larsen, R. (1978), "Thirty Years of Research on the Subjective Well Being of Older Americans," Journal of Gerontology, 33, 109-125.
- Lawther, K. (1978), "Social Integration and the Elderly Consumer: Unfairness Awareness, Complaint Actions, and Information Usage," AMA Educators Conference Proceedings, ed. S. C. Jain, Chicago: American Marketing Association, pp. 341-345.
- Layton, B. (1975), "Perceptual Noise and Aging," Psychological Bulletin, 82, 875-884.
- Loudon, D. L. (1976), "Senior Citizens: An Underdeveloped Market Segment," Proceedings of the Southern Marketing Association, pp. 124-126.
- _____, and Della Bitta, A. J. (1979), Consumer Behavior: Concepts and Applications, New York: McGraw-Hill Book Company.
- Martin, C. (1976), "A Transgenerational Comparison: The Elderly Fashion Consumer," in The Elderly Consumer, ed. F. E. Waddell, Columbia, MD: The Human Ecology Center, Antioch College, pp. 251-259.
- Mason, J. B., and Beardon, W. O. (1978), "Profiling the Shopping Behavior of Elderly Consumers," The Gerontologist, 18, 454-461.
- _____, and Smith, B. E. (1974), "An Exploratory Note on the Shopping Behavior of the Low Income Senior Citizen," Journal of Consumer Affairs, 8, 204-210.
- Media Decisions (1973), "Don't Write Off the Senior Citizen Market," 8, 64-67, 112-116.
- Morse, L. (1964), "Old Folks: An Overlooked Market?" Duns Review and Modern Industry, 83, 45-46, 86-88.
- Nation's Business (1971), "The Over-65 Set: A Bonanza for Business?," 59, 34-36.
- Neugarten, B. L. (1974), "Age Groups in American Society and the Rise of the Young Old," Annals of American Academy of Political and Social Sciences, pp. 187-198.
- _____, and Hagestad, G. O. (1978), "Age and The Life Course," in Handbook of Aging and the Social Sciences, eds. R. Binstock and E. Shanas, New York: Van Nostrand Reinhold Company, pp. 35-55.
- Phillips, L. W., and Sternthal, B. (1977), "Age Differences in Information Processing: A Perspective on the Aged Consumer," Journal of Marketing Research, 14, 444-457.
- Popper, K. R. (1968), The Logic of Scientific Discover, New York: Harper Torch Books.
- Rabbitt, P. (1965), "An Age Decrement in the Ability to Ignore Irrelevant Information," Journal of Gerontology, 20, 233-238.
- Reinecke, J. (1964), "The 'Older' Market - Fact or Fiction?" Journal of Marketing, 28, 60-64.
- _____ (1975), "Supermarkets, Shopping Centers and the Senior Citizen Shopper," Marquette Business Review, 14, 105-107.

Rosow, I. (1974), Socialization to Old Age, Berkeley, CA: University of California Press.

Samli, A. C. (1967), "The Elusive Senior Citizen Market," Dimensions, (November), pp. 7-16.

_____, and Palubinskas, F. (1972), "Some Lesser Known Aspects of the Senior Citizen Market - A California Study," Akron Business Review, 3, 47-55.

Schiffman, L. G. (1971), "Sources of Information for the Elderly," Journal of Advertising Research, 11, 33-37.

_____, (1972), "Perceived Risk in New Product Trial by Elderly Consumers," Journal of Marketing Research, 9, 106-108.

_____, (1973), "Social Interaction Patterns of the Elderly Consumer," in AMA Combined Conference Proceedings, eds. B. W. Becker and H. Becker, Chicago: American Marketing Association, pp. 445-451.

Schonfield, D. (1974), "Translations in Gerontology - From Lab to Life: Utilizing Information," American Psychologist, 29, 796-807.

Schramm, W. (1969), "Aging and Mass Communication," in Aging and Society, eds. M. Riley and J. Riley, NY: Russell Sage Foundation, pp. 352-375.

Stafford, J. E. (1966), "Effects of Group Influence on Consumer Brand Preferences," Journal of Marketing Research, 3, 68-75.

Tongren, H. N. (1975), "Consumer Credit and Over-65 Age Group," Journal of Consumer Credit Management, (Spring), pp. 117-123.

Towle, J. G. and Martin, C. R. (1976), "The Elderly Consumer: One Segment of Many?" in The Elderly Consumer, ed. F. E. Waddell, Columbia, MD: The Human Ecology Center, Antioch College, pp. 232-242.

U. S. Department of Commerce, Bureau of the Census (1973), Current Population Reports, Consumer Income, Money Income of 1972 of Families and Persons in the U. S., P-60, #87, (June), Washington, D. C. Government Printing Office.

Waddell, F. E. (1975), "Consumer Research and Programs for the Elderly - The Forgotten Dimension," Journal of Consumer Affairs, 9, 164-175.

Wethernick, N. (1966), "The Inferential Base of Concept Attainment," British Journal of Psychology, 57, 61-69.

Whittaker, J., and Meade, R. (1967), "Sex and Age Variables in Persuasibility," Journal of Social Psychology, 73, 47-52.

Young, M. (1966), "Problem-Solving Performance in Two Age Groups," Journal of Gerontology, 21, 505-509.

DIFFERENT INFLUENCES ON CONSUMER DECISION MAKING

E. H. Bonfield, Temple University

Abstract

Three papers dealing with diverse aspects of consumer behavior are discussed. While a few methodological problems are apparent, the more global concerns of lack of control/comparison groups, uncontrolled anchoring, and labeling in consumer research are addressed.

Introduction

The three papers presented for this discussion are diverse in subject matter. Each makes a contribution, and each leads to further questions for research. Two of the papers, those by Roberts, Wortzel, and Berkeley and by Burns and DeVere deal directly with family decision making. They contain primary research data. Meadow, Cosmas, and Plotkin present a literature review suggesting consumers described as elderly are different from other consumers.

If all three papers dealt with family consumer decision making, they would fit into the general model of intra-family influence Burns and DeVere briefly describe at the beginning of their paper. Burns and DeVere have explored two situational dimensions, while Roberts, Wortzel, and Berkely could be categorized as exploring a mothers-children relational dimension. While Meadow, Cosmas, and Plotkin correctly do not place their research in the family decision process domain, the individuals in the group they study are frequently actors in family decision making.

The discussion of each paper which follows contains little concern for the shortcomings found with the specific research projects. The exploratory nature of all three studies leads instead to a concern for suggestions for improving future research based on the present experiences.

Mothers' Perceptions

The Roberts, *et al* research is a positive case of "massaging data." The data from a current, large-sample, commercial study was apparently made available to the authors for additional analysis. This is enlightened behavior on the part of Needham, Harper, and Steers, and is to be commended.

Additional analysis of data gathered for other purposes is rarely as suited for testing additional hypotheses as studies designed for that specific purpose. That situation exists for the present research. Therefore, as the authors point out, their analysis is exploratory, and any findings must be taken as tentative.

"Attitudes"

Because of the exploratory nature of the study, the sets of scales labeled "attitude" are useful as indicators of such constructs as "concern about nutrition" and "cosmopolitaness." It should be clear, however, there is no basis, other than anecdotal, that these scales are valid measures of any attitude. Furthermore, the reliability measures of the "attitudinal" constructs suggest consistency in the way people mark scales expected to measure the construct. If there is only face validity of a construct, however, the analyst must question the reasons for these high interscale correlations. For example, are "concern about sex" (typical among ACR members) and "concern about violence" the same thing? Do they co-vary because of the admitted "middle America" bias of the sample, because all

respondents were women, because they were women with children,...? Is it possible the scales are marked the same, but the depth of concern--the involvement--is different? Some of these alternative hypotheses might be explored within the context of the present data set, while others would have to be tested with new research.

Labeling

After using factor analysis on the scales measuring perception of children's influence on brand choice, the author's labeled each factor despite the fact that only one, "gum," readily lends itself to such a label. Labeling can be a dangerous task in factor analysis, as well as in other areas, because it can be misleading. Are cheese and pasta children's foods? Are cold cereal and restaurants sweets or snacks?

A further problem stems from the sample used in factor analysis. If the authors used listwise deletion of cases as an option in factor analysis, only those respondents indicating use of all 35 products by marking the perceived influence scales could be included in the factor analysis. Considering pets alone, it is questionable whether a large proportion of the households owned, and therefore probably bought, both dog and cat food. The scales used forced respondents who did not use a product into a "least frequency" category. While there is legitimacy in utilizing this type of scale, inclusion of nonusers with users clouds interpretation.

Controlling for Children's Ages

Controlling for children's ages is difficult as the authors amply demonstrate. The problem is easily handled in one-child (and zero-child) households. Where more than one child is present, for any one product category, the mother might anchor on one child or somehow cognitively average her perception of the influence of her children on brand choice. An answer for this problem would be to instruct the mother to anchor on a specific child whenever more than one child is present. A randomization process for child selection could easily be developed to assure representativeness of the total sample, as well as adequate sample size for each major category of family composition.

Situation Dimension Effects

Burns and DeVere have developed a "general model of intra-family influence" which was used to guide their selection of the study presented here. They utilized written scenarios as stimuli differentiating four treatment levels. While it can be argued the written scenarios would not be realistic to the couples in each experimental group, their approach is a viable compromise to more expensive and more difficult to physically control alternatives. A useful alternative the researchers might have considered as a manipulation check would have been to add a control situation. This addition would have added some 40 additional couples to the total sample, but, given the sampling frame and data collection methods, the increased time and dollar cost would have been small. No scenario would have been provided the control group.

There is no quarrel with the data analysis as given. The use of figures, particularly when interaction effects are

involved, would make the results both easier for the researchers to interpret and for readers to follow.

Overstated Agreement

The use of means for husbands and wives are likely to increase the apparent level of agreement on the dependent variables. Since data on both spouses in the household is available, it is possible to compare responses on a household-by-household basis. Davis and Rigaux (1974) and Bonfield (1978) have found considerably greater disagreement when spouses' responses are compared on nominal scales similar to the disagreement outcome variable used by Burns and DeVere. The constant sum and discussion time estimates would show even greater disagreement unless some reasonable nominal scale were superimposed. A "scientific" approach might be to utilize standard deviation estimates to create nominal boundaries. An *ad hoc* approach of trying 10 to 20 percent differences in constant sum estimates and five-minute differences in discussion times might be even more useful.

Uncontrolled Anchoring

A problem which appears common to most, if not all, family decision process research can be labelled, "uncontrolled anchoring." That is, there is a lack of specificity as to what stimulus arrays respondents are reacting. For example, Burns and DeVere nicely imply a temporal anchor in their four scenarios. The respondents probably predicted what they thought would happen. In other research, however, it is impossible to determine whether respondents are predicting or reporting perceptions of what has happened in the past. If the latter is the anchor desired by the researcher, respondents should be asked to think about their last purchase occasion for the product.

Controlled anchoring means specifically describing products. Life insurance can be purchased on husbands' lives, wives' lives, children's lives, and in the case of mortgage insurance, jointly on both husbands' and wives' lives. A decision to buy a set of pots and pans is different from buying a replacement pot or pan or an additional pot or pan. Encyclopedias might be bought for household use or for children's use.

Finally, decisions should be more specifically described. While need verification is a relatively clear construct among family decision making researchers, "the need to buy one" may imply need recognition to respondents. "What model or type" may well mean different things to different respondents. Controlled anchoring might lead to the specification, for example, of "whether to buy term or whole life insurance," and "what size pot or pan," on future questionnaires. The model/type question might simply be deleted as inappropriate for encyclopedias.

An important aspect of having a model to guide research is the provision of a framework for exploring questions of differential anchoring. The questions addressed above suggest further exploration in the situational block of the Burns and DeVere model.

"What Do You Mean, 'I'm Elderly'?"

Meadow, Cosmas, and Plotkin provide a good literature review on an age category segment of the United States population. Their review is a useful addition to the consumer research literature in that it brings to this group a set of annotated sources in a single place. They add evaluation which cannot be found elsewhere. Two observations about the research and literature reviewed seem appropriate.

Labeling People "Elderly"

A label is a convenient, short-hand code used to indicate a group. When researchers label, the intent is to be able to generalize about whatever has been labeled. An important problem associated with labeling is the potential for labeling to lead to stereotyping. Stereotyping is the process of ascribing an attribute possessed by some members of a group to all members of that group.

Clearly, those people who have been labeled "elderly" are chronologically older than other age groupings of people. Some, even many or most, of the people who have been labeled "elderly" use dentures, are retired, and use less long-term credit. Labeling people "elderly" in the stereotyping sense, may lead to restricted granting of long-term credit--which, in turn, would make the group's limited use of long-term credit a self-fulfilling prophecy.

Some labeling of sub-segments among older age groups may result in mis-labeling. For example, Towle and Martin (1976) labeled a single psychographic sub-segment "Information Seeker" and "persuasive" apparently because the psychographic description of them was "kind" and "sincere." Whether one goes backward or forward in the logical process, no one label follows from any of the others.

Perhaps the most negative result of labeling is that some members of the group, in accepting the label, end up with lowered self images and reduced self esteem. For example, some individuals may decide they cannot exercise as much or continue in a job they enjoy and are capable of handling well simply because they have had a birthday.

Control/Comparison Groups

Consumer research, in general, has been characterized by nonuse of control groups. This generalization characterizes the research of those who have studied those in our population who are over 40, 45, or 65 depending on the study. Unless there is an appropriate comparison group, any characterization of a specific group as different is unsupported.

Conclusion

It was not the purpose of this paper to evaluate three exploratory studies. The discussion of each paper has led to general statements about problems with uncontrolled anchoring and labeling in consumer research, as well as a need for control or comparison groups in this research. The specific discussions suggested ways of combatting these problems.

KNOWLEDGE STRUCTURES, PRODUCTION SYSTEMS AND DECISION STRATEGIES

Merrie Brucks, Carnegie-Mellon University
Andrew Mitchell, Carnegie-Mellon University

Abstract

We suggest that in examining consumer decision making, more effort should be directed at understanding why consumers use a particular processing strategy instead of simply describing the resulting process. In addition, we suggest that more effort should be directed at understanding consumer decision making in "real world" situations. A conceptual model is presented which identifies the critical elements that affect the selection of a particular decision process. These elements include the task environment, problem perception, general goals and values, knowledge structures and production systems. Finally, knowledge structures and production systems are discussed in greater detail.

Introduction

Over the last five or six years, considerable research has been directed at understanding the dynamics of consumer decision making. Most of this research has used a process tracing methodology, a brand/attribute information format and a choice task (e.g., Bettman and Zins 1977, Payne 1976, Green, Mitchell and Staelin 1978). Under this research paradigm, subjects are asked to select a particular brand from a set of brands that differ along a prescribed number of attributes where this information is generally presented in either a verbal (e.g., "very good on decay prevention") or a numerical form (e.g., "gets 35 miles per gallon"). To record the order in which subjects examine the information, protocols (e.g., Payne 1976), eye movements (e.g., Russo and Rosen 1976), or information monitoring techniques (e.g., Bettman and Kakkar 1977) are generally used.

This research began, in retrospect somewhat naively, with the belief that consumers use one of the many hypothesized information integration rules (e.g., linear compensatory) when making choices between brands and that the use of a process tracing methodology would reveal which decision rule an individual used. In addition, it was generally assumed that the use of a particular decision rule would be relatively stable either within subjects over choice tasks or between subjects within choice tasks. It was soon realized, however, that a process tracing methodology reveals information search strategies -- not information integration processes. Although a relationship probably exists between search strategies and information integration processes, the relationship will not be perfect since subjects are able to store some information about the brands in long term memory.

Even though information integration processes cannot be directly inferred from search strategies, the results of the studies in this area seem to indicate that consumers do not use one of the hypothesized information integration rules (Mitchell 1978, Bettman and Park 1980). Although processing patterns can be identified (e.g., attribute vs. brand processing), these patterns tend to change frequently. These changing patterns have led Bettman to suggest that decision making under these conditions is frequently a constructive or a bottom up process (Bettman and Zins 1978). This means that the subjects are changing their processing strategy as they gather data during the task. The resulting processing strategies are, therefore, to a certain extent, data driven.

This research also indicates that consumers are very flexible in how they process information. Seemingly subtle changes in the information or its format can result in very different processing patterns (e.g., Bettman and Kakkar 1977). Finally, given a particular information set, there is considerable heterogeneity in how individuals process the information (e.g., Bettman and Jacoby 1976). In other words, each consumer seems to have his or her own idiosyncratic way of processing the information.

These results raise a number of interesting and inter-related problems concerning future research directions in this area. For instance, if individuals do not use one of the information integration rules, how do we define their processing strategies? If each individual has his or her own idiosyncratic way of processing information, how can we find generalizable principles that explain their behavior?

In this paper, we would like to present a preliminary discussion of these problems and a conceptual model that may prove useful in resolving them. The discussion and model represent our current thinking on these problems and we fully expect both to change as our research in the area evolves.

Research Goals and Problems

Although there are a number of possible goals for research on consumer decision making, we would like to suggest that the general goal for this research should be the prediction of how a given individual will make a decision in a particular situation and an understanding of why he or she used a particular processing strategy. We would also like to suggest that the general domain of phenomena studied in this area should be expanded beyond the information board paradigm. Since the task used in this paradigm appears to represent only a small percentage of the actual decision making tasks confronted by consumers, we believe that more effort should be directed at studying decision making in more "real world" environments.

In general, we believe that two important shifts need to occur in order to accomplish these goals. First, more emphasis needs to be directed at identifying the elements of the task and the individual that cause the between subject and task differences that we observe. This will require the development of a typology of task characteristics and procedures for measuring the individual level variables that affect the selection of a particular decision strategy. Second, more emphasis needs to be directed at understanding the cognitive activities (e.g., planning) that surround consumer decision making. In the past, we have been primarily concerned with describing how individuals processed or searched for information instead of investigating why they used a particular processing strategy. Both of these shifts require the integration of theories from cognitive psychology into conceptual models that provide an understanding of consumer decision making. We believe these shifts in emphasis will result in a better understanding of consumer decision making and may result in better models for predicting choice.

There are a number of problems, however, that need to be resolved before these goals can be accomplished. The first concerns the level of analysis. Process tracing procedures

produce a series of dense observations preceding final choice. Bettman and Park (1980), for instance, have decomposed protocols taken in a choice task using the information board paradigm into a typology of 70 different phrases and found that their subjects used an average of 35 phrases in making a choice. The fundamental problem, then, is how to develop measures of decision strategies that are causally related to elements of the task and the individual. One type of measure might be obtained by aggregating the processes measures. Alternatively, measures of general problem solving strategies such as "mean-ends analysis" and "operator subgoaling" (Newell, 1980) or "divide and conquer" and "define and successively refine" (Aho, Hopcraft and Ullman 1974) might be used.

In the same vein, there are many different conceptual dimensions of the variables that are thought to influence decision strategies. A number of studies, for instance, indicate that knowledge about a particular domain will affect how individuals process information (e.g., Edell and Mitchell 1978, Marks and Olson 1981) and their search strategies (e.g., Bettman and Park 1980, Johnson and Russo 1981). There are, however, many possible conceptual dimensions that one might use to identify differences in a domain of knowledge between two individuals. In addition, there are many possible conceptual dimensions that might be used to differentiate between two different tasks. In summary, the first problem involves the question of how to develop a typology of decision strategies and obtain measures of the task and individual so that causal relationships between the two may be identified.

The second, and related problem, concerns the indication that many decision processes are constructive or bottom up processes. How can we hope to predict how an individual will make a decision when he or she doesn't know how the decision will be made when he or she begins the task? This problem, of course, is related to the first. If, for instance, we use the proper level of aggregation for the dependent variable, we may be able to predict which decision strategy an individual will use in a given task based on specific elements of the task and the individual. This, of course, would be essentially a static analysis of a dynamic process. Alternatively, as discussed later in this paper, we may be able to develop procedures for understanding and predicting a constructive process.

This suggests that there are at least two different approaches for achieving the goals outlined earlier. The first involves the development of a typology of decision strategies which describes the process the individual used in making a decision. The second involves the prediction of the evolving dynamic process when it is constructive. At this point its difficult to assess which approach will prove more useful. Consequently, it would appear that fruitful research may proceed at both levels.

Finally, we would like to note that the consumer choice situation is essentially an ill-structured problem. Simon (1978) suggests that these problems differ from well-structured problems in three important ways. First, the criterion for determining if the goal has been achieved is poorly defined. Second, the problem instructions do not contain all the information required to solve the problem. Finally, the set of possible alternative moves at each stage of problem solving is not well defined. These differences, of course, increase the difficulty of understanding consumer decision making. Since there is no well defined goal to the problem, for instance, we need to understand how consumers decide when they have enough information about the alternatives to make a choice.

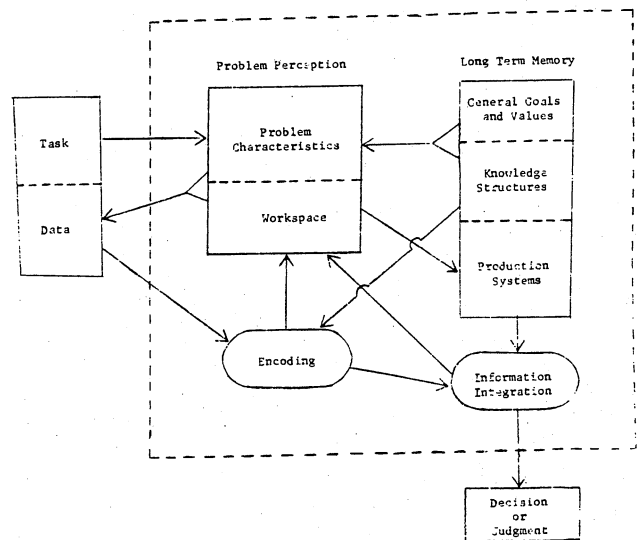
In the next section we present a simple conceptual model which we believe will be useful for studying decision strategies using either a static or a dynamic approach.

This model is heavily influenced by previous research examining problem solving in well-structured domains (e.g., Greeno 1978, Simon 1978). The critical individual level variables of this model are briefly discussed and then examples of how the model may be used to understand decision processes in both information board situations and more "real world" situations are outlined.

Conceptual Model

A conceptual model of the elements affecting the decision making process is presented in Figure 1. This model is a relatively simple one since little research has been directed at understanding the cognitive processes involved in decision making in ill-structured task environments (Simon 1978). We expect, however, that as research progresses in this area we will have a better understanding of these elements and cognitive processes.

FIGURE 1
CONCEPTUAL MODEL



Within this model, there are three different sets of critical elements that affect the decision making process. The elements of the first set exist in long term memory, are unique to each individual and affect all decision making processes. These are structural variables which include general goals or values, knowledge structures and production systems. The second set define the particular task and third are process variables which evolve during the decision process. These include, problem perception, information integration, and if the task involves the acquisition of external information, encoding processes. These specific elements will now be discussed in greater detail and then, two examples will be used to illustrate the model.

Elements of the Model

General Values and Goals. An individual's general values and goals are behavioral end states that the individual wants to achieve. These may include states such as professional success, wealth, and attractive physical appearance. These general values or goals will affect an individual's general approach to a purchase decision and the attributes that are used in evaluating the alternative brands. For instance, in purchasing clothes, an

individual who is concerned with having an attractive physical appearance will probably be concerned with the style of the clothes while someone who does not highly value this general goal may be primarily concerned with price.

Knowledge Structures and Production Systems. The remaining two elements of the first set are knowledge structures and production systems. These two elements represent the distinction that is frequently made between declarative knowledge and procedural knowledge (e.g., Anderson 1980). Declarative knowledge is our knowledge about concepts, objects or events. It is generally believed that this information is organized into packets of information or schemata (e.g., Rumelhart and Ortony 1977, Schank 1980). We may, for instance, have an organized set of information about fuel efficient automobiles, in general, and the Volkswagon Rabbit, in particular. We may also have stored organized information about events, such as visiting an automobile dealer to look for a new car, which are called scripts (Schank and Abelson 1977).

Production systems are representations of cognitive skills such as typing letters or solving geometry problems. The basic element of a production system is a condition-action statement. This statement allows the system to take a particular action when a particular condition is recognized. The particular condition may be an external stimulus such as a green traffic light or an internal state such as the lack of knowledge about a particular automobile. Similarly, the action may be a particular behavior such as stepping on the gas pedal of an automobile or a mental act such as adding two numbers together.

Problem Perception. Problem perception is the individual's internal representation of the problem. Considerable research on problem solving indicates that this internal representation will affect the strategy that an individual uses to solve the problem (e.g., Newell and Simon 1972, Greeno 1978, Einhorn and Hogarth 1981). Tversky and Kahneman (1981) and Simon and Hayes (1976) have shown that individuals may use different strategies for solving the same problem when the problem is presented differently (i.e., problem isomorphism).

A number of researchers have used different terms for this general notion. These include problem space (Newell and Simon 1972), decision frame (Tversky and Kahneman 1981), and problem frame (Wright and Rip 1980). We have selected a different term because we believe that our conceptualization differs from the definitions applied to these other terms. We currently view problem perception as containing two factors. The first are the general characteristics of the problem. These characteristics may include, for instance, the amount of time before a decision is required, the specific goals and subgoals for the problem, and the possible consequences of making a poor decision. The second factor, which we call the workspace, is the current state of the system with respect to the decision process. The current state of the system might contain, for instance, the number of automobile dealers visited and the amount of knowledge about the various alternative brands stored in long term memory.

This conceptualization differs from the definition of a decision frame as "the decision maker's conception of the acts, outcomes and contingencies associated with a particular choice" (Tversky and Kahneman 1981), or problem frames which Wright and Rip (1980) view as the information integration rules used to integrate information in making a decision. We find the former to be somewhat limiting in terms of the problems we wish to consider and we view the latter as a production system. Our notion of problem perception is probably closest to the concept of a problem space which was developed from the study of well-structured problems (Newell and Simon

1972). This concept contains two basic factors (a) a conceptual understanding of the problem consisting of the initial state, the goal state and the path constraints and (b) the different possible states of the space and the operators (Newell 1980). However, since our model is developed for ill-structured as opposed to well-structured problems, elements such as path constraints and operators may not be critical elements of problem perception. Finally, it should be noted that the cognitive limitations of humans imply that the cognitive elements of problem perception must necessarily be few in number (five to seven), unless some of this information is transferred to long term memory.

Encoding and Information Integration Processes. Encoding processes represent the transformation of external data into an internal symbolic code. These processes are influenced by knowledge structures and include both inferences and evaluative processing (i.e., counterarguing). For instance, an individual may infer that a particular automobile gets good gas mileage even though this information was not available to that individual. The information integration process involves the integration of encoded information from either external sources or from memory to form an evaluation or to make a choice.

Decision Making Examples

We will now discuss the conceptual model with respect to two different decision tasks. The first involves the standard information board task, while the second involves a more "real world" task. This latter example uses the task and model proposed by Hayes-Roth and Hayes-Roth (1979).

In the standard information board task, the subject is presented with an array of brands that vary along a number of different attributes. The information presented to the subject in the task (e.g., the product class and the brand names) and the relevant knowledge the individual has stored in long term memory (e.g., knowledge about the product category) result in a cognitive representation of the problem (i.e., problem perception). This cognitive representation for evaluating the alternative products may include the critical attributes that need to be considered. We have found, for instance, differences in how individuals process nutritional information when this information is associated with different foods (Brucks, Mitchell and Staelin 1980). This cognitive representation of the problem then triggers a particular set of productions that are used to start processing the information. In other words, the perception of the problem represents the condition portion of a production and the actual processing of information (e.g., encoding and information integration) represents the action portion. These productions may be at the level of the typology of phrases developed by Bettman and Park (1980) or they may be combinations of these phrases.

After a particular production has been executed the workspace is updated. This updated workspace, along with the subject's original understanding of the problem, then triggers a new production which then processes more of the information. These cycles continue until a choice is made.

What is critical, then, within this framework is understanding how an individual cognitively represents the problem and how the individual updates this representation at different points in time during the decision making process.

In more "real world" situations, the system becomes somewhat more complex; however, the structure remains the same. As mentioned previously, we use the task and the model proposed by Hayes-Roth and Hayes-Roth (1979) to

demonstrate this. The task involved having subjects develop a plan for completing a number of errands in an afternoon. The subjects were given the errands (e.g., meet a friend for lunch, pick up medicine for dog at the vet) and a map of a city with all the critical locations marked (e.g., veterinarian). The problem was designed so that it was very difficult to complete all the errands in a single afternoon.

The model developed for understanding the cognitive processes that the individual uses in developing the plan contains two basic elements: a number of different "specialists" and a "blackboard". The "specialists" are essentially production systems and their decisions are recorded on the "blackboard". The "blackboard" performs the function of recording the current state of the system with regard to the planning process. It is, therefore, similar to our notion of the workspace in problem perception. These different states trigger "specialists" which make decisions with regard to the planning process which, in turn, change the state of the system. For instance, knowledge that the individual is at a certain location in the town at a particular point in the planning process may cause a particular "specialist" to then select the errand that is closest to the current location. The process continues until a plan is formulated.

This, of course, is a simplified description of the model. In the actual model, the "blackboard" consists of five different planes called the plan, plan-abstraction, knowledge-base, executive and meta-plan. These five planes also contained different levels. An artificial intelligence program was developed based on this model, which provided a good representation of different subjects' planning processes.

This description is provided to indicate how our model might be used to understand actual "real world" decision processes. An obvious extension would be to understand how the formulated plan would change if it were actually implemented. An individual, for instance, might use more time completing a particular errand than expected (e.g., long time waiting for the vet) and then would have to revise the plan. The model, however, can also be used to describe the cognitive activity involved in revising the plan. With some adjustments this model might also be used to examine the strategy or plan that an individual uses in purchasing an automobile or a refrigerator. We are currently examining how the Hayes-Roth and Hayes-Roth model might be revised to study this task.

In summary, we are suggesting that in studying consumer decision making, more effort should be directed at understanding the cognitive activities surrounding the decision process. In other words, we should be more concerned with understanding why a consumer used a particular strategy instead of simply recording the actual process. We have presented a relatively simple conceptual model that might be used to provide a better understanding of these processes. In this model, the critical elements are problem perception, generalized goals or values, knowledge structures and production systems. In the next two sections we examine the latter two elements in greater depth.

Knowledge Structures

Knowledge structures contain our generalized knowledge about the external environment. Models of this generalized knowledge are generally conceptualized as network models, where the nodes of the network represent concepts and the links between nodes represent the relationship between concepts (e.g., Anderson 1976). As discussed in the previous section, it is currently believed that human knowledge is organized into packets of information called

schemata (Rumelhart and Ortony 1977). Individuals, for instance, may have a schema for fuel efficient automobiles that contains their generalized knowledge about these objects. Consequently, when we are presented with information about a fuel efficient automobile we activate this schema to process, encode, evaluate, and interpret the information. Ortony (1980) has recently suggested a dual layered theory of memory. The first layer contains our organized knowledge about specific concepts (e.g., schemata), while the second layer contains the relationships between schemata. Both of these layers are conceptualized as networks.

It is generally believed that our generalized knowledge will affect how we make decisions (e.g., Mitchell 1978, Olson 1978, Bettman 1979). Only recently, however, have researchers begun to actually examine how knowledge affects decision making (Bettman and Park 1980, Johnson and Russo 1981). Most of these studies, however, use only a single measure of an individual's knowledge of a particular domain. Frequently, this measure is based on the subjects' previous purchase behavior with respect to the product. Bettman and Park, 1980, for instance, use a subject's self reported experience with the product class as a measure of knowledge. Given the complexity of possible knowledge within a domain, multiple measures of knowledge should provide a better understanding of any related cognitive processes than a single measure. Objective measures of knowledge structures would also appear to be better than measures of previous behavior with respect to the product or self report measures. A teenager, for instance, who has never owned an automobile, may be more knowledgeable about different aspects of automobiles than an adult who has owned three automobiles.

Since there are many different possible conceptual dimensions of knowledge structures, we must identify which dimensions will provide the best explanatory power to explain the phenomena of interest. Kanwar, Olson and Sims (1981) have moved in this direction recently by proposing three different measures of cognitive structure -- dimensionality, abstraction and articulation. These measures are obtained through the use of free elicitation (Olson and Muderrisoglu 1979) and a modification of the repertory grid (Kelly 1955). The first measure, dimensionality, is the number of concepts activated by elicitation procedures within a particular content domain. The second, abstraction, refers to the level of abstractiveness that the related concepts within a domain are encoded. The third measure, articulation, is the number of levels that the individual uses to discriminate between objects within the domain.

The measures proposed by Kanwar, Olson and Sims (1981) are structural measures of an individual's knowledge within a particular domain. They essentially measure how information is organized within a domain. Alternative measures might be based on the content of knowledge within a particular domain. Although structural measures and content measures may be related, there may be important differences between these two types of measures. Kanwar, Olson and Sims (1981), for instance, found that a single measure of the content of knowledge within a particular domain (i.e., nutrition) was orthogonal to their structural measures.

Currently, we believe that content measures of a domain may provide a better explanation of the resulting cognitive processes within that domain than structural measures. In other words, we believe that what you know is more important than how much you know about a particular domain. It seems clear, however, that multiple measures are required to determine the extent of an individual's knowledge within a domain.

Some work on developing typologies of the content of knowledge within a domain has been done by educational psychologists (e.g., Bloom et al. 1956). We have recently developed a similar typology for classifying the knowledge that a consumer may have about a particular product category or purchase decision. The different types of knowledge within this typology are: (1) terminology (2) specific facts (3) relationships (4) criteria for evaluation and (5) procedural information.

Terminology refers to knowledge of terms used within a particular domain. These might include knowledge of the meaning of U.S. RDA percentages within the nutritional domain or the meaning of electronic fuel injection within the automotive domain. Specific facts are knowledge about objects within the domain. Examples of these are "apples do not contain much vitamin C" in the nutritional domain and "a Datsun B-210 gets 35 miles per gallon" in the automotive domain.

Causal relationships are knowledge about how different attributes of the objects within a domain affect the object's performance. Within the nutritional domain, for example, knowledge that if you don't get enough vitamin C you will get scurvy or within the automotive domain, knowledge that small four cylinder automobiles get good gas mileage are examples of this type of knowledge. The criteria for evaluation measure refers to knowledge that is used in evaluating objects within the domain. Examples of this measure might involve "cutoff" values for the different attributes. In the nutritional domain, a criteria for evaluation might be the knowledge that any fruit that has over 20% of the U.S. RDA of Vitamin A is "good" on that dimension or any small four cylinder automobile that gets over 40 miles per gallon is "good" on that dimension.

The final measure is procedures. This is knowledge concerning how the individual should behave with respect to objects within a domain. For instance, knowledge that you should have a green vegetable every day or that you should change the oil in your car every 5,000 miles are examples of procedures for the nutritional and the automotive domains, respectively.

In summary, an individual's knowledge about a particular domain will have a critical effect on how he or she processes related information. In order to understand the relationship between knowledge and processing activities, we need to develop procedures for measuring the different conceptual dimensions of knowledge.

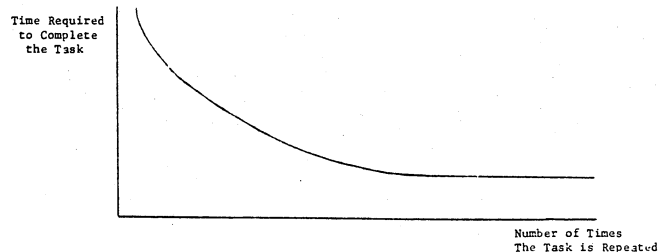
Production Systems

As mentioned previously, production systems are a representation of procedural knowledge or cognitive skill. Production systems have been used to represent cognitive skills in memory scanning (Newell 1973), language comprehension (Anderson, Kline and Lewis 1977) and solving geometry problems (Neves and Anderson 1981).

The basic element of a production system is a condition-action statement. If an individual recognizes a particular condition (e.g., a green traffic light), he or she executes a particular action (e.g., walk across the intersection). A production system contains four basic elements (Rychener and Newell 1978): (1) a production memory, (2) a working memory (3) a recognize-act cycle and (4) conflict resolution principles. The production memory contains the condition-act statements while working memory contains information about the present state of the system. This is the knowledge that triggers the condition element of the condition-act statements. Consequently, working memory is similar to our notion of the workspace in problem perception. Conflict resolution principles are evoked whenever the current state of the

system may trigger more than one condition-act statement. Finally, the recognize-act cycle executes a particular condition-act statement which, in turn, changes the state of the system. In general, network models have been used to represent procedural information in memory (e.g., Anderson 1976, Norman, et. al. 1975, Rumelhart and Ortony 1977).

FIGURE 2
RELATIONSHIP BETWEEN THE TIME REQUIRED TO COMPLETE THE TASK AND THE NUMBER OF TIMES ITS BEEN REPEATED



Learning obviously occurs within production systems. In many tasks, a power law type relationship has been found between the number of times a particular task has been executed and the amount of time that it takes to complete the task (Figure 2). These tasks include perceptual motor skills, perception, motor behavior, elementary decisions and problem solving (Newell and Rosenbloom 1981). In a pretest of an experiment, we found the same relationship between the number of times an individual evaluated different brands based on the same prescribed attributes and the amount of time that it takes him or her to make the evaluation.

Anderson (1980) has suggested that three basic stages occur in learning a cognitive skill. The first is a cognitive stage where the individual learns the steps that one must go through in performing the task. For example, in learning to drive a standard shift automobile, the individual learns that to shift gears the clutch must first be depressed, then the shift lever is moved from one position to the other, and finally the clutch must be released. At this stage, the information is stored in declarative form. The second stage is the associative stage. Here, the individual uses the declarative knowledge to execute the different steps required to perform the task. This causes a production system for the different steps to be formed. At this point in the learning process individuals generally need to devote all of their cognitive resources (e.g., attention) to executing the task. As they execute the task a large number of times, less and less attention is required. Finally, at the third stage, the autonomous stage, individuals can execute the task with little or no conscious attention -- the task has become automatic. A number of studies have examined automatic processes in both perception and reading (e.g., La Berge and Samuels 1974, Schneider and Shiffrin 1977) and research is currently being directed at understanding the factors affecting the development of automatic processes (Shiffrin and Dumais 1981).

Although the stages identified by Anderson (1980) provide insights into the learning of a cognitive skill, they do not provide an explanation of why the amount of time to execute a task seems to follow a power law. Two explanations for this phenomena have been presented. The first, by Newell and Rosenbloom (1981) uses a chunking hypotheses. Under this hypotheses, individuals learn to react to a pattern of stimuli in the environment instead of individual stimuli. For instance, in a task where an individual must give a particular response to a particular

sequence of lights, the individual may at first process the status of each light sequentially or serially. With practice, however, the individual may react to the entire sequence of lights. Evidence for this explanation comes from research examining novice and expert chess players. This research, for instance, indicates that expert chess players react to the general pattern of the pieces on the chess board (Chase and Simon 1973). Newell and Rosenbloom (1981) have shown mathematically that this form of chunking will result in power law type relationship between the time required to execute the task and the number of times the task has been repeated.

A second explanation comes from Neves and Anderson (1981). They hypothesize two mechanisms that may yield the power law type relationship. The first is proceduralization which occurs at the early stages in learning a cognitive skill. With this mechanism, declarative information that is required in executing a production is integrated into the production system. This, of course, reduces the need to retrieve the required information in executing the production which, in turn, reduces the amount of time required to execute the task. The second mechanism is composition which may occur at both early and late stages of learning. This mechanism combines two or more productions into a single production. Consequently, instead of a single condition activating a production, multiple conditions activate the production. This is similar, then, to the chunking notion of Newell and Rosenbloom (1981). Neves and Anderson (1981) have used simulations to demonstrate that these two mechanisms can result in the power law type relationship between the number of times a task is executed and the time required to complete the task.

It seems clear that understanding an individual's procedural knowledge is critical in understanding consumer decision making. To date, however, little research has been directed at understanding the development of the cognitive skills involved in consumer decision making or the measurement of these skills. We suspect that more research will be directed in this area in the near future.

Conclusion

In this paper we have suggested that the goal of future research on consumer decision making should center on predicting what decision strategy an individual will use in a particular situation. In addition, we have suggested that more research should be directed at understanding consumer decision making in "real world" situations. The achievement of these goals require the identification of the critical elements of the situation and the individual which cause the selection of a particular decision strategy. Once these elements have been identified and measured, they may be used to predict either global measures of the decision strategy or the evolution of the actual decision strategy when it is a constructive process. The former is essentially a static analysis while the latter provides an understanding of the dynamics of decision making.

These approaches require more emphasis on why a particular decision strategy was selected as opposed to simply a description of the resulting strategy. They also require a better understanding of the cognitive activity involved in decision making.

Whichever approach is used, and we believe that both will prove fruitful, effort must be directed at identifying and measuring the critical elements of the task and the individual that will affect the selection of a particular decision strategy. In this paper, we represent a simple conceptual model that identifies these elements. They include the task environment, problem perception, general

goals or values, knowledge structures and production systems. If we want to predict the selection of a particular decision strategy, future research effort will need to be directed at obtaining a better conceptualization of these elements and procedures for measuring the critical dimensions of these elements.

References

- Aho, A. J., Hopcroft, E. and Ullman, J.D. (1974), The Design and Analysis of Computer Algorithms, Reading, Mass.: Addison-Wesley.
- Anderson, J. R. (1980), Cognitive Psychology and Its Implications, San Francisco: W. H. Freeman and Company.
- Anderson, J. R. (1976), "Language, Memory and Thought," Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Anderson, J. R., Kline, P.J. and Beasley, C.M. Jr. (1980), "Complex Learning Processes," in R. E. Snow, P. A. Federico and W. E. Montague, eds., Aptitude, Learning, and Instruction: A Cognitive Process Analysis, Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Anderson, J. R., Kline, P.J. and Beasley, C.M. (1979), "A General Learning Theory and its Application to Schema Abstraction," in G. H. Bower, ed., The Psychology of Learning and Motivation, Vol. 13, New York: Academic Press.
- Anderson, J. R., Kline, P.J. and Lewis, C. (1977), "A Production System Model for Language Processing," in P. Carpenter and M. Just, eds., Cognitive Processes in Comprehension, Hillsdale, N.J.: Lawrence Erlbaum Associates, 271-311.
- Bettman, J. R. (1980), An Information Processing Theory of Consumer Choice, Reading, Mass.: Addison-Wesley.
- Bettman, J. R. and Jacoby, J. (1976), "Patterns of Processing in Consumer Information Acquisition," in B. B. Anderson, ed., Advances in Consumer Research, Vol. 3, Chicago: Association for Consumer Research, 315-320.
- Bettman, J. R. and Kakkar, P. (1977), "Effects of Information Presentation Format on Consumer Information Acquisition Strategies," Journal of Consumer Research, 3, 233-240.
- Bettman, J. R. and Park, C.W. (1980), "Effect of Prior Knowledge and Experience and Phase of the Choice Process on Consumer Decision Processes: A Protocol Analysis," Journal of Consumer Research, 7, 234-248.
- Bettman, J. R. and Park, C.W. (1980), "Implications of a Constructive View of Choice for Analysis of Protocol Data: A Coding Scheme for Elements of Choice Processing," in J. C. Olson, ed., Advances in Consumer Research, Vol. 7, Ann Arbor: Association for Consumer Research, 148-153.
- Bettman, J. R. and Zins, M.A. (1977), "Constructive Processes in Consumer Choice," Journal of Consumer Research, 4, 75-85.
- Bloom, B. S., Englehart, M.D., Furst, E.J., Hill, W.H. and Krathwohl, D.R. (1956), Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain, New York: David McKay Company, Inc.
- Brucks, M., Mitchell, A.A. and Staelin, R. (1980), "The Disclosure of Nutritional Information: An Information Processing Approach," Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA 15213.

- Chase, W. G. and Simon, H.A. (1973), "Perception in Chess," Cognitive Psychology, 4, 55-81.
- Edell, J. and Mitchell, A. (1978), "An Information Processing Approach to Cognitive Responses," in S. C. Jain, ed., Research Frontiers in Marketing: Dialogues and Directions, Chicago: American Marketing Association, 178-183.
- Einhorn, H. J. and Hogarth, R.M. (1981), "Behavioral Decision Theory: Processes of Judgment and Choice," Annual Review of Psychology, 32.
- Green, R., Mitchell, A.A. and Staelin, R. (1977), "Longitudinal Decision Studies Using a Process Approach: Some Results from a Preliminary Experiment," in B. A. Greenberg and D. A. Bellenger, eds., Contemporary Marketing Thought, Chicago: American Marketing Association, 461-466.
- Greeno, J. G. (1978), "Nature of Problem-Solving Abilities," in W. K. Estes, ed., Handbook of Learning and Cognitive Processes, Vol. 5, Hillsdale, N.J.: Lawrence Erlbaum Associates, 239-270.
- Hayes-Roth, B. and Hayes-Roth, F. (1979), "A Cognitive Model of Planning," Cognitive Science, 3, 275-310.
- Hayes-Roth, F., Waterman, D.A. and Lenat, D.B. (1978), "Principles of Pattern-Directed Inference Systems," in D. A. Waterman and F. Hayes-Roth, eds., Pattern Directed Inference Systems, New York: Academic Press, 577-601.
- Johnson, E. and Russo, J.E. (1981), "Product Familiarity and the Learning of New Information," in K. Monroe, ed., Advances in Consumer Research, Vol. 8, Ann Arbor: Association for Consumer Research.
- Kanwar, R., Olson, J.C. and Sims, L.S. (1981), "Toward Conceptualizing and Measuring Cognitive Structures," in K. Monroe, ed., Advances in Consumer Research, Vol. 8, Ann Arbor: Association for Consumer Research.
- Kelly, G. A. (1955), The Psychology of Personal Constructs, Vols. 1 and 2, New York: Norton.
- LaBerge, D. and Samuels, S.J. (1974), "Toward a Theory of Automatic Information Processing in Reading," Cognitive Psychology, 6, 293-323.
- Marks, L. J. and Olson, J.C. (1981), "Toward a Cognitive Structure Conceptualization of Product Familiarity," in K. Monroe, ed., Advances in Consumer Research, Vol. 8, Ann Arbor: Association for Consumer Research.
- Mitchell, A. A. (1978), "An Information Processing View of Consumer Behavior," in S. C. Jain, ed., Research Frontiers in Marketing: Dialogues and Directions, Chicago: American Marketing Association, 188-197.
- Neves, D. M. and Anderson, J.R. (1981), "Knowledge Compilation: Mechanisms for the Automatization of Cognitive Skills," in J. R. Anderson, ed., Cognitive Skills and Their Acquisition, Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Newell, A. (1980), "Reasoning, Problem Solving, and Decision Processing: The Problem Space as a Fundamental Category," in R. S. Nickerson, ed., Attention and Performance VIII, Hillsdale, N.J.: Lawrence Erlbaum Associates, 693-718.
- Newell, A. (1973), "Production Systems: Models of Control Structures," in W. G. Chase, ed., Visual Information Processing, New York: Academic Press, 463-526.
- Newell, A. and Rosenbloom, P.W. (1981), "Mechanisms of Skill Acquisition and the Law of Practice," in J. R. Anderson, ed., Cognitive Skills and Their Acquisition, Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Newell, A. and Simon, H. (1972), Human Problem Solving, Englewood Cliffs, N.J.: Prentice Hall.
- Norman, D. A., Rumelhart, D.E. and the LNR Research Group (1975), Explorations in Cognition, San Francisco: W.H. Freeman and Company.
- Olson, J. C. (1978), "Theories of Information Encoding and Storage: Implications for Research," in Andrew A. Mitchell, ed., The Effect of Information on Consumer and Market Behavior, Chicago: American Marketing Association, 49-60.
- Olson, J. C. and Muderrisoglu, A. (1979), "The Stability of Responses Obtained by Free Elicitation: Implications for Measuring Attribute Salience and Memory Structure," in W. Wilkie, ed., Advances for Consumer Research, Vol. 6, Ann Arbor, MI: Association for Consumer Research, 269-275.
- Ortony, A. (1978), "Remembering, Understanding, and Representation," Cognitive Science, 2, 53-69.
- Palmer, S. E. (1980), "Fundamental Aspects of Cognitive Representation," in E. Rosch and B. B. Lloyd, eds., Cognition and Categorization, Hillsdale, N.J., Lawrence Erlbaum Associates, 262-303.
- Payne, J. W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," Organizational Behavior and Human Performance, 16, 366-387.
- Rumelhart, D. E. and Ortony, A. (1978), "The Representation of Knowledge in Memory," in R. C. Anderson, R. J. Spiro and W. E. Montague, eds., Schooling and the Acquisition of Knowledge, Hillsdale, N.J.: Lawrence Erlbaum Associates, 37-54.
- Russo, J. E. and Rosen, L.D. (1975), "An Eye Fixation Analysis of Multi-Alternative Choice," Memory and Cognition, 3, 267-276.
- Rychener, M. D. and Newell, A. (1978), "An Instructable Production System: Basic Design Issues," in D. A. Waterman and F. Hayes-Roth, eds., Pattern-Directed Inference Systems, New York: Academic Press, 135-154.
- Shank, R. C. (1980), "Language and Memory," Cognitive Science, 4, 243-284.
- Shank, R. C. and Abelson, R. (1977), Scripts, Plans, Goals and Understanding, Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Schneider, W. and Shiffrin, R.M. (1977), "Controlled and Automatic Information Processing: I. Detection, Search and Attention," Psychological Review, 84, 1-66.
- Shiffrin, R. M. and Dumais, S.T. (1981), "The Development of Automatism," in J. R. Anderson, ed., Cognitive Skills and Their Acquisition, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Shiffrin, R. M. and Schneider, W. (1977), "Controlled and Automatic Human Information Processing: II Perceptual Learning, Automatic Attending and a General Theory," Psychological Review, 84, 492-512.

Simon, H. A. (1978), "Information Processing Theory of Human Problem Solving," in W. K. Estes, ed., Handbook of Learning and Cognitive Processes, Vol. 5, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 271-296.

Simon, H. A. and Hayes, J.R. (1976), "The Understanding Process: Problem Isomorphs" Cognitive Psychology, 8, 165-190.

Tversky, A. and Kahneman, D.(1981), "The Framing of Decisions and the Rationality of Choice," Science.

Wright, P. and Rip, P.D. (1980), "Product Class Advertising Effects on First-Time Buyers' Decision Strategies," Journal of Consumer Research, 7, 176-188.

AUTHOR INDEX

Aaker, David A.	616	Gottschalk, Ingrid	391
Acito, Franklin	313	Granbois, Donald	693
Albaum, Gerald S.	506	Green, Paul E.	309
Allen, Chris T.	398	Greenwald, Anthony G.	15
Allison, Neil K.	419	Gregory, Russell	624
Anderson, W. Thomas	292	Grønhaug, Kjell	83, 564
Arndt, Johan	564	Grossbart, Sanford L.	128
Arnold, Stephen J.	665	Grunert, Kaus G.	391
		Gutman, Jonathan	116
Bagozzi, Richard P.	195, 616	Hackleman, Edwin C.	477
Barak, Benny	602	Hansen, Fleming	367
Barnaby, David. J.	71	Hawkins, Del I.	713
Barnes, James H. Jr.	304	Heeler, Roger M.	728
Belk, Russell W.	170	Heimbach, James T.	474
Bennett, Peter D.	568	Hendrix, Philip E.	182
Berkely, Robert L.	730	Henion, Karl E.	624
Berkowitz, Eric N.	176	Hesselbart, Susan	570
Bernhardt, Kenneth L.	414, 452	Hilger, Harald	381
Blakely, Jerald W.	637	Hirschman, Elizabeth C.	100, 648
Bloch, Peter H.	61	Hoff, N. Gail	245
Bonfield, E.H.	748	Holbrook, Morris B.	35
Brinberg, David	48	Holman, Rebecca H.	187
Brisoux, Jacques E.	357	Homans, Richard E.	564
Brown, Stephen W.	437	Houston, Michael J.	677
Brucks, Merrie	750	Howard, John	9
Burns, Alvin C.	736	Howell, Roy D.	655, 671
		Hoyer, Wayne D.	299, 410, 416
Cacioppo, John T.	20	Huber, Joel	66
Cattin, Philippe	323	Hutton, R. Bruce	547
Chandran, Rajan	718		
Chapman, Randall G.	530	Jacobs, Eva E.	251
Chéron, Emmanuel J.	423	Jacoby, Jacob	299, 410, 416, 511
Clee, Mona A.	624	Johnson, Eric J.	151
Coney, Kenneth A.	713	Johnson, Michael E.	112
Cosmas, Stephen C.	742	Johnston, Wesley J.	192
Crosby, Lawrence A.	630	Jolly, David	405
Czerwonka, Christine	374		
		Kahne, Hilda	585
Danes, Jeffrey E.	57, 323	Kanwar, Rajesh	122
Darden, Donna K.	655	Kassarjian, Harold H.	6, 31
Darden, William R.	655	Kernan, Jerome B.	607
Dardis, Rachel	553	King, Robert L.	449
Deacon, Robert T.	517	Kinnear, Thomas C.	414
Dedler, Konrad	391	Kistler, Ernst	527
Della Bitta, Albert W.	161	Kohen, Janet A.	576
De Sarbo, Wayne S.	309	Krishnan, S.	338
Deshpande, Rohit	338		
DeVere, Stephen P.	736	Langmeyer, Lynn	612
Dholakia, Ruby Roy	362	Lannom, Linda B.	238
Dillon, William R.	209, 398	Laroche, Michel	357
Dolich, Ira J.	140	Lautman, Martin R.	434
Douglas, Susan P.	100	Leavitt, Clark	15
		Lehmann, Donald R.	233
Ekstrom, Ruth B.	580	Leigh, Thomas W.	317
Elrod, Terry	66	Leone, Robert P.	287
Engel, James F.	12	Lichtenstein, Sarah	497
		Linda, Gerald	88
Farley, John U.	233	Lowenhar, Jeffrey	718
Feldman, Shel	329	Lundsgaard, Niels Erik	367
Fenwick, Ian	723	Lusch, Robert F.	644
Ferber, Robert	238, 274, 545	Lynch, John G.	134
Finn, David W.	344		
Fischhoff, Baruch	497	MacKay, David B.	317
Fleischmann, Gerd	386	Mackenzie, Scott B.	708
Fritz, Wolfgang	381	MacLachlan, Douglas	178
		Maddox, R. Neil	564
Gardner, David M.	76	Malhotra, Naresh K.	306
Ghingold, Morry	442	Marks, Larry J.	145
Gill, James D.	630	Martin, A.J.	329
Golden, Linda L.	292		

Martin, Claude R.	182	Srivastava, Rajendra K.	106
Martin, Warren S.	77	Stanton, John	718
May, Frederick E.	564	Stewart, Michael L.	488
Mazis, Michael B.	414, 455	Stiff, Ronald	452
McCullough, James	178	Stremel, Julie	553
McNeill, Dennis L.	547	Sudman, Seymour	262
Meadow, H. Lee	742	Summers, John O.	317
Miaoulis, George	612	Sunenshine, Harry S.	332
Miller, Shirley J.	655	Swan, John E.	77
Mills, Michael K.	682	Swasy, John L.	349
Miniard, Paul W.	42		
Mitchell, Andrew A.	25, 750	Tabourel, Gerard J.	35
Mittelstaedt, Robert	270	Tan, Chin T iong	140
Mizerski, Richard W.	419	Thorelli, Hans B.	467
Moinpour, Reza	178	Tigert, Douglas J.	665
Molitor, Graham T.T.	458	Tybout, Alice M.	408
Monroe, Kent B.	161		
Mowen, John C.	405, 437	Unger, Lynette S.	607
Munch, James M.	349	Upah, Gregory D.	262
Munson, J. Michael	696	Uusitalo, Jyrki	559
		Uusitalo, Liisa	559
Nelson, Margaret C.	416		
Nevin, John R.	677	Vandenburg, Robert J.	484
Nickell, Gary S.	405	Vehorn, Charles L.	523
Noam, Eli M.	534	Velez, David A.	35
		Venkatesan, M.	503
Obermiller, Carl	15		
Oliver, Richard L.	88, 94	Warshaw, Paul R.	53
Olshavsky, Richard W.	313	Weinberger, Marc G.	398
Olson, Jerry C.	122, 145	Wenglorz, Gerhard	687
		Westbrook, Robert A.	94
Palda, Kristian S.	530	Westwood, Richard A.	332
Parsons, Leonard J.	203	Wheatley, John J.	166
Pearl, Robert B.	254	Wilkie, William L.	1
Percy, Larry	434	Williams-Jones, Joy	281
Perrien, Jean	423	Wilson, R. Dale	227
Peterson, Robert A.	287, 662	Winer, Russell S.	220
Petty, Richard E.	20	Wiseman, Fred	267
Pfaff, Martin	527	Wortzel, Lawrence H.	730
Philpot, John W.	71	Woodside, Arch G.	428
Plotkin, Andy	742		
Pommerchne, Werner W.	539	Yokum, J. Thomas	702
Raffée, Hans	381	Zaichkowsky, Judy L.	708
Rammohan, Balusu	128	Zaltman, Gerald	83
Reece, Bonnie B.	414	Zikmund, William K.	622
Reizenstein, Richard C.	71		
Rentz, Joseph O.	597		
Rethans, Arno J.	506		
Reynolds, Fred D.	596		
Roberts, Mary Lou	590, 730		
Rogers, Jerry D.	671		
Roshwalb, Irving	276		
Roth, Victor	665		
Rothschild, Michael L.	642		
Roupe, Don	713		
Russo, J. Edward	151		
Sabertehrani, Mohammad H.	287		
Schellinck, Tony	723		
Schiffman, Leon G.	602		
Schlinger, Mary Jane	432		
Schoppe, Gunter	374		
Schucker, Raymond E.	493		
Scott, Carol A.	408		
Sharpe, Louis K.	292		
Sheluga, David A.	410		
Sheth, Jagdish N.	355		
Shimp, Terence A.	702		
Shuptrine, F. Kelly	687		
Silberer, Gunter	381		
Sims, Laura S.	122		
Sirgy, M. Joseph	156		
Slovic, Paul	497		
Smith, Scott M.	637		
Spivey, W. Austin	696		

