

Norbert Schwarz / Seymour Sudman

Editors

---

Autobiographical  
Memory and the  
Validity of  
Retrospective Reports

---



Springer-Verlag

Schwarz  
Sudman  
Editors

Autobiographical Memory and the  
Validity of Retrospective Reports



*W. B. G. Gray*

*Dec 93*

**Autobiographical Memory and  
the Validity of Retrospective Reports**



Norbert Schwarz  
Editors

Seymour Sudman

# Autobiographical Memory and the Validity of Retrospective Reports

With 24 Illustrations



Springer-Verlag

New York Berlin Heidelberg London Paris  
Tokyo Hong Kong Barcelona Budapest

Norbert Schwarz  
Professor  
Institute for Social Research  
University of Michigan  
Ann Arbor, MI 48106-1248 USA

Seymour Sudman  
Professor  
Survey Research Laboratory  
University of Illinois at  
Urbana-Champaign  
Urbana, IL 61801-3883 USA

Library of Congress Cataloging-in-Publication Data  
Autobiographical memory and the validity of retrospective reports /  
[edited by] Norbert Schwarz, Seymour Sudman.

p. cm.

Papers presented at a conference.

Includes bibliographical references.

ISBN 0-387-94167-3. — ISBN 3-540-94167-3

1. Autobiographical memory—Congresses. I. Schwarz, Norbert, Dr.

phil. II. Sudman, Seymour.

BF378.A87A884 1993

153.1'2—dc20

93-31775

Printed on acid-free paper.

© 1994 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Hal Henglein; manufacturing supervised by Jacqui Ashri.

Camera-ready copy prepared by the editors.

Printed and bound by Edwards Brothers, Inc., Ann Arbor, MI.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-94167-3 Springer-Verlag New York Berlin Heidelberg

ISBN 3-540-94167-3 Springer-Verlag Berlin Heidelberg New York

# Acknowledgments

The conference on which the present volume is based was supported by the Walter Stellner Endowment in the Department of Business Administration of the University of Illinois at Urbana-Champaign, by the Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, Germany, and by grant SWF 0044-6 from the Bundesminister für Forschung und Technologie to Norbert Schwarz. In addition, preparation of the present volume was supported by a transatlantic cooperation grant from the Alexander von Humboldt Stiftung to Norbert Schwarz, Norman M. Bradburn, Hans-J. Hippler, and Seymour Sudman.

We also wish to thank the staff of the Robert Allerton Park and Conference Center of the University of Illinois at Urbana-Champaign for facilitating the interaction of the participants with their delightful midwestern cooking, and at the social hours and coffee breaks. We wish to thank Mary A. Spaeth for her careful editing and for formatting the material into camera-ready copy. Finally, we thank all of the conference participants for the fine chapters that they submitted and for their useful comments on each other's initial drafts that helped improve the final versions.

Ann Arbor, Michigan  
Urbana, Illinois  
April 1993

Norbert Schwarz  
Seymour Sudman



# Contents

Acknowledgments ..... v  
Contributors ..... xi

Introduction and Overview  
*Norbert Schwarz and Seymour Sudman* ..... 1

**PART I. PERSPECTIVES ON RETROSPECTIVE REPORTS**

Chapter 1  
Autobiographical Memory and Survey Research  
*William F. Brewer* ..... 11

Chapter 2  
The Validity of Retrospective Reports as a Function of the  
Directness of Retrieval Processes  
*Douglas J. Herrmann* ..... 21

Chapter 3  
Accuracy, Truth, and Meaning in Autobiographical Reports  
*Kurt W. Back* ..... 39

Chapter 4  
On Authenticating and Using Personal Recollections  
*Michael Ross and Roger Buehler* ..... 55

Chapter 5  
Affect and Memory in Retrospective Reports  
*Mahzarin R. Banaji and Curtis Hardin* ..... 71

## PART II. RETROSPECTIVE REPORTS OF BEHAVIORS

Chapter 6 The Recall of Physical Pain <i>Peter Salovey, William J. Sieber, Jared B. Jobe, and Gordon B. Willis</i> .....	89
Chapter 7 The Effects of Estimation Strategies on the Accuracy of Respondents' Reports of Cigarette Smoking <i>Barbara Means, Gary E. Swan, Jared B. Jobe, and James L. Esposito</i> .....	107
Chapter 8 Validity of Reports of Long-Term Dietary Memories: Data and a Model <i>Albert F. Smith and Jared B. Jobe</i> .....	121
Chapter 9 Errors of Experience: Response Errors in Reports about Child Support and Their Implications for Questionnaire Design <i>Nora Cate Schaeffer</i> .....	141
Chapter 10 Judgments of Behavioral Frequencies: Memory Search and Retrieval Strategies <i>Geeta Menon</i> .....	161
Chapter 11 On Providing Population Data To Improve Respondents' Estimates of Autobiographical Frequencies <i>Edward Blair and Kathleen Williamson</i> .....	173
Chapter 12 Retrospective Reports: The Impact of Response Formats <i>Norbert Schwarz, Hans-J. Hippler, and Elisabeth Noelle-Neumann</i> .....	187

## PART III. EVENT DATING AND TIME ESTIMATION

Chapter 13 Telescoping and Temporal Memory <i>Norman M. Bradburn, Janellen Huttenlocher, and Larry Hedges</i> ....	203
--	-----

Chapter 14  
 The Impact of Differing Memory Domains on Event-Dating  
 Processes in Self and Proxy Reports  
*John J. Skowronski, Andrew L. Betz, Charles P. Thompson,  
 W. Richard Walker, and Laura Shannon* ..... 217

PART IV. COMPARISONS OF SELF AND PROXY REPORTS

Chapter 15  
 Self and Proxy Reports of Everyday Events  
*David J. Mingay, Steven K. Shevell, Norman M. Bradburn,  
 and Carl Ramirez* ..... 235

Chapter 16  
 The Effect of Participation Level on Reports of Behavior  
 and Attitudes by Proxy Reporters  
*Seymour Sudman, Barbara Bickart, Johnny Blair, and  
 Geeta Menon* ..... 251

Chapter 17  
 Reconstruction of Relationship Memories:  
 A Mental Models Approach  
*Diane Holmberg and John G. Holmes* ..... 267

PART V. MEMORIES OF THE PAST AND JUDGMENT OF  
 PERSONAL AND SOCIAL CHANGE

Chapter 18  
 Biasing Effects of Retrospective Reports on  
 Current Self-Assessments  
*Leslie F. Clark, James E. Collins II, and Susan M. Henry* ..... 291

Chapter 19  
 Reconstructing Social Change through Retrospective Questions:  
 Methodological Problems and Prospects  
*Karl-Heinz Reuband* ..... 305

Chapter 20  
 Collective Memories in the United States and Lithuania  
*Howard Schuman, Cheryl Rieger, and Vldas Gaidys* ..... 313

References ..... 335



# Contributors

## Introduction

*Norbert Schwarz* is Professor of Psychology at the University of Michigan, Ann Arbor, and Research Scientist in the Institute for Social Research. At the time of the conference on which this volume is based, he directed the Cognition and Survey Research Program at the Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, Germany. His research interests focus on human judgmental processes, in particular the interplay of affect and cognition, and the application of social cognition research to survey research.

*Seymour Sudman* is Walter H. Stellner Distinguished Professor of Marketing, Professor of Business Administration and of Sociology, and Deputy Director and Research Professor at the Survey Research Laboratory, University of Illinois at Urbana-Champaign. He received his Ph.D. in Business from the University of Chicago. He is the author, co-author, or editor of over 15 books and 100 articles dealing primarily with survey methodology and response effects in surveys. His current research interests are related to cognitive aspects of proxy reporting in surveys.

## Chapter 1

*William F. Brewer* has appointments at the University of Illinois at Urbana-Champaign as Professor in the Department of Psychology and the Beckman Institute, Research Professor in the Institute of Communications Research, and Senior Scientist at the Center for the Study of Reading. He carries out research in the areas of knowledge representation, human memory, text structure, and knowledge acquisition.

## Chapter 2

*Douglas J. Herrmann* is Director of the Collection Procedures Research Laboratory, and Chief of the Behavioral Science Research group in the

Office of Survey Methodology Research, at the Bureau of Labor Statistics. He obtained his M.S. in 1970 and his Ph.D. in 1972 in Experimental Psychology at the University of Delaware. He taught at Hamilton College (Clinton, New York) until 1989. His current work investigates (a) ways to facilitate autobiographical and other kinds of memory performance and (b) the psycholinguistic bases of question comprehension.

### Chapter 3

*Kurt W. Back* is James B. Duke Professor Emeritus at Duke University. He obtained a Ph.D. in group psychology at MIT and has worked on applications of social psychology in sociopsychology, population control, housing, the life course, and social movements. Among his books are *Beyond Words: The Story of Sensitivity Training and the Encounter Group Movement* and *Family Planning and Population Control: The Challenge of a Successful Movement*.

### Chapter 4

*Michael Ross* received his Ph.D. in social psychology at the University of North Carolina, Chapel Hill. He is Professor of Psychology at the University of Waterloo, Ontario, Canada. His research interests are wide ranging, but he has an abiding fascination with autobiography and memory.

*Roger Buehler* obtained his B.A., M.A., and Ph.D. in psychology at the University of Waterloo, Ontario, Canada. He is currently Assistant Professor at Simon Fraser University, Burnaby, British Columbia, Canada. His research interests include autobiographical memory, social cognition, and the psychology of prediction.

### Chapter 5

*Mahzarin R. Banaji* is Assistant Professor in the Department of Psychology at Yale University. Her research examines systems of affect, self, stereotyping and prejudice. She is also interested in the theoretical analysis of method.

*Curtis Hardin* is a doctoral candidate in social psychology at Yale University. His research interests include the role of social identity processes in affect, memory, social judgments, and sense of self.

### Chapter 6

*Peter Salovey* received his Ph.D. in clinical psychology from Yale University, where he is Associate Professor and contributing to both the social/personality and clinical psychology programs. His major research interest is the function of human emotion, especially how emotions direct cognition, motivate social behavior, and color close relationships. He has recently been concerned with the influence of emotion on perceptions of

health, illness, and physical pain and has published findings in this area in the *Journal of Personality and Social Psychology*.

*William J. Sieber* is a doctoral candidate in clinical psychology at Yale University and is presently an intern at the Palo Alto Veterans Affairs Medical Center. His research interests concern the expression of physical symptoms and feeling states. His doctoral research concerns individual differences in attentional focus on somatic sensations versus emotional aspects of stressful events. His publications include articles on the influence of mood on self-focused attention in *Cognitive Therapy and Research*.

*Jared B. Jobe* is Director of the Collaborative Research Program at the National Laboratory for Collaborative Research in Cognition and Survey Measurement at the National Center for Health Statistics in Hyattsville, Maryland, where he has been since 1987. He directs an extramural program and conducts research in cognitive psychology and health surveys. He received his bachelor's, master's, and doctorate degrees in psychology from the University of Oklahoma in 1972, 1973, and 1976, respectively, specializing in learning and memory. With E. F. Loftus, he co-edited in 1991 a special issue of *Applied Cognitive Psychology*, devoted to cognition and survey measurement.

*Gordon B. Willis* is a survey statistician at the National Center for Health Statistics in Hyattsville, Maryland, where he has worked since 1987 in the Questionnaire Design Research Laboratory developing and testing health survey questionnaires. He received his Ph.D. from Northwestern University in 1985. His research program has focused on the application of cognitive psychology to early childhood mathematics education, the development of giftedness in adolescence, and survey methodology.

## Chapter 7

*Barbara Means* is Director of the Education and Human Services Research Program at SRI International in Menlo Park, California. Her research focuses on practical applications of cognitive psychology to areas such as surveys and education and training. Her publications have appeared in *Applied Cognitive Psychology* and in *Measurement error in surveys*, edited by P. P. Biemer et al. She earned her Ph.D. in educational psychology at the University of California, Berkeley.

*Gary Swan* is Director of the Health Sciences Program, SRI International. After receiving his Ph.D. in psychology from the State University of New York at Stony Brook in 1978, he joined the staff at SRI International and has since been conducting research in two primary areas: substance use (tobacco and alcohol) and aging (cardiovascular disease and premature

cognitive decline). He is an active contributor to this field and has over 50 journal publications.

*Jared B. Jobe*—see Chapter 6.

*James L. Esposito* is a research psychologist at the Bureau of Labor Statistics, working as part of a team that is redesigning the Current Population Survey—the federal government’s principal labor force survey. Prior to moving to BLS in 1990, he conducted sensitive questions research at the National Center for Health Statistics. In addition to his work on sensitive questions, his research interests include the development of a model/framework of the survey interaction process, the impact of confidentiality assurances on data quality, and survey methodological issues. He received a Ph.D. in social psychology from Temple University in 1986.

## Chapter 8

*Albert F. Smith* is a member of the Psychology faculty at the State University of New York at Binghamton. He received his M.S. from Columbia University in Biostatistics and his Ph.D. from Yale in Psychology. He conducts basic research on the perception and representation of stimulus structure and investigates the cognitive processes that subserve responding in a variety of survey contexts.

*Jared B. Jobe*—see Chapter 6.

## Chapter 9

*Nora Cate Schaeffer* is Associate Professor of Sociology at the University of Wisconsin–Madison. She received her Ph.D. from the University of Chicago in 1984. She has published in *Public Opinion Quarterly*, *Sociological Methods & Research*, and *Journal of the American Statistical Association*. Her current research focuses on how the interaction between the interviewer and respondent affect survey data and how to design standardized questions to improve the accuracy of self-reports.

## Chapter 10

*Geeta Menon* is Assistant Professor of Marketing at the Leonard N. Stern School of Business, New York University. She received her Ph.D. in Business Administration from the University of Illinois at Urbana-Champaign in 1991. Her current research interests include cognitive aspects of survey methodology (how people answer behavioral and attitudinal questions about themselves and others) and autobiographical memory.

## Chapter 11

*Edward Blair* is Associate Professor of Marketing at the University of Houston. He holds B.S. and Ph.D. degrees from the University of Illinois at

Urbana-Champaign. His papers on autobiographical frequency data in surveys have appeared in various journals.

*Kathleen Williamson* is a doctoral student at the University of Houston. She holds a B.A. degree from Rice University and an M.B.A. from the University of Houston.

## Chapter 12

*Norbert Schwarz*—see Introduction and Overview.

*Hans-J. Hippler* is Director of Methodology at the Regionalpresse, a media research center in Frankfurt, Germany. At the time of the conference, he was a Project Director at Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, Germany. His basic research interests focus on cognitive aspects of survey research and the impact of mode of data collection on data quality.

*Elisabeth Noelle-Neumann* is Director of the Institut für Demoskopie, Allensbach, Germany, which she founded in 1947. Since 1964 she has been Professor of Mass Communication Research at the University of Mainz. She has been a visiting professor in the Department of Political Science, University of Chicago, since 1978 and Georges Lurcy Visiting Professor since 1985. Her numerous publications focus on survey research methodology, international comparisons of value systems, public opinion theory, and the effects of mass media.

## Chapter 13

*Norman M. Bradburn* is the Tiffany and Margaret Blake Distinguished Service Professor, Department of Psychology, University of Chicago, and Director of Research at the National Opinion Research Center.

*Janellen Huttenlocher* is the William S. Gray Professor, Departments of Psychology and Education, University of Chicago.

*Larry Hedges* is a professor in the Department of Education, University of Chicago.

## Chapter 14

*John J. Skowronski* is Assistant Professor of Psychology at The Ohio State University at Newark. His M.A. and Ph.D. are from the University of Iowa. His research specialty is social cognition, with particular emphasis on the relationship between social memory and social judgment. Recent publications include an analysis of the causes of negativity and positivity effects in impression formation and an investigation of the attitudes underlying organ donation behavior.

*Andrew L. Betz* is a graduate student in social psychology at The Ohio State University. He received his B.A. Magna Cum Laude from Bowling Green State University and his M.A. from The Ohio State University. His research interests currently include the subliminal conditioning of attitudes and the effects of mental representations on social inference.

*Charles P. Thompson* is Professor of Psychology at Kansas State University. He received his Ph.D. from the University of Wisconsin and spent 3 years at the University of Wyoming. His research specialty is memory, with recent emphasis on autobiographical memory, voice identification, and extraordinary memory.

*W. Richard Walker* is an undergraduate psychology major at The Ohio State University. In addition to his work on memory and event dating, for which he received two honors grants from the Ohio State University, he has recently participated in studies investigating the impact of question order on judgments of personality traits, and in a series of studies investigating impression formation processes.

*Laura Shannon* is an undergraduate psychology major at The Ohio State University. In addition to her work on memory and event dating, she has recently participated in a series of studies investigating organ donation attitudes and in a series of studies investigating impression formation processes.

## Chapter 15

*David J. Mingay* obtained a Ph.D. in cognitive psychology at the MRC Applied Psychology Unit, Cambridge, in 1986. Since then he has worked in the field of cognitive aspects of survey methodology, first as a service fellow at the National Center for Health Statistics and since 1989 as a survey methodologist at the National Opinion Research Center. He has authored book chapters and journal articles on the cognitive interview and on the role of memory in survey reporting.

*Steven K. Shevell* is Professor in the Department of Psychology at the University of Chicago. He has an M.S. in engineering-economic systems from Stanford University (1973) and an M.A. in statistics (1975), an M.A. in psychology (1975), and a Ph.D. in psychology (1977), all from the University of Michigan. His primary research interests are in human vision and mathematical psychology.

*Norman M. Bradburn*—see Chapter 13.

*Carl Ramirez* obtained a master's degree in sociology at the University of Chicago in 1991. He has worked as an analyst for Penn and Schoen As-

sociates, a survey research organization, and currently works for the U.S. General Accounting Office as a social science analyst.

## Chapter 16

*Seymour Sudman*—see Introduction and Overview.

*Barbara Bickart* is currently Assistant Professor of Business Administration at Rutgers University. She received her Ph.D. in Business Administration from the University of Illinois at Urbana-Champaign in 1990. Her current research interests include context effects on consumer judgments and decision-making processes, the effects of socially shared cognition on household judgment and decision-making processes, and the development of methods to reduce measurement error in surveys.

*Johnny Blair* is Associate Director of the Survey Research Center, University of Maryland, College Park. His B.A. in political science and mathematics was obtained from the University of Illinois at Urbana-Champaign in 1969. His current research interests include experimental comparison of questionnaire pretest methods, sample design for special populations, and response effects in proxy reporting.

*Geeta Menon*—see Chapter 10.

## Chapter 17

*Diane Holmberg* received a B.A. in psychology from the University of Waterloo in 1989 and an A.M. in psychology from the University of Michigan in 1991. She is currently a Ph.D. candidate at the University of Michigan. Her research interests include memory in close relationships and challenges to the self-concept.

*John G. Holmes* is Professor of Psychology at the University of Waterloo, Waterloo, Ontario, Canada. He received his Ph.D. in social psychology in 1971 from the University of North Carolina. His current research interests include the appraisal process in close relationships, cognitive representation of significant others, and interpersonal conflict.

## Chapter 18

*Leslie F. Clark* obtained her Ph.D. in social psychology in 1985 and is currently Associate Professor at Purdue University. Her research interests are stress and coping, stereotyping, and autobiographical memory.

*James E. Collins II* earned a Ph.D. in social psychology in 1987 and is Associate Professor at Carson-Newman College, Jefferson City, Tennessee. He has published in *Social Cognition* and *Psychological Review*. His research interests are humor, autobiographical memory, and mental control.

*Susan M. Henry* is a graduate student at Purdue University, where she received an M.A. in social psychology in 1991. Her research interests include individuation processes, autobiographical memory, social identity and exemplar theory.

## Chapter 19

*Karl-Heinz Reuband* received his Ph.D. in sociology at the University of Hamburg and a habilitation at the University of Cologne. He is a research associate at the Zentralarchiv für empirische Sozialforschung of the University of Cologne and has been a Privatdozent for Sociology at that University since 1990. In 1991–92, he was a visiting professor at the University of Trier. His research interests include survey methodology, social and cultural change, and social problems and youth.

## Chapter 20

*Howard Schuman* is Professor of Sociology and Program Director in the Institute for Social Research, both at the University of Michigan. He is also currently the editor of *Public Opinion Quarterly*.

*Cheryl Rieger* is a Ph.D. candidate in sociology at the University of Michigan. With support from the University's Research Partnership Program, she is writing a dissertation on collective memory in Israel, Russia, and several other countries.

*Vladas Gaidys* is the Director of the Survey Research Center in Vilnius, Lithuania.

# Introduction and Overview

*Norbert Schwarz and Seymour Sudman*

Much of our knowledge about individuals' behavior, the state of society, or social change is based on direct verbal reports of research participants. From consumer behavior to health problems, and from the styles of parenting to the nation's unemployment rate or the prevalence of crime, social scientists and psychologists rely on respondents' reports as their major data base for testing theories of human behavior and for offering advice on public policy. Respondents may be asked how often they consume something, how frequently they see a doctor, if they have been mugged during the last two years, or how often their parents went to church when the respondent was a child. Given the importance of retrospective behavioral reports in basic as well as applied research, it is not surprising that researchers have often worried about the validity of these reports.

There are studies of reporting error that go back at least to the 1930s, and the survey literature in the 1950s and 1960s is already extensive and is summarized in Sudman and Bradburn's *Response Effects in Surveys* (1974). Most of this work was done by survey researchers who were not psychologists. This reflects, in part, the fact that psychological research on memory concentrated on the learning of highly artificial material under laboratory conditions and paid little attention to how individuals encode, organize, store, and retrieve information about their daily lives. It is only recently that cognitive psychologists have turned to a systematic exploration of people's memory for events in their lives, an area of research now known as *autobiographical memory*. As Martin Conway (1990, p. 16) observed, an initial interest in this area in the late 1800s and early 1900s was followed by "a gap in the research record between the 1900s and the early 1970s, in which only a handful of studies were reported." Following these "hundred years of silence" (Cohen, 1986), research into autobiographical memory has seen a strong revival during the last two decades (see M. A. Conway, 1990, for a textbook-length review of early and current research).

This revival of basic psychological research into autobiographical memory coincides with the survey research community's increasing interest in cognitive aspects of the survey process. How is it that respondents go about answering

retrospective questions? Presumably, the better we understand the cognitive processes involved, the better we shall be able to design questionnaires that facilitate respondents' performance and thus, we hope, increase the validity of the obtained reports. At the least, we should be able to get a better understanding of how well respondents are able to perform the tasks that they are given and what tasks it is reasonable to expect a respondent to perform.

These hopes have fueled much of the recent collaboration between cognitive psychologists and survey methodologists (see Jobe & Mingay, 1991, for a short history of this interdisciplinary development). In addition to research on context effects in attitude and opinion measurement (for reviews, see Schwarz & Strack, 1991, and Tourangeau & Rasinski, 1988; for recent contributions, see Schwarz & Sudman, 1992), much of the work done at the intersection of cognitive psychology and survey methods has concentrated on issues of retrospective behavioral reports (for reviews, see Bradburn, Rips, & Shevell, 1987; Jobe & Loftus, 1991; Schwarz, 1990a).

The present volume grew out of a conference that reflected this collaborative interest. The conference was held at the Robert Allerton Park and Conference Center of the University of Illinois at Urbana-Champaign in November 1990, with support from the Walter Stellner Endowment in the Department of Business Administration at the University of Illinois at Urbana-Champaign, the Survey Research Laboratory of the University of Illinois, and the Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim, Germany. The conference was the third in a series of conferences designed to foster the collaboration of cognitive psychologists and survey methodologists, who frequently address the same issues without being aware of the work done in the respective "other" field (see Hippler, Schwarz, & Sudman, 1987, and Schwarz & Sudman, 1992, for contributions to the previous conferences, which focused primarily on attitude and opinion measurement). In planning the conference, it was our goal to invite researchers whose work covered a wide range of issues related to autobiographical memory and the validity of retrospective reports, reflecting the diverse traditions in cognitive psychology and survey research.

## A Preview

### Perspectives on Retrospective Reports

The first five chapters of this volume provide different theoretical perspectives on retrospective reports, along with experimental evidence that supports these perspectives. In Chapter 1, William Brewer introduces a taxonomy that organizes different kinds of autobiographical memories in terms of frequency of exposure to an event or behavior, the type of input, and different forms of memory representations. He summarizes basic research on each of the different types of autobiographical memories and relates the emerging insights to the kinds of questions that respondents may be asked in a survey interview. The taxonomy

offered in this chapter is particularly helpful in identifying the nature of the specific task posed by different retrospective questions and the type of information that respondents are likely to draw on. Brewer also makes the important point that autobiographical memory contains information about location, people, and the actions of people but very limited information about time.

In Chapter 2, Douglas Herrmann contrasts different retrieval strategies. At one extreme, a memory may be difficult to retrieve and the report may be largely based on inference and reconstruction. At the other extreme, the memory may be directly accessible and little inference may be necessary. It has usually been assumed that most retrospective reports involve both kinds of retrieval strategies. Specifically, respondents are often assumed to have direct access to some pieces of relevant information and to use that information as a basis for additional inferences and reconstructions. Whereas much psychological research has focused on reconstructive processes, the studies reported in this chapter suggest that direct retrieval may be more common than has often been assumed: When asked to report on the retrieval process, subjects claimed that they recalled attributes of their experiences primarily by direct retrieval. The chapter also discusses the effects on the retrieval method used of such factors as the retention interval, familiarity, degree of firsthand experience, predictability of parts of an event, and the original emotional state. Inferential retrieval was used only when the episode possessed scriptlike qualities. Although the validity of such introspective reports about one's own cognitive processes is difficult to determine, the findings reported here raise numerous interesting issues for further research.

In Chapter 3, Kurt Back reminds us that autobiographical memories serve different functions depending on whether respondents report about themselves or about others and on whether the questions are mainly about the impact of society on individuals or of individuals on society. In survey research, respondents function as informants about society. In literary biography, on the other hand, the emphasis is on the uniqueness of the individual. Psychotherapeutic interviews are obtained to understand the impact of society on a specific individual, whereas in heroic myth, autobiographical data about an individual are transformed to the cultural stereotype of the hero. Back discusses how the meaning of accuracy differs for these different typologies.

The fourth chapter continues the general theme of the functions of autobiographical memory. Michael Ross and Roger Buehler ask, "When do people focus on their pasts, mining them for what might be of use in the present, and when do they disregard their histories?" Obviously, using our memories in the pursuit of our goals would be of little use if we distrusted their authenticity. Accordingly, Ross and Buehler explore how individuals determine if they can trust their recollections in the first place before turning to the use of recollections in making decisions and predictions. Whereas the findings suggest that people, by and large, trust their memories, they also indicate that we often fail to make use of these memories in predicting the future or planning our behavior.

In the final chapter in this section, Mahzarin Banaji and Curtis Hardin address the role of affect in autobiographical memory and explore if the valence

and intensity of an affective experience contribute to the memorability of the event that elicited it. Linking this issue to the recent distinction between explicit and implicit memory (see Schacter, 1987), they compare the impact of valence and intensity on direct measures of memory, such as explicit recall, as well as indirect measures, such as subjective evaluations. They conclude from their experiments and those of others that affective intensity produces superior memory for the information with which it is associated, and they discuss the implications of this finding for psychological theorizing as well as survey measurement.

## Retrospective Reports of Behaviors

The seven chapters of the second part of this volume focus specifically on retrospective reports of behaviors. Peter Salovey and his colleagues (Chapter 6) investigate the recall of the frequency and intensity of physical pain, Barbara Means and her colleagues (Chapter 7) the recall of the number of cigarettes smoked, Alfred Smith and Jared Jobe (Chapter 8) the recall of dietary habits, and Nora Cate Schaeffer (Chapter 9) the recall of child support payments. All four of these chapters demonstrate considerable response error and identify a wide range of factors that contribute to erroneous reports of the specific behaviors. What is especially important is that some of these factors are more generally related to memory for a range of behaviors. In addition, the authors document ways to increase report accuracy.

Specifically, Salovey, Sieber, Jobe, and Willis in Chapter 6 find that severity of pain at the time of recall significantly impacts on reports of past pain. Many earlier studies have shown that respondents' reports of earlier behavior or attitudes are influenced by current behavior or attitudes. The authors also find that subjective measures of pain and behavior measures of the impact of pain on daily activities and behaviors related to pain are about equally reliable but are usually not highly correlated with each other.

Means, Swan, Jobe, and Esposito in Chapter 7 tried four different methods—the standard question versus episodic recall, availability, and decomposition strategies—to obtain recall information on the number of cigarettes smoked, and they validated the data by having respondents save their cigarette butts and record each time that they smoked. The authors conclude that methods most closely resembling standard fast-paced interviews result in the most errors. Interestingly, asking respondents to think aloud while they answer the questions improves reporting, probably because respondents engage in greater cognitive activity.

A similar research strategy is reported in Chapter 8 by Smith and Jobe. They compare long-term recall of food items to diary records for 2- and 4-week reference periods. Reports were collected either immediately after the reference period or 2, 4, or 6 weeks later. They find from both their experiments and simulations that respondents use a combination of episodic and generic memory, with significant errors attached to both kinds of memory. The quality of data deterio-

rates rapidly as the time between the end of the reference period and memory test increases.

Schaeffer, in Chapter 9, compares resident mothers' reports of receiving child support payments with information from court records. The results indicate that errors are larger when receipt of child support is irregular and when the amount paid is variable. There also appears to be confusion between child support and alimony payments. In general, her data indicate that task and event characteristics have a more pronounced impact on the validity of retrospective reports than do respondent characteristics or other variables.

In Chapter 10, Geeta Menon, relying on the distinction between episodic and semantic memory, uses think-aloud protocols and latency measures to predict the counting and/or estimation strategies that respondents are likely to use for frequent behaviors that vary by degree of regularity and similarity. As expected on theoretical grounds, estimation strategies are most likely to be used for regular behaviors of high similarity and are least likely to be used for irregular behaviors of low similarity. This pattern of findings reflects that high similarity renders it difficult to recall separate episodes, thus forcing respondents to use estimation rather than counting strategies. At the same time, the use of estimation strategies is facilitated by the regularity of the behavior, resulting in the observed additive impact of similarity and regularity. Measures of response latency support these conclusions. Hence, the regularity and similarity of the behavior under study strongly influence the strategy adopted by respondents.

Exploring methods for improving estimation strategies, Edward Blair and Kathleen Williamson (Chapter 11) investigate the impact of providing population information on the accuracy of respondents' estimates. They find little evidence, however, that respondents use such estimates when provided. They conclude that population information is ignored because it is pallid compared with autobiographical information.

Finally, Norbert Schwarz, Hans-J. Hippler, and Elisabeth Noelle-Neumann (Chapter 12) report studies on the impact of open and closed response formats on retrospective behavioral reports that demonstrate that the obtained reports are strongly affected by the format used. They then discuss the effects on reported behavior of the order in which response alternatives are presented. These effects are complex because they also depend on mode of presentation and on the context created by previous items. They conclude with a discussion of the influence of numeric response alternatives on reports of behavioral frequencies. Their data indicate that the frequency range of the response scale influences respondents' understanding of the question and provides a frame of reference that respondents use in making their estimates.

## Event Dating and Time Estimation

In providing retrospective reports, respondents are typically asked to date the recalled event or to restrict their report to some specified reference period. Un-

fortunately, the date of an event is unlikely to be part of its representation in memory, and this is more the case if the event is mundane and has a high frequency of occurrence. As a result, respondents often include events that did not actually occur within the reference period or exclude events that did. The two chapters in the third part of the volume address the cognitive processes involved in event dating and time estimation.

Norman Bradburn, Janellen Huttenlocher, and Larry Hedges (Chapter 13) present a theoretical model that accounts for the temporal displacement of events known in the survey literature as “forward” or “backward” telescoping, that is, recalling an event as having occurred later, or earlier, than it did. The model assumes that respondents are uncertain about dates, with the level of uncertainty increasing as the elapsed time between the event and report increases. Then, although the estimates are unbiased, the greater variance for more remote events causes telescoping. Their model accounts well for the available data and suggests a number of straightforward practices that are likely to reduce event-dating error.

In Chapter 14, John Skowronski and his colleagues explore how respondents use multiple memories related to an event to determine its date. Among their important findings are that respondents who are asked to provide an exact date use more strategies and are more accurate than respondents who are asked to report how long ago the event occurred. They also find that women are better than men at exact dating, and they speculate about the reasons. Whereas all previous studies of event dating have been restricted either to events that respondents experienced themselves or to public events, Skowronski et al. extend their study to include the dating of events in the life of closely related others. Although self-reports are more accurate, proxy reporters use the same methods for date estimation but have less information available from which to make their estimates.

## Comparisons of Self and Proxy Reports

The fourth part of the volume extends Skowronski et al.’s discussion of the differences between self and proxy reports. In Chapter 15, David Mingay and his colleagues report a study that required college students to keep a daily record of the frequency with which they engaged in a number of activities. Subsequently, they tested the respondents’ memory for their own activities over a period of 5 weeks. To explore the accuracy of proxy reports, they also asked respondents’ roommates to report on the target’s activities. In general, underreporting was more prevalent than overreporting for selves as well as proxies. Moreover, proxy reports were as accurate as self-reports if the direction of error and its magnitude were taken into account, resulting in similar frequency estimates independent of the source of the report. In contrast, analyses of the absolute error scores, ignoring the direction of the error, indicated less error for self than for proxy reports. The authors conclude from this pattern of findings that proxy reports may not introduce systematic bias if first-order point estimates of

behavioral frequency are of key interest but may undermine the validity of higher order multivariate analyses.

In general, one should expect that the accuracy of proxy reports increases with the proxy's knowledge about the target's behavior. If so, accuracy should be higher for activities in which the proxy and the target participated jointly, as well as for activities that the target frequently discussed with the proxy. Seymour Sudman and his colleagues (Chapter 16) tested these hypotheses in a study with couples living in the same household and received consistent but moderate support for the validity of the hypotheses.

Completing the discussion of self and proxy reports, Diane Holmberg and John Holmes (Chapter 17) present a theoretical model of relationship memories. Drawing on theories of reconstructive memory, they show how spouses build mental models of their partners, operationalized as trust in partner, how these models change over time, and how they are used in explaining, predicting, and recalling relevant behaviors. Most importantly, they observe that changes in the mental model of one's spouse are accompanied by biases in recall. Male respondents who developed a more negative view of their spouse over the course of 2 years exhibited a negativity bias in their memories of the early stages of their relationship, when they once reported that they were happy. As the authors note, "Their new, more negative mental models are reinforced again by memory biases: Not only are things bad now, but they have always been bad." Hence, one's current state may strongly influence one's recollections of the past, as was also seen by Salovey and his colleagues in their study of the recall of pain (Chapter 6). The absence of this result for females is explained by concerns about self-presentation.

## Memories of the Past and Judgments of Personal and Social Change

Whereas the earlier sections of this volume address the individual processes involved in providing retrospective reports, the final part extends the discussion of autobiographical memories in different directions. In Chapter 18, Leslie Clark, James Collins, and Susan Henry explore the impact of autobiographical memories on individuals' assessment of their current life. Whereas the Holmberg and Holmes chapter demonstrates how one's current state biases one's memories of the past, Clark and her colleagues show how memories of one's past bias one's evaluation of the present. Extending previous research by Strack, Schwarz, and Gschneidinger (1985), the findings in this chapter indicate that the impact of past life events on respondents' current judgments of well-being depends on the vividness of their memories. Pallid memories—that is, those that are not emotionally involving and are generated by asking for an abstract causal account of an event—facilitate the emergence of contrast effects. Vivid memories—that is, those that elicit happy or sad feelings and are generated by asking for specific details—facilitate the emergence of assimilation effects. Hence, when we recall a negative event from our past, the present may seem bright by

contrast or may be tainted by sad feelings, depending on the vividness and detail with which we elaborate the memory.

Turning from personal memories to the assessment of social change on the basis of retrospective reports, Karl-Heinz Reuband (Chapter 19) compares data from two representative cross-sectional surveys conducted in Germany in 1959 and 1989 that report respondent's memories of how much influence and autonomy they had when they were teenagers. There have generally been significant increases in teenager influence and autonomy over the 30-year period. The 1989 survey, however, found a strong contrast effect. Similar age cohorts reported substantially less influence and autonomy in that survey than in the 1959 survey. Reuband speculates that the possible reasons for this contrast effect may be selective perception or the ambiguity of the words used in the questions.

So far, all of the chapters addressed individual memories without paying attention to how the social and historical location of individuals may contribute to the formation of autobiographical recollections. In the final chapter of this volume (Chapter 20), Howard Schuman, Cheryl Rieger, and Vidas Gaidys turn to this issue by addressing *collective memories*, that is, memories of the historical past that are shared by groups of people. Asking citizens of the United States and of Lithuania to report important events or changes that occurred over the last 50 years, the authors observe pronounced differences as a function of respondents' cohort and social group membership. Specifically, respondents recall as important those events and changes that occurred during their own adolescence and early adulthood. In Lithuania, Lithuanians report the "rebirth of Lithuania" as the most important event, whereas Russians living in Lithuania mention perestroika. Their findings nicely illustrate how individual memories of public events are shaped by life course, social structure, and group identity—variables that have been largely neglected by psychological research into the nature of autobiographical memory.

## Conclusion

In combination, the contributions to this volume, and to the conference on which it is based, demonstrate that we have made considerable progress in understanding the cognitive dynamics of retrospective reports. Whereas much of the early work into the validity of retrospective reports was atheoretical, recent developments in cognitive psychology have resulted in several succinct theoretical models and a large number of theory-driven investigations. The emerging insights allow for a coherent conceptualization of the underlying cognitive processes and permit the identification of variables that determine the accuracy of the obtained reports. Although there is much that remains to be learned, the progress documented in the chapters in this volume justifies considerable optimism for the development of this exciting area of research in the near future.

**Part I**  
**Perspectives on**  
**Retrospective Reports**



# 1

## Autobiographical Memory and Survey Research

*William F. Brewer*

In this chapter, I review some of the topics in the study of autobiographical memory that appear to be the most relevant for the area of survey research: (a) types of autobiographical memory, (b) types of data used in memory research, (c) information in personal memory, (d) forgetting in autobiographical memory, (e) event characteristics related to memory, and (f) the accuracy of autobiographical memory. For each topic, I first discuss the basic research from the area of autobiographical memory and then make a few comments about the relevance of the research for the area of survey research. My comments about survey research should be taken with a grain of salt, since I have no expertise in this area.

### Types of Autobiographical Memory

#### Autobiographical Memory

Some years ago John Pani and I (Brewer & Pani, 1983) proposed a structural account of human memory organized in terms of frequency of exposure, types of input, and forms of representation. In a more recent paper (Brewer, 1986), I argued that autobiographical memory is best thought of as the subset of human memory related to the self, and I proposed a classification scheme for autobiographical memory that is organized in terms of frequency of experience and imaginal properties of the representation to give four types of autobiographical memory (see Table 1.1).

My classification scheme was designed to try bringing some order into the wide variety of types of memory that are called “autobiographical memory” (e.g., Rubin, 1986). One of the major differences among types of autobiographical memory seemed to me to depend on the frequency with which something had been experienced. The other major difference seemed to be related to the experiential nature of memory representation. Some forms of autobiographical memory appeared to be expressed in the form of phenomenally experienced

TABLE 1.1. Forms of Autobiographical Memory

Frequency of Experience	Imaginal Properties	
	Imaginal	Nonimaginal
Single instance (episodic)	Personal memory	Autobiographical fact
Repeated with variation (generic)	Generic personal memory	Self-schema

images; other forms appeared to be accompanied by very little experienced imagery. In the next section of this chapter, I describe the four forms of autobiographical memory that result from the application of my classification scheme.

### *Personal Memory*

The first type of autobiographical memory to be discussed is personal memory, the form of memory with imaginal properties that results from experiencing a single event. This type of memory is a recollection of a particular episode from an individual's past. For example, I recollect sitting on a couch near a staircase at Allerton House reviewing my notes for a few minutes before the opening day of the conference on which this volume is based. This form of memory frequently appears to be a "reliving" of the original phenomenal experience and is almost always accompanied by reports of visual imagery (cf. Brewer, 1988). This type of memory is experienced as occurring at a specific time and place. Note, however, that this does not mean that the individual can assign an absolute date to the event, just that it is experienced as having occurred at a particular moment in the individual's life. Another characteristic of this type of memory is that it is usually accompanied by the belief that it is a veridical record of the originally experienced episode. Personal memory is very similar to Tulving's (1972) original *definition* of "episodic memory," but it is much narrower in scope than the way in which the term "episodic memory" has come to be used in current memory research (see Brewer, 1986, pp. 32–34, for a detailed discussion).

### *Generic Personal Memory*

The next type of autobiographical memory to be discussed is generic personal memory, the form of memory with imaginal properties that results from experiencing a series of similar events. For example, I have a generic personal memory of going into the basement of Allerton House and purchasing a drink from the soft drink machine there. I did this several times during the conference and had done it at earlier conferences that I had attended at Allerton House. I have a generic visual image of the old soft drink machine against the right-hand wall, but I do not have a specific personal memory of a particular episode (e.g., I cannot recall what type of soft drink I purchased or if there was anyone there with me). Several researchers (Brewer, 1986, p. 45; M. Linton, 1982, p. 79) have

proposed that memories of this type are produced by a “dual-process theory of repetition” (Brewer, 1988, p. 76) in which the repetition of similar events leads to the development of generic memories at the expense of the individual memories that were repeated. The next two types of memory to be discussed are the nonphenomenal forms of autobiographical memory.

### *Autobiographical Fact*

An autobiographical fact is a nonimaginal memory of a single instance of information relating to the self. For example, I recall the fact that I gave a talk at another conference in Chicago the day before the Allerton conference. I can recall this fact independently of any personal memories about either conference.

### *Self-Schema*

The last type of autobiographical memory to be discussed is the nonimage form of memory that results from experiencing a series of similar events. It is assumed that with repeated exposure to information related to the self, individuals come to organize the information about themselves (e.g., food preferences, places lived, etc.) into a self-schema (Epstein, 1973; Markus, 1980) just as they organize nonself information into schemata (cf. Brewer & Nakamura, 1984; Rumelhart, 1980). For example, from my experience in writing chapters for conference volumes over the years, I have come to know that I almost never get my first draft in by the deadline. This is now part of my self-schema (just ask any of my conference editors!), and I can recall this information about myself without having to recall each individual conference volume.

## Survey Research

It would appear that respondents answer most of the questions used in survey research by means of one or more of the forms of autobiographical memory discussed above. It seems to me that an analysis of which type of memory respondents are using to answer a particular type of question would lead to deeper understanding of the question-answering process.

In fact, there has recently been some progress along this line in the area of survey research. It appears that many researchers in this area (e.g., Sudman & Bradburn, 1974) had been making the implicit assumption that respondents were using instance memory (the top row of Table 1.1) to answer most survey questions. There is considerable potential for problems here if the survey respondents are actually using generic memories when the survey researchers think that the respondents are using the “recall-and-count” strategy. In fact, Blair and Burton (1987) have carried out a study of respondents’ self-reported memory strategies that suggests that answers for questions relating to events with frequencies greater than 10 are almost always answered through the noninstance forms of autobiographical memory (i.e., the bottom row of Table 1.1). This study opens up a variety of important topics: Are there differences in accuracy for information

based on instance memory versus generic memory? (for example, "How many cups of coffee did you drink yesterday?" vs. "How many cups of coffee do you drink in an average day?"). What is the relationship between the individual instances and the resulting generic memory structures? The work reported in this volume by Smith and Jobe (chap. 8), Menon (chap. 10), and Schwarz, Hippler, and Noelle-Neumann (chap. 12) supply some initial answers to several of these questions. The Blair and Burton (1987) study and the later work reported in this volume provide good examples of the possible interactions between the areas of autobiographical memory and survey research. The theoretical account of the different forms of memory deriving from an analysis of basic issues in autobiographical memory provides support for the distinction that workers in survey research want to make, and their data provide some new empirical evidence about factors that lead subjects to use one memory system versus another.

## Types of Data Used in Memory Research

### Autobiographical Memory

The study of autobiographical memory seems to require a discussion of the phenomenal properties of memory, yet this type of data has rarely been used in the study of memory in contemporary cognitive psychology. By the phenomenal properties of memory I mean the introspectively experienced aspects of the recall process such as mental imagery. In several papers (Brewer, 1986; Brewer & Pani, 1983), I have argued that the data from phenomenal experience deserve the same treatment as any other form of data and must be accounted for in any complete scientific psychology (cf. Ericsson & Simon, 1980; Hilgard, 1980; Nat-soulas, 1970).

### Survey Research

Although the issue of what forms of data are appropriate for the study of memory is still somewhat controversial, it seems to me that the researchers in the area of survey research should feel free to use a wider range of questions. It may be a while before full use can be made of these sources of information, since the basic research in this area is still in the early stages. For example, the relationship between visual imagery and memory accuracy has not been worked out (cf. Brewer, 1988). Several researchers (Brewer, 1992; Neisser & Harsch, 1992) have suggested that strength of memory imagery may be more closely related to the rememberer's belief about the accuracy of a memory than to the actual accuracy of the memory, but there are little clear data on this issue.

Blair and Burton (1987) are quite cautious in introducing data about respondent memory strategies, yet this is an aspect of mental performance that subjects have clear access to and therefore it should not be controversial. From my perspective, this is the type of work needed to provide a basic research foundation for the more applied problems of survey research.

## Information in Personal Memory

### Autobiographical Memory

The initial research on the topic of autobiographical memory (e.g., M. Linton, 1975) tended to focus on time as the core information of interest in this area. It was never clear how the time information was to be represented, but some investigators in the laboratory study of memory suggested that memory representations might contain time tags (Anisfeld & Knapp, 1968; Yntema & Trask, 1963).

My recent analyses of personal memory (Brewer, 1986, 1988) suggest that this form of memory typically contains information about location, people, and the actions of people but very limited information about absolute time. It appears that when asked questions about the absolute date of an event, people usually do not retrieve particular episodic information but generate the answer from other types of information in memory or from generic memory. For example, I recall having a conversation at Allerton House with Norbert Schwarz. I recall that we were standing in a doorway and that we talked about the German academic system. However, if I try to recall what the time was, I can only generate an answer by recalling that it was dark and that we were near a table with drinks; therefore, it must have been early evening.

### Survey Research

The analysis of the contents of personal memory seems directly relevant to survey research questions about the time of events (see Bradburn, Huttenlocher, & Hedges, chap. 13, and Skowronski, Betz, Thompson, Walker, & Shannon, chap. 14, this volume). Thus, when someone is asked, "When is the last time you ate broccoli?" many of the responses will actually be derived from generic memory—for example, "Humm, I think I had broccoli last week, and we usually have it on Wednesdays so it must have been last Wednesday." On this issue, my conclusions derived from recent work in autobiographical memory are in nice agreement with recent work in survey research (e.g., Blair & Burton, 1987; Sudman & Schwarz, 1989). It is not clear to me what the full implications of this issue are for the problems of survey research, but it does suggest that survey researchers' might want to compare the information obtained from questions that probe specific episodic memory with the information obtained from questions that probe generic memory (see Smith and Jobe, chap. 8, this volume).

## Forgetting in Autobiographical Memory

### Autobiographical Memory

The initial studies of the recognition of self-selected autobiographical events (M. Linton, 1978; Wagenaar, 1986) appeared to show little or no forgetting of such

events for up to a year. However, in these studies the items were selected to be the most memorable events that occurred in a given day.

In a recent study with undergraduate subjects and randomly sampled events (Brewer, 1988), I found a considerably faster rate of forgetting. These subjects showed recognition scores of roughly 70% after 5 months. With a recall task, the rate of forgetting for randomly selected items was even greater. On the cued-recall task, the number of correct recalls of actions varied from 20% to 60% depending on the type of cue used.

## Survey Research

It is not clear to me that the recent findings on rate of forgetting for autobiographical memory are news for researchers in the area of survey research. However, the finding that there is a fairly substantial rate of forgetting for personal memories over time does suggest that researchers should be sensitive to the issue when asking questions about personal memories that could have occurred quite some time in the past (e.g., "What was the name of the last magazine that you purchased at a newsstand?").

## Event Characteristics Related to Memory

### Autobiographical Memory

A number of researchers have attempted to discover which characteristics of events lead to superior memory. M. Linton (1979), Wagenaar (1986), and R. T. White (1982) all found that low-frequency (unique) autobiographical events were remembered better than high-frequency events. Wagenaar (1986) and White (1982) found that events with high emotional content were remembered better than neutral events. In my study of randomly sampled autobiographical events (Brewer, 1988), subjects showed good recall and recognition for low-frequency events and for events occurring in low-frequency locations. In that study, memory for subject's thoughts was best predicted by high scores on a scale of how "exciting" the thoughts were. Experiment 2 of that study examined the role of different types of information as cues to recall. The ranking of cues in terms of retrieval value (from best to worst) was (a) the action being carried out, (b) the thoughts occurring during the event, (c) the location and the time of the event (given together), (d) the location (alone), and (e) the time of the event (date plus day of the week plus time of day).

### Survey Research

The data on the role of event characteristics in forgetting suggest that a researcher who wants information about particular episodic events needs to take event frequency into account. Questions about an instance of a class of repeated events will have to be restricted to a short temporal horizon, whereas questions

about fairly infrequent events can cover a longer temporal span. As a practical matter, this would seem to work out well for the survey researcher, since for rare events the survey researcher has to ask questions that span long time intervals in order to obtain enough data for analysis, whereas for frequent events a short time interval should produce enough instances for analysis (cf. Sudman & Schwarz, 1989).

## Accuracy of Autobiographical Memory

### Autobiographical Memory

Theories about the accuracy of autobiographical memory have covered a rather extreme range of positions.

#### *Copy Theories*

In many early accounts of personal memory, theorists adopted the view that personal memories are more or less exact copies of the original experience. Thus, Hobbes (1651/1952) gave an account of memory images in terms of decaying sensations. He appeared to view personal memories as copies of the original sensations that had faded over time. The copy approach to personal memory continued in philosophy over the centuries. For example, Furlong (1951) proposed that personal memories were a representation of an individual's "whole state of mind of the past occasion" (p. 83). The copy approach was also indirectly supported by the reports of M. Linton (1978) and Wagenaar (1986), which showed little forgetting of personal memories over periods of a year or more.

#### *Strong Reconstructive Theories*

Recently, the copy theories have come under criticism by a number of researchers (cf. Neisser, 1982b). The basic line of attack has been to point out examples of reconstructive memory processes in other forms of memory, give some anecdotal examples of errors in personal memory, and then reject the copy approach. A strong form of the reconstructive position can be seen in the work of Barclay (1986), who stated that "memories for most everyday life events are, therefore, transformed, distorted, or forgotten" (p. 89).

#### *A Partially Reconstructive Theory*

In my discussions of the issue of the accuracy of personal memory, I have taken an intermediate position. In Brewer (1986), I reviewed the literature on this topic and concluded that there was little strong evidence to support either the copy position or the strong reconstructive view. I noted that support for the copy theory may derive from the theorists' strong beliefs in the veridicality of their own personal memories and pointed out that this was not a very powerful form of support. I also argued that the anecdotal evidence collected by the proponents

of the strong reconstructive position was not very convincing. I outlined a partially reconstructive view in which recent (days to weeks) personal memories are reasonably accurate copies of an individual's original phenomenal experiences but that schema-based reconstructive processes (Brewer & Nakamura, 1984; Rumelhart, 1980) operate on these memories just as they do in other forms of memory.

My empirical study of randomly selected autobiographical events (Brewer, 1988) allowed a direct examination of the issue of the accuracy of personal memories. In Experiment 2 of that study, subjects carried a random alarm device and wrote down a description of identifying information (location, time) and the events that were occurring (actions, thoughts, etc.) when the alarm device produced a signal. At several intervals over the next months, these subjects were asked to recall the original events in a cued-recall task. It was then possible to compare the subjects' recalls with their original descriptions of the autobiographical events. The data showed that subjects made many retrieval errors (i.e., given a retrieval cue, they recalled the wrong event). However, when subjects retrieved an appropriate event, the recalls were usually accurate with a small number (1.5%) of clear reconstructive errors. I interpreted these findings as support for the partially reconstructive view of personal memory.

## Flashbulb Memory

Recently, the issue of memory accuracy has been discussed in the closely related area of flashbulb memory. In this section, I briefly review these issues as they have been debated in this area. The construct "flashbulb memory" was introduced by R. Brown and Kulik (1977) to account for memories of the circumstances of hearing about a highly surprising and consequential event (e.g., "Can you remember where you were when you heard about the assassination of John F. Kennedy?").

### *Copy Theories*

The original Brown and Kulik (1977) paper was clear copy theory. They stated that flashbulb memories were produced by a special biological "now print!" mechanism that was "very like a photograph that indiscriminately preserves the scene" (p. 74). In another place, they stated that at the moment of hearing about the Kennedy assassination, "much of the world stopped still to have its picture taken" (p. 80). Brown and Kulik carried out a study of memories of the Kennedy assassination and reported that after 13 years only 1% of their subjects appeared to show forgetting of the circumstances in which they had heard about the event.

### *Strong Reconstructive Theories*

Neisser (1982b) developed a strong theoretical attack of copy theories in the area of flashbulb memory. However, the evidence that he used in this critique was predominantly anecdotal, and later observations by C. P. Thompson and

Cowan (1986) show that Neisser's own flashbulb memory report for the bombing of Pearl Harbor was much more accurate than he thought.

In a recent paper, Neisser and Harsch (1992) describe an important new empirical investigation of flashbulb memory. Roughly 24 hours after the explosion of the space shuttle *Challenger*, they collected data on memory of hearing about the event, and they obtained recalls of the circumstances of hearing of the event 2½ and 3 years later. They report that only 7% of the subjects showed perfect recalls and that 25% of the recalls were completely inaccurate. They take these data to support a strong reconstructive view of flashbulb memory.

### *A Partially Reconstructive Theory*

In a recent paper (Brewer, 1992), I reviewed the literature on flashbulb memory and opted for a partially reconstructive approach in this area as I had earlier in the area of personal memory. In the review of the flashbulb literature, I argued that there was little evidence, other than the intuitions of the investigators, to support the copy theory. For example, Brown and Kulik's (1977) data on the Kennedy assassination were gathered 13 years after the original event and therefore can provide no empirical support for the accuracy of these recalls. On the other hand, I argued that Neisser and Harsch's (1992) data do not support the strong reconstructive position. I suggested that many of the "reconstructive errors" in their study are actually retrieval errors and not true reconstructive errors. For example, Neisser and Harsch give an example of a subject who heard that something had happened during a class and then went back to her room and watched TV accounts of the explosion. Several years later she recalled that she *first* found out about the explosion from watching TV. I suggested that in her later recall she may have simply retrieved the wrong event (the emotional moment of seeing the first pictures of the explosion) and that her description of watching TV and calling her mother that night may well be an accurate memory. However, there are other aspects of the Neisser and Harsch data that do seem to be good evidence for true reconstructive errors. Therefore, I concluded that a moderate reconstructive approach gives the best account of the accuracy of flashbulb memories.

### Survey Research

Memory accuracy is obviously a crucial issue for survey researchers. If the copy theory or the strong reconstructive theory were correct, then there would be strong consequences for survey research. If the copy theory were correct, then survey researchers could have complete confidence in information obtained from questions that tap personal memory. If the strong reconstructive theory were correct, then the data obtained from questions about personal memory would be highly questionable.

However, if the moderate reconstructive view were the correct account of personal memory, there should be little impact on the survey area, since this view is consistent with what I expect is the working assumption in the area—

answers to survey questions give reasonably accurate accounts of the events, but they also include a moderate percentage of erroneous information. The degree of reconstruction will, of course, depend on a variety of factors, some of which have been discussed earlier. For example, the data in Brewer (1988) suggest that responses to questions about the time of an event's occurrence are much more likely to be reconstructed than are qualitative reports about the nature of an event that did, in fact, occur.

## Conclusions

It seems clear that basic research in autobiographical memory and survey research have much to offer each other. It is comforting to find a convergence of data and theory in the two areas. Of the topics discussed above, the one where I think survey researchers have the most to contribute to cognitive psychology is the issue of how subjects use episodic and generic information to answer questions. This seems like an important practical issue for survey research, and investigations of this topic should contribute to the basic issues of knowledge representation in human memory. As the areas of autobiographical memory and survey research continue to mature in the future, there should be much useful theoretical and empirical interchange.

## 2

# The Validity of Retrospective Reports as a Function of the Directness of Retrieval Processes

*Douglas J. Herrmann*

Most of what is known about autobiographical memory pertains to the encoding of a memory (M. A. Conway, 1990; Neisser & Winograd, 1988; Rubin, 1982). Much less research has examined the retrieval process separately from issues of encoding (Herrmann, 1985; Reder, 1987a). In this chapter, I address retrieval processes hypothesized to underlie autobiographical remembering, especially as a function of the attributes of an episode—for example, the time and the location of the event (Wagenaar, 1986) and feelings regarding the event (Baddeley, Lewis, & Nimmo-Smith, 1978; Bower, 1967; Robinson, 1976; Underwood, 1969; S. H. White & Pillimer, 1979).

Historically, two kinds of retrieval processes have been proposed (see Aristotle, c. 300 B.C./1905). For some memories, such as material that is well memorized, attributes of the material items may be directly retrieved. For other memories, such as a prose passage, a story, or an event, the attributes of the memory may be inferred or constructed (Bartlett, 1932; Hasher & Griffin, 1978; Means & Loftus, 1991; Neisser, 1967; Reder, 1987a). Direct retrievals occur automatically, whereas inferential retrievals are the result of an effortful process (Hasher & Zacks, 1979). Of course, retrieval of a particular memory most likely involves both direct retrieval and inference, such that certain attributes are directly retrieved and others are inferred. For example, some research has indicated that temporal attributes of memory are less well recalled (Brewer, 1988, and chap. 1, this volume), suggesting that the time of events will be directly retrieved less often than other attributes (such as location). Also, it might be expected that abstract attributes, such as the purpose of an event, would be retrieved inferentially more so than would a sensory attribute of the event. However, since research has not focused on memory for attributes per se, which attributes will be directly retrieved and which will be inferentially retrieved is a matter for empirical investigation.

In this chapter, I review a series of investigations concerning the mode of retrieval (direct, inferential, or both) of the attributes of autobiographical memories. These investigations examine retrieval as a function of the kind of attri-

bute that originally defined an episode (e.g., the location, time, and emotional impact of the episode) and as a function of some of the major characteristics that hypothetically might affect which mode of retrieval is used. For example, the present research investigated whether the likelihood of direct and inferential retrieval varied as a function of several characteristics—for example, the length of the retention interval, familiarity, degree of firsthand experience, predictability of parts of an event, and aspects of the original emotional state elicited by an event. I have reported some of these investigations in more detail previously (Herrmann, 1985). The rest have never been reported or have only been presented at conferences. Some of these investigations were conducted by my students; others were conducted independently by colleagues intrigued by the issue of attribute retrieval. This chapter provides the first integrative review of past and recent research on attribute retrieval for autobiographical memories.

All of the experiments reviewed here employed a common procedure. Before discussing the specific aims of each experiment, I shall explain this procedure in detail.

## Procedures Common to the Experiments

In each experiment, the subjects recalled an episode of their lives—for example, attending their high school prom or skipping school and going swimming. Then they indicated what attributes of the episode they would be able to retrieve if asked to do so. Note that the procedure does not require the subject to recall the attributes of the episode, only to indicate if they could retrieve them.

The memories recalled by subjects were selected for recall by the subjects themselves, their relatives, or the experimenters. Episodes generated by the subjects were described by them with a two-word label, sufficient to remind them of the memory. Episodes generated by relatives or the experimenters were expressed by one or two sentences.

For each episode, the subject indicated whether he or she could remember each of many attributes associated with it and how this remembering occurred, that is, directly or inferentially. The attributes used in the present research were selected to be typical of descriptions of autobiographical episodes in general (see Herrmann, 1985; also see Baddeley et al., 1978; Bower, 1967; Robinson, 1976; Underwood, 1969; Wagenaar, 1986; S. H. White & Pillimer, 1979). Subjects were not asked to recall the precise nature of the attributes that they remembered; thus, a subject might indicate an ability to retrieve the location of a prior experience without telling the experimenter what the location was (e.g., Gaithersburg, Maryland). Reports of attribute retrieval were made on response forms called the Trace Attribute Inventory (TAI; Herrmann, 1985) shown in Exhibit 2.1.

When using the TAI form, a subject records a few words at the top that describe the episode recalled. Then the subject mentally attempts to retrieve the attributes of the episode. Next the subject indicates which attributes were re-

**EXHIBIT 2.1. The Trace Attribute Inventory Form Used by Subjects to Report Whether They Were Able to Directly Retrieve or Inferentially Retrieve the Attributes of an Event or That They Did Not Know What the Attribute Had Been or That the Attribute Did Not Apply to the Memory Being Reported On**

---

The Trace Attribute Inventory

Name \_\_\_\_\_ Age \_\_\_\_\_ Sex \_\_\_\_\_

Memory Label - Please put 2 or 3 words meaningful to you.

R = Recalled; I = Inferred; DNK = do not know; NA = not applicable.

	Memory Label			
	R	I	DNK	NA
<b>I. Location of event</b>				
- indoors				
· exact place				
· general vicinity				
· appearance of room				
- outdoors				
· exact place				
· general vicinity				
· weather				
<b>II. Time</b>				
- age				
- grade in school				
- year				
- season				
- month				
- day of week				
- period of time (morning, afternoon, evening)				
- time of day (approximate hour)				
<b>III. Social Context</b>				
- alone				
- people present				
· number of people				
· characteristics of people (sex, age)				
· relationship of people (family, friends, sweetheart, acquaintance)				
<b>IV. Feelings</b>				
- kind of feeling (neg. - positive)				
- strength of feeling (strong - weak)				
- attentiveness (attentive - relaxed)				
<b>V. Nature of Behavior</b>				
- activity level (active - passive)				
- involvement (participant - observer)				
- goal directed (goal - no goal)				
<b>VI. Nature of Event</b>				
- setting of event (accident, romantic, recreation, work, etc.)				
- unusualness of event (common - unusual)				
- kind of event (physical - emotional)				
<b>VII. Related Events</b>				
- prior events				
- subsequent events				
- consequent events				

---

trieved and how this retrieval occurred as a function of the seven attribute categories (e.g., time) and the attributes within each attribute category (e.g., for time: grade, year, season) that are listed in the left-hand column of the form. To the right of each attribute within an attribute category are four response blocks, identified at the top of the page as R, I, DNK, and NA (representing *recalled*, *inferred*, *do not know*, and *not applicable*, respectively). For each attribute of a particular episode, the subject checks one of the four blocks to indicate whether he or she directly (i.e., immediately) retrieved the memory attribute (the R block), inferred the memory attribute (the I block), did not retrieve the memory attributes (the DNK block), or felt that the attribute was not applicable for the memory (the NA block). An example of an episode that calls for an NA response would be one that occurred indoors only, making all outdoor attributes irrelevant.

Thus, a subject might have directly retrieved the general vicinity of place of an outdoor episode and inferentially retrieved the weather but failed to retrieve the exact place of the episode. Responses on the TAI form were reduced in two ways for the analyses. First, responses were summed over the attributes in each attribute category. Thus, each attribute category was represented by a total number of responses, including all direct retrievals, inferences, “do not knows,” and “not applicables.” These totals were converted into percentages of successful retrievals. Since the episode occurred outdoors, the indoor attributes were not applicable to this episodic memory.

Analyses were conducted on the percentages of direct and inferential retrievals as a function of the attribute categories and other salient aspects of episodic memories described below for each experiment. The percentage of direct retrievals was computed by dividing the total number of attributes retrieved directly by the total number of attributes that the subject might have retrieved ( $R/[R+I+DNK]$ ). Similarly, the percentage of inferred attributes was computed by dividing the total number of attributes retrieved inferentially by the total number of attributes that the subject might have retrieved ( $I/[R+I+DNK]$ ). Since the attributes identified as not applicable could not be retrieved, the NA attributes did not enter into the computation of the percentages. Additionally, the analyses did not address the DNK percentages, since they represent the difference between 100% and the sum of the R and I percentages.

Since autobiographical memory episodes are well known to be flawed, direct and inferred retrieval estimates generated from the TAI represent the attributes that subjects believe to be true of the memories that they recall.

## Experiment 1

The first experiment, conducted with Jonathan Schooler, used the Trace Attribute Inventory to investigate the retrieval modes by which people recollect memories that differ in recency. Six subjects recalled an experience that occurred “yesterday,” another that occurred “one week ago,” and yet another that oc-

curred on the "last day of classes," which involved an interval of 5 months (the last day of classes was in December and the experiment took place in late April). As described above, when subjects recalled their experiences, they wrote a two-word label to represent the experience. Subsequent to episode recall, the subjects transferred the labels onto TAI sheets and then completed the inventories.

The percentage of attributes retrieved for each memory was computed. The means of the percentages of direct and inferential retrievals are shown in Figure 2.1 as a function of the attribute categories, retrieval mode (direct or inferential), and retention interval. (The percentage of attributes that subjects classified that they did not know is not shown in the figure; however, it may be calculated by subtracting the sum of direct and inferential retrievals as shown in the figure.) An analysis of variance found overall main effects for attributes, retrieval mode, and retention interval. As shown in Figure 2.1, retrieval mode interacted with attribute category and especially with retention interval. Specifically, direct retrievals (in the left-hand panel of the figure) occurred far more often than inferential retrievals (right-hand panel), and the proportion of direct retrievals varied over attribute categories in a more pronounced manner and in a somewhat different pattern than did the proportion of inferential retrievals. Most importantly, the percentage of attributes that were directly retrieved increased as the retention interval became shorter. In contrast, the percentage of attributes that were inferred was unaffected by the length of the retention interval except that the attribute of feelings on the last day of classes was reported to be successfully retrieved with the aid of inference more than were the other attributes.

The results of Experiment 1, although manifesting some clear differences, were based on relatively short retention intervals, no longer than 7 months. Consequently, it seemed possible that a different pattern of retrieval strategies might be found if substantially longer retention intervals were used. For example, it might be expected that subjects did not use inferential retrieval for relatively recent events (such as those that occurred in the last year), since they might believe that inferential retrievals are appropriate for very long retention intervals (where direct retrievals are not very likely). Accordingly, a second experiment examined the retrieval processes reported for intervals much longer than those used in Experiment 1.

## Experiment 2

The second experiment used the Trace Attribute Inventory to investigate the retrieval modes by which students recollect memories from different periods of their lives (cf. Crowitz & Schiffman, 1974; M. Linton, 1986; Rubin, 1982). Sixty undergraduates from Hamilton College (in Clinton, New York) served in the experiment. Each subject began the experiment by recalling 10 experiences. For 20 subjects, it was required that these experiences occurred in their early childhood (defined as less than 8 years old). For another 20 subjects, the experiences were

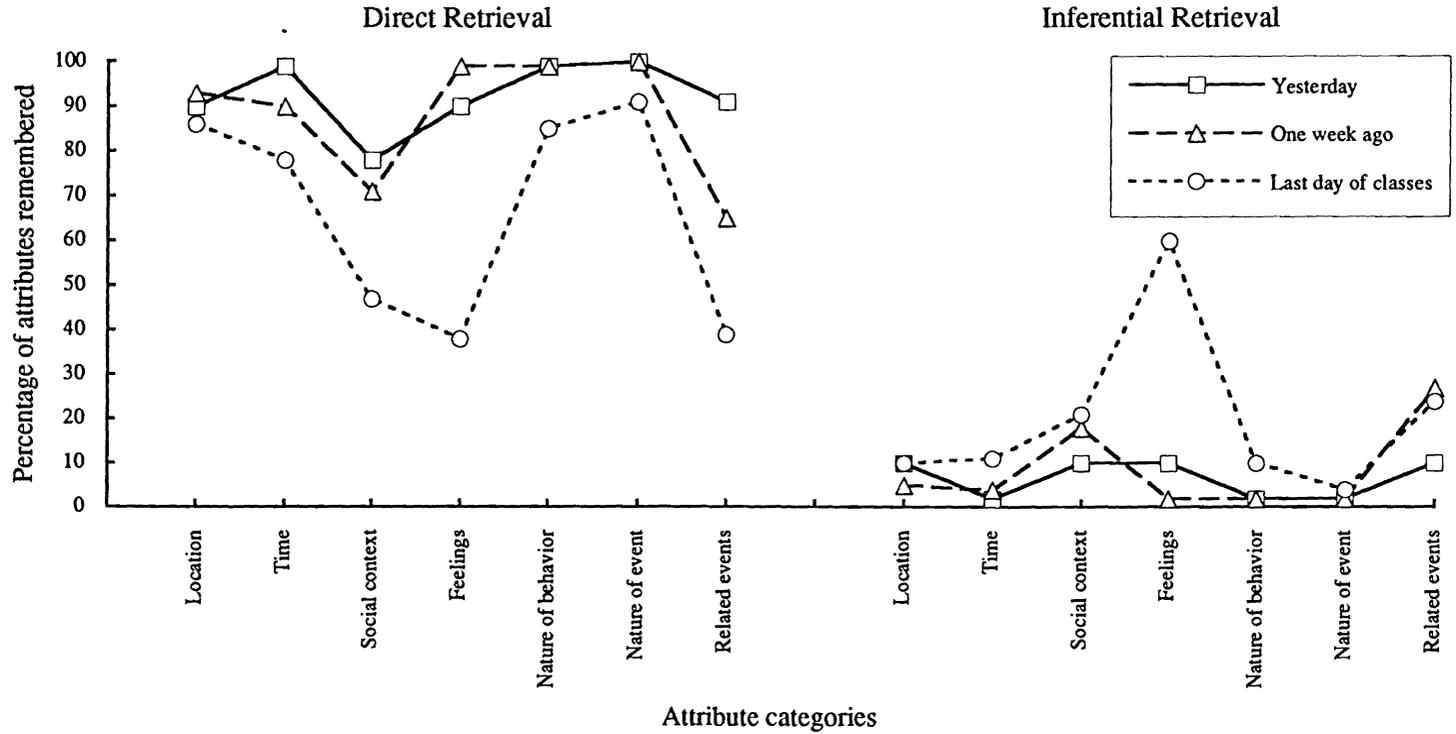


FIGURE 2.1. Mean Percentage of Attributes Remembered for Direct and Inferential Retrievals as a Function of Attribute Category and Length of Retention Interval: Experiment 1

to have occurred in “adolescence” (defined as from 8 to 15 years of age). And for yet another 20 subjects, the experiences were to have occurred in “adulthood” (defined as 15 years of age through the present).

The results for the study, reported in Herrmann (1985), conformed to a pattern similar to that found in Experiment 1 (as exhibited in Figure 2.1). Specifically, direct retrievals occurred far more often than did inferential retrievals, and the proportions of direct and inferential retrievals varied over attribute categories. Selected comparisons revealed that the direct retrievals increased with decreases in retention interval for two attributes (time and related events) and that the likelihood of inferred attributes increased with increased retention interval for the related events category.

Thus, as in Experiment 1, Experiment 2 indicated that autobiographical retrievals involve primarily direct retrievals and that such retrievals increase in use as the retention interval decreases. Thus, the length of the retention interval apparently does affect what kind of retrieval strategy is used. However, as is always the case when interpreting the effects of retention interval, the results of these two experiments may not be due simply to the passage of time. Instead, the effects of retention interval may be due to the influence of the length of the retention interval on some aspect of memory representation that in turn affects the use of retrieval strategies. An obvious variable mediating the effects of retention interval on retrieval strategy is familiarity, since it has long been known that familiarity decreases as retention interval increases (e.g., Rubin, 1982). Accordingly, a third experiment examined the retrieval processes reported for memories that varied in familiarity.

### Experiment 3

In Experiment 3 (reported in Herrmann, 1985), the parents or guardians of 15 undergraduates mailed me 12 brief descriptions of episodes from their children’s lives. These descriptions were to be one or two sentences long and refer to events that varied in familiarity to their children such that four episodes were to be very familiar, four moderately familiar, and four vaguely familiar. Parental descriptions dealt with a wide range of topics, for example, “the Christmas morning that Frank, who still believed in Santa Claus, got up too early and found us [the parents] putting presents under the tree.” On receipt of the episode descriptions, each was written on a Trace Attribute Inventory sheet; the sheets were then randomly ordered and given to the appropriate subjects.

The results of Experiment 3 also conformed to a pattern similar to that found in Experiment 1 (and Experiment 2 as well) except that direct retrieval decreased in usage (from approximately 80% to 30% overall) as familiarity decreased. In contrast, inferential retrieval remained constant across decreases in familiarity. These results are consistent with the familiarity explanation of retention interval effects in Experiments 1 and 2. All three experiments indicate that the use of direct retrieval is very common in the remembering of a wide variety

of autobiographical memories, more common than is normally held to be the case. Nevertheless, these experiments do not preclude the possibility that there may be conditions that elicit inferential retrieval more than occurred in these three experiments.

## Experiment 4

Having found substantial use of direct retrieval under different conditions, considerable effort was given to identifying some variables that would dispose subjects to make substantial use of inferential retrieval. Inference is hypothesized to facilitate retrieval because the use of reason to determine what attributes might be held by a memory also generates cues that are able to access genuine memory attributes. Thus, inferential retrieval ultimately leads to a direct retrieval of memory attributes that are normally not accessible to direct retrieval.

Inferential retrieval may not have been used much in the first three experiments because the memories that they addressed may not have possessed the characteristics that permit inference. Thus, the fact that a retention interval is long, or that the degree of familiarity is low, does not indicate that a particular memory contains the information that might aid inferential retrieval. Apparently, for people to see inference as useful, it must be apparent to them that they have information that will enable inference to guide retrieval to legitimate attributes of memories or, at least, that the memory might contain such information.

One characteristic of memories that subjects might find suggestive of the utility of inferential retrieval is the origin of memories. For example, it has been shown that people remember public events that they experienced (such as witnessing an event or at least witnessing news reports) better than they remember events that they simply read about (Larsen, 1988). Similarly, people sometimes forget an experience but later learn about it secondhand (i.e., from someone else, such as a parent or friend). Secondhand memories presumably are represented somewhat differently than firsthand memories (i.e., when a person retains the original memory of the episode; Herrmann, 1985). Thus, Experiment 4 investigated whether inferential retrieval might be more evident for secondhand memories than for firsthand memories, based on the hypothesis that knowing that a memory was secondhand might lead subjects to be confident that inference would be helpful.

In Experiment 4 (reported in Herrmann, 1985), the parents or guardians of 12 undergraduates completed a brief questionnaire that asked the parents or guardians to generate one- or two-sentence descriptions of episodes that they guessed would be remembered in a firsthand or secondhand way by their children. On receipt of the descriptions, each description was put on a Trace Attribute Inventory sheet, which was then completed by the subject.

The results for Experiment 4 conformed once again to the pattern found in the previous experiments. These data showed a tendency for certain attribute categories (feelings, nature of behavior, nature of event) to be inferred more for

secondhand memories than for firsthand memories. However, the proportions of inferential retrievals of firsthand and secondhand memories were small and not significantly different.

## Experiment 5

Although Experiment 4 did not provide a clear example of the kind of autobiographical memory that might foster inferential retrieval of attributes, the results did suggest that indeed there were kinds of memories that might make inferential retrieval useful. An examination of secondhand memories showed that they tended to be more stereotypical than firsthand memories.

Since secondhand memories appeared to elicit a somewhat more inferential retrieval strategy, it was hypothesized that memories that are more stereotypical of common everyday experience might elicit greater use of inferential retrieval than would memories that are idiosyncratic. Idiosyncratic memories do not follow a script (Rumelhart & Norman, 1983; Schank & Abelson, 1977); therefore, they may not provide as much of a basis for inference. For example, a person may have a better sense of what might be legitimately inferred about the event for idiosyncratic memories than for stereotypical memories. Alternatively, those episodes of our lives that follow a script (e.g., birthdays, religious and national holidays) make the use of inference a potentially effective retrieval strategy.

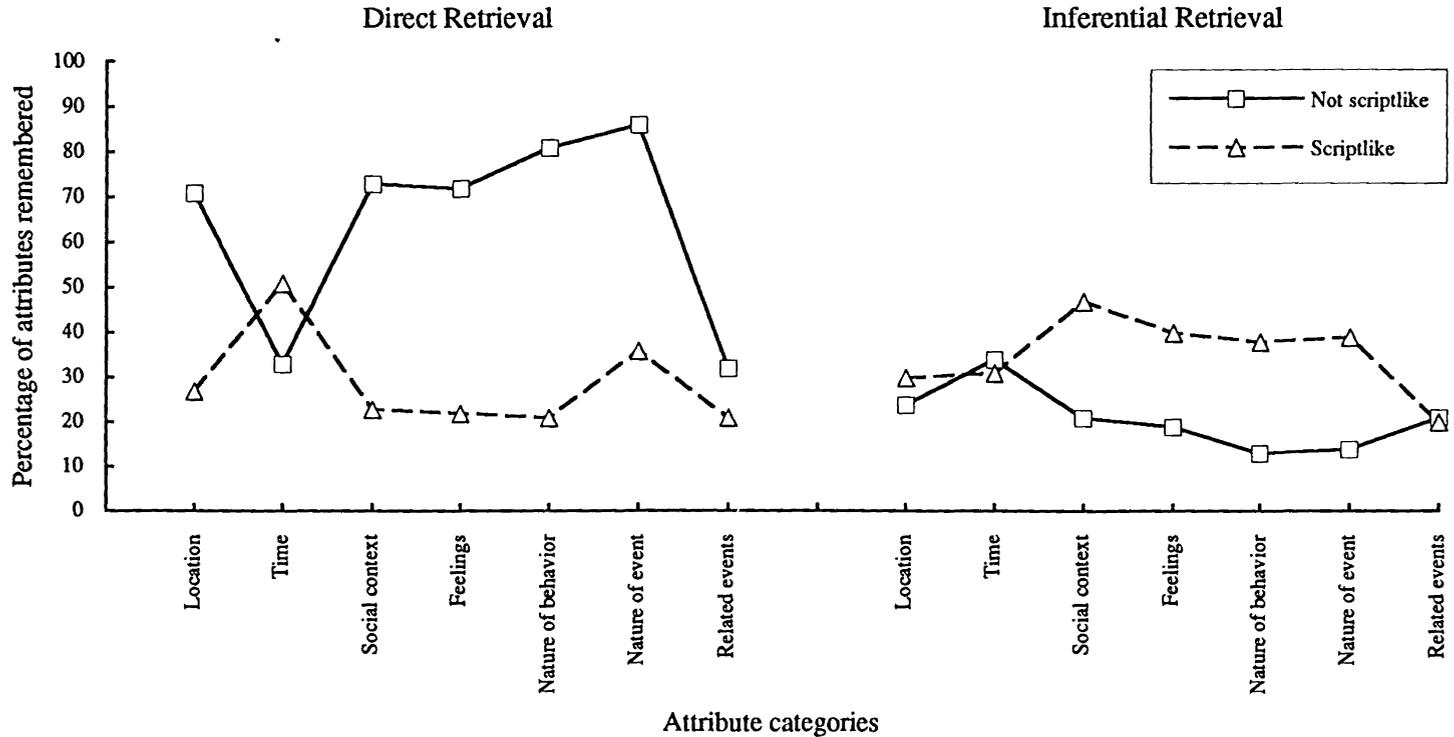
Experiment 5 investigated whether inferential retrieval might be substantially higher for memories of events that conform to a script than for memories of events that are not scriptlike. Eighteen undergraduates were asked to recall spontaneously four past experiences (presumed to be idiosyncratic) of their choosing and four experiences of a scriptlike nature (the subjects' 6th and 7th birthdays, Christmas or the first day of Hanukkah, and the fourth of July when they were 7 years old).

Figure 2.2 presents the results of Experiment 5. Inspection of the figure shows that inference was higher for scriptlike than for nonscriptlike episodes for most of the attribute categories. Indeed, the overall level of inferential retrieval was slightly higher than that for direct retrieval.

## Experiment 6

The results of Experiment 5 confirmed the hypothesis derived from Experiment 4 that the likelihood of using inferential retrieval increases when the event described by a memory is predictable or scriptlike. The scriptlike nature of memories might be described as pertaining to the cognitive basis of memories. It might be expected that the emotive nature of memories would also affect the likelihood of inferential retrieval.

One aspect of an event's emotionality that might be expected to influence the mode of retrieval is its affective valence, that is, whether the event recorded in memory tends to elicit a positive or negative emotion (Conway, 1990; Sorren-



Adapted from Herrmann (1985).

FIGURE 2.2. Mean Percentage of Attributes Remembered for Direct and Inferential Retrievals as a Function of Attribute Category and Whether the Memory Was One That Conforms To a Script or Was Not Scriptlike: Experiment 5

tino & Higgins, 1986). Following classical Freudian notions about the effect of emotionality on memory retrieval, it might be expected that positive events (being ego enhancing) would elicit more direct retrievals than would negative events (being ego deflating). Alternatively, if it is correct that people have difficulty accessing the attributes of negative events, then it might also be expected that inference would be used more for negative events than for positive events. An experiment that examined this hypothesis was reported by Neale, White, Herrmann, and Chaffin (1982).

This experiment asked 15 subjects to generate examples of five kinds of events that were presumed to vary in valence: two positive events (a successful event, a pleasurable event), two negative events (a failure, an embarrassing event), and a first date (which could be expected to be either positive or negative depending on the individual). The results for the study are shown in Figure 2.3. As in the previous experiments, direct retrievals occurred far more often than did inferential retrievals, and the proportions of direct and inferential retrievals varied over attribute categories. Generally, the positive events led to somewhat more direct retrievals than did the negative events. However, contrary to expectations, the percentage of inferences was unaffected by the affective valence of the kinds of events retrieved.

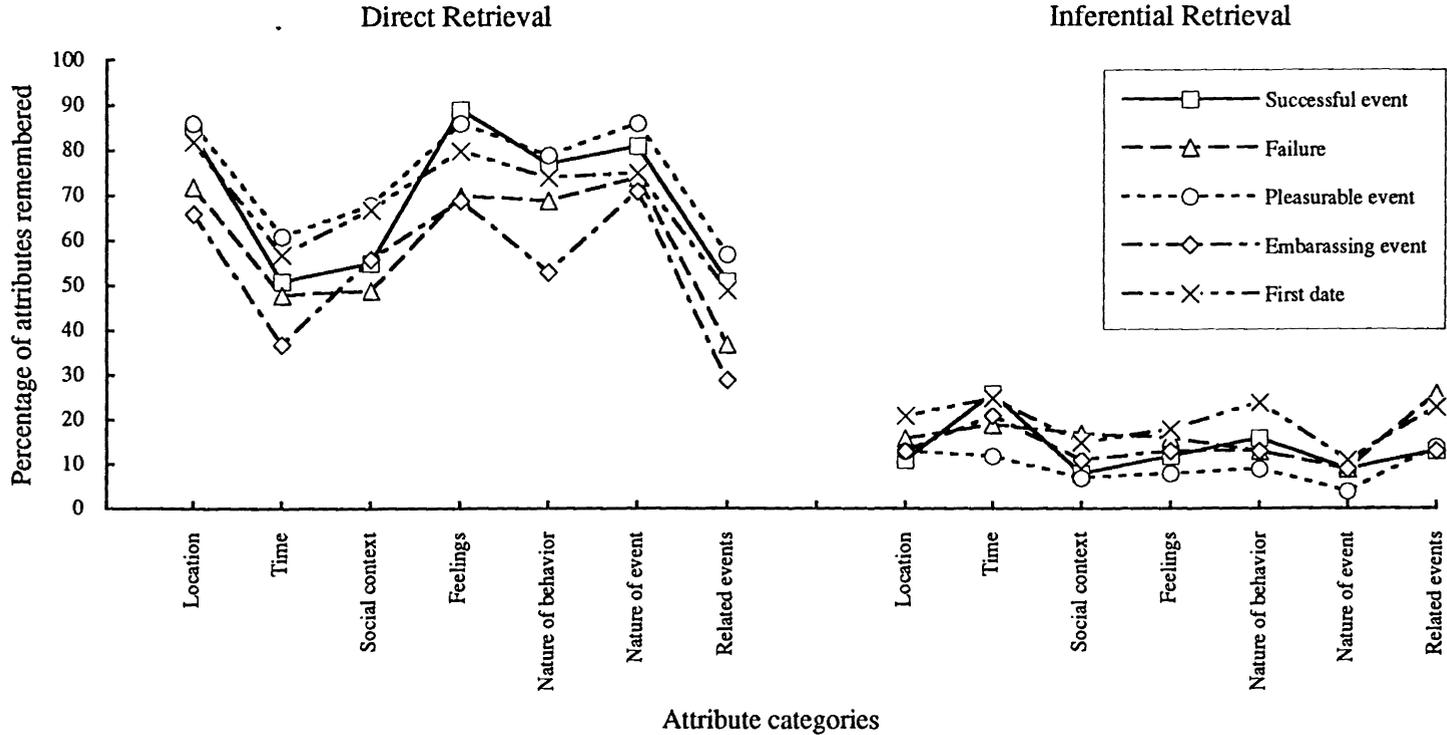
## Experiment 7

Experiment 6 failed to find greater use of inferential retrieval for negative vis-à-vis positive events. On reflection, it was easy to conceive of explanations for the interaction between the mode of retrieval and the kind of event (M. A. Conway, 1990; Sorrentino & Higgins, 1986). For example, if there was a general disinclination to infer attributes of memories, there might be a floor effect on the use of inference as a function of the event's valence. Consequently, it was hypothesized that it might be possible to demonstrate an effect of valence if the degree of such valence was manipulated more systematically than was done in Experiment 6.

Experiment 7 asked subjects to generate memories that represented either a moderate or extreme degree of positive or negative affect (C. M. Moore, Herrmann, & Vessot, 1989). Fifteen undergraduates from Hamilton College served in this experiment. The results for the study, presented in Figure 2.4, showed that the extremity of affect influenced direct retrieval. Moderately positive or negative events elicited more inferential retrievals than did extreme positive or negative events. Interestingly, this effect reversed itself in the direct retrievals; that is, direct retrieval was more prevalent for the events of extreme affect.

## Summary of Findings

The seven experiments that I have described provide a coherent pattern of results. When recalling episodic memories, subjects claimed that they retrieved



Adapted from Neale, White, Hermann, and Chaffin (1986).

FIGURE 2.3. Mean Percentage of Attributes Remembered for Direct and Inferential Retrievals as a Function of Attribute Category and the Positive and Negative Affect of an Event at the Time That the Memory Was Acquired: Experiment 6

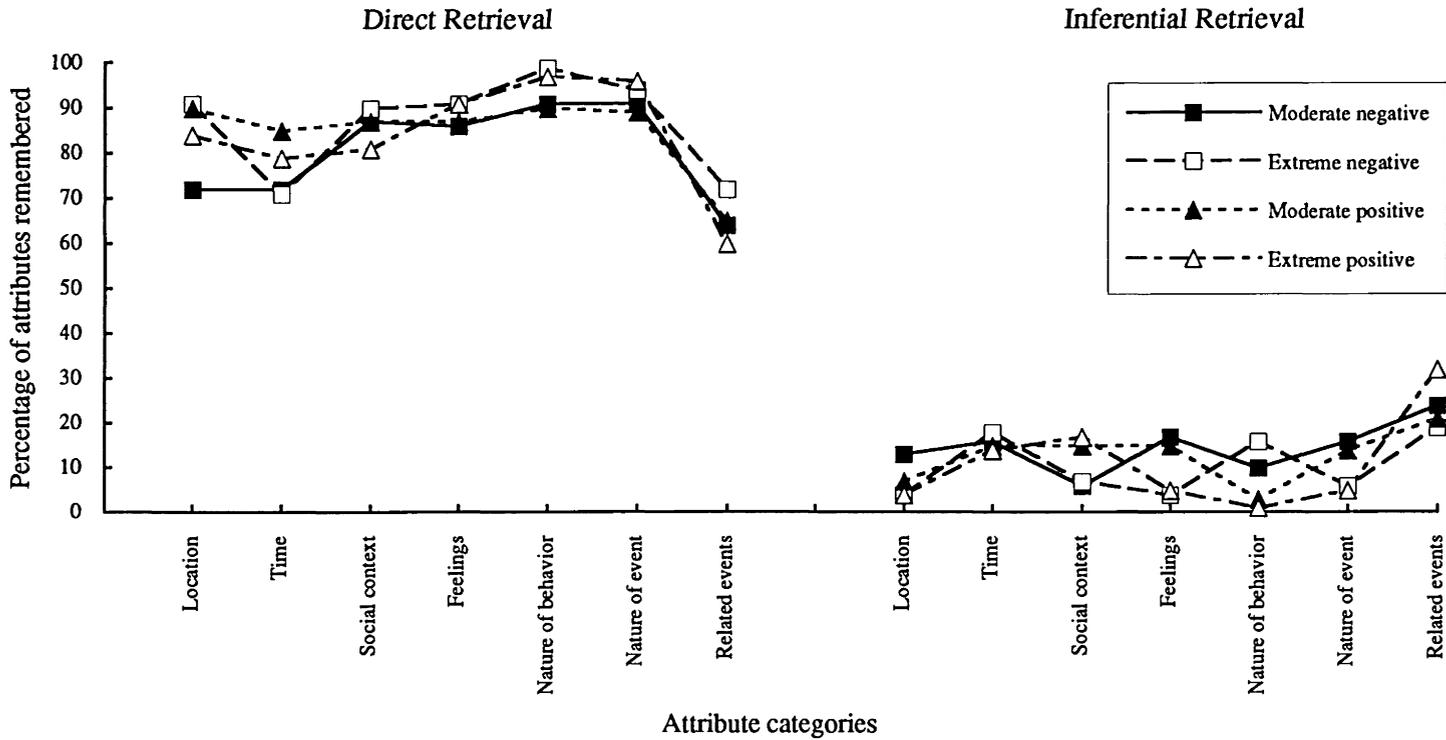


FIGURE 2.4. Mean Percentage of Attributes Remembered for Direct and Inferential Retrievals as a Function of Both the Valence and the Intensity of Affect Attribute of an Event at the Time That the Memory Was Acquired: Experiment 7

attributes of the experience primarily by direct retrieval. The likelihood of direct retrieval was affected by the length of the retention interval, the familiarity of the memories, and their emotionality. The results suggest that direct retrieval is more likely if the memory is spontaneously recalled and if the memory is experienced firsthand.

Inferential retrieval was reportedly used infrequently, except when the episode possessed scriptlike qualities. The results suggest that subjects are inclined to infer attributes more when the memory is scripted than when it is not scripted. Spontaneously generated memories are probably not going to elicit inference often, since such memories are likely to be ones that people know well. Inferential retrieval appears to be used more if the event was learned about second-hand, was ego deflating, and (at the time of occurrence) elicited moderate emotion.

Constructive processes are sometimes seen as adaptive in learning, since they avoid the unnecessary storage of needless detail. However, there may be equally good reasons for not wanting to employ constructive (i.e., inferential) processes at retrieval. Additional hypotheses remain about why a person might choose not to infer any more attributes than are necessary to satisfy the demands of the current memory task. Each inference of an attribute presumably generates information that may also be stored along with the original record of an event. Careless inference can be expected to generate false attributes that, when added to the event record, make successive retrievals less and less accurate. In the case of autobiographical memory, as studied here, the need to keep the record "clean" is especially important, since these memories are central to the self (Neisser, 1988).

The low frequency of inferential processes may also originate in the confidence that people have in direct and inferential retrieval. Although inferential retrieval was defined to subjects as one that eventually resulted in retrieval of the desired memory (as contrasted with guessing), the subjects in these experiments may have felt that inferential retrieval was more error prone than was direct retrieval. Hence, use of inferential retrieval was low because the subjects may not have trusted this process.

The relatively infrequent use of inference contrasts with research in several areas that recognizes constructive processes as common: for example, in the retention of prose (Bartlett, 1932; Neisser, 1967; Spiro, 1980; however, cf. Gauld & Stephenson, 1967), in the acquisition and retention of emotionally charged episodic memories (Greenwald, 1980), in the access of relatively old episodic memories from a well-known context (e.g., a person's early schooling; Williams & Hollan, 1981), and in the incidental retention of objects in a person's recent immediate environment (Brewer & Treyens, 1981). The reasons for the discrepancy between the findings reviewed here and other literature are not clear. One reason is that the literature has been primarily concerned with constructive processes in encoding, although the research paradigms used do not preclude constructive processes occurring after encoding and during retrieval (Alba & Hasher, 1983). The present research made no attempt to assess the

influence of constructive processes during encoding but focused exclusively on constructive processes during retrieval. Thus, constructive processes may be largely the result of encoding and not of retrieval.

On further consideration of past and present findings, the present findings may not seem so discrepant with other literature. For example, the present data show that conditions can be developed that foster more inferring. This finding is consistent with other literature in that the best demonstrations of constructive processes in such literature are those using scripted stimuli. In addition, the present results may supplement the literature by suggesting that previous demonstrations of constructive processes may have been due to constructive processes during retrieval as well as during encoding.

Although the present results support hypotheses about when inference occurs, they do not explain what an “inferential” process is. The directions used in the research reported here specified that an inferred memory was one that was (a) not direct, (b) retrieved as a result of a chain of reasoning, and (c) recognized by the subject as a genuine memory. It is not known how much the inferential process is truly based on a chain of reasoning, reconstruction, or free association (Barsalou, 1988; M. A. Conway, 1990; M. Linton, 1986; Norman & Bobrow, 1979; Reiser, Black, & Kalamarides, 1986; Rumelhart & Norman, 1983; Whitten & Leonard, 1981; Williams & Hollan, 1981; Wagenaar, 1986). The procedure also does not address all kinds of retrieval processes. For example, it does not include those retrievals that occur after an interval of unconscious thought (in a Gestalt manner, after a period of incubation). The present procedure also does not distinguish between different categories of “do not know” responses. For example, some people who cannot recall a memory can guess correctly some of the attributes of the memory (R. Brown & McNeill, 1966).

## Validity of Retrieval Reports

The research reviewed above provided few indications about the validity of the memories retrieved. As noted earlier, the variation in the likelihood of direct retrieval with the familiarity or firsthandness of memories produced by the parents or guardians of subjects suggests that the direct retrieval likelihood varies with the validity of remembered attributes. Also the research reviewed above revealed a remarkable consistency across most conditions in both direct and inferential retrieval (Figures 2.1–2.4), suggesting substantial agreement in the mode of retrieval reported for different memories involving similar characteristics.

An investigation into the reliability of retrieving the same memory on two separate occasions (separated by a 2-week period) found almost total agreement in the relative proportion of attributes retrieved on the two occasions (Panzer, 1981). Thus, although reliability does not ensure validity, the substantial reliability of retrieval reports at least makes it possible that these reports would turn out to be valid.

Establishing the validity of autobiographical memories is a difficult and important topic. One fundamental issue that was investigated early was whether the attributes on which subjects reported were indeed appropriate for the memories retrieved. Gelfand (1983) examined the intercorrelations among the attributes on the Trace Attribute Inventory and found that these intercorrelations were generally low and nonsignificant; thus, the attribute reports were independent as assumed for the different kinds of attributes. Brewer (1988; see also Brewer, chap. 1, this volume) obtained reports of events as they occurred and measured retrieval for these events later. It is notable that the attribute most likely to be retrieved least accurately in Brewer's research was the same as the attribute reported to be retrieved least often in the present research, that is, the temporal attribute.

Consequently, my colleagues and I have sought evidence of the construct validity of retrieval reports. Roger Chaffin and his students (Hedlund, Meckel, & Remz, 1981) hypothesized that if the attributes retrieved by individuals possess at least some validity, the attributes retrieved by married individuals should compare well with the attributes reported by their spouses. The results showed that the subjects were indeed able to predict above chance levels whether their spouses would recall particular attributes of certain events directly or by inference.

Three studies have examined issues of construct validity. Adler (1984) measured the latency of direct and inferential retrieval reports. As would be expected from the definitions of these two kinds of retrieval, the latency of direct retrieval was faster than that of inferential retrieval. Adler (1984) also found developmental differences in use of inference that are consistent with general findings showing that young children are less aware, and less inclined, to use strategic memory processes. Second graders rarely reported that they used inference; additionally, reported use of inference increased across the 2nd, 4th, and 7th grades. Finally, Gelfand (1983) has conducted many analyses of the intercorrelations of attribute retrieval. These analyses have indicated that the retrieval of any particular attribute category is essentially independent of the retrieval of any other attribute category.

Only one study has tested the validity of retrieval reports with respect to the attributes of the original experience as it is recorded in memory. Campell (1981), in collaboration with Chaffin, presented subjects with prose passages and subsequently asked them to recall both the attributes of these passages and the mode of retrieval for each attribute retrieved. This study also examined the "do not know" category in that it also allowed subjects to guess and to indicate when they had done so. Consistent with the interpretation of direct and inferential retrieval, the results showed that directly retrieved attributes tended to be more accurate than inferred attributes. Finally, Brewer (1988; see also Brewer, chap. 1, this volume) concluded that his research on autobiographical recall involved little in the way of constructive processes, and his data indicate that inferential processes at retrieval were infrequent.

## Conclusions

The research reviewed here should be recognized as having made just an initial foray into the topic of modes of retrieval. This research equated direct retrievals with immediacy and inferential retrievals with delay. It is important to acknowledge that direct retrievals may occur after a delay (as often occurs with the resolution of memory blocks and after the feeling that a desired memory is on the "tip of the tongue"). It is equally important to acknowledge that inferential retrievals may in some cases occur immediately (such as occurs when a particular retrieval task is attempted often, sharpening the reasoning processes that tend to lead to a successful retrieval). Additionally, with sufficient practice rapid inference may even turn out to be outside awareness. However, distinguishing empirically between fast direct retrievals and rapid unconscious inferences will likely be very difficult.

Regardless of the difficulty underlying the identification of retrieval mode, the present research indicates that people are reluctant to use inference. The relative infrequency of inference is especially interesting, since so much research has proposed that inference is a pervasive, common process (Baddeley et al., 1978; Bartlett, 1932; M. A. Conway, 1990; M. Linton, 1986; Norman & Bobrow, 1979; Rumelhart & Norman, 1983; Spiro, 1980; Wagenaar, 1986; Whitten & Leonard, 1981; Williams & Hollan, 1981; Williams & Santos-Williams, 1980; however, see Brewer, 1988, and chap. 1, this volume; Reiser et al., 1986).

It is unclear why, in the light of the present review, inferential processes have been reported as so common. Perhaps it is because inferential processes are more interesting than direct retrievals to psychologists who have investigated retrieval processes. Of course, regardless of how frequent inferential processes are, the nature of how they occur is worth investigating. Alternatively, it appears that the time has come to devote more effort to investigating the—genuinely common—direct retrievals.

*Acknowledgments.* I very much appreciate the work of several individuals that made possible the overall understanding of the issue of directness of retrieval; these include, among others, Jill Campell, Roger Chaffin, Cathy Dippo, Jim Esposito, Harold Gelfand, Joy Hedlund, Douglas Meckel, Cathleen Moore, Mark Palmisano, Paula Panzer, Cathy Puskar, Nancy Remz, Jonathan Schooler, Norbert Schwarz, Chris White, and Peggy Vessot.



### 3

## Accuracy, Truth, and Meaning in Autobiographical Reports

*Kurt W. Back*

Memory about ourselves should be the simplest memory possible: If we do not know about ourselves, what can we know about? If we cannot remember about ourselves, what can we remember? Thus, we can be sure about our own past, but we are less certain about anything else.

In contrast with this naively obvious view, the cognitive psychologist's input-output model of the human organism looks at the person as a transmitter and a storer of impulses. He or she can reproduce these impulses and what they represent, but it is extremely difficult to process data about the mechanism itself.

Both models are, of course, exaggerated extremes, but they express the ambiguous nature of autobiographical data. Survey research, perhaps more than any other field of human action, depends on answers to questions, that is, on a human input-output model consisting of perception, memory, and response. These functions are not purely cognitive, but what is perceived, retained, and transmitted depends on a dynamic framework. These functions are necessarily selective: Within a number of units, which are at each step too large to be considered, a choice must be made, and this will depend on intentions or motivations, that is, on extra-cognitive factors.

Both the commonsense and the psychological perspectives have some validity. Because of these conflicting aspects of autobiography, the collection of data in this field faces a special set of difficulties. On the one hand, we have the cognitive mechanisms that govern the production of specific autobiographical units; on the other hand, we have the direct experience of identity that puts these different units into context, and specifically a context that constitutes a direct experience of the subject. The methods of experimental psychology include a specific setting that presents specific questions and demands answers. A double context must be considered: the context of the situation in which the memories are produced and the context of the autobiographical facts in a person's whole life. Survey research uses these techniques in a more realistic social setting than the laboratory, but it still tries to abstract each item of information from its context.

Thus, the use of survey research as a principal tool for social research has several implications for the nature of the data to be collected that reflect on the meaning and importance of the accuracy of the data. Two important ways—the role of the interviewee and the autobiographical content—distinguish the collection of biographical data in survey interviews from that in other situations. One can use the two ways to identify these different situations and to construct a typology.

## A Typology of Autobiographical Occasions

### Informants and Respondents

In order to define the role of the interview, it will be useful to return to Shannon and Weaver's (1948) familiar model of the communication process to sketch in the position of autobiographical data in data collection.

The model of the communications process describes how data in one place (or in possession of one person) are transmitted to another place (or into the possession of another person). This process consists of several steps:

1. The original data (the source)
2. The translation of the data into a form that can be transmitted (i.e., the coder)
3. The transmission of the coded material
4. The translating of the coded material into personally usable form (i.e., the decoder)
5. The assembly to reconstruct the original data at the source

This model can be used to represent the interviewing process. Here our attention is concentrated on the first two steps: the source and the coder.

People as sources are the measurable elements of a society; they may be studied as representatives of a social grouping. Thus, a person is not only the unique source of information about individuals, but through the medium of individuals, a person also becomes the main source of data about society. This is the case in the survey interview. The interviewee is told, "We are only interested in aggregates and your own answer will not be identifiable in any reports."

The person of the interviewee has, moreover, an ambiguous position as a source of data. The interviewee is either seen as an informant, a transmitting device for data outside the person (Figure 3.1a), or as a respondent, who can give data about his or her own psychological data, especially about attitudes, opinions, or other personal perspectives (Figure 3.1b). In autobiographical interviews, this precarious position is underlined by the fact that the process of autobiographical memory itself can be modeled in the same way and the steps in information transmission correspond to perception, storage, and retrieval. In turn, retrieval, the last step in the memory model, becomes the first step in the interview.

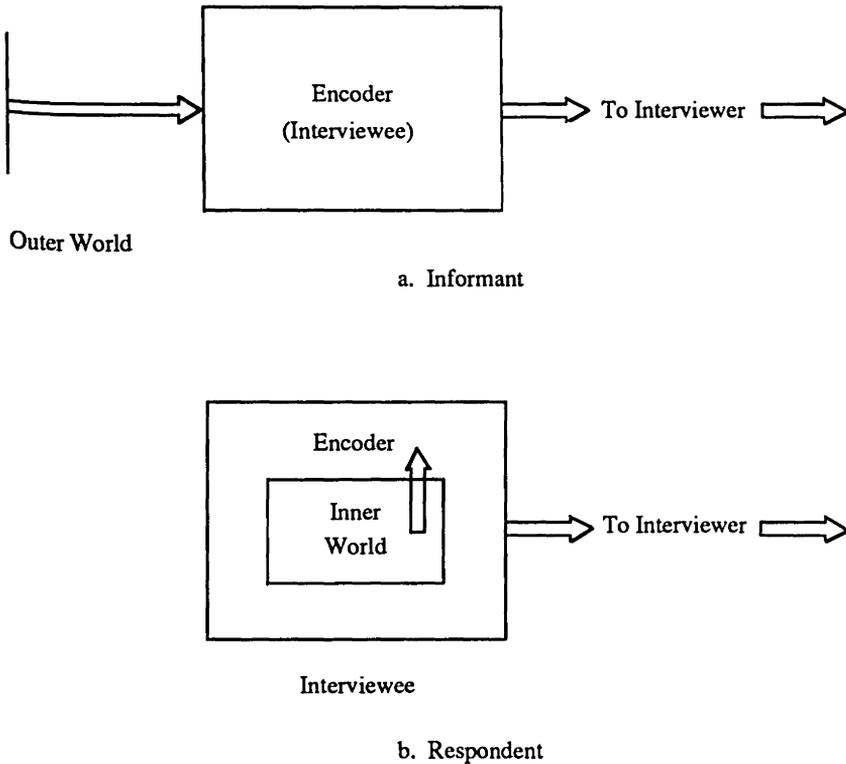


FIGURE 3.1. Informants and Respondents

Because these multiple functions are placed in one person, we must elaborate the model. It is possible that the interest in the gathering of data is concentrated in interviewees and not in the data that they can provide about outside events. Here even the distortion of transmission can be the object of inquiry; the extreme example here is a projective test, such as the Thematic Apperception Test, where we are not interested in the description of the picture but in peculiarities of the respondent's approach. Under these conditions, the interviewee gives internal data, which cannot be said to be either right or wrong. Even if the answer is intentionally a lie, it still represents a particular way of answering; the fact that the respondent chose to answer in this way represents an attitude, perhaps of fear, hostility, or a misplaced sense of humor. The work of the researcher is then to find a design and analysis program that can separate the kind of data needed for the problem of the particular study. The respondent is not only the encoder but also the source of the data. A person who combines the two stages of the communication process is called "a respondent." The data obtained from an informant could presumably have been obtained from another source, whereas the respondent is the unique source for this set of data.

The collection of autobiographical data represents a peculiar mixture of these two roles. Here the interviewees act as informants reporting presumably objective data, but the data concern the individuals themselves. The methodological problems of collecting these data combine those of cognitive distortion (the lens may not be a perfect transmitter) and distortions due to the personal meaning of the facts (the data can be arranged in different ways.) Thus, the "world" in Figure 3.1 is internal but can either be viewed like an external event or as uniquely individual. The interviewee is here at the same time the transparent lens through which the fact is viewed and the locus of these facts; he or she can be either an informant about his or her own experiences or a pure respondent.

### Input and Output

If one wants to construct a model of society from the data that one collects, then one is interested in those types of experiences that are common to everybody in society; the distribution of categories in these experiences becomes all important. These are mainly the data that describe the input of society on the individual, the ways in which social norms affect the individual life course. These will describe the social structures that affect the individual, such as parents' background, education, marriage and offspring, occupation, and income. Many other aspects of an individual biography cannot be captured in this way, and these are often the subjectively most important events; especially individually distinctive, but rarely occurring, incidents cannot be captured in this way. These form the social output, the influence of the individual on society.

When social output is the primary consideration, the interviewee is used as a respondent about the individual life and is selected for the importance of the distinctive events. The normative events are seen more as a framework for the other activities. The influences of the individual on society, that is, the social output, will be the center of attention. Thus, the biographical data will be the spontaneous activities, not determined by the social order, such as ideas, beliefs, traumatic or conversion experiences, conflicts, or creative acts.

### A Typology of Autobiography

The two classifications by role of interviewee and by content of autobiographical data can be used to organize the variety of interviewing situations and to characterize the survey interview. This is outlined in Figure 3.2.

#### *Survey Interview*

Survey researchers use interviewees purely as informants about society: They collect data about common events and use them for describing social units. This combination defines one type of biographical interviewing situation.

Here one collects an array of facts from an ensemble of alternatives that have been previously defined, through either conceptual analysis or previous re-

Content	Role of Interviewee	
	Informant	Respondent
Social Input	Survey Interview	Psychotherapeutic Interview
Social Output	Heroic Myth	Literary Biography

FIGURE 3.2. Types of Autobiographical Situations

search; the task in data collection is then to question precisely in order to select the desired facts.

### *Literary Biography*

An opposite perspective emphasizes the meaning of the separate facts in the whole course of life; the same demographic data, such as migration or selection of an occupation, may be embedded differently in the interviewee's life. This aim leads necessarily to an emphasis on the uniqueness of the interviewee. This uniqueness implies concentration on the whole life, where the individual incidents are put into context. Accuracy here does not mean the exact fidelity to individual events but the reproduction of a whole life course, where all of the events mentioned have a comprehensible place. A good example of this type is the collection of data for a literary biography: It uses the interviewee as respondent and concentrates on the social output, the unique contributions.

Literary life-course descriptions deny the existence of isolated facts but take the whole understanding of the self as their starting point; the different events in a person's life receive their meaning only within the framework of the understanding of this life experience. Even this is not necessarily accomplished by the respondent alone but evolves from a joint effort of respondent and interviewer (or writer and analyst). This generation of points of interpretation, of facts, and of language and the progress from particular phrases to an integrated picture can lead in the extreme to such a multiplication of relationships that analysis and understanding become practically impossible (Denzin, 1989).

The literary biography shows up in the opposite corner from the survey interview and provides a contrasting approach. It is no wonder that these two approaches, different in their attention to specific data and general aims, have little to say to each other about their methods. Hard, isolated facts and textual construction seem incompatible; proponents of each perspective stay within their own circle, ignoring or denigrating the other point of view as irrelevant for their purpose.

### *Psychotherapeutic Interview*

Another type of interview combines the normative effects on the person—the social input—with particular attention to the individual. A representative of this type is the psychotherapeutic interview. Here the kind of data that are collected do not represent common experience, but the occasion is designed to give the unusual, idiosyncratic aspects. However, the data are collected purely to understand the individual and not for their impact on society.

The important context here is the interviewing situation and not the narrative of the whole life. The interpretation by the skilled therapist makes different incidents important for the current purpose, but this procedure does not make for comprehensive narrative. Steven Marcus (1984) has shown this effect in comparing two of Freud's case studies, that of Dora and that of the Ratman. Dora was Freud's great failure in therapy but helped Freud in his own development of theory. The study is often quoted, in great part because it is an impressive story. It could easily have been a successful novel in its time. The Ratman is an account of a successful therapy, but it is a collection of fragments from the interviews, giving a series of incidents but no connected biographical narrative.

### *Heroic Myth*

The opposite combination represents a typified, almost standardized representation of personal data that tries to approach the ideal of important roles within the society. This is the transformation of a life experience into a heroic myth; aspiring heroes at times created them almost consciously. It is a way of accentuating social norms in biographical data that is different from producing means from individual data. Here the retrospective data are transformed into the cultural stereotype of the hero. As some scholars, for instance, Raglan (1936) and Propp (1968), have indicated, this may even be a general human ideal. Here the immediate audience consists of the current followers and competitors who are being encouraged to accept the legitimate position of the hero. The intended audience may be the future generations of society, whose allegiance to their social group will be reinforced through the myth of the pertinent culture hero. Here the interviewing context is practically suppressed: The story, as it were, creates itself, but each incident is only meaningful in its total context.

## Accuracy

The four types show a range of variation within the basic communications model: They differ in their definition of what is transmitted, the source of the data (whether they come from an informant or a respondent), the definition of signal and noise, and the intended audience, the recipient of the information in the communications process.

## Facts and Context

The principal distinction in regard to accuracy is the relation of particular items to the context; this context may be cross sectional, referring to the interviewing situation, or it may be longitudinal, referring to the life history.

Thus, the four types of situations represent ways in which accuracy can be defined in autobiographical reports. The interest in truth or accuracy is not necessarily the criterion for the quality of autobiographical reports. In recalling the main motives for producing these reports, we can see that some, like entertainment, can easily sacrifice truth for style, and that others, like self-understanding, have a different definition of truth than fidelity in transmission.

The meaning of accuracy changes will be defined here by the emphasis put on the fidelity of transmission of units against the consideration of the context, in the interview or in the temporal sequence. One might say that this is the difference between facts and meaning.

## The Role of the Audience

We may note that the distortions are not always due only to cognitive factors. Very often memories are requested to help the interviewer in some task. Common examples are criminal investigations, legal proceedings, or intelligence operations (Back, 1956). In these cases, motivation is intentionally introduced by the interviewer, sometimes in a spectacular manner through physical and mental pressures. These pressures induce compliance with the interviewer's intentions as far as the respondent can guess them. By the same token, they can also lead to wild distortions. Here I shall concentrate on purely cognitive factors, assuming an ideal situation in which the interest of interviewer and interviewee are united in producing accurate data, as far as possible.

Even with this proviso, a variety of self-generated motives are possible. Among the objectives that are listed by Ross and Buehler (chap. 4, this volume) for recalling personal memories are "to entertain"; "to sway opinions"; "to establish, maintain, and justify social relationships"; and "to formulate and justify current decisions and judgments." Research situations form an unusual addition to this list, and other reasons could be added. These objectives certainly influence accuracy of transmission. They may be difficult to ascertain, but an indirect measure is given by the social situation and especially the audience to which the story of these memories is addressed. The audience gives the context, sometimes for individual reports but often for the setting of the whole life. The audience is therefore an important part of autobiographical cognition itself, not only of the reporting of these memories. The recent concern with reader response in the theory of literary criticism treats only the latter in mass communication (Tompkins, 1980); 30 years ago social psychologists proposed a circular effect in which the audience influences the respondents' cognitive processes (Bauer, 1964; Schramm and Danielson, 1958).

The study of autobiographical memory is therefore aided by considering the

audience, which represents the social environment during the retrieval process and becomes an integral part of the investigation of autobiographical memory. The general audience is determined by the culture of the whole society; in different societies, the audience is socialized to expect different autobiographical communications and has its own definitions of truth and its importance.

The definition of accuracy—what is signal and what is noise—in autobiographical and biographical data is strongly influenced by general cultural conditions. We might assume that it is natural for people to want to talk about themselves; however, the point in history at which individuals could distinguish their selves from the rest of the world remains a philosophical and historical problem. Attention to accuracy of autobiographical data has an even more restricted past.

## The Four Types of Autobiographical Situations and Accuracy

### *Survey Interview*

In the survey interview, accuracy is defined as fidelity of transmission of individual facts, and attention to context is seen as noise or interviewer error. Survey researchers want to obtain the truth about life in a society by constructing typical life courses based on statistical manipulation of individual data. In this way, they construct a model representing society but not any particular individual. The traditional model of survey research and of the experiment sees the interviewees as pure informants about themselves. The data to be obtained are supposed to be objectively given (i.e., existing outside the interview), and thus there is a standard to which the responses in the interview can be compared. The general facts are defined, and these definitions can be presented to the informant. The use of language is to be refined as a pure medium of communication, and the text of the interview is not seen as problematic. In fact, research shows that for this purpose an impersonal situation is often superior to a personal interview (Sudman & Bradburn, 1974). The experiences in a retrospective interview will be evoked and reported in a standardized fashion and can be reliably coded and tabulated.

This description of hard science may look overly simple, but it tries to accentuate the strengths of this approach, as well as the weaknesses; these can be exposed by questioning the assumptions of survey research, such as the existence of an objective truth, and by looking for alternative examples. However, both the oversimplification and the possible attack on the assumptions decrease in importance if we look at the functions of the interview. The data are autobiographical but are not designed to obtain individual biographies. On the contrary, they are designed to obtain data about a whole population, reporting on standard lives, with applicable variations, as shown in dispersion measures. Small variations from these standards, due to idiosyncratic experiences, are errors in this research and considered negligible (or statistically insignificant) compared with the main trends that are established. Well-established mathematical and statistical methods determine how the autobiographical data are treated. One can argue about specific procedures or even question if they make any sense, but once one accepts this kind of method, the path to conclusions is determined.

It is unlikely that there is any individual whose life experiences correspond exactly to the data derived from surveys. Thus, even the accumulation of the data result in a hypothetical construction; this is, of course, what the technique is designed to do. It is an extremely useful convention for the combination of isolated facts; it is not designed to give meaning.

This analysis indicates that other definitions of accuracy and truth are possible. The definition of accuracy in the survey interview is determined by the need for objectively true data about specific events. One can discern here an implicit assumption that the respondent has the same aims but also that the respondent does not care about the audience. The immediate audience is the interviewer, and the best interviewer remains practically anonymous and is hardly remembered by the interviewee. The ultimate audience consists mainly of professionals who have little concern with the individual. Although the respondent is encouraged to maintain this perspective, it may be difficult to do so consistently because this perspective is different from the one that is usually taken in talking about oneself.

### *Literary Biography*

The opposite extreme, recounting a complete autobiography, is also a relatively recent phenomenon and shows accuracy in a new light. In the collection of data for literary biographies, the unity of the person is often taken as the important starting point. Biographers, such as Leon Edel (1978) in a "manifesto" on biography, have considered biographical work as fiction, but fiction that does not contradict any known facts, that is, "novel under oath." Similarly, Lytton Strachey (1918), often called the founder of modern biography, insisted that unessential facts should be omitted from good biographies. Of course, the biographer defines what is unessential, and this definition is based on the guiding theory of the biography.

The interviewing situation and the meaning of the whole life course become here legitimate parts of the communication process: They are the signal and not the noise. Biographers add these contexts to the transmission of individual facts. The definition of the dividing line between the two is left to the biographer. This personal contribution reminds us that these definitions do not come from a mathematical algorithm but depend on the biographer, the historical moment of the writing of the biography, and the particular theory of the analysis. Here the definition and the importance of accuracy have changed. Accuracy is not necessarily equal to truth.

Although interest in life stories seems to have an ancient heritage, it is questionable whether for most of this time biographical or autobiographical interest in our sense of the word was understood or even existed. Even Plutarch, considered to be the father of biography, was mainly concerned in presenting his characters as representing types of political heroes and villains, almost as the best examples of Platonic ideals; that is, biography for him was mainly a forceful way to make a moral point. Similarly, St. Augustine, whose autobiography is considered to be first in this field, wrote it for moral and religious purposes.

Writing autobiographies for their own sake needed a recognition of the uniqueness of the individual and the importance of documentation. Rousseau stands out as the example of writing an autobiography to tell the truth about himself—*The Confessions*. During his time, the 18th century biography and autobiography started to become a major field of literature and to find their audience.

For a long time, accuracy of life stories seems to have been a minor concern. Even 19th century biography was written more for edification than for accuracy. The motives for describing one's life did not lead to claims for the accuracy of particular incidents. The stories that were put down were supposed to be instructive or openly tendentious (Cockshut, 1974). In each case, some purpose was served, sometimes leavened with entertaining decorations. These purposes undermined attention to accuracy of details: The aim was to gain the attention of the audience and to influence it.

The purveyors of this fanciful material would not have thought of themselves as liars, but they might have said that a different kind of truth—about people's characters or about issues of morality—was served (Barclay, 1986). In fact, interest in pure transmission of objective truth, and even the belief that such objective truth exists, is a quite modern idea. Further, accumulating particular autobiographical facts and averaging them over a number of people is a peculiar development, coming from a new orientation. One can follow this new perspective from censuses and analyses of vital statistics through social surveys to contemporary survey research. However, one has to recognize the novelty of this concern with a particular kind of accuracy. It is not very deep or widespread in the population, and this new emphasis can still profit from insight into other types of autobiographical data. These other types of data collection show important variations in definitions of truth.

### *Heroic Myth*

Mythical biographies show how little truth has to depend on accuracy. Accuracy in myths seems almost a contradiction. Mythical biographies exhibit the triumph of topical context over individual facts. It is almost as if the social context of the life story were the signal and individual items the noise. In fact, the original aims of biographical and autobiographical studies were to show models of action, and they encouraged the creation of heroes who did not correspond to real people. Scholars have documented the consistency of a sequence of incidents in the lives of heroes that seem to be prerequisites for becoming a culture hero (Propp, 1968; Raglan, 1936). If the incidents did not exist or if contradictory incidents existed, they were adjusted to fit the story. The life story appeared almost spontaneously out of the heritage of a people, and the context of data collection was played down. It seems almost sacrilegious to assign these mythical biographies to an author. In its adjustment of particular events to social norms of the heroic life course, heroic myth stands in contrast to the survey interview; in its intentional neglect of the author, heroic myth stands in contrast to the literary biography. As we shall see, heroic myth differs from the clinical

interview in overlooking the interview situation. Historically, heroic myth is documented as the oldest form and may represent a basic aspect of autobiographical memory. Certainly, the process of creating these myths corresponds to Bartlett's (1932) classical studies of remembering.

The modern interest in accuracy and documentation of details has opposed this procedure of collective memory; hero worship has fallen into disrepute. However, society still needs its mythical heroes, and the brute facts are often changed, or at least reemphasized, to conform to a general norm about an exemplary person. Especially the need for an epiphany—a sudden conversion experience leading to the assumption of leadership—seems to be prominent in would-be leaders of nations and social movements (Back, 1988). For instance, Margaret Sanger (1938) describes vividly how she watched a patient dying from a self-induced abortion and decided after a bitter night watch to devote her life to the spread of contraception; her rival, Marie Stopes, had no clinical experience but only laboratory training. However, she reported a similar experience told to her by a schoolmate, which prompted her concern with family planning (Stopes, 1923). Both stories are still quoted in many sources, although biographers are very skeptical about their accuracy (Kennedy, 1970; Reed, 1978).

Studies of autobiographical material of current culture heroes show the mythical process occurring even now as accuracy in details is sacrificed for a consistent image. This consistent image follows the cultural image of the hero. Introduction of particular facts does little to change the whole picture accepted by society. Kai Erikson (n.d.) has analyzed the way in which the public rejects the facts if they contradict the accepted picture of a person of any notability. Currently we can see how any factual additions to the biographies of such popular heroes as John F. Kennedy or Martin Luther King, Jr., are rejected or at least deprived of any importance. In fact, here we have a social definition of truth and accuracy: We have to be impressed by the social meaning of a biography; its accuracy rests in the function that it has for society. One might even say that culture heroes of any kind have a social duty to adjust their autobiography to a scheme that promotes their role in establishing viable myths for their followers. Thus, candidates for higher office in the United States feel compelled to reconstruct in their biographies the popular "born in a log cabin" myth, if only by constructing a poverty-stricken background.

### *Psychotherapeutic Interview*

Like survey research, the psychotherapeutic interview is an product of modern social science, but other contrasts between the two are pronounced. The clinical interview stresses the context of the interview situation as well as the context in a particular interpretation of the life story. Clinical interviewers do not expect that the individual facts will be accurately presented, and they try to inject their interpretation beyond the answers that are presented. Ironically, clinical interviewers are praised for showing insight when they interpret and guide the answers in ways that, for survey interviewers, would be errors. In the general

discussions of the survey interview, from Herbert Hyman (1954, especially pp. 52-69) on, clinical insight is considered to be interviewer bias.

Thus, in the psychotherapeutic interview, the personal creativity of the interviewer is valued highly; in this the clinical interview is closer to the literary end of the spectrum. In fact, clinical interviews have become a successful novelistic device, as in Philip Roth's *Portnoy's Complaint*, and actual case studies have been regarded as literary works—for instance, Robert Lindner's *Rebel Without a Cause* and *The Fifty-Minute Hour*. However, unlike pure works of fiction, research in psychotherapeutic technique tries to find a systematic method to discover a pattern of presentation, omission, distortion, and inconsistency that will lead to a defensible interpretation of a person's problems (Alexander, 1990). In any case, and in contrast to the survey interview, the interviewer's contribution is supposed to be active; the interviewee is treated as respondent but as not able to express the facts about himself or herself without aid and interpretation (that is, the interviewer's assessment about the "real" meaning of the answers).

Actually, the validity of reproduction of a single event is not very important: There are still controversies about whether Freud's hysteric patients were subject to childhood seduction or whether they just imagined it. For the clinician, the same consequences resulted. From this we can learn that accurate interpretations are possible even if we do not believe in the facts.

In order to reach significant overall conclusions, formal training of interviewers and instructions to them try to develop their ability to look for things that the patient does not say. These interviewers probably receive the most extensive instruction to perfect this skill. Truth is seen here as neither absolute nor subjective but as an agreement between the two participants, who agree on a framework, an assumptive world as it has been called, in which they can then agree on a joint definition of truth; that is, the interview is the setting for the relation between single responses and the total life story. One can also say that the relationship between the respondent and informant role is negotiated here.

A comparison with the survey interview is instructive. Both are mainly interested in the input factors, in identifying the conditions that made the respondents what they are. However, the survey interview starts with categories that are accepted within a social unit, which already has its own definition of facts. Clinical interviewers are interested in the individually constructed categories that vary from general social norms; they have to build a definition of truth that is acceptable in this unique framework. The relations of intended audience and meaning of accuracy are summarized in Figure 3.3.

## Conclusion: Truth and Meaning

To a certain degree, any autobiographical recall includes some features of all these types, and perhaps even more, but tend to be mainly influenced by one set of individual constraints on the recall situation. There are, however, some examples that stay ambivalent among the different definitions of truth. The meth-

Type	Immediate Audience	Ultimate Audience	Definition of Truth
Survey Interview	Interviewer	Professional Users	Accuracy of Each Unit
Literary Biography	Biographer	General Public	"Novel under Oath"
Heroic Myth	Followers of Hero	Future Generations	Current Social Values
Psychotherapeutic Interview	Therapist	Therapeutic Team (and Colleagues)	Useful Reconstruction

FIGURE 3.3. Types, Audiences, and Accuracy

ods of interpreting documents of this kind represent well the paths of research in this field.

The definition of truth in autobiography becomes especially questionable when data are analyzed that are produced by individuals who clearly lied in their autobiographical statements but who lived these lies quite consistently. Their statements become a mixture of the different types of autobiographical communication. They show distortion in individual incidents, in episodes, and in assembly of the whole life; the skilled analyst here must extricate the type of situation for each part of the available data. Two examples of these extreme cases will show these problems and their possible solutions: the discussion of Edmund Backhouse by Trevor-Roper (1978) and of Trebitsch-Lincoln by Bernard Wasserstein (1988).

Edmund Backhouse was an Englishman who, early in his life, went to China, first as an employee for a trading firm. He was also involved with political figures during the times surrounding the Boxer Rebellion early in this century. After this he started a career of "discovering" documents, some of which were supposed to be originals and some his own translations, and establishing himself as an expert in Chinese history. During World War I, he was engaged in negotiating armament deals for the Chinese Government, claiming useful contacts in England to expedite these deals, none of which ever came to fruition. Afterward he stayed in Peking collecting and producing a variety of historical and literary materials (some of which he donated to his University) in addition to volumes of autobiographies. Backhouse died in Peking in 1943. On investigation, it was found that most of his material was faked, his business dealing fraudulent, and his autobiography, with its exalted connections, a fabrication.

Ignacz Trebitsch-Lincoln was born in Hungary and had a varied life around the world. He made several attempts to emigrate, first to Montreal, where he, a Jew, managed a Christian mission to Jews. After financial difficulties, he landed finally in England. There he managed to get the confidence of a wealthy manufacturer, who gave him a good position and also supported him for a seat in the House of Commons. He used this position to engage in some dubious financial

dealings in Eastern Europe. After several scandals, he quit the Commons, was imprisoned, and after World War I engaged prominently in several extremely right-wing, proto-fascist movements. Here, too, he ran into troubles. After a number of adventures, he appeared in China, where he converted to Buddhism and started a Buddhist movement, which—besides its religious aspects—promised to eliminate British influence in that part of the world. He died in Shanghai during the Japanese occupation in 1944, still trying to promote political and financial schemes. He, too, left partial autobiographies, mostly invented.

These two cases are extreme examples where the creation of autobiographical data became a major part of the individual's life work. The biographers faced a difficult problem. They decided on a rule of not accepting any fact in the autobiographical documents that was not confirmed by independent evidence, using the autobiographies themselves only as psychological evidence; that is, they used the autobiographical data for meaning, not facts. With this technique, Trevor-Roper and Wasserstein wrote impressive biographies.

The technique meant that the autobiographical data were used purely as psychological documents. However, they were not produced for that purpose: They claimed to represent facts about the person, in the manner of a factual interview. The hidden purpose was to show the accomplishment of the individuals, in the manner of myth fabrication or, sometimes, an artful biography. The data that these two men had produced attempted to give a unified picture of a person, sometimes to achieve an immediate effect (such as starting an intricate swindle) but sometimes as an expression of their current point of view. The myth that they had created was accepted by their contemporaries, at least intermittently. Backhouse was acclaimed as a scholar and benefactor of Oxford University, Trebitsch-Lincoln as a religious leader. The documents that they left included some factual data, some myths that fitted well into the social context, and some evidence of psychological disturbance. The biographers gave clinical interpretations of the contradictions of self-reports with outside reality, leading them to diagnoses of psychoses.

These are, of course, unusual examples. However, they add another aspect to the analysis of autobiographical memory, namely, that the memory does not just transmit events in life but that it may itself determine some the events, making a kind of loop in the communications process. Backhouse and Trebitsch-Lincoln really tried to live as if their constructed memories were correct. Thus, to a certain degree, people live their autobiographies.

As Janet Malcolm (1990) has put it, biographies have some consistency; but putting one's life in the form of a novel is a sign of an abnormal mental state and therapy consists in destroying this novel. However, trying to arrange one's life into a systematic story is not necessarily pathological. It may be a general human tendency but be more pronounced among certain people and under certain conditions. Robert Butler (1968) has noted a widespread pattern of reviewing one's whole life on reaching decennial birthdays. Moreover, according to studies by David Phillips (Phillips & Feldman, 1973), prominent people tend to live

beyond the months of their birthdays, and the more prominent they are, the stronger this effect. It has been noted in biographies that they try deliberately to fit their life into a pattern. Thus, the German poet Goethe called his autobiography *Poetry and Truth*, and his life showed definite efforts to fit the sequence of events to his responsibilities as a genius (Boyle, 1991). Autobiographical memory can be seen as an attitude leading to action, and it has a definite function in the life course.

A systematic understanding of the dynamics of the different conditions makes it possible to exploit all data according to their value for accurate understanding. From the point of view of survey research, we can find the way to use additional data productively. Practically all autobiographical data can be used, but they must be used with reference to the constraints of the situation.



## 4

# On Authenticating and Using Personal Recollections

*Michael Ross and Roger Buehler*

If there is one topic that binds the various subdisciplines of psychology together, it is memory. The social psychologist examining attitudes or interpersonal conflict, the cognitive psychologist studying learning or decision making, the developmental psychologist researching the growth of cognitive or social capacities, the clinical psychologist investigating maladaptive perceptions or behaviors, all consider how the cognitive or emotional residues of individuals' past experiences influence their current thoughts, feelings, or behavior.

This concern with recall is presumably not restricted to the ivory tower. In this chapter, we examine personal or autobiographical memories, a form of recollection that can be important in people's everyday lives. In previous work on personal memories, psychologists have tended to focus on the mechanisms underlying accurate or inaccurate recall (e.g., Neisser, 1982a; M. Ross, 1989; Rubin, 1986). Recently, investigators have also begun to enumerate the social-psychological functions of personal memories. This shift is significant, since accuracy is not the rememberer's only goal. Personal recall often occurs in a social context and can facilitate a number of interpersonal objectives. People recount episodes from their pasts to convey information, to entertain audiences, to sway opinions, and to establish, maintain, and justify social relationships (Brewer & Lichtenstein, 1982; Higgins, 1981; Neisser, 1988; M. Ross & Holmberg, 1990).

People may vary their accounts of the past in order to achieve their social goals. For instance, rememberers may report their experiences in a way that is relevant or interesting to their present audience or that enhances their public image. Such adjustments to personal stories can occur in at least three ways. First, rememberers may simply communicate the subset of their memories that seems to satisfy their current purposes, ignoring other relevant recollections. Second, they may adapt the meaning or interpretation of events to suit their

needs. Third, rememberers may offer embellished accounts of the past, rewriting history in order to provide a more satisfying story (M. Ross & Holmberg, 1990).

Personal memories serve individual as well as interpersonal functions. People can reflect on their histories to examine and derive support for their beliefs and values, to formulate and justify current decisions and judgments, and to generate predictions about the future. Most importantly, perhaps, memories aid in the construction of self. People can study the past to learn about their preferences, abilities, and so forth. Psychological theorists have claimed that the relation between self-perception and memory is reciprocal: Memory influences self-perception and self-perception influences memory. The latter causal sequence assumes particular importance for those theorists who maintain that personal recall is often an active, constructive process that is guided by people's self-perceptions, especially their feelings and beliefs at the time of retrieval (e.g., Bartlett, 1932; Greenwald, 1980; Mead, 1964; M. Ross, 1989).

Analyses of the social and personal functions of memory rest on the assumption that people value their recollections and use them to achieve their goals. Is this assumption valid? Do people make extensive use of their pasts as a functional analysis would seem to imply? The answer to this question is surely "yes." Much of people's social behavior, for example, is guided by their past experiences and learning. We shake hands on meeting a person for the first time because we have learned that this is an accepted ritual in Western culture. In this chapter, we examine a slightly different usage of the past. We ask whether people consciously choose to examine their past experiences in order to inform their current judgments and decisions. Specifically, when do people focus on their pasts, mining them for what might be of use in the present, and when do they disregard their histories?

It is not self-evident that people dwell on their pasts to any great extent. In response to our questioning, many acquaintances and colleagues claimed hardly ever to think about their histories; they asserted that their attention is focused on the present or on their plans for the future. Moreover, people sometimes are critical of those who seem to focus on the past. For example, Oscar Wilde reputedly was unsympathetic when his wife described the travails of her youth: "People who went back to their childhoods for their tragedies bored him" (Eilmann, 1988, p. 235).

To date, the psychological literature has contained little analysis of the conditions that influence whether people use personal memories in the pursuit of individual or social goals. In the present chapter, we describe some of our own attempts to investigate this issue. Our ultimate aim is to determine when and how people use their personal memories. A more immediate question, however, is whether people even believe their own memories. Presumably, people would not use their memories in the service of personal goals if they deem their recollections to be unreliable. Thus, we first consider the ways in which people evaluate the authenticity of their own and others' recollections. We then review our own research on whether and when people refer to their past experiences.

## Assessing Authenticity

People's subjective beliefs about the accuracy of a memory are likely to be related to their use of information contained in that memory. People may be more attentive to retrospective accounts that seem trustworthy. This concern with validity raises an important psychological question: How do people evaluate the accuracy of a personal memory—either their own or someone else's?

We suspect that typically people do not engage in extensive evaluation of their own recall. Instead, they tend to accept their recollections uncritically. A belief in the accuracy of many of one's own memories is almost certainly crucial to mental well-being. People's sense of their identity, of who they are, is intricately bound to their personal memories.

Similarly, people generally believe what others tell them. A fundamental assumption underlying social discourse is that speakers communicate the truth (Grice, 1975); people who habitually question the veracity of others' statements are labeled "paranoid." Dozens of published experiments in social psychology journals provide compelling evidence that people tend to accept another's behavior and words at face value. Individuals are seen as acting consistently with their own beliefs even when their responses are obviously constrained by external circumstances (Gilbert & Jones, 1986; Jones, 1979; L. Ross, 1977). For example, a target who is required to make a speech praising Castro is evaluated as being more favorable toward Castro than is a target who is required to make an anti-Castro speech (Jones & Harris, 1967). In everyday life, movie actors are seen as possessing character traits associated with their roles, and hence they are concerned about typecasting. Also, advertisers and salespeople owe their success, in part, to their audiences' credulity. While shopping, people may accept advice from a salesperson without appreciating fully that she or he may have an ulterior motive.

Studies of deception provide further evidence of people's gullibility. Even research participants who are informed that a speaker may deceive them tend to believe the speaker. Observers notice deception at only slightly better than chance levels (DePaulo, Stone, & Lassiter, 1985; Fleming, Darley, Hilton, & Kojetin, 1990; Krauss, 1981).

In studies of lie detection, observers are typically informed that a target may be deceiving them, and then they are required to judge the veracity of the target's utterances. In everyday life, a rememberer's audience would not normally be poised to detect lies. People in social settings are often prepared to enjoy a well-told story (M. Ross & Holmberg, 1990). They may willingly suspend disbelief, just as they do while viewing a film or play.

Although individuals are trusting across a wide variety of social contexts, there may be limits. The negative stereotype of used-car salesmen, for example, probably induces wariness in even the most ingenuous consumer. It seems reasonable to suppose that people are more trusting in some situations than in

others, just as there are likely to be individual and cultural differences in credulousness.

Suppose individuals are motivated to assess the truth of a personal memory. How might they do so? Although people often lack objective criteria for judging the validity of a personal memory, this is not always the case. Dates, times, and other details of people's stories can sometimes be verified. More often, perhaps, people can seek proxies for such objective criteria. For example, researchers sometimes assess whether people's initial, relatively immediate representation of an event (e.g., as recorded in a diary) corresponds with a subsequent recollection of the episode (see Chapters 6–8 of this volume for diary-based research). A basic premise of this research is that recollections are accurate to the extent that they agree with the initial representation. This interpretation is plausible, although not necessarily valid. For instance, two representations could reflect different, but equally compelling, depictions of reality (Mead, 1964). It is tempting to suggest that private diaries written shortly after the target event are quite accurate when the material being recorded is comparatively straightforward or inference free (e.g., what the respondent ate for dinner as opposed to a description of a marital dispute), but we are not aware of data that directly support this contention.

Consensus constitutes a second proxy for objective reality. If people's recollections of an event agree, they and others tend to assume the validity of the memories. The rule of consensus is often applied in legal settings to establish the veracity of the accounts of various witnesses. Yet agreement is no guarantee of truth. The history of science reveals that consensually accepted truths of today are often overturned tomorrow. By the same token, disagreement is no proof of falsity. Conflicting accounts may both be "true" in the sense that each captures the person's original or current understanding of events.

Disagreement serves an important purpose, however. It prompts individuals to evaluate, rather than simply to assume, the validity of memories. Although each account may be perfectly plausible on its own, the presence of a discrepancy leads people to adopt a more skeptical orientation toward the stories. Common sense suggests that when two memories conflict, at least one is likely to be wrong. People's reactions to conflicting memories thus provide a context in which to examine the criteria that individuals adopt in judging accuracy.

In the absence of objective standards, what cues do people use when disagreement motivates them to evaluate the accuracy of their own and others' memories? Although previous investigators have not examined this question directly, there are several pertinent theoretical and empirical analyses. In one relevant line of work, Johnson and her colleagues examined how people distinguish memories of actual episodes from memories of imagined events. Johnson and Raye (1981) reasoned that because recollections of external events result from perception, such memories should include more sensory detail and contextual information than do memories of imaginary events. A rememberer can, therefore, use the amount of sensory or contextual information in a memory as a basis for inferring whether the recollection refers to an imagined or a real event. In

addition, Johnson and Raye suggested that individuals may evaluate the reality of a memory by considering whether it is consistent with their other knowledge. For example, a memory of sliding into home plate at Yankee Stadium might be attributed to fantasy if the rememberer cannot recall playing professional baseball. Similarly, most normal adults would relegate to the realm of imagination a memory of flapping their arms and flying over Toronto. Research by Johnson and her colleagues, as well as others, has provided support for the Johnson and Raye analysis (e.g., Bell & Loftus, 1989; Johnson, 1988; Johnson, Foley, Suengas, & Raye, 1988; Johnson & Raye, 1981; Schooler, Gerhard & Loftus, 1986).

A number of other authors have considered the thought processes and standards that people use to assess truth (e.g., Bruner, 1986; Spence, 1982). A form of deliberation that Bruner has labeled "narrative thought" is particularly useful in the current context. Possessing the qualities of a well-constructed story, high-caliber narrative thought is (a) vivid and detailed rather than sketchy, (b) coherent in that the events are sequenced and connected in an intuitively plausible manner, and (c) characterologically consistent, with characters' actions seeming to stem from their personalities, intentions, and motives. A gripping story can be narratively "true" even though the characters and events are fictional.

Bruner's (1986) analysis appears applicable to people's personal recollections. Personal memories often take the form of extended narratives that include a chronological recounting of events and the actions of various individuals. Following Bruner, we should expect people to judge their stories to be true to the extent that the recalled narratives are vivid, detailed, coherent, and characterologically consistent. Together with the consensus rule, the various standards discussed by Johnson and by Bruner serve as criteria by which individuals may assess the validity of their memories.

Although we have focused on individuals' evaluations of their own memories, people may use the truth criteria to assess the veracity of another's recollections. Consider the following accounts of a married couple's first encounter. The husband's description of this episode was approximately as follows: "I was sitting in \_\_\_ airport waiting for my flight to be announced, when this lady came into the room. She chose to sit beside me even though many other seats were available. We talked, hit it off, and the rest is history." The wife's account was identical in every aspect to her husband's, except for one. She recalled the waiting room as being entirely full, except for the seat beside her future husband! Both spouses are aware of the discrepancy in their recall, and they have discussed it on a number of occasions. Both firmly claim to believe their own recollection.

When memories conflict, rememberers may scrutinize the alternative accounts, evaluating the degree to which each fulfills various truth criteria. Whereas rememberers may well suppose that they are impartial in their deliberations, there is reason to suspect that they will typically favor their own accounts. Rememberers have a major psychological stake in accepting the truthfulness of their own recollections. Moreover, their subjective experience may serve to

confirm this belief. In recalling an episode, a rememberer may have the experience of hearing or seeing the event unfold once again. While reciting the first encounter story, the husband may “see” the empty seats in the airport. His spouse’s denial of their existence cannot eliminate his sensory experience. On the other hand, the husband does not have direct access to the vivid, sensory experience that his wife possibly derives from her memory. As a result of such differential access, people may often suppose that their own memories are more vivid and contain more sensory information than does another person’s recall of the same event.

Rememberers are also likely to view their own accounts as more compatible with other truth criteria. Neither verbal account is likely to incorporate all of the relevant features that the storyteller could provide, but rememberers can mentally attach further details to their own accounts when considering the merits of each version of the story. As a result, the rememberer’s subjective experience is that his or her own portrayal is more detailed. Moreover, the rememberer may recall other memories involving parallel events and similar actions by the leading characters. Rememberers would, then, find their own depiction to be congruent with their additional memories. Rememberers are also likely to construct a story in which the protagonists’ actions are consistent with the rememberers’ current understanding of their attitudes and dispositions. Alternative portrayals of the same episode may seem less characterologically consistent to the rememberer. Finally, rememberers are likely to view their own accounts as more coherent, partly because they created the accounts and partly because they are more familiar with these accounts. All of us who lecture or write have had the experience of generating puzzlement when we offer what we consider to be perfectly lucid explanations. In sum, people may tend to rate their own personal memories as more plausible than alternative versions because their own accounts seem to fulfill the various truth criteria more adequately.

Rememberers may present their conflicting stories to an audience. What criteria do audiences use to analyze the truthfulness of competing accounts? Like rememberers, audiences find detailed and vivid accounts more compelling than pallid ones (Bell & Loftus, 1989). Unlike rememberers, audiences lack direct access to the subjective experience of remembering. Although a rememberer privately experiences his or her own memories as more vivid, an audience may detect no difference in the vividness of the two verbal accounts. Also, audiences may often know the protagonists in the stories less well than do the rememberers and, therefore, be less able to evaluate whether story characters’ actions are consistent with their dispositions.

Because audiences cannot base their validity judgments on the subjective experience of remembering, they may use more explicit truth criteria, such as the presence or absence of corroboration. Audiences should judge story elements on which rememberers agree as highly accurate and should be skeptical about conflicting details. Audiences may also consider whether the purported events or the actions of the participants are consistent with the audiences’ general knowledge, for example, their normative expectations.

Rememberers should be less dependent than audiences on these explicit criteria. Rememberers may often consider their coherent and vivid recall to be factual rather than a story to be evaluated in terms of corroborating evidence or normative expectations. Rememberers may interpret another person's confirmation of elements in the rememberer's story as an incidental byproduct of the fact that the story accurately depicts reality rather than as *prima facie* evidence of the accuracy of the recollection. By the same token, another person's rejection of elements in the rememberer's own account may be insufficient grounds for the rememberer to deny the validity of his or her coherent and vivid recall.

Some of these ideas have been tested in a study conducted by James Karr (1990) at the University of Waterloo for his Master's thesis. In this study, undergraduates wrote accounts of a personal incident for which their memory conflicted with somebody else's. They described both their own and the other person's memory of the incident. Participants were encouraged to provide as much detail as possible and to be equally detailed in both accounts. Next, they evaluated the historical accuracy of each line in both accounts on 7-point Likert scales. In a final section of the questionnaire, they assessed the subjective quality of their own and the other person's recall.

The study also included a group of "observers." Each observer read transcripts of the memories that one participant had reported. Observers then assessed the accuracy of each transcript statement and evaluated characteristics of the two accounts on the same rating scales used by the participants.

Most of the participants were able to recall examples of conflicting memories. Only 13% of those who were approached claimed to be unable to think of such an instance. Almost all of the reported incidents involved memory for everyday events. The example provided in Table 4.1 is representative of the type of conflicting memories reported by participants. As in the example, participants provided approximately the same amount of detail in each account. There was considerable variability in participants' reports of when the episode occurred: The range was from less than a month to 26 years ago, with a mean of 4.5 years.

Each statement in both accounts was coded by research assistants as "similar," "conflicting," or "extraneous." A statement was coded as similar if there was a corroborating statement in the other account, conflicting if there was a contradictory statement in the other account, and extraneous if there was no related statement in the other account. On average, 51% of the statements were similar, 40% dissimilar, and 9% extraneous.

Participants' and observers' evaluations of the accuracy of the similar and dissimilar statements revealed that consensus is an important cue. Rememberers and observers judged statements that were corroborated across accounts to be more accurate depictions of reality than were statements that were contradicted across accounts. Rememberers also exhibited a significant bias in favor of self, but only for dissimilar statements. They judged statements in their own accounts that conflicted with statements in the others' accounts to be reasonably accurate; in contrast, rememberers deemed statements in the others' accounts that challenged statements in their own accounts to be inaccurate. Observers exhibited

TABLE 4.1. Example of Conflicting Memory Accounts

---

**Own account:**

I was 7 years old.  
 I woke up at 7:30 AM in preparation for school.  
 I put my winter outerwear on, i.e., hat, scarf, snowpants.  
 I walked with two of my friends.  
 No one was at school because it was a Professional Development day.  
 I returned home happily when I realized that school had been cancelled.

**Other's account (sister):**

My sister claims that I was 10 years old.  
 I woke up at 4:30 AM for school.  
 She saw me put my snowpants and winter clothes on.  
 I walked alone.  
 No one was at school because I was 3 hours early.  
 I returned home upset that she had set the alarm 3 hours ahead of time.

---

no systematic favoritism: They considered statements in the accounts of the rememberer and other to be equally valid.

A direct comparison of rememberers' and observers' ratings is revealing. Both rememberers and observers evaluated statements that were corroborated across accounts as highly accurate, with average ratings exceeding 6 on a scale where 7 represented completely accurate. On the other hand, rememberers and observers deviated in their ratings of conflicting statements. Rememberers rated their own conflicting statements as more accurate, and the others' conflicting statements as less accurate, than did observers. The data imply that observers used the consensus rule even-handedly, whereas rememberers seemed to ascribe more importance to corroboration in judging the validity of the others' conflicting statements than in judging their own.

Consistent with our theoretical analysis, rememberers also judged their own memories of the events to be more vivid, narratively coherent, and characterologically consistent than the others' memories. Participants could have reported that their own memories were highly vivid and correct, while the other people's memories were equally vivid but wrong. Instead, they seemed to deny the phenomenological reality of the others' experience. In contrast, observers noted no difference in the perceived clarity, coherence, and characterological consistency of the conflicting memories. Apparently, the rememberers' sense of a qualitative difference in recall was not communicated in the written record.

Karr's (1990) study provides evidence that people assert the accuracy of their own personal memories. This finding has important methodological and theoretical implications. First, let us briefly consider methodological implications. Social scientists are often interested in obtaining valid retrospective reports from their respondents. Research surveys and interviews impose strong demands for accuracy on rememberers. This accuracy goal may well reduce respondents' tendencies to self-present or engage in deliberate falsehood. However, self-

presentation and revision of the past need not be intentional. Individuals' recall may be biased by their present feelings, beliefs, and self-perceptions (M. Ross, 1989). The individual who currently views himself as in the pink of health, for example, may underestimate the extent to which he has experienced various maladies. Similarly, the individual who views herself as highly cultured may inflate the number of times that she has attended the ballet or theater; such cultural experiences play a major and exaggerated role in her view of herself.

Respondents would not necessarily either intend or recognize their alterations to history. They simply report tales consistent with their self-concepts and various psychological goals. This altered past is presented to the researcher as "truth." For the rememberer, it may well constitute a form of truth, but it is not the type of historical accuracy that the researcher has in mind.

As we have already noted, researchers are often prompted to offer an operational definition of accuracy, such as the degree of correspondence between respondents' diary records and subsequent recollections. The validity of these operational definitions is typically not evaluated. Another possible approach would involve asking rememberers themselves to evaluate the accuracy of their memories. The researcher would then only treat seriously those recollections in which rememberers expressed a high degree of certainty. Although some previous research suggests that people's confidence in their memories is at least moderately predictive of accuracy (Brigham, 1988; Tulving & Thompson, 1971), the results of Karr's (1990) study are troubling in this regard. Respondents assumed that their own recall was quite accurate, even when it conflicted with another person's. Karr had no direct means of proving them wrong; however, it seems likely that respondents' accounts of the episode were not as historically accurate as they appeared to suppose. Karr's respondents claimed vivid recall, but research on flashbulb memory (Neisser, 1982b) and eyewitness testimony (Wells & Loftus, 1984) reveals that it can be an error to equate detailed and vivid memory with accuracy. A recent longitudinal study of flashbulb memory provides evidence that rememberers can be very confident of the accuracy of their recall, report extremely vivid memories, and yet contradict their earlier descriptions of the target episode (Harsch & Neisser, 1989).

The various truth criteria that people may use in their everyday lives to evaluate the credibility of recollections also offer no assurance of success. Researchers have identified qualitative differences between people's descriptions of accurate memories and inaccurate memories, with accurate recall including more sensory detail and contextual information (Schooler et al., 1986; Schooler, Clark, & Loftus, 1988). The distinctions are subtle, however, and observers' ability to discriminate between the two kinds of descriptions is typically only slightly better than chance. Telling observers exactly which cues to attend to improves their accuracy somewhat, but their overall classification performance remains unimpressive (Schooler et al., 1986). By invoking various truth criteria, individuals may become suspicious of some accounts or may believe others, but the criteria will rarely provide them with a definitive basis for accepting or rejecting people's stories of their pasts.

Research on assessing the authenticity of personal stories seems to offer no simple panacea for the accuracy problem. If anything, it emphasizes both the magnitude and the importance of the problem. We believe that it is incumbent on researchers concerned with historical accuracy to verify their respondents' recollections. Such validation may necessitate the use of prospective studies and direct behavioral observation. These are often time-consuming and costly procedures, but the alternative is even more disconcerting. We cannot hope to build a science of human behavior on data that may be rife with bias and error.

Finally, consider the implication of Karr's (1990) data for a functional analysis of memory. People apparently have considerable faith in the accuracy of their own recollections. Consequently, if individuals fail to use their past experience in the pursuit of social and personal goals, it is probably not because they lack confidence in their recall. We now turn to the second major focus of this chapter, an examination of some of the conditions that prompt people to study or to ignore their pasts.

## The Effect of Performance Level on Whether People Dwell on Past Activities

As a first approach to the issue of the relationship between people's performance level on past activities and whether they think about those activities, consider the effects of winning or losing a competitive game on spontaneous recall of the game. Which outcome is likely to prompt you subsequently to think more about the game? Remembering a win could serve to make you feel good about yourself; dwelling on a loss could promote discontent. Thus you might be more inclined to recollect the details of a successful experience. On the other hand, pondering the particulars of a loss might help you to understand the basis of your failure and arm you against its future recurrence. There is reason to argue either way; moreover, a variety of personal (e.g., self-esteem, depression) and situational variables may well influence whether people dwell more on a failure or a success.

In a recent study conducted for a honors thesis at the University of Waterloo, Bonnie Turner (1991) examined the impact of one potentially important situational variable. Participants who had either won or lost a competitive game were led to believe either that they would shortly play the same game again or that they would play a different game. They were then left alone in a room for several minutes and asked to speak aloud into a tape recorder any thoughts that went through their heads. On listening to the tapes, Turner discovered that participants discussed many different topics, including the paint on the ceiling of the room, the weather, their university courses, and their love lives. Almost all of them also discussed the game that they had just played. Turner recorded the amount of time that participants spent talking about their experiences in the game. When they anticipated playing a *different* game in the immediate future, winning or losing did not affect the amount of time that participants spent talk-

ing about the game. In contrast, when participants expected to play the same game again, losers dwelled on the game for a significantly longer time than did winners. Apparently, losers were more likely than winners to think about the game only when such thoughts could be of benefit in the immediate future. Confident of their prowess, winners may have felt little need to mine their past experience even when they anticipated playing the game again.

Turner's experiment suggests that people can choose to use or to ignore the past. Note, however, that Turner's participants had suffered only a minor failure or victory. Powerful negative past experiences are banished less readily from people's thoughts (Silver, Boon, & Stones, 1983). In addition, we next consider a series of studies on prediction that indicate that people's use of the past is sometimes less felicitous than Turner's results would seem to suggest.

## Do People Use Past Experience as a Basis for Predicting Their Own Future Behavior?

Our research on prediction addressed two related questions that have surely puzzled most of us. Why do we frequently seem to underestimate how long it will take to complete a task? Why don't we learn from past experience and adjust our estimates accordingly? To answer these questions, we explored the relations among people's predictions of when they will finish tasks, their actual completion times, and their reports of relevant past experiences.

We began by assessing people's ability to predict when they would complete various activities. For example, we asked fourth-year psychology students when they realistically expected to finish their partially completed honors theses. The students' predictions were overly optimistic, as revealed by objective records of when the theses were actually finished. Only 30% finished their theses by the date that they had expected to. On average, students took 55 days to complete their theses, 22 days longer than they had anticipated. In subsequent research, we examined students' ability to predict when they would finish everyday tasks around their homes, as well as school assignments. On average, respondents took almost twice as long as they had predicted, even though they exhibited considerable confidence in their predictions.

After documenting the robustness of this optimistic bias, we explored the psychological processes underlying people's predictions. In particular, we examined whether people used their memories of past experiences in formulating their task completion estimates. As a first step, we surveyed university students about their past experiences in finishing tasks. The respondents were instructed to recall occasions when, prior to completing an assignment or task, they had tried to estimate when they would finish. They reported that the majority of these tasks were actually completed after they had anticipated ( $M = 68\%$ ) rather than by the time that they had anticipated.

Although the survey data were consistent with our earlier findings, we cannot determine whether respondents' memories were veridical. Global, retrospective

frequency estimates of this kind are subject to a number of distortions and biases. These frequency estimates are nonetheless psychologically important: They reflect people's beliefs about their past performance. Presumably, people's subjective representation of their past, rather than the past itself, can potentially be used to inform their predictions.

The survey data also reveal an apparent paradox. When asked explicitly about their past performances, people acknowledge that they often finish tasks later than they anticipate. Yet, respondents in our prediction experiments were confident of the accuracy of their overly optimistic predictions. We inferred that respondents in our prediction studies exuded confidence because they failed to take into account their past experiences. When people make predictions, they may focus on the future rather than on the past. To test this hypothesis, we conducted a study that included a think-aloud procedure. While predicting when they would finish an upcoming project, respondents were instructed to say aloud every thought or idea that came into their minds. Research assistants coded typed transcripts of these verbalizations for whether each statement reflected a focus on the future or the past. Consistent with our hypothesis, the majority of respondents' thoughts were directed toward the future ( $M = 74\%$ ). For the most part, respondents optimistically planned how and when they would complete the target activity. Only about 7% of their thoughts about the future involved a consideration of possible problems or impediments. Moreover, just 7% of their total thoughts were devoted to past successes or failures. These data suggest, then, that people rarely refer to their past experiences when generating their task completion estimates.

People's failure to make use of the past could reflect a number of related processes. Conceivably, most people totally ignore their past experiences when generating time estimates. By its very nature, prediction elicits a focus on the future rather than on the past. An inclination to think ahead may tend to preclude other types of thought, in particular thoughts about the past.

On the other hand, people's apparent failure to invoke their past experiences need not always result from complete neglect. People might sometimes consider their past experiences but fail to incorporate this information into their predictions. The connection between past experiences and a specific prediction task is not straightforward. The person must select an appropriate comparison, a past experience or class of experiences similar in relevant respects to the activity at hand. Often it will be difficult to identify a relevant class of past experiences; the various instances seem so different from each other that they cannot be compared meaningfully (Kahneman & Tversky, 1982). Adults have completed thousands of projects: Which ones provide a relevant standard for comparison? On what dimensions should relevance be determined? For instance, should people assess similarity on the basis of task content (e.g., English assignments) or on the basis of life circumstances (e.g., how busy they are)? Even if people were inclined to use their pasts, they might encounter difficulties in abstracting relevant information.

Along with the ambiguity inherent in identifying a class of relevant past ex-

periences, people might actively process information about the past in a way that diminishes its relevance to the future. The meaning of a prior behavior depends largely on an individual's explanation of why it occurred (Heider, 1958; Jones & Davis, 1965; Kelley, 1967, 1973; M. Ross & Fletcher, 1985). Certain types of attributions will have the effect of linking a past event to present and future events; other kinds of attributions will serve to isolate the past from the present and future. A past episode will seem generally diagnostic to the extent that it is attributed to a cause that could be present across a variety of situations (global attribution) rather than a cause tied directly to a specific instance (specific attribution). Similarly, an experience will seem more diagnostic if it is seen to be caused by stable factors as opposed to unstable or transient factors. Clearly, if people tend to see past events as resulting from *specific* and *unstable* causes, they will be less likely to use those experiences to inform their present judgments.

We have examined people's attributions for their past experiences in several studies. For example, we asked university students to recall an occasion when they had failed to complete a task by the time that they had originally anticipated and then to recall a similar prediction failure experienced by a roommate or a close friend. Respondents were asked to explain why each of the two tasks was not finished by the expected time. Respondents' open-ended responses were coded by research assistants, who rated the stability and globality of respondents' attributions.

Respondents tended to attribute impediments to meeting their own predictions to transitory and specific causes. The reasons that respondents reported for their own lateness were rated as significantly less stable and less global than the reasons that they provided for similar tardiness by close acquaintances. People interpret their lateness in a manner that makes it seem relatively unique and unlikely to recur, an assessment that might well lead individuals to discount past experiences when making judgments and predictions.

We have thus identified one mechanism that could contribute to people's failure to incorporate their memories of their past experiences into their task completion estimates. People judge each past impediment to task completion as relatively exceptional and unlikely to repeat itself (e.g., their personal computer blew up while they were typing their English essay, or they encountered a family emergency). They may well be correct in this judgment of uniqueness. What people seem insensitive to, however, is the sheer number of problems that could possibly arise. There exists a vast, perhaps infinite, number of potential impediments. Although most of these impediments have a relatively low probability of occurrence, the probabilities are additive. The likelihood that *some* unexpected problem will arise is actually quite high. This is one lesson that people apparently fail to derive from their own past experiences.

The tendency to discount the relevance of past experiences while making predictions is not limited to task completion estimates. For example, we have all watched individuals careen from one disastrous romance to another, seemingly not learning from their past relationships. Canadian divorce statistics tend to

support this impression: The likelihood of divorce is greater in second than in first marriages (McKie, Prentice, & Reed, 1983). A readiness to discount the past probably extends beyond personal experiences to the prior fortunes of others. People may often have little difficulty disavowing the relevance of others' experiences to their own situation. Such denial has been documented in a variety of historical contexts. For instance, Lyndon Johnson and his advisors considered the French experience in Vietnam to be largely irrelevant to predicting how the United States would fare: It was richer, more powerful, more united, and not in the habit of losing wars (Neustadt & May, 1986). Similarly, in 1914 the Europeans distinguished their conflict from previous wars. They detached the past from the present, apparently believing, among other things, that war had become too expensive to last long (Neustadt & May, 1986).

People's failures to incorporate information from either their own or others' past experiences into their predictions may seem at odds with our earlier proposition that people use the past to achieve current goals. In our studies, participants apparently neglected information from their pasts that could have helped them to achieve their goal of generating an accurate prediction. People's failure to use history could serve other objectives, however. It could help people to maintain a positive self-image and a sense of personal control (C. R. Snyder & Higgins, 1988; Taylor & Brown, 1988).

Consider the issue of control. People in our culture are socialized to believe that they ought to command their own destinies. A prescription that links the past to the future implies that the future is foreordained. To assert their sense of control, individuals can neglect the past or deny its implications. By disclaiming the relevance of history, individuals can maintain the belief that they control their futures, that they have the power to achieve desired outcomes.

People are probably most inclined to deny the relevance of history when they dislike its apparent implications (for example, that a project will take longer than hoped). On the other hand, individuals may pay greater attention to interpretations of the past that suit their present purposes. Victory in Vietnam was consistent with the strategic objectives of the United States in Asia. For this reason, perhaps, Johnson and his advisors chose to focus on the differences and to ignore the parallels between the French and American incursions into Vietnam (Neustadt & May, 1986). Similarly, according to various newspaper reports, Saddam Hussein's understanding of the Vietnam War led him to believe that Iraq might prevail against the much stronger forces of the United States and its allies in the Persian Gulf war of 1991.

We have conducted a preliminary test of the notion that people are more inclined to deny the relevance of past experiences that have negative rather than positive implications. In one of our studies, participants were first asked to predict when they would finish an assignment and subsequently to explain why they finished it before or after their estimates. The participants themselves then assessed the stability and globality of the reasons that they reported. Participants reported significantly less stable and less global reasons for finishing late than for finishing early. This pattern of attributions is consistent with the view that

individuals are inclined to “make excuses” for negative past experiences (C. R. Snyder & Higgins, 1988), a tendency that could result in subsequent inattentiveness to those events.

## Summary and Conclusions

People are likely to use their personal histories in the service of decisions or predictions only if they are reasonably confident of the accuracy of their memories. We have presented research that suggests that this initial prerequisite is typically fulfilled. Individuals appear confident of the authenticity of their own recollections, expressing considerable certainty in their memory of a past event even when others recall the same occasion quite differently.

Our research also suggests that although people’s confidence in their memories is probably a necessary condition, a number of additional conditions determine whether they will actually invoke their memories. For instance, people’s tendency to recall their performance on a competitive game depended on two factors: their degree of success and whether they expected to play the game again. Our prediction studies indicated that people may fail to attend to the past even when a focus on past experience could facilitate their current judgments. If individuals would incorporate their memories of past performances into their task completion estimates, they would generate more conservative and hence more realistic predictions. Our research indicated, however, that people rarely consider past experiences when predicting how long their current tasks and activities will take to complete. Instead, they appear to dwell primarily on the future, optimistically planning how and when they will complete the task under consideration. We suggested that people’s neglect of the past may be motivated, in part, by a concern for personal control, and we reported preliminary findings consistent with this position. We described, however, a number of additional contributing factors. Most importantly perhaps, we reasoned that it is often difficult for people to know a priori exactly which aspects of past experience they should attend to. For example, perhaps it is only obvious with hindsight that Johnson and his advisors should have taken the French experience in Vietnam more seriously.

Previous research on memory has focused almost exclusively on the causes of accurate and inaccurate recall. In contrast, we have considered the role of the past in people’s social and personal lives by exploring some of the conditions that determine whether, when, and how people make use of their personal recollections. Such research has the potential to enhance our understanding of judgment and decision making, as well as of the reciprocal relation between memory and social and self-perception.

*Acknowledgments.* Preparation of this chapter was supported by a research grant and doctoral fellowship from the Social Sciences and Humanities Research Council of Canada.



## 5

# Affect and Memory in Retrospective Reports

*Mahzarin R. Banaji and Curtis Hardin*

Retrospective reports have long served as the warhorses of experimental and nonexperimental psychologists, although in both classic and contemporary discussions of method, the validity of retrospective reports to understand human thought has been questioned (Ericsson & Simon, 1980; James, 1890/1950; Nisbett & Wilson, 1977). Psychological data are often obtained as verbal reports from subjects about an event that occurred in their past, for example, "Whom did you vote for in the last election?" or "How friendly is the person described earlier in the experiment?" Implicit in the research enterprise that characterizes contemporary psychology is the assumption that retrospective reports are informative about mental processes and the actions that they guide. Verbal, retrospective self-reports have served as the tool to understanding human values, beliefs, attitudes, attributions, emotions, perceptions, thought, memory, personality, motives, and goals and as indicators of past and future behavior. As such, retrospective report data have defined the central epistemological questions about psychological knowledge: What can the subject *know* about the past? How can the subject *report* about it?

Investigators whose primary sources of data are questionnaires or surveys have been most cognizant of problems concerning the veracity of the responses produced by their subjects. Recently, systematic investigations have begun to identify the properties of cognitive functioning that not only influence the judgments that psychologists require of their subjects but also demonstrate the mutual practical and intellectual goals of survey methodologists and experimental psychologists interested in the accuracy of retrospective reports (Hippler, Schwarz, & Sudman, 1987; Jabine, Straf, Tanur, & Tourangeau, 1984; Loftus, Fienberg, & Tanur, 1985; Schwarz, 1990a; Tanur, 1992; Tourangeau & Rasinski, 1988). A consequence of this collaboration has been the application of information-processing theory to the study of survey research (Hastie, 1987; Ostrom, 1987), which has provided a framework for posing testable research questions about the processes that produce retrospective reports. Such research

efforts have proved invaluable, and undoubtedly the accumulation of empirical findings will continue to benefit the development of theory about the retrospective report process.

Of the various stages of the information-processing sequence involved in the production of a retrospective report, memory mechanisms have received the most attention (Blair & Burton, 1987; Loftus, Smith, Johnson, & Fiedler, 1988; M. Ross, 1989; Strube, 1987). Factors influencing memory for past events that are examined concern characteristics of the subject (e.g., retrieval strategies, physiological arousal), stimulus (e.g., meaningfulness, frequency, imagery), or situation (e.g., time delay, the source of the stimulus). The role of such factors in the accuracy of retrospective reports of the kind that occur in the course of everyday life has grown as a focus of interest.

Quite recently, an interest in the affective factors that influence retrospective reports of the sort collected in surveys has also emerged (Salovey, Sieber, Jobe, & Willis, chap. 6, and Clark, Collins, & Henry, chap. 18, this volume; Ottati, Riggle, Wyer, Schwarz, & Kuklinski, 1989). In this chapter, we present experimental research on the influence of affect on memory, particularly as it pertains to the collection of retrospective reports. Several investigations have demonstrated that affective states, affective judgments, and the affective properties of the stimulus can influence the way in which information is learned and remembered. Although mainstream coverage of the study of memory largely continues to ignore the role of affect (but see Baddeley, 1990), the research presented here represents a thriving interest among some social and cognitive psychologists who have identified a variety of influences of affect on memory. Conceptions of affect vary as a function of the location of affect (e.g., as a mental or physiological state or a feature of the stimulus) and also the investigators' theoretical and methodological goals, but each approach contributes to an understanding of the affect-memory relationship as well as to the larger question of the processes underlying retrospective reports.

We identify three approaches to the study of affect and memory. Separate reviews of each literature exist, and it is not our intention to provide an exhaustive review of any of these literatures (Blaney, 1986; M. S. Clark, Milberg, & Erber, 1988; Rapaport, 1942/1971; Revelle & Loftus, 1990; Zajonc, 1980). Instead, by treating these traditionally separate investigations concurrently, it is possible to observe the contribution of each approach in relation to others in understanding the larger question of the role of affect and memory in retrospective reports. We focus in particular on an aspect of the affect-memory relationship that can be traced to the earliest experimental research in psychology: What are the influences of the affective properties of an experience on memory for that experience? Do we remember pleasant events with greater facility than we do unpleasant events, or are events associated with greater affective intensity, both pleasant and unpleasant, remembered better? We discuss data that address this longstanding and controversial question in psychology, focusing on two variables that characterize the affective experience, its valence and its intensity. Finally, we speculate about the role of affect on explicit and implicit measures of

past events based on our current knowledge of demonstrable affect–memory findings.

## Experimental Approaches to Investigating the Affect–Memory Relationship

We recognize three ways in which the variable “affect” has been conceptualized and operationalized in contemporary experimental examinations of the affect–memory relationship: *affect as mood state*, *affect as arousal state*, and *affect as evaluation*. Both the mood and arousal approaches locate affect as an ambient state within the individual, although the former has focused on valence and the latter on the intensity of the mood state. Some research on evaluation has viewed affect as a feature of the judgment task (e.g., How much do you like this stimulus?), allowing a comparison of the sensitivity of affective and cognitive judgments to events that have occurred in the past. More typically, research on evaluation has viewed affect as a property of the stimulus event, that is, most commonly as a property of the verbal information for which memory or its context is tested.

### Affect as Mood State

The mid-1970s witnessed a large-scale effort to investigate the effects of mild positive and negative moods on memory. In fact, the mood–memory approach to understanding the influence of affect and memory has been so dominant that the term “affect” is often used synonymously with “mood.” The experimental strategy has involved manipulating mood states at encoding and retrieval and testing memory for material that is evaluatively congruent, incongruent, or neutral relative to the mood state. This basic procedure has allowed investigators to test four kinds of effects: (a) effects of mood states on learning, (b) effects of mood states on retrieval, (c) effects of the match between mood states at learning and retrieval, and (d) effects of the match between mood and material valence. In the large body of research that has accumulated, there is some evidence to support each of these postulated relationships between mood and memory (for reviews, see Blaney, 1986; Bower, 1981; Isen, 1984; Singer & Salovey, 1988), although some effects have proved unreliable and others are known to occur only under limited conditions (Blaney, 1986; Bower & Mayer, 1985).

Two distinct responses to the mixed findings on the influence of mood on memory have emerged, both of which appear quite promising. One approach has capitalized on implicit memory measures to study the mood–memory question. Implicit memory refers to memory effects obtained on tests that do not require the subject to refer to the earlier learning episode (Schacter, 1987). For example, following initial exposure to stimuli (e.g., words), subjects are asked to perform a task on an apparently unrelated set of stimuli, which contains, perhaps in modified form, both previously seen (old) and new items. Facilitation or inhibition on

old items compared with new items on tasks such as perceptual identification, word-fragment completion, free association, and evaluative judgments are interpreted as evidence for implicit memory. The surprising result obtained in several programs of research is that implicit tests reveal effects of prior exposure in the absence of the subject's ability to recall or recognize those items (Richardson-Klaven & Bjork, 1988; Roediger, 1990).

Preliminary research by Tobias, Kihlstrom, and Schacter (1992) compared explicit and implicit measures of memory for material learned and retrieved under varying mood conditions. They found that mood did not influence memory as measured by traditional free and cued recall. However, when subjects were asked to write down the first words that came to mind (under the guise that words had been presented subliminally during mood induction), results revealed the influence of mood on memory. Specifically, better memory for mood-congruent words was found when mood was matched at learning and retrieval compared with unmatched conditions. Thus, although mood did not affect explicit memory, it did produce mood-dependent effects on the implicit measures for mood-congruent words. These findings represent the first attempt to identify mood effects on implicit memory measures, and if these results survive further empirical scrutiny, they will demonstrate that mood can influence memory when subjects do not explicitly remember the learning episode.

The second approach, which has received much empirical support, is offered by Schwarz and his colleagues (Schwarz, 1990b; Schwarz & Clore, 1988). They found that a global mood state influences judgments about seemingly unrelated events. For example, subjects unknowingly misattribute the effects of rainy weather to the quality of their life; that is, they rate the quality of their life as being worse when questioned during rainy weather. For bad moods caused by rainy weather, this misattribution effect is eliminated if subjects rate the weather before performing the quality-of-life judgment. The theoretical interpretation of these findings hinges on the informational value of mood; that is, mood at the time of judgment may be mistakenly used as a cue unless it is obvious that the source of the mood is irrelevant to the task at hand.

This analysis is similar to recent demonstrations by Jacoby and his colleagues (Jacoby & Kelley, 1987; Jacoby, Kelley, & Dywan, 1989) of unconscious influences of the past on memory. They have shown that there are strong misattribution effects on judgment, caused by perceptual familiarity from previous exposure, in the absence of episodic memory for the stimulus. For example, subjects exposed to the names of nonfamous people were later presented these names and new nonfamous names, as well as the names of famous people. In a task that required them to identify the names of famous people, subjects were more likely to falsely judge old nonfamous names as famous compared with new nonfamous names. This increased false alarm rate for familiar nonfamous names occurred only when the subject no longer retained explicit memory for the name but nevertheless retained some degree of perceptual fluency with the name. In other words, familiarity with the nonfamous names was misattributed to fame. Just as Schwarz (1990b) demonstrated the effects of feelings-as-

information biases in judgment, Jacoby and his colleagues have documented how the quality of memory acts as a source of information that leads to misattribution biases in judgment. The relevance of this line of research for retrospective reports is that mood at the time of retrieval may affect judgments that are causally unrelated to the mood. Further, because mood effects on judgment occur in the absence of evidence of explicit mood-congruent memory, these findings add to the small literature demonstrating, under some conditions, the independence of explicit memory and judgment.

Although the approaches of Tobias et al. (1992) and Schwarz (1990b) are quite distinct in purpose and design, they are similar in what they reveal about the role of consciousness in judgment. In the research on mood-implicit memory, subjects were not asked to refer to the learning episode but rather to generate the first word that came to mind. Likewise, subjects in the research on mood as information were deliberately not asked to attend to the mood present at the time of judgment. In both cases, effects of mood on memory or judgment were obtained when subjects did not or could not make the correct attributions for a particular psychological experience (e.g., a physiological state, a memory, a mood). Such implicit effects are gaining increased attention in the study of social behavior, where misattribution effects in memory (e.g., due to familiarity through exposure) or affect (e.g., due to the informational value of mood valence) may exert more pervasive effects on social memory and judgment than previously recognized (see Brody, 1987).

To investigators interested in the validity of retrospective reports, unambiguous findings of mood effects on memory could prove very useful. If what can be remembered through an explicit attempt to recall a past event is a function of the mood state at learning and/or retrieval, or the match between mood and material valence, then these variables would be implicated in predictions of the direction and strength of retrospective reports obtained in contexts with known affective properties. Although an understanding of mood and memory will be necessary for a complete understanding of the various influences on retrospective reports, at the present time the empirical uncertainty of several mood and explicit memory effects does not allow clear predictions.

Findings reported by Schwarz (1990b) and the preliminary results of Tobias et al. (1992), on the other hand, suggest an alternative to the current emphasis on retrospective reports that require the subject to refer to a prior target episode. It is possible that implicit measures (including preference judgments of the kind obtained by Zajonc, 1980) are more sensitive to past events and may occur reliably even in the absence of an explicit retrospective report of the event. As observed in Banaji and Greenwald (1991) about the measurement of stereotypes and attitudes, more reliable effects may be obtained when the feature of the stimulus driving the attitude is not the direct focus of judgment. Implicit measures of past events, as a particular form of indirect measures, may be particularly useful if social-desirability concerns are likely to compromise accuracy in self-reports (Dovidio & Fazio, 1992) or if subjects are unaware of internal or external influences on their thoughts, emotions, and behavior.

## Affect as Arousal State

As an alternative strategy to examining the influence of valenced mood state on memory, some investigators have viewed affect as the intensity dimension of an arousal state. In some studies, the arousal is generated by a verbal or pictorial stimulus or stimulus context (e.g., neutral words presented in the presence of arousing words or an arousing film), through an ambient arousal manipulation (e.g., white noise), or by physically involving the subject (e.g., exercise). In some studies, the primary interest is in a feature of the test situation, such as the time delay between learning and test, whereas in others the interest is in the effects of arousal-state-dependent memory. In research on arousal, measures of memory have exclusively been explicit tests, and the stimuli have typically consisted of affectively neutral items.

Research on arousal and memory gained attention as a result of an intriguing result reported by Kleinsmith and Kaplan (1963, 1964). They found that affectively neutral items (e.g., digits, nonsense syllables) linked to affectively neutral control words in a paired-associate task showed the expected forgetting curve when recall for the items was measured at various time delays, ranging from 2 minutes to a week. However, items linked to high-arousal words (measured by an ipsative analysis of each subject's galvanic skin responses) showed poorer retention on an immediate test compared with items associated with neutral words. After a delay, however, items linked with high-arousal words showed better retention compared with the neutral condition. More surprisingly, items associated with high-arousal words were better remembered after a delay than immediately. Both findings are of interest. Why is performance in the arousal condition poorer than (a) the control condition after a short delay and (b) the delayed recall condition?

The explanation offered for the interaction hinges on the notion of differential rates of consolidation for material learned under conditions of high and low arousal. This interpretation, offered by Kleinsmith and Kaplan (1963), assumes that the greater the "consolidation of the neural trace," the better the resulting memory; and while consolidation is ongoing, retrieval of that information is inhibited. Items linked to high-arousal material benefit from the higher "reverberation" of such items, which ultimately leads to their superior consolidation as supported by the results at delay. However, if retrieval is attempted while consolidation is ongoing, the consolidation process interferes with successful retrieval. Although no satisfactory evidence for this hypothesis itself exists, and the hypermnnesia in the high-arousal condition is not always obtained, there is considerable evidence from a wide variety of procedures and materials that supports the interaction of arousal and time delay on memory (cf. Baddeley, 1990; Eysenck, 1982; Revelle & Loftus, 1990).

More recently, some investigators have become interested in whether arousal produces state-dependent memory effects (M. S. Clark, 1982; M. S. Clark, Milberg, & Erber, 1984; M. S. Clark, Milberg, & Ross, 1983). Manipulating arousal at learning and at retrieval (e.g., through physical exercise or sexually

explicit films), Clark and her colleagues found consistent support for arousal-state-dependent memory. Material learned in an autonomically aroused state was recalled better when subjects were also aroused at retrieval; likewise, material learned in a nonaroused state was recalled better when subjects were nonaroused at retrieval. These findings suggest that the relative fragility of mood-state-dependent memory effects (see Bower & Mayer, 1985) may be a function of the fragility of mood valence but not of mood intensity. Recently, E. Eich and Metcalfe (1989) reported limited support for this hypothesis. In their experiments, the valence of the mood alone produced state-dependent effects on memory for internally generated events, although the largest discrepancies in *both* mood and arousal between learning and retrieval events resulted in the poorest recall, suggesting that arousal may be a factor in obtaining the strongest state-dependent effects of affect on memory.

In addition to showing the effects of arousal on memory, Clark and her colleagues have demonstrated that arousal biases perceptions of the emotions of others and other social judgments (see M. S. Clark et al., 1988). This suggests that arousal is a theoretically important variable, too often ignored in research on affect and memory. In particular, Clark et al.'s (1988) research suggests that biased retrospective reports can be expected when obtained under naturally occurring conditions of heightened or depressed states of affective arousal produced through mood states, drug or alcohol states, psychological depression, or situational factors such as accidents or examinations. The more robust findings of arousal compared with mood-valence effects on memory are especially striking when arousal-state-dependent memory effects are obtained with quite distinct manipulations of arousal at learning and retrieval—for example, exercise and sexually explicit films (M. S. Clark et al., 1983).

For retrospective reports, the finding that memory is enhanced for neutral material associated with arousal-producing information is relevant. The consolidation slope for arousing material (compared with neutral material) is lower, but the asymptotic value is higher. Thus, material associated with arousal-producing information should be remembered better over time and could potentially influence judgments if the judgment requires explicit memory for the critical information. The finding of an interaction between arousal and time delay has serious implications for retrospective reports of arousing experiences—for example, those that may constitute eyewitness testimony in a courtroom.

### Affect as Evaluation

The preceding discussion indicates that investigations of affect and memory have often operationalized affect as a mood or arousal state in which material is learned and retrieved. Affect takes the form of mood valence or arousal intensity at learning and retrieval, or the match in stimuli and mood valence. This conceptualization of affect, however, represents a relatively recent approach to investigating affect. Historically, the construct "affect" has referred to a property of the information to be remembered and, occasionally, as a property of the

judgment task. Social information is distinctly evaluative in quality. As Osgood, Suci, and Tannenbaum (1957) pointed out, the evaluative component of information accounts for one-half to three-quarters of the variance in extracted meaning. Judging from the nature of the stimuli and dependent variable measures, the evaluative dimension of information is inherent in the fabric of social-psychological phenomena: self and person perception, attributional processes, stereotypes, intergroup perception, and, most obviously, attitudes. The evaluative properties of social knowledge distinguish social-psychological phenomena, although little research has expressly tested the effects of evaluative items on memory. Evidence for the influence of affect as evaluation on memory may be discussed in two parts. First, there are demonstrations of the sensitivity of affective over cognitive judgments as indicators of past events. Second, there are demonstrations of the effects of information valence and intensity on memory.<sup>1</sup>

### *The Sensitivity of Affective Judgments*

Zajonc and his colleagues (Kunst-Wilson & Zajonc, 1980; Moreland & Zajonc, 1979; Zajonc, 1980) demonstrated that prior exposure to neutral stimuli increases the likelihood of preference discriminations (i.e., judgments of liking) in the absence of recognition for the stimuli (but see Brooks & Watkins, 1989, for a critique). Based on this apparent dissociation between judgments of affect and cognition on previously seen stimuli, Zajonc (1980) argued that affect and cognitive measures contribute independent effects to the processing of information. In subsequent research, Mandler, Nakamura, and Van Zandt (1987) demonstrated that after exposure to affectively neutral information, a nonaffective judgment such as a brightness rating was also a more sensitive measure of exposure than of recognition.

Such effects of mere exposure can be interpreted within the more recently available framework of explicit and implicit memory. The recognition test, an explicit measure of memory, is less sensitive to the effects of prior exposure than is the liking judgment, which represents an implicit measure of memory because it does not require the subject to refer explicitly to the earlier exposure episode. Thus, the privileged status of affect in sensitivity to prior exposure exists in much the same way that other implicit measures have been shown to have greater sensitivity compared with explicit measures under certain encoding-retrieval conditions (Roediger, 1990). This interpretation of the mere exposure effect should not detract from the importance of the finding of greater sensitivity of an evaluative measure compared with a recognition measure of previous exposure. However, until comparable measures of affect and cognition

---

<sup>1</sup>Research on the effects of the evaluative properties of information on perception has a long and controversial history (Erdelyi, 1974; Greenwald, Klinger, & Liu, 1989; Kitayama, 1990). Although we do not wish to draw a strict distinction between processes of perception and memory, based on considerations of space and scope we restrict the present discussion to research conventionally considered to be on memory.

are available (e.g., two otherwise similar implicit measures, one of affect and the other of cognition), the hypothesis of the greater sensitivity of affective judgments to prior exposure ought to be treated with caution.

The dissociation between affect and cognition observed in research by Zajonc bears resemblance to the findings reported by Schwarz (1990b; Schwarz & Clore, 1988). In both lines of research, an evaluative judgment appears to be more sensitive to factors such as mood or previous exposure without the mediation of explicit memory or when the explicit memory measure reveals chance-level performance. Such effects are important because they demonstrate that retrospective reports, especially those involving evaluative judgments, can be influenced by factors about which subjects are unaware.

### *Effects of Affective Valence and Intensity*

In addition to research that examined the effects of affective (preference) versus cognitive (recognition) judgments, the affect–memory relationship has been investigated most by examining the influence of the affective properties of information, specifically valence and intensity, on memory. The tests of memory are of the traditional explicit variety, primarily free recall and recognition, and the focus is on the relative effects on memory of information of varying affective valence and intensity. We turn now to an account of the history of this research and to our own research on the affect–memory relationship.

The first major theoretical articulation of the relationship between affect and memory in psychology is found in Freud's well-known theory of repression (Freud, 1900/1965). Simply stated, ego-threatening information is relegated to an unconscious store from which it cannot be easily retrieved into conscious awareness. Thus, memory for ego-threatening information is poor compared with nonthreatening or ego-enhancing information. The intellectual excitement surrounding psychoanalytic theory and Freud's (1901/1960) own claim that the mechanism of repression operated in ordinary, everyday behaviors created in the experimental psychologists of the day an interest in tests of the repression hypothesis. Through the 1930s and 1940s, an empirically tractable form of the hypothesis motivated much experimental research on the influence of affect on memory (Rapaport, 1942/1971).

Two distinct methodological strategies were used in this research, and the issues that separated them remain today. The first placed a premium on the ecological realism of the events for which memory was tested. Thus, memory was tested for events that had occurred in the life of the subject outside the laboratory. For example, Jersild (1931) asked subjects to record all pleasant and unpleasant experiences of the most recent 3 weeks. When asked 3 weeks later, subjects remembered more pleasant than unpleasant experiences that they had generated in the prior session. In an early influential review, Meltzer (1930) reported that experimental evidence favored the finding of superior memory of affectively pleasant compared with unpleasant everyday experiences. He further argued that the self-relevance of the event was central to tests of repression,

which necessarily involved the investigation of events that had a continuing reality in the subject's life. The finding of superior memory for pleasant everyday events continues to be reported (M. Linton, 1975; Matlin & Stang, 1978).

Despite the apparent success of this method, it posed several threats to the internal validity of the findings. In particular, the lack of control over the initial encoding of the events and the previously established affective value of the to-be-remembered material led some to question the validity of the findings obtained from experiences that had occurred outside the laboratory. Such challenges to interpretation of the findings led Barret (1938) to suggest that memory for experiences over which experimental control could not be exerted shed no light on the affect-memory relationship. To establish better control over the encoding and retrieval of the affective experience, pleasant, unpleasant, and neutral words were learned and tested under controlled conditions. In contrast to studies inspired by Meltzer (1930), the predominant conclusion from studies in this tradition was that memory for affectively charged material, whether pleasant or unpleasant, was superior to memory for affectively neutral material; that is, the intensity of the affective experience, not valence, predicted memory. This method, however, was not without its problems. Besides eliciting some questions about the generalizability of these findings to typically occurring affective events, methodological problems plagued these studies as well. Properties of words such as affective intensity, frequency, and imagery, which are now well established as predictors of memory (Pavio, 1969; Rubin, 1980), were uncontrolled, rendering ambiguous the effects of affect on memory.

In summary, two empirical findings emerged from the early research, each corresponding to the use of a particular methodology. With an emphasis on ecological validity of method, examinations of memory for episodes that occurred outside the laboratory largely supported the hypothesis of *affective asymmetry*, illustrated in Figure 5.1. In contrast, an emphasis on experimental control over the encoding experience led to examinations of memory for episodes created in the laboratory. These experiments largely supported the hypothesis of *affective intensity*, illustrated in Figure 5.2.

In the 1950s, procedures to study the influence of affect on memory involved tests of affectively neutral stimuli such as nonsense syllables, attached to pleasant or unpleasant experiences such as success or failure (Zeller, 1950). Such studies represented an improvement over methods used in previous decades of research because memory was tested for material with no prior affective value. Although some studies reported affective asymmetries in memory, touting support for the repression hypothesis, others did not (e.g., Aborn, 1953); and by the mid-1960s, the approach was abandoned, leaving some believers and others unconvinced of empirical support for repression (cf. Erdelyi & Goldberg, 1979; D. Holmes, 1972; Kubie, 1952; Weiner, 1966).

In a monograph on the relationship between affect and memory, Dutta and Kanungo (1975) provided evidence to support affective intensity and offered an interpretation of empirical effects of affective asymmetry in terms of affective intensity. In one of their experiments, subjects learned lists of positive and neg-

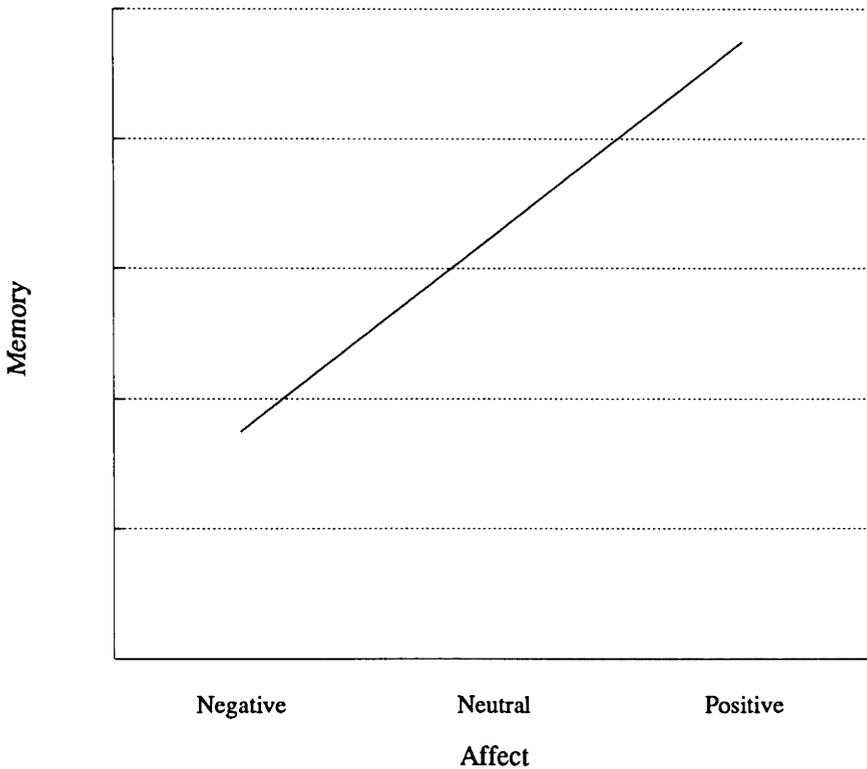


FIGURE 5.1. Affective Symmetry

ative adjectives that were either attributed to an ethnic in-group or a fictitious ethnic out-group. Dutta and Kanungo found that positive adjectives attributed to the in-group were remembered better than negative adjectives. However, for adjectives attributed to the out-group, negative adjectives were remembered better than positive adjectives. Importantly, ratings of the intensity of adjectives revealed that for adjectives attributed to the in-group, positive adjectives were more affectively intense than negative ones. For adjectives attributed to the out-group, negative adjectives were more intense than positive ones. From findings of this sort, Dutta and Kanungo argued that affective asymmetries, whether positively or negatively biased, are actually a product of affective intensity.

In response to the tension between experimental control and ecological validity of the task in tests of the affect-memory relationship, one of us (Banaji, 1986) developed a procedure designed to retain the advantages of both. Two issues of method that had posed problems with the early research were addressed. First, experimental control over the encoding of the affective experience was established. Second, the primary measure of memory allowed a test of the influence of affect uncontaminated by previously established affective value

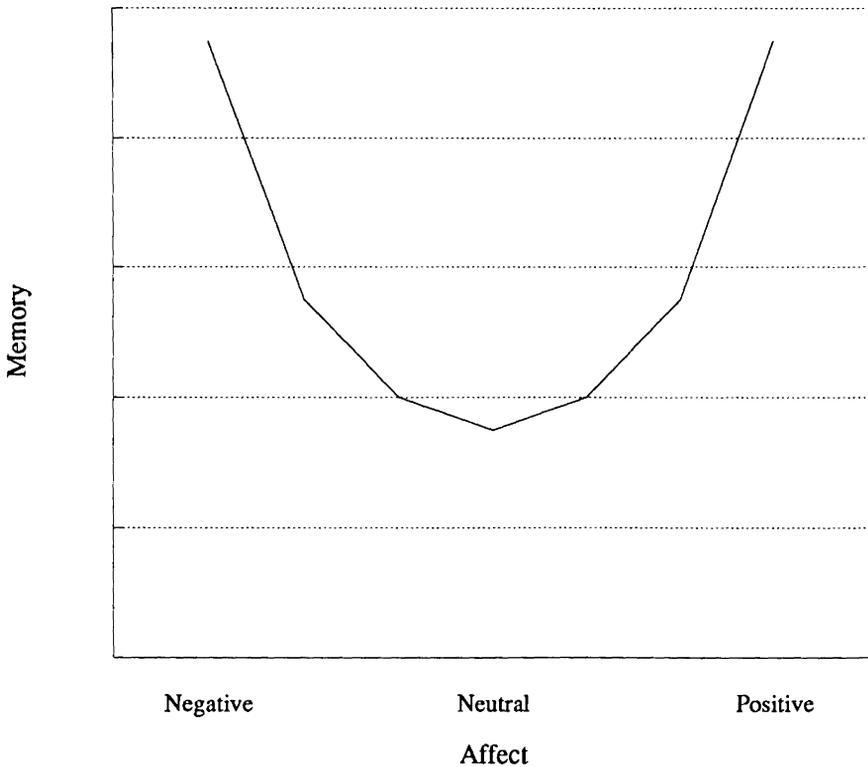


FIGURE 5.2. Affective Intensity

of the material. To establish control over the initial encoding of the affective experience, subjects described ordinary but imagined pleasant, unpleasant, and neutral episodes. The specific episodes had not actually occurred in the subject's life, although they were very plausible future events (e.g., losing a valued item, going to Disneyland). Following Meltzer's (1930) claim that self-relevance of the episode is critical to tests of the repression hypothesis, each episode involved the subject as the agent of the action. To create a retrieval measure not susceptible to the interpretative confounds of previous research, memory was tested for an affectively neutral target word embedded in the episode (e.g., window, magazine). Further, because targets were randomly assigned to the three affect conditions for each subject, differential memory across subjects could be attributed more directly to the affective quality of the events in which they were embedded and not to properties of the words themselves.

The procedure involved subjects in a sentence construction task, each sentence representing a specific event. In constructing each event, subjects (a) incorporated themselves as the agent of the action, (b) described a hypothetical event whose affective valence was specified by the experimenter to be pleasant,

unpleasant, or neutral, and (c) incorporated a neutral target noun provided by the experimenter. Pleasant events were happy or exciting, and unpleasant events were unhappy or depressing. For example, a pleasant event using the target word refrigerator was, "I won a brand new refrigerator with a year's supply of food in the Ohio State lottery." An unpleasant event using the same target was, "I found cockroaches all over the refrigerator when I returned home from my vacation." Likewise, a neutral event was, "I opened the refrigerator approximately seven times today." Subjects found the task to be involving and performed it easily. Incidental learning was measured by giving subjects a surprise free-recall test for the affectively neutral target words embedded in the events. After free recall, subjects rated each event on a 7-point scale of +3 through -3 on the degree of affective valence and intensity of each event. This rating allowed a check on the experimental manipulation of affective valence and provided additional data about the affective intensity of each valenced event.

Using this procedure, consistent support for the hypothesis of affective intensity was demonstrated (Banaji, 1986). Free recall for neutral target nouns embedded in affectively intense sentences was better than for targets embedded in less affective and neutral sentences, regardless of valence, reflecting the affective-intensity effect. The intensity effect in memory was obtained when subjects' own ratings of affective intensity were used as the predictor variable. In a variant of the basic procedure, the affective-intensity effect was also obtained for target words embedded in affective events generated by other subjects; that is, when subjects were presented with the events generated by another group of subjects, memory for targets was found to be a function of the affective intensity of the events. The intensity effect was also obtained when an imagined other was used as the agent of the action. In Hardin and Banaji (1990), we replicated the finding of affective intensity when recall and recognition measures were obtained after a 24-hour delay.

We further explored the influence of affect on memory by exploiting the critical role of self-evaluation in the affect-memory relationship. One of us (Banaji, 1986) had identified a relationship between self-esteem and the influence of affective valence on memory. Compared with subjects with moderate self-esteem, subjects with high self-esteem remembered more targets embedded in pleasant events after a 24-hour delay. This result suggested that self-evaluation may moderate the affect-memory relationship. In Hardin and Banaji (1990), we examined this hypothesis more directly by manipulating the importance of events to self-evaluation in the construction of events. We accomplished this by asking subjects to construct hypothetical events in domains empirically identified as important to self-evaluation among college students (e.g., academic performance, social relationships) or unimportant to self-evaluation (e.g., bird-watching, coupon clipping). Instructions requested pleasant events that would reflect positively on the self and unpleasant events that would reflect negatively on the self in both domains. Twenty-four hours after constructing imaginary events in domains that varied in importance to self-evaluation, tests for recognition revealed two findings. When the domain was unimportant to subjects, the

affective-intensity effect was obtained, replicating the finding obtained in several previous experiments. When the domain was identified a priori as important to self-evaluation, target words embedded in affectively positive and neutral events were remembered better than those embedded in affectively negative events. Thus, under conditions in which the information was highly ego-involving, an affective-asymmetry effect was obtained.

## Discussion

In this chapter, we identified three approaches to the study of the affect-memory relationship. Differing significantly in the location of affect and in method, each nonetheless demonstrates that affect does influence memory. These demonstrations qualify affect as a variable of importance in the retrospective report process, because such reports often involve memory for events to which affect is associated. Not only is affect quite obviously a feature of memory for information about affectively charged events such as an abortion bill or an automobile accident, but it is also more subtly present in imagined events featuring oneself or in encountering a familiar face. What makes affect a critical variable here is that it is an intrinsic and often unnoticed feature of everyday social events about which retrospective reports are elicited.

Retrospective reports have, by definition, referred to measures that require the subject to remember explicitly an earlier episode about which memory or judgments are obtained. To recall a visit to a dentist, to judge the difficulty of an exam, to evaluate the honesty of a friend, all require the subject to recall the time and place of target episodes and their contents to produce a response. Yet, some of the data that we have discussed (Schwarz, 1990b; Tobias et al., 1992; Zajonc, 1980) suggest that it is when subjects do not refer to the earlier episode that strong effects of affect are obtained. In the Tobias et al. research, it is the match between the mood conditions at encoding and retrieval that produces superior implicit (but not explicit) memory for material matched in valence; in the research reported by Schwarz (1990b), a mood state at retrieval influences a judgment on a causally unrelated item; and in research reported by Zajonc (1980), an implicit measure of memory, that is, the evaluative judgment, proves to be a more sensitive measure of memory for the prior episode than does an explicit measure of memory.

Findings such as these lend support to the growing literature on the power of implicit memory measures as reliable indicators of memory for past events. The sensitivity of implicit measures suggests that they ought to be introduced as additional or alternative measures when situations justify and permit their meaningful use. For example, implicit techniques may provide a better index of memory for episodes that are not available to explicit recall or when social-desirability factors in responding may pose a threat. Among the many classes of retrospective reports that are obtained, a widely used category represents measures of attitudes. Attitude methodologists have long recommended the use of

indirect or unobtrusive measures (Webb, Campbell, Schwartz, & Sechrest, 1966), which may allow more accurate measures of the attitude. Such measures have not as yet become well established in practice, although more recent programs of research by investigators such as Dovidio and Fazio (1992) alert social psychologists not only to the value of indirect measures but also to ways in which the practice of obtaining such measures may become a future reality. For example, Dovidio and Fazio argue that response latency measures of attitude are more valid estimates of socially sensitive topics, and while acknowledging the difficulty of implementing such indirect assessment measures, they point to the success of using portable computers to record latency data in field settings.

It is not a trivial task to identify how implicit affect-memory measures may be adapted to aid survey investigators. However, a feature of the survey situation may encourage the use of such measures. Among the factors that often threaten the validity of memory measures obtained on surveys is the lack of control over the encoding conditions of the episode to be remembered; that is, information is often requested from subjects about events that occurred in contexts about which little or nothing is known. Transfer-appropriate procedures to study implicit memory (see Roediger, 1990) emphasize the degree of overlap in the operations that were performed at learning and test. For example, material that is conceptually processed (meaning based) should be better remembered if the test is conceptually driven (recall, recognition); likewise, perceptually processed material should show savings on data-driven (perceptual identification, affective identification) tests. Thus, implicit measures (whether they are affect based or not) can provide an additional measure of memory, and if differential sensitivity of the explicit and implicit test is observed, perhaps they can also divulge the nature of the original learning event.

Affect as a property of the information-processing environment in which retrospective reports are obtained may also acquire greater power when explicit retrieval of information is hindered. Research by Schwarz & Clore (1988) suggests that when the judgment task does not easily allow retrieval of a specific episode (e.g., a judgment of the quality of life), the mood state at retrieval may be used as the basis of an evaluation. Likewise, when memory for prior information is not amenable to explicit memory measures, as in the mere-exposure research, an affective judgment (e.g., "How much do you like X?") is sensitive to the perceptual familiarity of the items. The affective features present in a situation may not be easily amenable to conscious verbal report and may therefore prove to be particularly powerful as implicit influences and implicit measures that, by definition, require the subject to be unaware of the prior episode at retrieval or judgment. In these early years of research on the role of affect in implicit memory and judgments, it is difficult to identify when the presence of affect as a mood or as an evaluative judgment will facilitate or interfere with memory or judgment. For example, in the research by Schwarz (1990b), mood (affect) was shown to produce a *biasing* effect on judgment, whereas the implicit evaluative (affect) measure in the mere-exposure research demonstrated the greater *sensitivity* of the affective measure.

Although important aspects of the affect–memory relationship are revealed through implicit measures, much of the data that we have presented demonstrates that affect influences memory on direct, explicit memory measures. Some findings indicate the more influential role of affective intensity in memory (Banaji, 1986; Dutta & Kanungo, 1975), and other findings suggest the prominence of pleasant over unpleasant events in memory (Hardin & Banaji, 1990; M. Linton, 1975; Matlin & Stang, 1978; Meltzer, 1930). Such findings, we know, often emerge when memory for events outside the laboratory is tested, and questions of interpretation raised by Barret (1938) remain even today. When encoding conditions are unspecified, when time delay between the event and memory test is uncontrolled, and when other variables that may be correlated with pleasant and unpleasant events are unidentified (e.g., rehearsal), there is a tendency for pleasant events to be reported more than unpleasant events. We treat this finding with some skepticism because it is unclear whether the causal factor is affective valence or a multitude of other variables correlated with valence. For those investigators whose goal is to identify the causal factors that produce asymmetry, research must continue to examine the correlates of affective valence and the mechanisms by which affective valence guides memory. However, if the goal is to predict memory for everyday events in the life of a subject—for example, when a survey requires retrospective memory reports of affective events—both intensity and valence ought to be considered as agents of influence.

The important effect of affect on memory, in our judgment, is affective intensity. We know that the match in arousal at encoding and retrieval shows better memory for affectively neutral information (M. Clark et al., 1988), and arousal at encoding also produces better memory for information associated with it after a delay (Revelle & Loftus, 1990). In addition, our own findings suggest that the affective intensity of an event influences immediate and delayed memory for the event. In the widely differing procedures used in investigations of arousal (Revelle & Loftus, 1990) and evaluation (Banaji, 1986), there is evidence that intensity of affect manipulated at the time of encoding alone is sufficient to produce benefits in memory. Together, these findings strongly suggest that affective intensity produces superior memory for information to which it is associated.

We emphasize the importance of the finding on affective intensity because the dimension of intensity appears to have been ignored when compared with affective valence in investigations of the affect–memory relationship. Affective intensity is particularly important because it produces robust and reliable effects on memory across a variety of experimental tasks. For survey investigators, these findings suggest that affective intensity, as an aroused state or as a property of the information event, can influence retrospective reports such that memory for events that have occurred under some conditions of affective intensity will show superior memory.

*Acknowledgments.* We thank Paula Niedenthal, Alex Rothman, Norbert Schwarz, and Seymour Sudman for comments on a previous draft.

**Part II**  
**Retrospective Reports**  
**of Behaviors**



# 6

## The Recall of Physical Pain

*Peter Salovey, William J. Sieber, Jared B. Jobe,  
and Gordon B. Willis*

Pain, especially chronic pain, has been called the most universal form of human stress (Turk, Meichenbaum, & Genest, 1983). Millions of Americans suffer from pain-related problems, often resulting in partial or complete disabilities. The medical and scientific communities, as well as government agencies such as the Public Health Service, all recognize the need for better quality data on the prevalence and severity of chronic pain in the population. In order to address the personal and societal consequences of chronic physical pain, accurate data regarding pain must be collected.

The collection of such information generally relies on retrospective reports of pain sufferers about the nature and intensity of their pain and its treatment. As addressed by many of the chapters in this volume, autobiographical accounts of all kinds are likely to be inaccurate in a variety of ways. Respondents can have difficulty recalling objective events on surveys, such as the years that they served in the Armed Forces, the price paid for their homes, or the birth dates of their children. At least with objective information, however, survey data can be confirmed by external observers; often it is merely a matter of public record. On the other hand, memory for subjective experiences such as pain may be especially vulnerable to error; and because the "truth" is only known to the respondents themselves, the quality of such data is often difficult to ascertain by external observers (see C. F. Turner & Martin, 1984, vol. 1, for a more thorough discussion of these issues).

This chapter describes several factors influencing accurate judgments about and recall of pain and painful experiences. Four specific factors associated with judgments about and recall of pain are explored: (a) the role of a respondent's mood state in the recall of past experiences with pain; (b) the impact of a respondent's present pain severity on the recall of the intensity and quality of past episodes of pain; (c) the relative susceptibility to cognitive biases in recall of pain intensity, pain behaviors, and changes in daily activities; and (d) difficulties in the use of language by respondents to describe pain and painful experiences.

## The Reporting of Pain on Surveys

### The Recall Accuracy Literature

Despite its pervasiveness, accurate statistical accounts of the personal and societal impact of pain have been plagued by problems of language and recall accuracy (e.g., Fienberg, Loftus, & Tanur, 1985). Most self-report instruments and surveys suffer from difficulties in matching descriptive language to the perceptual qualities of pain experiences. Further, recall of pain experiences, episodes, and intensity may not be accurate. However, the degree (and even the direction) of bias has been the source of considerable controversy. Although the majority of researchers claim that individuals generally inflate the amount of pain experienced at a given time when asked to report about it later (e.g., Jamison, Sbrocco, & Parris, 1989; S. J. Linton & Melin, 1982), others claim that individuals can accurately report their pain experiences, at least for several days after the original episode (e.g., Hunter, Philips, & Rachman, 1979).

#### *Retrospective Overestimation of Pain*

A closer examination of some of the studies cited above reveals a fairly consistent pattern of difficulties in accurately reporting past pain experiences. The Hunter et al. (1979) study is often cited as an example of the accuracy of memory for pain. For example, S. J. Linton and Melin (1982, p. 282) claimed that the Hunter et al. results demonstrated that the recall of pain is "surprisingly accurate." Yet, Hunter et al.'s own data do not support this view. They divided 16 headache patients into two groups. One group reported pain after 5 days and the other after 1 day and then again after 5 days. The intensity and quality of pain reported after 5 days were similar to those reported by the patients at the time of their initial interview while experiencing the headache. However, 5 days is not a long time interval during which much decay in vivid pain-relevant memories should be expected. Yet, even after just 5 days, there was only a 70% overlap in pain terms used when patients described their pain on the McGill Pain Questionnaire (MPQ), a measure of pain that tries to separate and quantify three distinct components of pain: sensory-discriminative, motivational-affective, and cognitive-evaluative (see Byrne et al., 1982; Leavitt, Garron, Whisler, & Sheinkop, 1978; McCreary, Turner, & Dawson, 1982; Turk, Rudy, & Salovey, 1985). Further, 5 of the 16 patients, labeled "shifters," recalled only 30–50% of their initial pain ratings.

Hunter et al. (1979) considered these accuracy rates as evidence for the lack of bias in retrospective pain ratings. They noted that this level of accuracy was "similar to that for the recall of incidental material" (p. 43) such as the interviewer's name. However, a pain experience, such as an intense headache, is much more salient, vivid, personally relevant, and affectively charged than incidental information such as the interviewer's name and, hence, should be recalled more accurately. The fact that it was recalled only with approximately

equal accuracy to incidental information is not as positive a finding as Hunter et al. claim. Further, over 30% of their subjects exhibited dramatic shifts in pain reporting over the 5-day period. When considered in terms of a national health survey, major inaccuracies by 30% of the respondents would have a serious impact on the quality of the survey results.

Similar studies using the MPQ as the measure of pain have yielded low correlations between initial reports and later memories. After only 7 days, reports of pain by rheumatoid arthritis patients on the MPQ were only modestly correlated with initial reports, although reports of a single experience of ischemia pain were more accurately recalled (Roche & Gijbsers, 1986). In general, however, individuals who are more disabled by their chronic pain problems are more likely to overestimate their pain when asked to recall its intensity (Jamison et al., 1989). It may be that implicit theories about how pain has changed their lives motivate individuals, in a sense, to rewrite their personal histories—in this case, to remember the original painful episode as extremely severe (cf. M. Ross, 1989).

Other studies also provide evidence for the idea that retrospective recall of pain is often inaccurate and usually exaggerated. S. J. Linton and Melin (1982) studied 12 back and joint pain patients undergoing a 3- to 11-week treatment and found that ratings of pain at the initiation of treatment were higher when estimated at termination compared with ratings actually made prior to the beginning of the program. Of course, patients may have been motivated to inflate their estimates of pretreatment pain in order to feel that the effort expended in the treatment program was justified (recall the idea of M. Ross, 1989, of the rewriting of personal histories, discussed above). In a similar study, S. J. Linton and Göttestam (1983) had patients rate their pain on both a verbal scale (0 to 5 with each point labeled) or a visual analogue scale (100-mm line). After 4 to 9 weeks, 12 of the 15 patients recalled their pain at baseline as being more intense than they had actually reported it to be, and the biggest discrepancies were noted on the visual analogue scale. This latter effect is probably due to the retrieval cue provided by the verbal content of the 5-point scale. Similar effects have been obtained for recurrent pain, such as headache or menstrual cramps (Rachman & Eyrl, 1989).

Kent (1985) asked dental patients to rate pain expected prior to a dental procedure and actual pain experienced as a result of the procedure and then to recall 3 months after the dental appointment the amount of pain that they had experienced. Only modest correlations between recalled and actual pain were obtained ( $r = .42$ ), indicating that factors other than the initial pain experience accounted for most of the variance in pain recall after 3 months. Interestingly, among individuals who were not anxious about dental work, the recalled-experienced correlation was much higher ( $r = .79$ ), but there was virtually no correlation between recalled and experienced pain among highly anxious individuals ( $r = -.11$ ). In general, recall drifted in the direction of anxiety; that is, highly anxious individuals remembered the pain experience as much more severe than it actually was.

*Retrospective Underestimation of Pain*

Compared with the findings of the studies reviewed above, there are occasions when retrospective accounts of pain may be underestimates of actual pain. This situation seems to result when the consequences of pain produce considerable positive affect. For example, an athlete straining to win a gold medal at the Olympics might complete a marathon despite intense pain and then later report that she had not experienced much pain at all. Perhaps the prototype of this sort of pain is childbirth. Labor is thought to be quite painful, yet most mothers are unlikely to dwell on the intensity of such pain after the child is born. Norvell, Gaston-Johansson, and Fridh (1987) found that pain ratings on visual analogue scales made during three phases of labor revealed considerably more intense pain than did retrospective ratings made 2 days postpartum. It may be that when pain intensity is low—for example, several days after the birth of a child—memory for the original experience of pain is underestimated; but that when existing pain intensity is high or can be high, as in a chronic pain patient, memory for the original pain experience is augmented (E. Eich, Reeves, Jaeger, & Graff-Radford, 1985). Alternatively, the affective state of the mother during labor (fearful, anxious) and her affective state after the birth of the child (joy, relief) are so incongruent that there may be considerable interference with recall of material in the latter state that was encoded during the former, that is, there will be no state-dependent memory facilitation (cf. Bower, 1981; Pearce et al., 1990).

We shall develop more fully below these two themes, that recall and judgment about prior pain may be influenced by the respondent's emotional state and that they are also affected by the intensity of present pain. It is clear, however, that ongoing moods and emotions and the intensity of present pain experiences are the two most frequently mentioned variables thought to explain findings in studies in which considerable recall inaccuracy is obtained.

### Emotional Arousal and Pain Recall

Although often overlooked, there is increasingly strong evidence that mood affects the manner in which information is encoded, organized in memory, and ultimately retrieved (for reviews, see Blaney, 1986; Singer & Salovey, 1988). Mood has an especially strong impact on the processing of affectively laden information. Painful experiences are certainly affectively charged events, and memories about them may be organized around their affective qualities. The role of moods and emotions in the accuracy of judgments about and recall of painful experiences has not been studied directly (although for an exception see E. Eich, Rachman, & Lopatka, 1990) despite the fact that many researchers (e.g., E. Eich et al., 1985; Hunter et al., 1979; Kent, 1985, 1989; Norvell et al., 1987; Roche & Gijsbers, 1986) have suggested it as the key variable mediating the accuracy of pain recall.

The most frequently cited causes for inaccuracy in the recall of pain episodes

are affect related. As described earlier, respondents who experienced fear of going to the dentist remembered dental pain as more severe than it actually was. Mothers experiencing the joy of giving birth to a new baby tended to under-report the intense pains of labor after the baby was born. On the other hand, individuals experiencing depressed moods, perhaps because their pain problems had not been alleviated, tended to overestimate the intensity of previous painful experiences. Each of these situations exemplifies the important role played by affect in the recall of pain. Despite its common endorsement as an important factor resulting in inaccurate pain reporting, survey researchers have tended to give scant attention to affect in the design of surveys and in the interpretation of their results. Yet, in the past decade, cognitive and personality/social psychologists have become intrigued by the role played by moods and emotions in the processing of information (Blaney, 1986; Isen, 1987; Mayer & Salovey, 1988; Schwarz, 1990b; Singer & Salovey, 1988).

There are three different ways in which ongoing affective states might bias the recall of pain experiences. One could be called "mood-congruent pain reporting," in which individuals' current mood states, perhaps by influencing the accessibility of positive versus negative memories, directly influence ratings in a direction consistent with the mood (cf. Bower, 1981; Isen, 1984). The second bias may occur when mood at the time of the pain rating does not match the individuals' initial mood during the pain experience. This mismatch in affective context does not allow the individual to experience the memory benefits of state-dependent learning and recall. These two roles for affect in biasing pain recall correspond to what has been termed "thought congruity" (or "mood-congruent recall") and "state-dependent recall" in the memory literature (cf. Bower & Mayer, 1989; Gilligan & Bower, 1984). Thought congruity can be described as the phenomenon whereby respondents' thoughts, free associations, fantasies, interpretations, and judgments are thematically congruent with their mood state. State-dependent recall concerns the superior memory that occurs when the recall mood state matches the learning mood state. The third way in which emotional arousal can bias retrospective pain recall is through its impact on attentional focus. Depressed moods may affect allocation of attention, turning it from the environment on to the self, the so-called "depressive self-focusing style" (Ingram & Smith, 1984; Pyszczynski & Greenberg, 1986, 1987; T. W. Smith & Greenberg, 1981; T. W. Smith, Ingram, & Roth, 1985).

Associations between mood and recollections about previous painful experiences can also be understood from the perspective of Schwarz's (1990b) "feelings-as-information" model. When asked how much pain they experienced previously, individuals may use their present emotional feelings as a source of information about this judgment. "Well, I feel pretty good right now; I guess my pain wasn't so bad," they may think to themselves. So long as individuals are not specifically led to understand that their feelings are not relevant to this memory judgment (e.g., by getting subjects to notice how their present mood is actually a function of, say, the weather outside), they may use present feelings as the basis for interpretations about the past.

*Mood-Congruent Pain Reporting*

Let us examine associations between mood and pain ratings first. An observation made by mental and physical health care professionals alike is that patients who report psychological distress also complain of a variety of physical symptoms (Katon, 1984). In particular, complaints of diffuse aches and pains are especially likely among patients experiencing dysphoric moods. Overall, perceived health status varies directly with degree of dysphoria (Tessler & Mechanic, 1978). One question stimulated by the co-occurrence of psychological distress and physical symptoms is the direction of this relationship. Although it is probably true that chronic and irritating physical symptoms may lead to depressed mood (perhaps because of interference with daily activities and the induction of helpless thoughts and feelings), mood may also affect perceptions of pain.

Croyle and Uretsky (1987) reported a study in which they induced happy and sad moods in the laboratory and noted that sad subjects perceived themselves to be less healthy following negative mood induction. More directly relevant to the present chapter is a study, reported in Salovey and Birnbaum (1989), in which 66 individuals suffering from influenza were asked to experience either a happy, sad, or neutral laboratory-induced mood. We later assessed the aches, pains, and other symptomatic discomforts experienced by these individuals. Two relevant findings emerged. The first was that mood had its most powerful impact on measures of aches and pains compared with other symptoms of the flu (e.g., nasal congestion, GI distress, sleepiness). The second was that reports of aches and pains varied depending on subjects' assignment to mood condition. Subjects induced into mildly sad affective states in the laboratory reported considerably greater pain than did neutral mood (control) subjects. Conversely, happy subjects reported fewer aches and pains.

Results consistent with these have been reported in the pain-recall literature. For example, consider the subjects described by Hunter et al. (1979) as "shifters," those subjects whose recall of headache pain was most biased after 5 days. Shifters tended to use significantly more negative affective words to describe their pain than did other patients. Similarly, Kent (1985) noticed the most memory distortion for dental pain among his subjects who reported the greatest dysphoric affect associated with dental procedures. His subjects tended to distort their recall for dental pain in a direction consistent with their anxiety.

In our own work on mood and pain recall, we asked healthy adults to participate in a laboratory experiment in which happy, sad, and neutral moods were induced (Salovey et al., 1989). Subjects were asked to recall and rate a painful incident from the past, to read stories about painful episodes and rate the likely intensity of such pain, and to estimate the number of days during the previous year on which they themselves experienced several different kinds of pain. Among female subjects, there was some evidence for the expected mood-congruent pattern of pain reporting. Women tended to rate previous painful experiences and hypothetical painful episodes as more frequent and intense,

respectively, when made to feel sad, and their ratings were attenuated when made to feel happy. Males did not show this pattern of pain reporting at all.

### *Mismatch in Affective Context*

The second way in which mood may affect pain-related memory is that accurate recall is promoted when mood at the time of initial encoding matches mood at time of recall, the state-dependent memory effect described earlier (Bower, 1981). In any kind of state-dependent memory phenomenon (e.g., J. E. Eich, 1980; J. E. Eich, Weingartner, Stillman, & Gillin, 1975), contextual factors (in this case, mood) serve as discriminatory cues such that when learning and recall contexts match, retrieval is facilitated, and when the two contexts differ, retrieval is inhibited. Several studies have indicated, although the effect is not always obtained, that mood can act as such a contextual cue in learning and recall (Bower, Monteiro, & Gilligan, 1978; Schare, Lisman, & Spear, 1984).

Studies of mood-state-dependent memory lead to the hypothesis that when mood at the time of the initial experience of pain matches mood at the time of pain recall, recall should be more accurate than when there is a mismatch. Thus, for example, in the Hunter et al. (1979) study, higher levels of dysphoric affect at assessment were more strongly associated with "shifting" than was any other variable. Similarly, the positive affect associated with the birth of a new baby does not match the negative affect associated with the actual pain of labor, and postpartum recall of labor pain can be expected to be quite poor (Norvell et al., 1987). Moreover, the poor recall of preoperative pain following surgery for rheumatoid arthritis among Roche and Gijsbers's (1986) patients may be due to their much improved affective state following successful surgery. Many other studies showing poor recall of pain often have relied on initial ratings of pain at a time of heightened dysphoric affect and recall when individuals were feeling much better (e.g., S. J. Linton & Götestam, 1983; S. J. Linton & Melin, 1982).

### *Changes in Allocation of Attentional Resources*

Moods may also be associated with perceptions of physical symptoms and beliefs about health, because changes in mood are associated with shifts in attentional focus onto or away from oneself. In Salovey and Rodin (1985), we proposed that during all strong emotional experiences, but especially when they are negative, there is a tendency for individuals to focus their attention on themselves rather than on the external environment (see also Salovey, 1992; Salovey, O'Leary, Stretton, Fishkin, & Drake, 1991). In a variety of correlational studies, increased attentional focus on the self has been associated with depressed moods (e.g., Ingram & Smith, 1984; T. W. Smith & Greenberg, 1981; see also a recent review of this literature by Ingram, 1990). Sadness induced self-focusing has also been demonstrated experimentally (J. V. Wood, Saltzberg, & Goldsamt, 1990). Pyszczynski and Greenberg (1987) have proposed that there is a self-focusing style that plays a role in the onset, maintenance, and exacerbation of depression. This theory builds on earlier work (Carver & Scheier, 1981; Duval

& Wicklund, 1972) suggesting that allocation of attention onto the self serves a regulatory function and helps the individual to maintain goal-directed behavior.

Focusing attention on the body increases perceptions of symptoms and sensations (Pennebaker, 1982). For example, individuals who live in unstimulating environments—presumably those that do not provide enough competition for internal cues—report more physical symptoms than do individuals in stimulating environments. In experimental research, when subjects are instructed to exercise while attending to bodily cues (heart beat or breathing), they are more likely to report symptoms (Pennebaker & Lightner, 1980).

The comorbidity of depression and pain reporting may be understood because both appear to be related to self-focused attention. We would expect that when sad moods produce body-oriented, self-focused attention, symptoms should be more likely noticed and, indeed, experienced more intensely. The greater salience of somatic cues subsequently may be associated with judgment about both present and future pain.

### *Summary*

Judgments about present pain and recall of prior pain episodes are, at times, associated with mood. In particular, survey responses regarding past pain may be rendered less accurate when respondents are experiencing fairly intense moods and emotions or when their current moods are quite different from their dominant affect at the time of the painful experience. The impact of mood on recall of pain is best examined under fairly controlled circumstances. The results of such experiments, however, have implications for the design and interpretation of health surveys: Mood at the time of the survey should be assessed as part of the survey and then taken into account in the interpretation of survey results. Otherwise, it is difficult to know whether the respondent who answers, “Yes, I experienced intense pain at least 5 of every 7 days last year” in fact had painful episodes of that frequency or was quite depressed at the time of the survey and consequently prone to recall unpleasant painful episodes.

### Present Pain and Recall of Past Pain

We described above some of the mood-related consequences of asking individuals to recall pain experiences after the pain itself had terminated. Accuracy in this situation should be lower than if recall is solicited during a similar pain experience, because these pain experiences may be associated with mismatched versus matched mood states at the two points in time. Pain researchers have long been concerned that retrospective ratings of pain made after pain has been relieved (or changed in intensity in other ways) are most likely to be inaccurate (e.g., Melzack, 1975). The biasing effects of present pain levels on recall of past pain may result from two nonmood factors as well. We shall label the first factor “cuing” and the second “anchoring” and describe each in turn.

### *Cuing*

In the same way that happy versus sad moods facilitate the recall of positive versus negative material from memory, ongoing pain experiences may have similar cuing and assimilative effects on memory. E. Eich et al. (1985) have, in fact, noted that because affect is an integral component of pain behavior and experience, pain may produce assimilative effects on memory that parallel those engendered by emotions. This line of theorizing suggests that prior pain is remembered as being more severe than it actually was when the intensity of present pain is high but that it is remembered as being less severe when the present pain is low.

Eric Eich and his colleagues asked 57 headache patients to maintain pain diaries during a treatment program by having them record hourly ratings of pain on a 10-point scale. In addition, during weekly scheduled appointments, they were asked to rate their present pain intensity on a visual analogue scale. They were then asked to recall the "maximum," "usual," and "minimum" levels of pain experienced since their last visit. Patients' ratings of present level of pain were strongly associated with their recall of maximum, usual, and minimum pain levels since the last visit. When present pain was high, patients' recalled pain levels were higher than their pain diaries indicated. When present pain was low, their recalled pain ratings were less severe than indicated by the diaries. Eich et al.'s results suggest that studies of pain recall in which currently pain-free subjects are asked to recall past pain episodes should result in underestimations of past pain (e.g., Hunter et al., 1979), but that when subjects who are still experiencing pain are asked to recall past pain, they should overestimate it (e.g., S. J. Linton & Melin, 1982).

We replicated the work of Eich and his colleagues in a study involving 80 patients at a pain clinic in Pittsburgh, Pennsylvania (Sieber, Salovey, Jobe, & Willis, 1990). Before this heterogeneous group of subjects began treatment, they were asked to keep hourly pain diaries for 2 weeks. A month later, they were asked to recall the usual amount of pain that they had experienced during the diary period. As indicated by the multiple regression analysis presented in Table 6.1, patients' level of pain at the time of recall accounted for a significant portion of the variance in their recall of pain during the diary period, over and above their actual level of pain during the diary period. We also controlled for variability in original pain ratings in the analysis because low variability may artificially inflate memory accuracy. In other words, pain at the time of recall significantly influenced memory for pain in the expected positive direction.

### *Anchoring*

Another way in which current pain experiences may bias pain recall is by providing an anchor on which memory for pain might be based. Anchoring has been described as a type of confirmatory bias whereby initial estimates based on easily obtained information serve as the basis for subsequent judgments (Nisbett

TABLE 6.1. Predicting Recalled Intensity of Pain from Actual Average Pain during Diary Period, Variability of Pain during Diary Period, and Current Level of Pain

Predictor	$\beta$	$p$
Actual average pain	0.51	.0001
Variability of pain	0.16	.08
Current pain intensity	0.28	.001

$R^2 = .73, F(3, 76) = 55.42, p < .0001$

& Ross, 1980; Turk & Salovey, 1986); that is, appraisals and decisions are heavily influenced or “anchored” by the initial appraisal. Thus, in regard to pain, one’s present experience of intense pain may anchor judgments of past pain episodes and, when pain is presently intense, produce overestimations of their intensity.

Although studies of the effects of anchoring on the recall of pain episodes have not yet been carried out, the anchoring affect has been demonstrated in other contexts (Tversky & Kahneman, 1974). Individuals tend to give excessive weight to initial, highly salient information that then serves as a template against which other judgments are made. A similar anchoring problem could result if a respondent experiences pain at the time that he or she is surveyed about past painful episodes, although, clearly, present pain is not a randomly provided initial estimate and may have some validity as an index of prior pain. This problem should be especially likely if the present pain is made salient. Suppose that a survey first asks, “On a 10-point scale, how much pain are you experiencing right now?” and then the next question is, “On how many occasions during the last year did you have such severe pain that you stayed home from work?” Responses to the second question could be anchored by answers to the first one.

### *Summary*

Present pain may affect survey responses in two ways. Respondents may tend to use easily available information about their present pain as a basis for judging past pain episodes (anchoring), and present pain may make memories for previous painful situations more available (cuing). If present pain does influence survey responses, health surveys may wish to include questions that assess present levels of pain but are less likely to anchor subsequent responses or, better yet, measure it unobtrusively. Then, response biases that can be traced to the impact of present pain can be corrected.

### Recall of Pain Experience versus Pain Behaviors

Weaknesses and biases in self-report instruments have led pain researchers to suggest two other approaches to eliciting pain-relevant information from pain sufferers. The first approach is to evaluate pain based on its impact on the life of

the pain patient and those around him or her (e.g., Kerns, Turk, & Rudy, 1985). The second is to abandon self-reports of pain altogether and instead assess only observable pain behaviors (e.g., Fordyce, 1976; Keefe & Block, 1982; Richards, Nepomuceno, Riles, & Suer, 1982; Turk, Wack, & Kerns, 1985). Finally, researchers have combined these two approaches into what is being called a "comprehensive assessment" (Turk & Rudy, 1987). Preliminary research on the usefulness of these approaches to gathering accurate statistical information about the experience of pain will be discussed in more detail below. In this section, we present the idea that the recall of pain episodes might be more accurate if the impact of pain on daily activities is assessed or pain-related behaviors are observed rather than if the respondents are asked to report on past levels of pain intensity and quality.

The West Haven–Yale Multidimensional Pain Inventory (WHYMPI) is one attempt to assess pain by examining its impact on the life of the patient rather than by relying on pain language (Kerns et al., 1985). The WHYMPI consists of three sections. The first contains the following measures: (a) pain severity and suffering; (b) pain-related life interference (interference with family and marriage, work, and recreational activities); (c) dissatisfaction with present levels of functioning; (d) appraisal of support from spouse, family, and others; (e) perceived life control, problem-solving ability, and feelings of mastery and competence; and (f) affective distress. The second part of the WHYMPI was designed to evaluate patients' perceptions of the range and frequency of responses by significant others to displays of pain and suffering and contains three scales: punishing responses, solicitous responses, and distracting responses. The third part is a set of 30 common domestic activities, household chores, social activities, and recreational activities, for which individuals indicate their level of participation. Kerns et al. (1985) reported adequate internal consistency and stability for all of the subscales. At present, no empirical work has examined the accuracy of recall for information reported on the WHYMPI, especially the scales in Parts 1 and 3, compared with recall of the intensity and quality of experienced pain.

A second approach to reducing inaccuracies in self-reports of pain intensity and quality is to rely on ratings of observable pain behaviors instead. The interest in pain behaviors is usually traced to Fordyce's (1976) influential work on operant learning factors in the pain experience. Fordyce proposed that patients display a range of pain behaviors that serve to communicate to others that they are experiencing pain and suffering. According to Fordyce, these might include verbal complaints, paraverbal sounds (e.g., moans), body posturing and gesturing, display of functional limitations or impairments, and medication use and other pain-attenuating behaviors. Other studies of pain behavior have been reported in the literature as well. For example, Turk, Wack, and Kerns (1985), using multidimensional scaling and cluster analysis, identified two primary dimensions around which pain behaviors could be organized: audible–visible and affective–behavioral. Four clusters of pain behaviors were superimposed on these dimensions and labeled (a) distorted ambulation and posture, (b) negative affect, (c) facial/audible expressions of distress, and (d) avoidance of activity.

The accurate observation and measurement of pain behaviors by health care professionals and other interviewers are somewhat complicated. According to Turk and Flor (1987), the most systematic approach to quantifying pain behaviors is that of Keefe and his colleagues. For example, Keefe and Block (1982) developed an observer coding system for five pain behaviors typically displayed by back pain patients. Patients were videotaped while performing a structured task, and the frequency of their pain behaviors was counted. These behaviors seem to be observed reliably, are correlated with patients' and observers' subjective pain ratings, and are specific to pain patients—compared, for example, with depressives and normals (Keefe, Brantley, Gerdenio, & Crisson, 1985; Keefe & Gil, 1986; Keefe & Hill, 1985; Keefe, Wilkins, & Cook, 1984; Keefe, Wilkins, Cook, Crisson, & Muhlbaier, 1986). Other pain behavior rating systems have been developed that do not require videotaping and can be more easily performed by lay observers (e.g., Cinciripini & Floreen, 1983; Richards et al., 1982).

The questions that have concerned our research are (a) whether recall for the daily life consequences of pain and pain-related behaviors is better than for the intensity and quality of pain itself and, consequently, (b) whether pain is more accurately assessed if questions are focused on the behavioral and life consequences of past pain experiences. In fact, Fordyce et al. (1984) noted that there is little relationship between patients' self-reports of pain and their capacity for physical activity and suggested that it would be desirable to pay greater attention to observable pain behaviors and not to rely exclusively on self-reports of pain intensity. At present, however, the issue of whether pain behaviors capture the experience of pain more accurately than do intensity ratings is largely unaddressed. Researchers still debate whether observed pain behaviors and self-reported pain ratings are highly correlated (e.g., Keefe & Block, 1982; McDaniel et al., 1986) or nearly independent (e.g., Teske, Daut, & Cleland, 1983).

We conducted two studies to address this issue. In one, 40 adults with chronic pain problems of at least 6 months duration came to the laboratory twice, 1 month apart. On each occasion, they provided self-reports of pain intensity and interference with daily tasks, and their overt pain behaviors were rated by two judges. These ratings appeared relatively stable over time, with 1-month reliabilities of  $r = .74$ ,  $.86$ , and  $.71$ , respectively. In a second study, 107 pain patients at a clinic kept one of four kinds of daily diaries for a month. They (a) recorded their daily history on a set of pain behaviors (e.g., took aspirin, used a heating pad), (b) rated average daily pain intensity, (c) recorded both behaviors and pain intensity, or (d) recorded neither. All subjects were called 30 days later and asked to recall the frequency of these pain behaviors during the diary period and the usual level of intensity of their pain. Recall for both pain intensity and pain behaviors appeared to be quite accurate and showed no effect of daily record keeping (Sieber et al., 1990). We were encouraged at the especially high stability in reports of how pain interferes with life tasks. Because these life tasks seem more easily remembered than pain intensity (and because they represent a con-

crete consequence of pain), they may be especially useful as a mode of inquiry into pain problems on surveys.

## The Language of Pain and Pain Recall

In addition to problems with the recall of painful episodes, investigators also have been concerned with the accurate measurement of an individual's phenomenological experience of pain. Enhancing our understanding of the descriptive language of pain cannot help but facilitate the construction of surveys and the understanding of survey data. Traditionally, only the intensity of pain (e.g., mild, moderate, severe) was elicited from respondents, but an influential study by Melzack and Torgerson (1971) suggested that (a) the English language offers a rich vocabulary for describing pain experiences; (b) there is high agreement on the dimensions underlying particular classes and subclasses of these words; and (c) despite individual differences in background and pain problems, most of these pain words are used similarly across individuals. In a series of studies using the scaling techniques available at the time, Melzack and Torgerson identified three general aspects of pain experience: sensory (temporal, spatial, punctate pressure, incisive pressure, constrictive pressure, traction pressure, thermal, brightness, dullness), affective (tension, autonomic, fear, punishment), and evaluative (intensity).

This effort led to the development of a multidimensional pain inventory and the ability to assess pain on more than a simple intensity dimension. The widely used McGill Pain Questionnaire mentioned earlier (Melzack, 1975) was thus developed, based on these aspects of pain experience. Despite the rich pain vocabulary provided by the MPQ, we cannot really know another person's experience of pain directly. What we can do more easily is describe in language that is as objective as possible episodes that might generally be expected to result in pain (e.g., having one's finger slammed in a car door). The critical question then seems to be whether people share common descriptive terms for painful episodes of this sort. It is quite possible that relative experts when it comes to pain (e.g., individuals who experience pain chronically) use pain-relevant language quite differently than do pain researchers, survey designers, and others who have had considerably less experience introspecting about pain and describing pain to others.

In work in our laboratory, two different groups of subjects, one suffering from chronic pain and the other a group of healthy volunteers, made direct similarity ratings among 15 pain terms from the short form of the MPQ. With the aid of several analytic procedures to identify structure in multivariate data, we attempted to identify clusters of terms that have common meanings and to identify dimensions of meaning along which these clusters vary. We also compared the nature of the pain lexicon for the two groups of subjects. Both groups generated four-dimensional multidimensional scaling solutions. Most striking about these data was the substantial overlap in the solutions generated by the two different

groups. It seems that healthy people, who do not think often about pain, and pain patients, who must consider pain on a daily basis, mentally represent the pain lexicon in remarkably similar ways. Moreover, the 15 terms chosen from the short form of the MPQ and used in this study seemed to capture most of the possible variance in the use of pain terms, and their meanings (relative to each other) seemed to be consensually agreed on. At this point, we might be bold enough to suggest that when pain is questioned on health surveys, these 15 descriptors might be good ones to use compared with the much larger corpus of pain adjectives available.

We also conducted a more open-ended, spontaneous exploration of pain language. One purpose of this study was to determine whether descriptions of standardized pain episodes evoke common linguistic responses. Additionally, we explored the language used to define and describe various pain episodes and whether these descriptions varied depending on whether one had personal experience with the particular pain problem. The subjects were 21 native speakers of English randomly selected from the New Haven area and 22 pain patients from a pain clinic. In a single session, subjects completed two tasks. First, they were asked a series of questions designed to elicit spontaneous pain descriptions. Such questions included the following: "Describe the most painful experience you have ever had?" "What did it feel like?" "Specifically, what was the pain like?" "Can you generate a set of 5 words that really captures for you the painful experience?" "Describe your most recent painful experience." Another question, which was asked in order to define an upper limit on painful episodes, was, "Describe the most painful situation imaginable."

In addition, for the pain patient subjects, a "think-aloud" component was added to the interview. In the course of these interviews, subjects were asked to rate their pain for that day. They were then asked to describe what was going through their minds as they made these ratings. After describing their thoughts, they were asked a series of more structured questions such as the following to obtain additional information concerning how pain ratings are made: (a) "How did you go about making your pain rating?" (b) "Did your pain vary much during the course of the day?" (c) "If so, how did you arrive at a final number?" (d) "Imagine a day in which your pain was excruciating in the morning but then gradually improved during the day until it was barely noticeable; how would you go about rating your pain for that day?" (e) "What if your pain was bad, you took some medication, and it got better; how would you go about rating your pain for that day?" We also presented all subjects with the same subset of 15 MPQ terms used in the scaling study and asked them to describe painful experiences that could best be characterized by the provided terms. Thus, for example, subjects might be asked, "Describe a painful experience in which the pain was *cramping*" or "Describe a painful experience in which the pain was *heavy*."

Data collected from the two groups of subjects were tabulated and content analyzed. When asked to describe a recent painful incident, the healthy subjects came up with 18 discrete types of episodes (e.g., cut, headache, burn) and the pain patients with 10. We then asked subjects to generate 5 words to describe

the pain. The healthy group generated 74 different words, and the pain group generated 60, an enormous and heterogeneous assortment of terms. The words most frequently generated by the healthy subjects were “sharp,” “throbbing,” and “annoying”; those most frequently mentioned by the pain patients were “sharp,” “aching,” and “severe.”

In the next section of the interview, only the pain patients described how they use numbered scales to make different kinds of pain ratings. In brief, some of the observations that can be culled from these data were: (a) patients are more likely to use the odd numbers on a 0-to-9 pain scale, especially 7; (b) patients have great difficulty introspecting about how they chose from among scale alternatives (e.g., the most frequent reason is, “I just considered how much pain I was in”); (c) when pain shifts from excruciating to barely noticeable in the course of a day, for whatever reason, patients overwhelmingly choose the number 3 to represent their average amount of pain for that day; but (d) they have little understanding about how or why they did this.

The final section of the interview asked subjects to generate pain episodes that might involve pain described by a particular term from the short form of the MPQ. Table 6.2 summarizes the most common pain episodes generated in response to these cues. What is most interesting about these results is that there is considerable diversity in the application of these pain terms to painful episodes. With the exception of a few items in which semantic associations are obvious (e.g., splitting-headache), agreement in term application is not very high. Compare this observation, however, with the results of the scaling study described earlier. That study demonstrated quite clearly that pain patients and healthy subjects understand the relative meanings of these pain terms. The scaling solutions were stable and replicable. However, this more open-ended study indicated that individuals have greater difficulty mapping these terms onto specific pain instances.

## Summary and Recommendations

Pain may not be recalled with great accuracy. In the chronic pain populations in which pain recall is often studied, retrospective accounts of pain generally yield elevated intensity ratings. On the other hand, in certain special populations who experience acute pain of shorter duration, especially when followed by positive affect or relief from this pain, retrospective recall of pain can actually produce attenuated pain estimates. The literature suggests several sources of systematic bias in pain recall: (a) a respondent’s affective state at the time that the pain rating is made, (b) whether or not a respondent continues to experience pain at the time of the retrospective rating, (c) whether pain is assessed verbally through self-report measures or by rating pain-related behaviors and pain-related life changes, and (d) whether respondents with different levels of experience in dealing with and reporting pain organize pain-relevant language differently and whether a common pain language is used to report painful experiences.

TABLE 6.2. Most Frequently Generated Painful Incidents in Response to Pain Terms

Term	Healthy Sample ( <i>N</i> = 21)	Pain Sample ( <i>N</i> = 22)
Throbbing	Headache	Headache
# of nominations	9	9
# of different incidents	9	9
Shooting	Muscle	Abdominal
# of nominations	4	5
# of different incidents	13	11
Stabbing	Stabbing, injection, cramp	Back
# of nominations	3 each	5
# of different incidents	13	10
Sharp	Stomach, head, bang	Back
# of nominations	3 each	5
# of different incidents	11	11
Cramping	Menstrual	Leg
# of nominations	7	7
# of different incidents	6	9
Gnawing	Dental	Dental
# of nominations	8	6
# of different incidents	7	10
Hot	Burn	Burn
# of nominations	13	7
# of different incidents	7	11
Aching	Exertion	Head
# of nominations	7	5
# of different incidents	8	10
Heavy	Muscle, head, crush	Abdominal
# of nominations	3 each	5
# of different incidents	12	10
Tender	Bruise, blister, joint	Muscle
# of nominations	4 each	4
# of different incidents	12	14
Splitting	Headache	Headache
# of nominations	14	21
# of different incidents	7	2
Tiring	Fatigue	Fatigue
# of nominations	5	7
# of different incidents	11	8
Sickening	Nausea	Nausea
# of nominations	10	16
# of different incidents	9	6
Fearful	Stomach, surgery, heart, bone	Nerves, illness
# of nominations	2 each	3 each
# of different incidents	16	15
Punishing	Beat/torture	Beat, accident
# of nominations	5	4 each
# of different incidents	12	11

The designers of surveys that are based on the retrospective recall of autobiographical memories must consider each of these sources of bias in the construction of their instruments and in the analysis of the data generated by them. Preliminary results from our research program suggest the following conclusions and recommendations with regard to the retrospective reporting of painful experiences:

1. Although transient (induced) mood states have a systematic effect on various kinds of judgment tasks explored in other contexts, evidence for their associations with pain reports is mixed.
2. One systematic source of bias in pain ratings is created by the severity of pain at the time of recall. Controlling for original levels of pain and the amount that pain fluctuates during the applicable time period, greater pain at recall is associated with overestimating prior pain experience. Survey researchers asking questions about prior experiences with pain may also wish to include questions about current levels of pain.
3. The severity of prior pain, its impact on daily activities, and behaviors related to the pain problem are all recalled approximately equally well and seem to be equally stable over time among individuals with chronic pain problems. Survey researchers who seek more informative data than those provided by mere intensity ratings should feel comfortable querying respondents about these other pain-related behaviors (or should even consider rating respondents' observable pain behaviors).
4. The pain lexicon is enormous and relatively vague. There is considerable variability in how pain descriptors are assigned to painful experiences. Yet, there seems to be considerable consensus in the understanding of the meaning of a subset of pain descriptors across individuals with varying levels of experience with pain. Fifteen descriptors that show considerable stability in the way in which they are organized mentally are *throbbing, shooting, stabbing, sharp, cramping, gnawing, hot, aching, heavy, tender, splitting, tiring, sickening, fearful, and punishing*. Although we have no way of knowing whether these descriptors are more or less stable than any other set of 15, we feel confident that the relative meaning of these words is widely shared by most potential survey respondents.

*Acknowledgments.* We would like to thank Albert F. Smith and Dennis C. Turk, who served as consultants to our research on memory for painful experiences, some of the results of which are reported here. Chloe Drake, Stephanie Fishkin, and Sasha van der Sleen assisted in the collection and analysis of data for this project, and we gratefully acknowledge their help.

This research (supported by DHHS contract 200-88-7001 from the National Center for Health Statistics to Peter Salovey) was part of a larger project entitled "National Laboratory for Collaborative Research on Cognition and Survey Measurement," conducted by NCHS under grant SES-8612320 from the National Science Foundation. Peter Salovey is also supported by NIH Biomedical Research Support Grant S07 RR07015,

National Cancer Institute grant CA42101, and a Presidential Young Investigator Award from the National Science Foundation (BNS 9058020).

A full report of the findings from this research program has been published by NCHS (Salovey et al., 1992). Some of the material presented in this chapter appears in that report along with a more complete description of our research program on the recall of physical pain.

# 7

## The Effects of Estimation Strategies on the Accuracy of Respondents' Reports of Cigarette Smoking

Barbara Means, Gary E. Swan, Jared B. Jobe,  
and James L. Esposito

Many survey questions ask about the frequency of specific behaviors (e.g., consumption of a particular food, use of a product, voting). Until recently, the assumption has been that responses to these items are formed through a process of episodic recall, in which the respondent mentally retrieves all pertinent incidents and then tallies them (e.g., Bradburn, 1983). Survey research focused on potential sources of response error associated with episodic recall (i.e., retrieval failures and telescoping effects).

It is now generally agreed that respondents use a variety of response heuristics to answer survey frequency questions (e.g., Blair & Burton, 1986, 1987; Burton & Blair, 1991; Menon & Sudman, 1989; L. Ross, 1984; Schwarz, 1990a; Strube, 1987). This appears to be particularly true in cases where the target behavior is high in frequency.

### Studies of Frequency

Blair and Burton (1986; Burton & Blair, 1991) explored the issue of alternative frequency estimation strategies in a series of studies. In one (Blair & Burton, 1986), they administered survey questions about frequency of behaviors, such as dining in a restaurant or using an automatic teller machine. Immediately after answering the frequency question, the respondent was queried about the method used to arrive at the response. *Episodic recall* was the most frequently reported strategy when there were three or fewer incidents in the event category, but use of this strategy declined rapidly as event frequency rose. For those who ate in a restaurant 10 times or more, for example, none reported basing his or her frequency estimates on episodic recall. Instead, responses were based on what Blair and Burton called "direct" estimates or on procedures for decomposing the reference period and computing rates. For a very frequent, commonplace behavior, writing checks, nearly two-thirds of their subjects reported using direct estimates under normal interviewing procedures. Later studies (Burton & Blair, 1991)

confirmed that respondents often use nonepisodic processes to answer frequency questions and found that task conditions such as the number of events to be reported and the time used in response formulation affect the processes that respondents use. Similarly, Menon and Sudman (1989) found that their respondents used episodic recall for frequencies of less than 5 and an estimation strategy for frequencies of more than 5.

The psychological literature suggests that the use of direct estimates are subject to biases quite different from the omission and telescoping errors found in episodic recall. Tversky and Kahneman (1982) argue that a commonly used heuristic for producing these estimates is the *availability judgment*. Easy-to-retrieve events are assumed to have a high rate of occurrence. Because factors other than frequency affect retrievability, this heuristic can lead to bias. For example, in one Tversky and Kahneman study, subjects were asked to consider a letter (K, L, N, R, or V) and to judge whether the letter was more likely to appear in the first position or in the third position in English words. Although these letters are in fact more likely to appear in the third position, it is easier to retrieve words starting with a given letter than it is words in which the letter appears in the third position. Each of the letters was judged by the majority of subjects to be more frequent in the first position of words than in the third position, and the bias was quite strong. The mean estimate for the ratio of the two frequencies was 2:1. Given the relationship between event salience and retrievability, and the low salience of individual smoking episodes, the Tversky and Kahneman work suggests that using an availability strategy would tend to lead to underestimates of smoking rate.

On the other hand, a study reported by Bruce and Read (1988) suggests that however biased direct estimates based on availability may be, they are better than estimates based on attempted episodic recall for high-frequency events. In that study (conducted with F. Craik), subjects saw lists of words from different semantic categories (e.g., animals). After viewing a list, subjects were asked first to give an initial frequency estimate (presumably based on availability) for each category (e.g., "How many animals were on the list?"). This first frequency estimate was followed by cued recall: The experimenter provided category labels and instructed the subject to recall as many instances of that category from the word list as he or she could. This cued recall was followed by a second frequency estimate. These second estimates were typically less accurate than the initial estimates, the interpretation being that although subjects had failed to recall all instances, they used their partial recall as a guide in making the second estimate.

Another approach to estimating frequency reported by subjects interviewed by Blair and Burton (1987) and in a similar study reported by Bradburn, Rips, and Shevell (1987) is *decomposition*. This entails decomposing the reference period into temporal segments, computing a frequency for one or more parts of it, and then multiplying or adding part frequencies to arrive at an estimate for the whole. Menon and Sudman (1989) assessed the accuracy of decomposition compared with other strategies for reporting frequencies. For an irregular be-

havior (snacking), decomposition proved to be more accurate than a strategy that encouraged respondents to note exceptions to their normal rate of frequency (stability and change). For a regular behavior (hair washing), decomposition proved to be less accurate than stability and change. Additional research on the relative accuracy of frequency estimates derived in these and other ways is required.

## Surveys of Smoking

Surveys are the most frequently used method of obtaining population information about smoking. Survey questions on smoking require respondents to retrospectively report their smoking level. Both current smoking levels and long-term exposure are important for epidemiologic purposes. Respondents who retrospectively report their smoking levels show a strong bias toward reporting in units of 10 (Vesey, Saloojee, Cole, & Russell, 1982; Warner, 1978). Data suggest that retrospective reports are often underestimates of the actual amount smoked and that many current smokers deny their smoking status (see Warner, 1978). In one study, 35% of those who claimed in a telephone interview that they had quit smoking reported in a subsequent home interview that they smoked (Luepker, Pallonen, Murray, & Pirie, 1989). According to Sillett, Wilson, Malcolm, and Ball (1978), up to 25% of respondents may falsely report their smoking behavior.

Validation of self-reported smoking levels may be assessed by various biochemical markers, of which cotinine (a chemical byproduct of nicotine) appears to be the most accurate (Jarvis, Tunstall-Pedoe, Feyerabend, Vesey, & Saloojee, 1987); by the collection of butts (e.g., Robinson & Young, 1980); by direct observation (e.g., Orleans & Shipley, 1982); or by self-reported records (e.g., diaries). (For a review of various approaches, see Orleans & Shipley, 1982.)

The smoking-frequency item appearing on most smoking surveys, "How many cigarettes do you smoke a day?" clearly places heavy demands on the smoker trying to episodically recall each cigarette smoked during a given day. Given not only the frequency with which smokers light a cigarette but also the almost unconscious nature of this activity, it seems reasonable to assume that few subjects answer smoking-frequency questions through exhaustive episodic recall of individual incidents. If not episodic recall, then what processes are respondents using and what is the relative accuracy of different frequency estimation strategies?

In the two studies reported here, we sought to (a) identify the strategies that people use to answer questions about smoking frequency and (b) assess the relative accuracy of different frequency estimation strategies. To permit assessment of recall accuracy in subjects, both studies collected behavioral evidence of their smoking behavior for 4–5 days. Afterward, they were given an unexpected question about the number of cigarettes that they had smoked on one or more of the days during that assessment period. Using a four-group, between-subjects

design, one group of subjects answered the frequency question using any method that they chose, whereas the other three groups were instructed to employ a particular frequency estimation strategy. These experiments are summarized below.

## Experiment 1

### Method

In Experiment 1 (Means, Swan, Jobe, Esposito, & Loftus, 1989), we obtained a convenience sample of 25 subjects—10 males and 15 females—by local advertisements soliciting smokers to participate in the research for a \$50 fee. The experimental sessions were conducted at the offices of SRI International in Menlo Park, California. Potential subjects were told that the study concerned different methods for measuring nicotine dosage, including biochemical analyses of saliva samples, and were screened to make sure that they were not actively trying to quit smoking.

Subjects participated in an initial face-to-face interview and then collected two kinds of behavioral evidence of their weekday smoking for a period of 4 or 5 days (depending on the interview schedule). The behavioral evidence used in this study consisted of (a) a canister with each day's cigarette butts and (b) a written tally of each time a cigarette was lit.

After collecting the behavioral evidence of their smoking, the subjects brought their canisters and tallies to the laboratory and participated in a second interview. The critical question in this interview concerned the number of cigarettes that they had smoked on one of the days during the assessment period (the target day). Depending on their treatment group, subjects either were trained in a specific frequency estimation strategy and directed to use it in answering the question or were allowed to compute a smoking frequency in any way that they liked. The three specified strategies—*availability*, *additive decomposition*, and *episodic recall*—are described in Table 7.1.

In the *free-strategy/think-aloud* condition, subjects were not told how they should answer the smoking-frequency question but were directed to think aloud as they formulated their answer (see Ericsson & Simon, 1980, 1984). All subjects gave a confidence estimate for the smoking frequency that they reported. They then provided a saliva sample for biochemical analysis to test for saliva cotinine.

### Results

#### *Self-Reports Prior to Participation*

Subject responses to the telephone screening question about smoking level were reviewed to get a sense of the precision of their prior knowledge about how

TABLE 7.1. Frequency Estimation Strategies

**Availability:**

The subject is instructed to answer the smoking-frequency question quickly with a "gut reaction" without trying to think of specific instances of smoking.

**Additive decomposition:**

The typical weekday is divided into different types of activities such as "commuting," "in the office," and "after dinner." The subject is asked to estimate the number of cigarettes consumed in each of these portions of the day. After all separate estimates are made, they are summed to arrive at a total day's estimate.

**Episodic recall:**

The subject is instructed to take a selected weekday and to work through the day recalling all instances of smoking. The subject is then asked to use the recalled instances to arrive at a daily total.

much they smoke. Slightly more than half of the subjects (13, or 52%) described their smoking in terms of packs rather than number of cigarettes. Of those subjects who responded in terms of cigarettes, 3 gave a range, 2 spoke of "more than 20" or "less than 20," and the remainder gave a number that was a multiple of 10 (most, in fact, were multiples of 20). Hence, it appears that subjects have a rough idea of their own smoking level but do not demonstrate precise knowledge. The digit bias observed in other studies of smoking (Vesey et al., 1982; Warner, 1978) is apparent here as well.

*Quality of Behavioral Evidence*

Two forms of behavioral evidence—the number of cigarette butts saved and the subject's cigarette-lighting tally—were compared for each day within the assessment period. A high level of agreement was found ( $r = .91$ ). For the target day, the number of cigarettes on the tally record was identical to the number of cigarette butts in the canister for 15 of the 25 subjects. Where differences did occur, they were small in magnitude, averaging just 0.64 cigarette. For simplicity's sake, we report here only those analyses using the cigarette-lighting tally (which was more likely to have the higher number) as the "true" smoking frequency.

*Smoking Levels*

The mean number of cigarettes smoked during the days when subjects kept records was 22.94 and the median was 20.20. The lightest smoker averaged 7.60 cigarettes per day, and the heaviest averaged 58.00. The first row in Table 7.2 shows the mean daily smoking by treatment condition: The means were 23.17 for the free-strategy/think-aloud group, 22.67 for availability, 20.17 for additive decomposition, and 28.57 for episodic recall. Although the mean of the episodic-recall group was somewhat higher relative to the other groups, this difference was not significant.

TABLE 7.2. Mean Level of Smoking in Behavioral Records and Self-Reports, by Condition: Experiment 1

Measure	Strategy Condition			
	Free-strategy/ Think-aloud	Availability	Additive Decomposition	Episodic Recall
Target day behavioral record	23.17 (8.57)	22.67 (10.46)	20.17 (6.59)	28.57 (17.44)
Behavioral record – report	1.17 (1.60)	3.50* (1.87)	-3.50 (5.89)	-1.57 (5.59)
Absolute value of record – report	1.50* (1.22)	3.50* (1.87)	4.17 (5.34)	4.43* (3.36)
(Record – report)/ Record	0.05 (0.09)	0.15* (0.05)	-0.16 (0.27)	-0.12 (0.23)

Note: A minus sign indicates overreporting. Standard deviations appear in parentheses.

\*  $p < .05$ .

### Accuracy of Frequency Reports

The second row in Table 7.2 shows the mean difference between the number of cigarettes reported and the number on the behavioral record for the target day in each treatment condition. We tested whether each of these means was significantly different from 0. The only significant difference was found in the availability condition,  $t(6) = -4.58$ ,  $p < .01$ , in which reports averaged 3.50 cigarettes lower than the behavioral record. In the free-strategy/think-aloud condition, reports averaged an insignificant 1.17 cigarettes below the behavioral record. In contrast to these two conditions, subjects in the additive-decomposition and episodic-recall conditions tended to overreport their smoking for the target day, but the differences were not significant.

The third row in Table 7.2 shows the mean absolute difference between the number of cigarettes reported and the number in the behavioral record ("gross error"). Gross error was largest for the episodic-recall and additive-decomposition conditions (4.43 and 4.17 cigarettes, respectively) and smallest for the free-strategy/think-aloud condition (1.50). Gross error rates were significant (different from 0) in all conditions except additive decomposition. When tested through a one-way ANOVA, however, the effect of condition on gross error rate was not significant.

A third measure of accuracy was created in order to correct for the fact that individuals with different levels of smoking had recall tasks of different difficulty. This measure, (Behavioral Record Count – Reported Frequency)/Behav-

ioral Record Count, is shown in the fourth row of Table 7.2. An ANOVA testing for the effect of treatment condition on this measure was significant,  $F(3,21) = 3.70$ ,  $p < .05$ . Comparisons between conditions revealed that performance in the availability condition differed significantly from each of the other conditions. Thus, reported frequencies in the availability condition were larger relative to the corresponding record frequencies than in any of the other strategy conditions.

### *Strategies Used in the Free-Strategy/Think-Aloud Condition*

The strategies that subjects reported using in the free-strategy/think-aloud condition were examined. Three subjects said that they used "knowledge about my usual daily smoking"; this is a strategy based on semantic rather than episodic memory. Three other subjects reported complex strategies that involved dividing the day up into segments (decomposition) and then either recalling the instances of smoking within each segment (episodic) or using the first number that came to mind for each segment (availability). A piece of converging evidence for these self-reports was obtained by having a researcher independently review the verbal protocols for a sample of 19 subjects. Of these, 15 protocols were sufficient to permit categorizing the subject's strategy into 1 of 6 classifications. Among these, 11 of the experimenter classifications (73%) matched the subject's self-reported strategy.

### *Cotinine*

Saliva samples were analyzed for cotinine, a nicotine byproduct. Cotinine levels in all subjects were in the range normally associated with smokers. Correlations between cotinine level and the various forms of behavioral evidence of smoking ranged between .35 and .40 and compared favorably with reported correlations between cotinine and self-reported smoking in the literature, which range from .24 to .43 (Abrams, Follick, Biener, Carey, & Hitti, 1987; Benowitz, Kuyt, Jacob, Jones, & Osman, 1983; Haley, Axelrad, & Tilton, 1983; Hill, Haley, & Wynder, 1983; Pierce et al., 1987). However, although cotinine is the preferred biochemical measure for distinguishing between smokers and nonsmokers (Jarvis et al., 1987; Pechacek, Fox, Murray, & Luepker, 1984), it does not appear to be calibrated accurately enough to replace self-reports as the primary measure of smoking frequency among smokers.

### *Confidence*

Subjects rated their confidence in their frequency estimates on a 5-point scale, with 1 = "off by more than 10 cigarettes" and 5 = "exactly correct." The correlations between confidence level and the absolute difference between reported and recorded smoking were  $-.33$  for Wednesday smoking,  $-.30$  for Monday, and  $-.21$  for Saturday. Table 7.3 shows the confidence ratings by treatment condition. These data suggest that subjects are least confident about frequency esti-

TABLE 7.3. Mean Subject Confidence, by Treatment Condition: Experiment 1

Condition	Mean Rating
Free-strategy/think-aloud	3.50
Availability	3.33
Additive decomposition	3.67
Episodic recall	2.29

Key to confidence ratings:

- 1 = Off by more than 10 cigarettes
- 2 = Off by no more than 10 cigarettes
- 3 = Off by no more than 5 cigarettes
- 4 = Off by no more than 1 or 2 cigarettes
- 5 = Exactly correct

mates derived through attempts to recall all smoking events during the target day (episodic-recall strategy) and most confident about frequencies derived through additive decomposition. It is interesting to note that the pattern of differences in the confidence ratings is not congruent to that for recall accuracy, where the trend was for better recall in the episodic-recall condition than in the additive-decomposition condition.

## Experiment 2

The second study (Means, Swan, Jobe, & Esposito, 1991) also investigated the accuracy of reports of current smoking levels. The results of the first experiment led us to the conclusion that the free-strategy group (in which the interview was slow paced and think-alouds were requested) was not a good approximation of the procedures used in standard fast-paced smoking interviews. Therefore, in the second experiment, we dropped the requirement for the free-strategy group to think aloud. The other major change in the second experiment was that behavioral records were maintained by subjects using a hand-held computer.

### Method

Subjects ( $N = 128$ ) were recruited as in Experiment 1, and all experimental sessions were conducted at SRI's offices. The same four-group, between-subjects design was used. Subjects in three of the groups were asked to use one of three specific strategies: *availability*, *additive decomposition*, or *episodic recall*. Subjects in the fourth group, the *free-strategy* condition, were simply questioned as they would be on a standardized national survey.

All subjects collected behavioral evidence of their smoking by using a hand-held computer to record each cigarette smoked. Use of the hand-held computer required some effort but was quite feasible for most subjects. The computer was approximately the size and weight (255 gm) of the hand-held computers popular

in the 1970s. Respondents reported very high compliance with the computer logging procedures. On average, subjects made a mean of only 2.4 entries on their error cards over the 5 days for which data were analyzed. There were some cases in which a subject left the electronic diary where he or she did not have continuous access to it for part of the day or failed to consistently record smoking for some period. Two subjects stated that they got tired of entering cigarettes and stopped logging all of them on the final day. In such cases, the day's recorded and reported smoking were eliminated from the data set. In total, the data for 1 day was excluded for eight subjects, and data for 2 days was excluded for two subjects.

At the end of the 5-day record-keeping period, subjects were unexpectedly asked how many cigarettes they had smoked on the preceding day, on the previous Monday (3 days prior), and on the previous Saturday (5 days prior). A saliva sample was taken to measure saliva cotinine, a biochemical marker for nicotine.

## Results

The mean number of cigarettes smoked during the days when subjects kept records was 19.94. The lightest smoker averaged 2 cigarettes per day and the heaviest averaged 54. For each subject, we computed the size of the deviation between the number of cigarettes recalled for each target day and the recorded number of cigarettes for that day (the computer record as amended by any notations on the error card).

Table 7.4 shows the mean number of cigarettes recorded by subjects in each condition for each target day and the deviation (both net and gross) between this number and the number of cigarettes reported. For all of the groups and each target day, mean reports were higher than the mean number of cigarettes recorded, but the deviations were not large (the largest being 3.77 cigarettes for reports of Saturday's smoking in the availability group). For reports of the preceding day's smoking, net deviations appear smallest in the free-strategy condition (0.17 cigarette). However, this apparent advantage is not significant and disappears at the longer lags (i.e., reports of smoking 3 and 5 days prior). We again tested whether each net deviation was different from 0 and found significant differences only for reports of the previous Saturday's smoking in the availability group, the previous Monday's smoking in the additive-decomposition group, and the preceding day's smoking in the episodic-recall group. To measure the accuracy of individual reports, we examined the absolute value of the difference between the recorded and reported level of smoking. All of these measures were significantly different from 0. The mean difference between reported and logged smoking frequency appears largest for the free-strategy and availability conditions, at least over longer retention intervals. For example, recall for the previous Saturday's smoking was off by an average of 7.00 and 6.63 cigarettes in the free-strategy and availability conditions, respectively, compared with mean deviations of 4.70 and 4.87 cigarettes for the additive-decomposition and episodic-recall conditions. For the frequency question most

TABLE 7.4. Mean Level of Smoking in Behavioral Records and Self-Reports, by Condition: Experiment 2

Measure and Target Day	Strategy Condition			
	Free-strategy/ Think-aloud	Availability	Additive Decomposition	Episodic Recall
Behavioral record:				
Preceding day (Wednesday)	20.30	18.68	20.32	19.70
3 days prior (Monday)	19.63	19.97	20.20	21.52
5 days prior (Saturday)	22.68	21.27	20.73	21.94
Behavioral record – report:				
Preceding day (Wednesday)	-0.17	-1.48	-1.16	-2.87*
3 days prior (Monday)	-1.03	-1.55	-2.40*	1.13
5 days prior (Saturday)	-1.93	-3.77*	-1.83	-1.77
Absolute value of record – report:				
Preceding day (Wednesday)	4.57*	4.13*	4.06*	3.73*
3 days prior (Monday)	5.50*	6.13*	3.87*	4.48*
5 days prior (Saturday)	7.00*	6.63*	4.70*	4.87*
(Record – report)/				
Record:				
Preceding day (Wednesday)	-0.08	-0.07	-0.19*	-0.18
3 days prior (Monday)	-0.13	-0.18	-0.13*	-0.10*
5 days prior (Saturday)	-0.13	-0.20*	-0.22*	-0.03

\*Differs from 0 at  $p < .05$ .

likely to be asked on smoking surveys—the number of cigarettes smoked yesterday—differences were in the same direction but much smaller. Subject reports were off by an average of 4.57 in the free-strategy group, 4.13 in the availability group, 4.06 in the additive-decomposition group, and 3.73 in the episodic-recall group.

The average number of cigarettes smoked over the 5 days was computed for each subject and the sample was divided at the median (19.40 cigarettes per day). Table 7.5 illustrates the relationship between smoking level and frequency report accuracy. For the question about the preceding day's smoking, lighter smokers were off by an average of 2.95, compared with 5.26 for heavier smokers.

### Cotinine

Saliva samples were again analyzed for cotinine. Correlations were performed between subject cotinine levels and a number of the measures of smoking frequency. Correlations were very similar whether cotinine was correlated with

TABLE 7.5. Mean Difference between Reported and Logged Smoking Frequency, by Smoking Level and Recall Interval: Experiment 2

Smoking Level	Recall Interval		
	Preceding Day (Wednesday)	3 Days Prior (Monday)	5 Days Prior (Saturday)
Lighter (< 19.40 per day)	2.95	3.73	4.14
Heavier ( $\geq$ 19.40 per day)	5.26	6.19	7.27

reported or logged smoking. As in Experiment 1, cotinine levels in all subjects were in the range normally associated with smokers. Correlations with measures of the preceding day's smoking were .31 for cigarettes logged onto the computer (as corrected by the error card) and .30 for self-reported smoking. Because cotinine has a half-life of 17 hours, a correlation was also performed between cotinine level and the number of cigarettes recorded on the computer in the 17 hours preceding the interview. As in Experiment 1, this correlation, .44, compares favorably with reported correlations between cotinine and self-reported smoking in the literature.

### Confidence

Table 7.6 shows the confidence data by treatment condition. (Note that the verbal anchors for scale values were modified somewhat from those used in Experiment 1.) The confidence scores are very similar across the four strategy conditions, suggesting that subjects' sense of their estimation accuracy was not affected by the method by which they had derived the estimate. Contrary to intuition, the day for which subjects were providing confidence estimates had little effect. Subjects tended to evaluate their degree of error as either 2-3 or 4-5 cigarettes, regardless of whether they were providing an estimate for the preceding day or 5 days prior. (Results might have been different, of course, if day had been a between-subjects rather than a within-subjects variable.)

## General Discussion

Although the methods differed slightly between the two experiments, both suggested that methods closely resembling the fast-paced survey interview resulted in the most errors by subjects. In Experiment 1, this group was the availability group, and in Experiment 2, this group was the free-strategy group. This result confirms results of other investigators (Burton & Blair, 1991; Marquis, Marquis, & Polich, 1986) that the normal fast pace of interviews often leads to less accurate data.

The large difference between the free-strategy/think-aloud group in Experiment 1 and the free-strategy group in Experiment 2 in terms of gross deviation

TABLE 7.6. Mean Subject Confidence, by Treatment Condition and Recall Interval: Experiment 2

Condition	Recall Interval		
	Preceding Day (Wednesday)	3 Days Prior (Monday)	5 Days Prior (Saturday)
Free-strategy/think-aloud	3.10	3.47	3.37
Availability	3.26	3.19	3.29
Additive decomposition	3.29	3.54	3.11
Episodic recall	3.53	3.31	3.25

Key to confidence ratings:

- 1 = Off by more than 10 cigarettes
- 2 = Off by no more than 10 cigarettes
- 3 = Off by 4 or 5 cigarettes
- 4 = Off by 2 or 3 cigarettes
- 5 = Off by no more than 1 cigarette

between recorded and reported smoking suggests that instructions to think aloud while estimating frequencies has a large, beneficial effect on estimate accuracy. In fact, the free-strategy/think-aloud group in Experiment 1 performed better than any of the strategy groups in Experiment 2, where the smallest deviation in reporting the preceding day's smoking was 3.73 for the episodic-recall group, compared with 1.50 for the free-strategy/think-aloud group in Experiment 1. Although differences between the methodology and recruiting practices in the two studies suggest caution in interpreting this finding, it seems reasonable to propose that there is an advantage in letting subjects select their own frequency estimation strategy, provided that they are stimulated to emphasize accuracy in choosing and carrying out a strategy. Requiring subjects to think aloud appears to have this effect, leading them to use more effective strategies or to implement their strategies more diligently than when they are queried under normal survey conditions.

From a psychological perspective, these results suggest that the methodology used in fast-paced surveys with short questions where the respondents are given little time to think are not conducive to accurate reports of high-frequency, low-salience events such as cigarette smoking. Alternative methodologies that encourage episodic-recall or additive-decomposition approaches to estimating high-frequency behavior appear to be more promising. Although individual subject's reports were not accurate under the free-strategy condition at the longer retention interval, when we examine the net error for the group as a whole (rather than as the mean absolute value for the errors as reported above), subject's errors tend to cancel each other out; that is, some subjects gave reports that were too high, whereas others gave reports that were too low. Epidemiologic researchers who are interested in the relationship between the amount smoked and health status should have the same concerns with accuracy of individual reports as do

psychologists and survey researchers who are interested in studying the relationships between self-reports and other variables (e.g., Marquis et al., 1986).

It should also be noted that these studies utilized subjects who were admitted smokers. They responded to advertisements requesting smokers for a research study. Thus, it is unlikely that deliberate underreporting of smoking levels was a factor in these studies, and deliberate underreporting of smoking status (i.e., actual smokers who denied that they were smokers) was definitely not a factor. Thus, these results can only be generalized to smoking subjects who have no motivation to conceal their cigarette consumption.

*Acknowledgments.* This research was conducted under DHHS contract 200-88-7074 between the National Center for Health Statistics and SRI International and is part of a larger project conducted by NCHS under grant SES-8612320 from the National Science Foundation.

We wish to express our thanks to Katherine Habina, Anna Carroll, Alison Dame, and Lisa Jack for their diligent work in conducting interviews, coding data, and running analyses.



## 8

# Validity of Reports of Long-Term Dietary Memories: Data and a Model

*Albert F. Smith and Jared B. Jobe*

Health surveys ask respondents to report on a variety of behaviors and activities in which they have engaged during specified and sometimes extended periods of time. These behaviors and activities range from the salient to the mundane, from behaviors that are quite common to those that occur rarely, and from behaviors that are homogeneous to those that are highly variegated.

Dietary intake consists of a set of frequent, mundane, and variegated behaviors that is believed by epidemiologists to have significant consequences for health status (e.g., Willett, 1990). The typical respondent to a Western health survey eats and drinks frequently, consuming a variety of foods without devoting very much attention either to the identity and quantity of what he or she eats or to the frequency with which he or she eats most items. Questions about dietary intake over extended periods of time strike one as being difficult to answer accurately. Consider, for example, such questions as "How many apples did you eat last year?" and "How many cups of coffee did you drink last month?" Nevertheless, nutritional survey procedures ask people to recall what they have eaten during extended periods and to report the frequency with which they have eaten specific food items during defined periods (for reviews and examples, see Block, 1982; Rizek & Pao, 1990; Willett, 1990; Young, 1981).

Two questioning procedures dominate the survey collection of self-report information about dietary intake: dietary-recall and food-frequency questions. In this chapter, we consider each of these procedures and describe experimental data that were collected to illuminate how respondents carry out the tasks presented by these methods. Analyses of respondents' performance permit preliminary theorizing about the cognitive processes used and, hence, an understanding of the validity of this sort of retrospective report. In particular, we consider the distinction between the implicit demands of a nutritional survey and how people actually respond to such a survey. Survey questions imply that respondents should retrieve and report memories of specific experiences, but respondents may instead use general knowledge to formulate their answers. As we shall discuss later, the apparent validity of responses may stem to a considerable ex-

tent from sources of knowledge other than those intended by the askers of survey questions.

Nathaniel Hawthorne (1850/1962), in his introductory sketch to *The Scarlet Letter*, described an individual with seemingly remarkable memory for his dietary intake. According to Hawthorne, the 80-year-old Custom House Inspector was distinguished by

his ability to recollect the good dinners which it had made no small portion of the happiness of his life to eat. . . . His reminiscences of good cheer, however ancient the date of the actual banquet, seemed to bring the savor of pig or turkey under one's very nostrils. There were flavors on his palate that had lingered there not less than sixty or seventy years, and were still apparently as fresh as that of the mutton chop which he had just devoured for his breakfast. [Hawthorne had] heard him smack his lips over dinners, every guest at which, except himself, had long been food for worms. It was marvellous to observe how the ghosts of bygone meals were continually rising up before him . . . . A tenderloin of beef, a hind-quarter of veal, a sparerib of pork, a particular chicken, or a remarkably praiseworthy turkey, which had perhaps adorned his board in the days of the elder Adams, would be remembered; while all the subsequent experience of our race . . . had gone over him with as little permanent effect as the passing breeze. (pp. 49–50)

Although such memory for dietary intake seems implausible, the Inspector, if his memories are accurate, represents the ideal respondent to a nutritional survey. Designers of health survey questions assume covertly that the Inspector's ability to report about specific past dietary experiences is a general ability.

The Inspector, in his ability to recall specific dietary experiences, is likely to be exceptional. (This, of course, is what makes him an interesting literary character.) Many individuals asked to report their dietary intake might fail to remember specific experiences in the detail provided by the Inspector. Nevertheless, the typical respondent need not be at a complete loss. An alternative plausible source of accurate reports about dietary intake is the general knowledge that an individual might have about his or her diet. Using such knowledge, an individual might correctly report a dietary experience without having a specific recollective memory of that experience. An individual might correctly report having eaten turkey on Thanksgiving Day in 1977; a member of a household in which Wednesday is Prince Spaghetti Day might correctly report having eaten spaghetti on the third Wednesday of January 1982; Cookie Monster might report correctly that he had chocolate chip cookies as a snack on Big Bird's birthday in 1983. Certain meals can be described accurately not because they are specifically remembered but because generic knowledge can be used to make plausible statements about them.

In this chapter, we discuss, in turn, data from a study that involved free report of dietary intake for a specified reference period and data from a companion study in which subjects estimated the frequency with which they ate each of a set of items during the reference period. After presenting the results of the free-report study, we propose some hypotheses about the nature of the memory structure that subserves performance in the free-report task and present simulation

data that confirm the plausibility of these hypotheses. We later show that the same model is capable of accounting for the frequency judgments.

## Free Report of Dietary Intake

In a dietary-recall procedure, a respondent is asked to recall and report everything that he or she has eaten and drunk during a specified period. The procedure is somewhat analogous to free-recall procedures used to study memory. In the simplest free-recall experiment, a list of unrelated items is presented to a subject and, at some time following the presentation of the last item, the subject is asked to report the items that were on the list. The subject's report is then scored for accuracy, for the degree and nature of the organization of reported items, or for other characteristics of theoretical interest. Such experimental procedures afford the investigator complete control over the to-be-reported material and the conditions under which it is acquired by the subject (see Tulving, 1968; Underwood, 1983). In dietary recall, the "list items" are ones that the subject has eaten and drunk, reports cannot typically be scored against a veridical record of the "list," and the investigator lacks control over both the to-be-reported material and the acquisition conditions.

Dietary-recall procedures differ in several additional respects from free-recall experiments conducted in psychological laboratories. First, the survey respondent who engages in a dietary-recall procedure is typically unaware, until the time that he or she is asked to report his or her dietary intake for a specified period, that a list of to-be-reported items has been experienced. Any report of these items must be based on incidental learning (rather than intentional learning). Second, insofar as the items that the respondent ate during a reference period constitute a list, the list is presented over an extended period, intermixed with other events in the respondent's life. Third, all of the items on the to-be-reported list are exemplars of a single semantic category. We shall further elaborate the special significance of this last point below.

We conducted a study (Smith, Jobe, & Mingay, 1991) that investigated the accuracy and content of free reports of dietary intake for specified reference periods. We manipulated acquisition, retention, and retrieval conditions in order to collect a set of data that would afford development of theoretical hypotheses about the memory representation accessed by respondents during dietary-recall procedures and the processes that operate on that representation.

Subjects were asked to report all food and beverage items that they had consumed during either a 2-week or 4-week reference period. Reports were collected either immediately following the reference period or 2, 4, or 6 weeks later. A well-established fact about memory is that reporting performance for a target item deteriorates as the time between acquisition and test increases. Retention interval was varied to determine whether this characterizes dietary reports. The length of the reference period was varied to assess the rate at which to-be-reported items are accumulated and to study possible tradeoffs in reporting

between reference period length and the degree to which sampled items are representative of individuals' diets (longer periods are more representative). An interaction of reference period length and retention interval length on reporting performance could be highly informative concerning the cognitive representations and processes that subservise dietary reporting. To help reveal the retrieval strategies used by respondents during the report task, subjects were asked, when they reported, to use one of three sets of cues—*temporal*, *meal*, or *food group*—to organize their reports.

Subjects' reports were scored against diaries that they had kept during the reference period. In Smith, Jobe, and Mingay (1992), we establish that reports of dietary intake by subjects who have kept food diaries do not differ from reports by subjects who have not recorded their food intake. The absence of an effect of diary keeping on reports permits the generalization of the results described both in that paper and in this chapter to individuals who report their dietary intake without having kept diaries.

In the next sections, we review the procedure, results, and conclusions of our study (Smith et al., 1991) of free-dietary reports. We then describe a simulation of our hypotheses about long-term dietary reports that yields results qualitatively similar to the empirical results.

## Method

Participating in the experiment were 170 subjects (115 of whom were female), recruited in the vicinity of Binghamton, New York. Subjects ranged in age from 18 to 80 and had the full range of educational backgrounds. Subjects kept diaries of their food intake for either 2 weeks or 4 weeks, submitting their daily records by mail every 3 or 4 days on a prescribed schedule. At a specified time following the end of the diary-keeping period, subjects came to the laboratory and were asked to report everything that they had eaten and drunk during the diary-keeping period.

The subjects were assigned to eight experimental groups defined by a combination of length of the reference period (2 or 4 weeks) and length of the retention interval, the time from the end of the reference period to the memory test (immediately or 2, 4, or 6 weeks). At the time of the memory test, subjects were given one of three sets of instructions that were intended to guide but not to constrain reports: Approximately one-third of the subjects were encouraged to report temporally, starting from the end of the reference period and thinking back toward the beginning (reverse chronological); one-third were cued with the names of meals (e.g., breakfast, snack); and one-third were cued with the names of food groups. (For additional details of the procedure, see Smith et al., 1991).

## Results

We describe first the procedure for scoring reports. We next analyze the impact on performance of retention interval and reference period length, and then discuss the consequences for performance of reporting instructions.

### Scoring Reports

Subjects' diaries and reports were scored using an ever/never criterion: An item was treated as having been recorded if it was ever recorded and as having been reported if it was ever reported. Thus, to score a subject's reported items against those that he or she had recorded, each set of items was edited to eliminate multiple occurrences of items and to combine different names for items with multiple names: For example, *green salad*, *tossed salad*, and *lettuce and tomato salad* would have been consolidated into a single entry. The recorded and reported items were then classified into three groups: *matches* (reported items that had been recorded), *intrusions* (reported items that had not been recorded), and *omissions* (items that were not reported that had been recorded). From these counts, two performance indices were calculated: The *match rate*,  $p(\text{Mtc})$ , is the number of matches divided by the number of recorded items; and the *intrusion rate*,  $p(\text{Int})$ , is the number of intrusions divided by the number of reported items. (These indices, their reliability, their definitional and empirical relationships, and the rationale for their use are described in detail in the Appendix to Smith et al., 1991.)

### Retention Interval and Reference Period Length

Table 8.1 shows the mean match and intrusion rates for subjects in each of the eight experimental conditions defined by reference period length and retention interval. Three aspects of the results are readily apparent. First, for reference periods of the lengths studied in this experiment, subjects' reports about dietary intake were far from accurate: The average match rate, over conditions, was 0.38, and the highest condition mean match rate was only 0.49. Second, subjects exhibited high intrusion rates: On average, approximately one-third of the reported items were scored as intrusions, and even the lowest condition mean intrusion rate was 0.27. Third, reporting performance appears to have deteriorated as retention interval was lengthened: Subjects who reported after longer retention intervals tended to have lower match rates and higher intrusion rates than did subjects who reported after shorter retention intervals.

A MANOVA was used to assess the dependence of performance on the lengths of the reference period and the retention interval. The tested hypothesis concerning retention interval was that the deterioration of performance over time is negatively accelerated (see, e.g., Ebbinghaus, 1885/1964). The contrast on the levels of retention interval that captures this hypothesis was significant,  $F(2,161) = 46.82$ ,  $p < .0001$ , and the residual was not, which confirms that the deterioration in reporting performance is indeed a negatively accelerated function of retention interval length. This pattern was seen in match and intrusion rates.

The MANOVA also showed a significant main effect of reference period length,  $F(2,161) = 4.28$ ,  $p < .02$ , and a significant interaction of reference period length and the contrast on retention interval,  $F(2,161) = 4.00$ ,  $p < .05$ . Follow-up univariate analyses revealed that these effects were due solely to differences among conditions in intrusion rates. Analysis of the interaction showed that

TABLE 8.1. Performance Measures for Report of Dietary Intake

Retention Interval (Weeks)	Reference Period Length					
	2 Weeks		4 Weeks		Mean	
	p(Mtc)	p(Int)	p(Mtc)	p(Int)	p(Mtc)	p(Int)
0	0.49	0.27	0.46	0.29	0.48	0.28
2	0.42	0.39	0.37	0.28	0.39	0.34
4	0.28	0.40	0.34	0.34	0.31	0.37
6	0.30	0.42	0.32	0.33	0.31	0.38
Mean	0.38	0.37	0.37	0.31	0.38	0.34

Note: This table is adapted from A. F. Smith, Jobe, and Mingay (1991). p(Mtc) is match rate; p(Int) is intrusion rate.

after the mean intrusion rates shown in Table 8.1 were adjusted for the main effects, the pattern displayed by the residuals was an *increase* over retention intervals for subjects who reported on 2-week reference periods and a *decrease* for subjects who reported on 4-week periods. The implications of this pattern will be elaborated below.

### Reporting Instructions

Although a reporting strategy was recommended to each subject, our primary interest was in obtaining maximum recall. Subjects were told, therefore, that they were always free to report whatever came to mind and were given substantial opportunity at the ends of their memory tests to report additional items that came to mind. Because subjects were free to use their preferred reporting strategies, we anticipated that instructed strategy would have relatively little effect on reporting performance. However, instructions had a decided influence on performance,  $F(4,290) = 4.40$ ,  $p < .01$ : Table 8.2 shows mean match and intrusion rates for subjects given each type of instruction (averaged across retention interval and reference period length). Performance of subjects given food-group cues differed from that of subjects given reverse chronological and meal instructions,  $F(2,145) = 7.61$ ,  $p < .001$ ; performance did not differ significantly between subjects assigned to these latter instructions,  $F(2,145) = 1.40$ ,  $p > .25$ . The instruction manipulation affected only intrusion rates, which were significantly higher when subjects were cued with food groups than when they were instructed to report reverse chronologically or by meal. None of the interactions of reporting instructions with the other manipulated variables was significant.

### Discussion

If validity of report is understood to mean accurate characterization of dietary intake during the reference period, then the retrospective reports of dietary intake of subjects in this study were far from valid. Substantial error rates were

TABLE 8.2. Performance Measures for Report of Dietary Intake, by Instruction

Instruction	p(Mtc)	p(Int)
Reverse chronological	0.37	0.31
Food group	0.39	0.38
Meal	0.36	0.32
Mean	0.38	0.34

*Note:* This is table adapted from A. F. Smith, Jobe, and Mingay (1991). p(Mtc) is match rate; p(Int) is intrusion rate.

observed: Subjects failed to report exhaustively the items that they had eaten and drunk, and they evidently reported many items that they had not consumed during the reference period.

Two questions may be raised at this point. First, should the failure of subjects to provide valid reports be a cause for concern to users of the reported information? Second, to what extent can the obtained results inform the development of a theory of the memory information available to subjects and the processes that subjects use to operate on that information? We consider first the theoretical implications of the results but shall address the practical implications—including whether the reporting inaccuracies should be a source of concern—later in this chapter.

Subjects in this experiment were instructed explicitly to retrieve and report specific memories about their dietary intake. Indeed, the results suggest that specific memories contribute to long-term dietary reports: The deterioration of measured memory performance as retention interval was lengthened may have been due, in part, to a decrease in reports of specific memories as time elapsed from the end of the reference period. However, two aspects of the data implicate generic knowledge as a substantial basis for subjects' reports: the high intrusion rates and the pattern of intrusion rates over experimental conditions.

Intrusion rates were substantial even at the shortest retention interval—reports on the day after the last day of the reference period. This indicates that subjects do not base their responses exclusively on their dietary experiences during the reference period. Reported items that were scored as intrusions were likely to be items that these subjects had eaten at some time during their lives—they had just not eaten them during the specified reference period.

The observed intrusion rates represent a distinct difference between the results of this experiment and the results of typical free-recall experiments. Intrusions tend to be rare in free recall of lists of nominally unrelated items (see, e.g., Erdelyi, Finks, & Feigen-Pfau, 1989), although they occur regularly in recall of lists of categorically related words (Cofer, 1967; Cofer, Bruce, & Reicher, 1966; Rabinowitz, Mandler, & Patterson, 1977). Intrusions in recall of such lists tend to be items from list-relevant categories. An experimental subject who remembers that a number of words from a list were exemplars of a particular category may report items from that category even without being certain that those items

were on the list. Such items would, indeed, be good candidates for report. For subjects who reported their dietary intake, there was a single list-relevant category, "things that I ate during the reference period." Subjects may have sampled items from the category of personal dietary items but failed to discriminate accurately whether the sampled items had actually been eaten during the reference period, that is, whether they had been on the "list" (see Wagenaar, 1986, concerning the crudeness with which temporal information about autobiographical events is coded).

The observed interaction of retention interval length and reference period length on intrusion rate is consistent with the hypothesis that subjects rely on generic knowledge when they report their dietary intake. The informative index is the residual of the condition mean intrusion rate after it has been adjusted for the main effects. For subjects who reported on 4-week reference periods, these residuals decreased as retention interval length increased, whereas for subjects who reported on 2-week periods, the residual intrusion rates increased over retention intervals.

This pattern of intrusion rates could result if subjects simply described their typical diets and in addition, at short retention intervals, reported some specific recent memories. A 4-week period is likelier than a 2-week period to contain any item that an individual routinely eats, so that subjects who reported on 4-week periods were considerably freer to report routine elements of their diets without concern for precisely when those items were eaten. When report is delayed, subjects who report about longer reference periods may exhibit lower intrusion rates than subjects who report about shorter periods, not because they remember better but because what they report better matches what they are supposed to report. If subjects restricted their reports to items that they had ever eaten, then, if the reference period were made sufficiently long, intrusion rates would tend to zero and match rates would depend only on the number of items reported.

The conjecture that reports consisted of reports of generic knowledge supplemented by reports of specific memories is consistent with results of a variety of laboratory experiments on memory. In experiments in which subjects are asked to recall lists of categorically related words, intrusion rates increase as retention intervals are lengthened (Cofer, 1967; Cofer et al., 1966; Rabinowitz et al., 1977). Similarly, in recall of more complex material, the influence of memory schemas on retrieval increases as retention interval is lengthened: Graesser (1981) proposed that after some amount of time has elapsed following the presentation of to-be-remembered material, virtually nothing specific is remembered, and that after that point, subjects respond entirely on the basis of a relevant schema. Reder (1987b) distinguished between direct retrieval and plausible inference as strategies for question answering and showed that as the amount of time since the learning of information relevant to a question increased, plausible inference supplanted direct retrieval as the dominant strategy of question answering.

The effect of reporting instructions on performance that we observed (Smith

et al., 1991) is consistent with this interpretation of the data. The intrusion rates of subjects cued with food-group terms were higher than those of subjects given reverse chronological or meal instructions. If subjects' general knowledge about food is organized according to food groups, presentation of food-group cues may facilitate the retrieval and report of items from memory, with reports consisting of items that are category exemplars rather than items from specific meals. Food groups are a plausible basis of organization of general knowledge about food and personal diet, and the effect of reporting instructions on performance was consistent with the use by subjects of such knowledge when they reported.

## Simulation

A core aspect of this account of dietary-reporting performance is that subjects rely on generic knowledge during dietary-recall procedures. A subject asked to report dietary intake for a specified reference period should search memory for items that are members of the category "things that I eat" *and* that were eaten during the reference period. The correctness of a response depends on the combination of category and contextual (temporal) information. However, although items that are members of the relevant category may be accessible in memory, subjects may find that items that were eaten during the reference period are difficult to discriminate from items that were not. Temporal information may be generally ineffective as a basis for a search of memory for experiences over an extended time (see Brewer, chap. 1, and Bradburn, Huttenlocher, & Hedges, chap. 13, this volume; Wagenaar, 1986). Given the availability of plausible responses, but lacking certainty regarding their correctness, subjects may report some items with the hope that they will happen to be correct.

To evaluate this hypothesis, we implemented a partial simulation of the dietary-report task. The limited purpose of the simulation was to evaluate whether the observed pattern of performance over reference period lengths and retention intervals was consistent with the conjecture that dietary reports consist of generic knowledge supplemented, at short retention intervals, by reports of specific memories. We describe first the cognitive structures and processes assumed for the model, and then we relate these to current theories of memory representation and the recall process.

### *Model*

The model assumes that an individual asked to report his or her dietary intake for a specified period samples candidate items from a categorically organized memory and, with some probability, reports sampled items. By postulating categorical organization, we assume that the individual's search of memory is effectively restricted to items from the relevant category (i.e., dietary items). We assume a memory that contains two kinds of elements, permanent and episodic, that may be labeled to represent items. The population of permanent elements reflects the long-term base rates with which items are eaten, whereas the epi-

sodic elements are labeled during particular episodes of dietary intake. The probability of sampling an element with a particular item label thus depends both on the base rate at which it is eaten and on the frequency with which it has been eaten during the recent past. We assume that a single episodic element is labeled for each occasion on which an item is eaten. Thus, the impact of an eating episode on the probability of sampling a rarely eaten item will be much greater than the impact of such an episode on the probability of sampling a frequently eaten item, although the probability of sampling the rarely eaten item will still be low. Conditional on an element being sampled, we assume that its item label is reported with some probability that is constant for permanent elements but is a decreasing function of time since the episode for episodic elements.

For each simulated subject, a list of to-be-reported items (i.e., the memory set) is generated by sampling from a specified number of unique items. For each sampled item, an episodic element is labeled; functionally, the probability of that element being reported, given that it is sampled, begins immediately to decline. To simulate reporting, the number of items to be reported is determined from the number of unique items experienced during the acquisition phase (i.e., the number "recorded") and the empirical relationship between numbers of recorded and reported items. The reporting mechanism samples elements from memory—the population of permanent elements augmented by labeled episodic elements—and if the item has not yet been "reported," "reports" the item with the appropriate probability. Note that in this simulation, because the numbers of recorded and reported items are specified beforehand, the question of interest is the relationship between these two sets of items—in particular, the nature of any overlap between them. Such overlap is indexed by match and intrusion rates.

This model is consistent, in several respects, with features of prominent current theories of memory. For example, the notion that a new element is labeled for each episode on which an item is eaten is consistent with the class of theories of memory that assume that multiple experiences with a single item are represented by labeling multiple memory elements rather than by strengthening a single element (e.g., Hintzman, 1988).

The assumptions of categorical organization and the decay of episodic elements, as well as the separation of the sampling and report processes, are compatible with Raaijmakers and Shiffrin's (1980, 1981; Gronlund & Shiffrin, 1986) theory of probabilistic search of associative memory (SAM). In the SAM model, long-term memory is populated by images that may be retrieved when memory is probed with a cue. Memory images are supposed to be complex information structures that may contain item information, category information, and information about the context in which an item was experienced. SAM is a multiple-trace model in that new images may form when an item is experienced in new contexts. What we describe as an episodic element is analogous to the contextual information in an image in SAM. Raaijmakers and Shiffrin (1981) note the possibility of memory images that contain no contextual information.

We postulate that memory is categorically organized and that search, during the dietary-recall task, is functionally restricted to items in the category "things

that I eat.” Equivalently, we assume that memory elements are labeled not only with item names but also with category names and that search can be carried out efficiently on the basis of category labels. We assume a class of episodic elements with probabilities of report, conditional on the sampling of the element, that depend on the time since these elements were labeled. This is equivalent to postulating that memory elements have contextual labels and that the similarity of context at test to context at the time that an episodic element was labeled decreases over time, thereby decreasing the effectiveness of the test context as a retrieval cue.

### *Simulation Method*

For the simulation run reported here, we assumed a pool of 200 items with a common probability structure for all simulated subjects. Data were simulated for 24 subjects in each of four conditions defined by crossing orthogonally two reference period lengths (2 weeks and 4 weeks) with two retention interval lengths (immediate test and delayed test).

For each simulated subject, the number of unique items to be experienced during the acquisition phase was specified by sampling from a distribution with mean and variance equal to those of actual subjects. Sampling of item episodes during the acquisition phase was from this restricted subset of items in which the relative probabilities of items were equal to their relative probabilities in the total set of items. For each simulated day of the reference period, the number of items to be sampled was drawn randomly from a distribution that was uniform over the range 8 to 12. To reiterate, sampling an item (i.e., experiencing an item) involved labeling and dating an episodic memory element.

From the number of unique items experienced by a simulated subject and the empirical relationship between numbers of items recorded and reported, the number of unique items to be reported by that simulated subject was calculated. The reporting process was simulated as follows: On each cycle, an element was sampled from the set of permanent elements augmented by labeled episodic elements. Since every element had an equal chance of being sampled, the probability of sampling an element representing any item depended on the total number of elements—permanent and episodic—labeled for that item. If a sampled element was labeled as an item that had already been reported, it was not considered further (because the simulation was programmed to run until some number of unique items had been reported). If a sampled element was labeled as an as-yet-unreported item, the item was added to the report list either with probability that depended on the recency of the episode for episodic elements or with fixed probability for permanent elements. The probability of reporting an episodic element, conditional on its having been sampled, was assumed to decline exponentially over 14 days. An episodic element was treated as inaccessible after 14 days and, if sampled, was not reported. The report probability of a permanent element, if sampled, was equal to the report probability for 14-day-old episodic elements.

For each simulated subject, the list of "reported" items was scored against the list of items sampled during the acquisition phase, and match and intrusion rates, defined above, were calculated.

### *Simulation Results and Discussion*

Table 8.3 shows the results of one simulated experiment using the two performance indices introduced previously. Note that although the numbers of unique recorded and reported items were fixed by design, no constraints were placed on the overlap relationship between these two sets. It is this overlap relationship, and particularly the form of its dependence on reference period length and retention interval, that are of interest: The nature of this relationship stems from the probability structure of the items in memory and the probabilities of reporting items conditional on the sampling of labeled elements.

The simulation data were subjected to the same analysis as the actual data, and key aspects of the simulation results accord with the empirical results. Although the exact levels of the match and intrusion rates differ from those observed empirically, the general pattern of results mimics closely that of the data from the experiment described earlier. Match and intrusion rates, considered jointly, depended on retention interval length, reference period length, and their interaction. As was true for the actual data, follow-up univariate analyses showed that each performance measure depended significantly on retention interval length—for match and intrusion rates,  $F(1,91)s=121$  and  $166$ , respectively. However, only intrusion rates depended significantly on reference period length and on the interaction of reference period length and retention interval. Overall, match rates did not differ for the two retention intervals,  $F(1,91)=1.42$ ,  $p > .20$ , but intrusion rates were significantly higher when test was delayed than when it was immediate,  $F(1,91)=2.85$ ,  $p < .001$ . As was true for the actual data, the mean intrusion rates for both reference period lengths were higher for tests that followed long delays than for immediate tests. However, the difference was considerably larger for simulated subjects who reported on a 2-week period than for simulated subjects who reported on a 4-week period,  $F(1,91)=10.86$ ,

TABLE 8.3. Performance Measures for Simulated Report of Dietary Intake

Interval to Test	Reference Period Length					
	2 Weeks		4 Weeks		Mean	
	p(Mtc)	p(Int)	p(Mtc)	p(Int)	p(Mtc)	p(Int)
Immediate	0.56	0.22	0.57	0.19	0.57	0.21
Delayed	0.39	0.48	0.34	0.35	0.37	0.42
Mean	0.48	0.35	0.46	0.27	0.47	0.31

Note: p(Mtc) is match rate; p(Int) is intrusion rate. Each cell represents the mean of 24 simulated subjects.

$p < .001$ . Consistent with the actual data, the interaction of reference period length and retention interval on match rate,  $F(1,91) = 2.45$ ,  $p > .10$ , was not significant.

The model that has been described and simulated is intended only as a partial account of free reports of dietary intake. Certain features of the model are admittedly ad hoc and were implemented for convenience. For example, it is overly simplistic to assume that a single distribution of base rates over a particular number of items is appropriate for all subjects. Arbitrary assumptions were made about the relative sizes of the populations of permanent and episodic memory elements. In addition, the size of the acquisition subset was restricted artificially, although acquisition is not a psychological process and no model of it is required. In contrast, the number of items actually reported by subjects is presumably a consequence of psychological processes (see, e.g., Raaijmakers & Shiffrin, 1980), but in this simulation, the number of items to be reported was specified prior to report. Nevertheless, for the limited objectives of the simulation, these shortcuts seem tolerable.

### *Summary*

The simulation of report involved two basic features: (a) random sampling of memory elements and (b) reporting of corresponding labels that depended on element "age" for episodic elements and was constant for permanent elements. The primary intent of the simulation was to investigate whether this model would yield a pattern of performance measures, given variation in reference period length and retention interval, that resembled that of actual data. (We did not attempt to model the effect of the reporting instruction manipulation.) The model was successful in that it yielded data that resemble those of real subjects. We recognize that the simulation almost surely fails to capture important aspects of the performance of real subjects: Real subjects probably behave less randomly than does the simulation, and they likely exercise some judgment during the reporting process (see Gronlund & Shiffrin, 1986; Walker & Kintsch, 1985). Nevertheless, the achievement of the simulation is that its results confirm the plausibility of the hypothesized memory representation and reporting process and provide a point of departure for further exploration of free reports of dietary intake and psychologically similar tasks.

## Reports of Frequency of Dietary Intake

The results of the free-report experiment described in the first part of this chapter indicate that subjects' reports are far from accurate. The data, along with a set of well-established principles about memory, served as a basis for a model of how subjects carry out the dietary-recall task. However, neither the described task nor the reported analyses are representative of the data collection strategies of nutritional epidemiologists: Free reports of dietary intake for extended peri-

ods, collected after long retention intervals and scored using an ever-eaten criterion, are simply not typical of the methodology of epidemiology.

For investigations of dietary intake over periods as long as one month, epidemiologists are typically interested not only in whether an item was ever eaten but also in how often it was eaten. Food-frequency questionnaires ask respondents to provide either count or rate information concerning their intake of each of a set of food and beverage items during a reference period that might range from one month to one year. An example of a count question is, "How many artichokes did you eat last year?" An example of a rate question is, "How often, during last year, did you eat an artichoke?" Equivalent responses to these two questions are "104" and "twice per week," respectively.

Studies of frequency judgments have shown that people are highly sensitive to variations among events in their frequencies of occurrence (e.g., Hasher & Zacks, 1984; Hintzman, 1988; Jonides & Naveh-Benjamin, 1987; Underwood, Zimmerman, & Freund, 1971). Such studies usually involve the presentation of a relatively small number of unique events over a short time span, relatively little variation over items in frequency of occurrence, and the administration of a memory test shortly after the set of target events has been presented (but see Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; Underwood et al., 1971). In contrast, the reference events of a food-frequency questionnaire are many in number, have occurred with a wide range of frequencies over an extended time period, and constitute but a small fraction of the entire set of events experienced by the respondent during the reference period.

A companion study to the free-report experiment described earlier in this chapter (Smith, 1991, Experiment 4) investigated the validity of food-frequency estimates. Subjects were asked to indicate how often they had eaten (or drunk) each of a set of items during a specific reference period. Since the subjects were a subset of those who participated in the free-report experiment, they experienced the reference period length and retention interval manipulations that have been described. The impact on the quality of food-frequency estimates of manipulating retention interval length may further illuminate the nature of the memory representations and cognitive processes used in responding to food-frequency questionnaires. In this section, we make predictions about food-frequency estimation performance and summarize the experiment reported by Smith (1991). We then assess the food-frequency estimation performance of the simulation model that we described in the previous section.

Suppose, to make frequency estimates, that memory is sampled for labeled elements that correspond to a designated target item and that frequency estimates are some nondecreasing function of the number of retrieved, appropriately labeled elements. Suppose further that the memory structure that is sampled is that discussed earlier: a representation of generic knowledge about personal dietary intake augmented by representations of specific recent dietary experiences. Frequency estimates for 2-week periods should correspond closely to actual frequencies when judgments are collected immediately after the end of the reference period, because recently labeled elements will contribute substan-

tially toward estimates about target items. As the amount of time between the end of the reference period and the test increases, sampling may be restricted to the permanent elements in which generic knowledge is represented, so that the correlation between estimated and actual frequencies will be lower. For an individual, the variability in dietary experiences among 4-week reference periods should tend to be less than the variability among 2-week reference periods, because 4-week periods encompass a larger sample of experiences. Consequently, the dietary intake for a just-ended 4-week period will be more similar to a typical 4-week period than will that for a just-completed 2-week period be to a typical 2-week period. Although the retrieval of memories of recent episodes will improve the correspondence of estimated and actual frequencies when frequencies are estimated immediately after the end of the reference period, there should be less deterioration over time in the relationship between actual and estimated frequencies for 4-week periods than for 2-week periods.

## Method

Participating in this experiment were 128 of the participants in the free-report experiment. The design of the experiment paralleled that of the free-report experiment: Subjects served in conditions defined by crossing two reference period lengths (2 and 4 weeks) with four retention intervals (immediate test and 2-, 4-, and 6-week delayed tests). (Recall instruction was irrelevant to this experiment.)

After completing the free-report procedure, each subject was asked to indicate how many times, *during the reference period*, he or she had eaten each of a list of items. (Smith, 1991, Experiment 5, showed that administering a recall test prior to collecting frequency estimates has no discernible effect on the frequency estimates.)

For each subject, a special food-frequency questionnaire was constructed, based on the items that he or she had recorded during the reference period and on the frequency with which those items had been recorded. The questionnaire included items with a broad range of frequencies of recorded occurrence. Recorded items were counted, grouped if appropriate (e.g., *chocolate chip cookies* and *ginger snaps* would have been combined into a class labeled "*cookies*"), and sorted by frequency. Then, up to 25 items were selected for inclusion on the questionnaire, with the constraints that the 5 most frequently recorded items be tested and that approximately equal numbers of items be sampled from each of the four quartiles of the remainder of the distribution.

## Results and Discussion

Because subjects who made food-frequency judgments about a 4-week reference period ate their most frequently eaten items substantially more often than did subjects who made judgments for a 2-week period, recorded counts and frequency estimates were transformed logarithmically to equate the variances of recorded and reported frequencies over conditions. For each subject, the corre-

lation between log-transformed estimates and log-transformed counts was calculated. The analysis of variation in performance over conditions was carried out on  $z$ -transformed correlations. (For further discussion of the choice of the correlation coefficient as the performance index for this experiment, see Smith, 1991, pp. 16–17; for a complete discussion of measures of memory for frequency of occurrence, see Naveh-Benjamin & Jonides, 1986.)

Table 8.4 shows the mean correlation, over subjects, for each experimental condition defined by a combination of reference period and retention interval lengths. For subjects who reported frequencies for 2-week periods, the mean correlation decreased monotonically as retention interval lengths increased, whereas for subjects who reported frequencies for 4-week periods, the mean correlation was highest for subjects who reported immediately after the end of the reference period and lower, but approximately constant over retention intervals, for subjects who reported after a delay. A test of this pattern on the means of  $z$ -transformed correlations was significant,  $F(1,120) = 28.16$ ,  $p < .0001$ ; the residual was not significant,  $F(6,120) = 1.04$ ,  $p > .25$ . These results are entirely consistent with the predictions described earlier.

### Simulation

A frequency estimation experiment was simulated to investigate further the adequacy of the hypotheses concerning memory structure and reporting processes that were proposed to account for the free-report data. Data were simulated for each of 24 artificial subjects in each of the four conditions yielded by crossing orthogonally two levels of reference period length (2 weeks and 4 weeks) and two levels of retention interval (immediate test and delayed test). The simulation of item acquisition during the reference period was identical to that described earlier. To estimate frequencies, the memory representation described earlier was probed for a target item some specified number of times, the sampled memory element was retrieved with a specified probability, and if a retrieved memory matched the probe item, a counter was incremented. These counts of

TABLE 8.4. Mean Correlations of Log(Frequency Estimates) with Log(Number of Recorded Occurrences)

Retention Interval (Weeks)	Reference Period Length		Mean
	2 Weeks	4 Weeks	
0	.86	.84	.85
2	.79	.76	.77
4	.70	.77	.74
6	.68	.77	.73
Mean	.75	.78	.77

*Note:* This table is adapted from A. F. Smith (1991, Experiment 4). Analysis was conducted on  $z$ -transformed correlation coefficients.

retrieved elements that matched the probe item were taken as proxies for frequency estimates.

To simulate frequency judgments that were made with a delay interposed between the reference period and the test, we assumed that the number of samples of memory depended on the length of the reference period. Since the probability that a sampled element would be retrieved (and hence counted) was the same for all memory elements, the functional number of samples per day of reference period was the same for each reference period length. In contrast, to simulate frequency judgments that were made immediately after the end of the reference period, we did not assume any difference in the number of samples as a function of the length of the reference period. We assumed that the respondent would perceive judgments for the 2-week reference period as a more plausible task, and therefore we made the effective number of samples per day of reference period larger for 2-week reference periods than for 4-week reference periods.

Table 8.5 shows mean correlation coefficients, over simulated subjects, between the logarithms of experienced and estimated frequencies. Although the absolute levels of the correlations are rather different from those observed with actual subjects, particularly with a delayed frequency test, the qualitative pattern is the same as that observed in the data on real subjects (see Table 8.4). For simulations of frequency judgments for each reference period length, the mean correlation decreased as retention interval increased, but the decrease was sharper for simulated frequency judgments for 2-week reference periods than for 4-week reference periods. An analysis of variance confirmed the presence of a significant interaction of reference period length and retention interval on the *z*-transformed correlation measure,  $F(1,92) = 15.02, p < .01$ .

### Comment on the Validity of Frequency Judgments

The high degree of linear relationship between reported and experienced frequencies suggests that, in some respects, frequency judgments are valid. However, the correlation of reported and experienced frequencies reflects but one aspect of

TABLE 8.5. Mean Correlations of Log(Simulated Frequency Estimates) with Log (Simulated Number of Recorded Occurrences)

Interval to Test	Reference Period Length		
	2 Weeks	4 Weeks	Mean
Immediate	.75	.72	.74
Delayed	.50	.62	.56
Mean	.64	.68	.66

Note: Each cell represents the mean of 24 simulated subjects. Analysis was conducted on *z*-transformed correlation coefficients.

validity of frequency judgments, and this is probably not the most significant aspect (see Naveh-Benjamin & Jonides, 1986). The manipulations of this experiment mandated the use of the correlation coefficient as the performance index, but such aspects of frequency judgments as the slope of the regression of estimated on experienced frequencies and percentage error are important as well. In addition, the analyses of these data were concerned with the relationship of estimated and experienced frequencies within individuals over food items: The high correlations show that subjects report more experiences with items that they recorded more frequently and report fewer experiences with items that they recorded less frequently. Epidemiologists are concerned with analyses of frequencies within food items over individuals, that is, with whether individuals who eat a particular item more often report larger frequencies than do those who eat the same item less often. A study of food-frequency estimates (Smith, 1991, Experiment 5) showed that the quality of relative frequency estimates (within individuals over items) is high but that the quality of absolute estimates (within items over individuals) is poorer. High-quality absolute estimates require that all respondents be calibrated to the same frequency scale (i.e., that absolute values of judgments have the same meaning to all subjects), but this appears not to be the case. Development of calibration procedures would surely enhance the value of frequency judgments to survey data users.

## General Discussion

### Summary

In this chapter, we have described two experiments that were designed to investigate the cognitive structures and processes that subservise long-term dietary reports. The general research strategy was to manipulate the conditions under which to-be-reported information was acquired, retained, and retrieved and, by studying the impact of these manipulations on performance, to refine hypotheses about the relevant representations and processes.

Experiments were conducted to investigate essential aspects of each of the two predominant methods of collecting information about nutritional intake: dietary-recall and food-frequency questionnaires. In the free-report experiment, subjects exhibited negatively accelerated deterioration in reporting performance as the amount of time between the end of the reference period and the memory test increased. This was manifested both as a decrease in match rates and as an overall increase in intrusion rates. However, after intrusion rates were adjusted for the main effects of reference period length and retention interval, it was clear that over retention intervals of increasing length, intrusion rates increased for subjects who reported on relatively short reference periods (2 weeks) but decreased for subjects who reported on relatively long reference periods (4 weeks). Reports of food frequency were evaluated by calculating correlations between experienced and reported frequencies. For reports on 2-week reference periods,

the quality of reports was a decreasing function of retention interval length, whereas for reports on 4-week periods, nonimmediate reports were worse than immediate reports but otherwise did not depend on the length of the retention interval.

On the basis of the results of the free-report experiment, we developed a model of the memory structure that subserves dietary reports. We proposed that individuals report their dietary intake by addressing a memory that consists of two classes of labeled elements. When these elements are sampled during a search of memory, their labels, which denote items, are reported with class-specific probabilities. A population of permanent elements, labeled with category and item information, reflects the base rates at which an individual consumes the routine elements of his or her diet. A set of episodic elements is available to be labeled with category, item, and contextual information when items are experienced. Item labels of permanent elements are reported with probability that is independent of time; item labels for episodic elements are reported with probability that decreases as a function of time since the episode represented by the memory elements. A simulation of the central parts of this account yielded data patterns that were qualitatively similar to those of the results of each of the described experiments.

## Practical Implications for Health Survey Methodology

The results of the experiments showed that reports of dietary intake are far from valid. We suggested that an individual's generic knowledge about his or her dietary intake contributes significantly to reports about dietary intake during specific reference periods, and we demonstrated the compatibility of this idea with other results in the literature on memory. A simulation of the account was able to generate a pattern of data qualitatively similar to that yielded by real subjects.

The experiment in which frequency estimates were collected indicated that subjects are quite capable of reporting frequencies that are consistent within individuals. The pattern of results that measured relative frequency judgment performance was well mimicked by the simulation program that assumed sampling of generic knowledge about diet augmented by memories of recent episodes. The experiment reported here did not address absolute frequency estimates; but Smith (1991, Experiment 5) reported that the quality of absolute estimates of food frequency is lower than that of relative estimates and suggested that this was due, in part, to a lack of frequency calibration among individuals.

Should the use of generic knowledge in dietary reports and food-frequency estimates be of concern to users of nutritional survey data? To the extent that users of nutritional survey data require precise estimates of dietary intake for extended periods of time, these results should be of concern. However, respondents may be able to report creditably about their average (i.e., routine) dietary intake. For example, if the items reported by respondents that were scored as

intrusions tend to have more dietary significance than do the items that were eaten by respondents that went unreported, free reports supplemented by calibrated frequency estimates could yield adequate portraits of dietary intake. If average dietary intake is what people can report about, perhaps it is what should be asked about. Asking for frequency estimates as a follow-up to free reports could improve considerably the efficiency of nutritional surveys, which often involve asking respondents about many items that they rarely, if ever, eat.

### Concluding Remark

Recently, investigators of memory have engaged in considerable discussion of the value of studies of memory in which investigators relinquish some of the control over conditions afforded by the psychological laboratory (e.g., Banaji & Crowder, 1989; M. A. Conway, 1991). Banaji and Crowder (1989) criticized the “everyday” approach to memory as being overly preoccupied with ecological relevance, endorsed the laboratory as the locus in which generalizable principles are discovered and theories are tested, and faulted the “everyday” approach for its failure to produce new theories or discover new principles.

Roediger (1991) observed that many of those who responded to Banaji and Crowder’s (1989) criticisms of the everyday approach to the study of memory failed to appreciate the authors’ main point—that experimental control is essential to understanding the phenomena of memory. We agree. What remains to be settled is the extent to which some control might be relinquished in order to investigate the scope of generality of a principle or theory. In the studies reported in this chapter, we relinquished considerable control over the conditions of acquisition of the to-be-reported material: Subjects were responsible for recording the list of items that they would subsequently report on, and there was considerable variability, over subjects, in the length and diversity of these lists. On the benefit side, however, we showed that important principles that have been studied in the laboratory are operative under substantially less controlled conditions. Although the results of these experiments perhaps yield no new principles of memory, our confidence in and respect for the old principles are strengthened by finding that they are relevant even when the controlled conditions under which they were discerned are relaxed (see also Morton, 1991).

*Acknowledgments.* This research was supported by DHHS contract 282-87-0039 between the National Center for Health Statistics and the Research Foundation of the State University of New York, under grant SES-8612320 from the National Science Foundation to NCHS.

We appreciate the assistance of Joseph Belluck, Adrian Clark, Elsa Issa, David Manzer, and Kathryn Murphy in carrying out this research; the equipment resources loaned by Peter Campos; the very helpful comments of Mahzarin Banaji, Ralph Miller, David Mingay, David Payne, and Deborah Prentice on a previous draft of this manuscript; and the special contributions of Snack Smith to this research effort.

## 9

# Errors of Experience: Response Errors in Reports about Child Support and Their Implications for Questionnaire Design

*Nora Cate Schaeffer*

Reporting errors in surveys are partly shaped by the structure of respondents' experiences. This happens because respondents often use estimation strategies to supplement memory, and these strategies use information about the frequency and other characteristics of events (e.g., Blair & Burton, 1987; Burton & Blair, 1991; Menon & Sudman, 1989). When events have a simple structure, estimation strategies may give accurate results; but when events are complex, estimation is more difficult and may fail. Consider the example of reporting annual earnings, a task commonly requested of respondents. Some respondents will know their previous years' earnings. Others, however, will estimate them, using information about the frequency and other characteristics of the event "getting paid." A person who is continuously employed at a constant monthly salary can estimate earnings very accurately by multiplying the amount of the most recent paycheck by 12. In contrast, estimation is more difficult, and probably less accurate, for someone who is employed erratically. Such a respondent must use a decomposition strategy and may recall different employment spells and perhaps different wage rates to calculate total earnings for the previous 12 months. Thus, these two employment experiences may lead to different reporting errors.

These two employment experiences may also give rise to different "true" values of earnings, since the person who is continuously employed may earn more than the person who is not. If respondents whose actual earnings differ also differ in the accuracy of their reports, errors and true values may be correlated. A correlation between errors and true values violates an important assumption of many analytic models and may bias resulting estimates (Bohrnstedt, 1983; Duncan & Hill, 1985; Duncan & Mathiowetz, 1985).<sup>1</sup>

---

<sup>1</sup>Other important assumptions about errors may also be violated. Errors may not have a zero mean, may not be normally distributed, and may be correlated with the errors in other variables as well as with the true values of variables in an analytic model (e.g., Duncan & Hill 1985; L. C. Hamilton 1981; Wyner 1980).

Designing instruments to improve the accuracy of self-reports requires understanding the complex interrelationships among the structure of events in respondents' lives, the estimation strategies that they use to supplement memory, and the resulting errors. Although recent research has begun this task (e.g., Bradburn, Rips, & Shevell, 1987; Burton & Blair 1991; Menon & Sudman, 1989), we still lack a general framework for systematically analyzing what features of experience may influence respondents' choice of estimation strategies or the resulting reporting errors. This chapter builds on the work of others (e.g., Duncan & Mathiowetz, 1985; Mathiowetz & Duncan, 1988; Sudman & Bradburn, 1974), to contribute to the development of such a framework.

The analysis presented in this chapter examined two key variables in the study of contemporary family relationships: child support awards and payments. Child support is money that a nonresident parent pays to the parent who lives with their children for the children's support. Questions about child support implicitly ask respondents to summarize characteristics of a class of discrete events, and the analysis examined how the characteristics of events affect reporting errors. There are clear parallels between child support and other topics encountered in surveys. For example, questions about earnings also ask respondents to summarize one characteristic (the amount) of a host of discrete events that fall into the class of income transfers. The analysis attempted to identify important structural features shared by many events implicitly asked about in surveys that may affect the success of respondents' estimation strategies and the consequent accuracy of their reports.

The data analyzed are self-reports obtained during telephone interviews in 1987 with a sample of divorced Wisconsin mothers who resided with their children. Records from the Wisconsin court system about child support awards and payments provided the criterion used to estimate reporting errors.

## Data and Sample

To describe the characteristics of child support awards and payments that may affect reporting errors, I first present a brief overview of the child support system in Wisconsin, which is the source of the data that I analyzed. Variations on the most common experience with the system are noted later, when I describe the measures used in this study. I then give a brief description of the Parent Survey, which obtained self-reports about child support awards and payments.

### The Child Support System in Wisconsin and the Court Record Data

Child support awards in Wisconsin are made by the court, and most awards involve monthly cash transfers from the nonresident parent (who does not live with the children) to the resident parent. In most cases, either the nonresident parent's employer or the nonresident parent is required to send support pay-

ments to the Clerk of Courts. The Clerk of Courts records that the payment was made and then forwards it to the resident parent. The court may order that payment be withheld from the nonresident parent's pay if he (or she) is delinquent in making support payments. In addition to cash transfers, the court may obligate the nonresident parent to provide the children with other benefits, such as payments on a mortgage or health insurance. Such transfers are typically not made through the courts; the nonresident parent may make these payments directly (e.g., insurance premiums may be deducted at work or mortgage payments made to the mortgage holder) or to the resident parent. The Court Record Database (CRD), based on abstracts from court records, includes information about awards, regular support payments that nonresident parents make to the county Clerk of Courts, and other case activity. (See Garfinkel et al., 1988, for details of the study design.)

### The Parent Survey

The Parent Survey (PS) attempted to operationalize "child support" in the same way as the CRD does: regular cash transfers from the nonresident to the resident parent. In the PS, the resident parent was identified based on the report of the first parent interviewed about where a randomly selected child actually lived. The PS questions distinguished among parents with court-ordered agreements for regular support, those with informal agreements for regular support, and those with no agreements for regular support. The questions asked for the amount of regular support owed under either a formal or an informal agreement and asked respondents to exclude other financial assistance. To simplify reporting the amount of support paid, the PS used a series of questions based on the National Survey of Families and Households (Sweet, Bumpass, & Call, 1988). The series asked whether all, some, or none of the payments owed in 1986 were received. If only some payments were made, the respondent was asked the amount of the payments. If all payments were made, the amount of support received was assumed to be equal to the amount owed. (Exact question wording is available in Schaeffer, Seltzer, & Klawitter, 1991.) To augment the court record, the PS also asked about circumstances that might affect the accuracy of the CRD for that parent.

### The Sample

The data were provided by a sample of court cases in 20 Wisconsin counties. The CRD cases used as the sample frame for the Parent Survey included child-support-eligible family court cases (i.e., those with minor children who have two living parents) that entered the court system between 1984 and 1986. The PS sample included 1,390 divorce cases from the CRD in which the court indicated that the mother had physical custody during most of 1986. Thus, CRD information on who was the resident parent was used to define the sample.

Beginning in the late spring of 1987, telephone interviews were conducted with 964 of the mothers in these court cases. Of the mothers in the sample, 122

refused to be interviewed, and 296 could not be located despite extensive locating procedures that included searching records of the Wisconsin Department of Transportation, welfare agencies (AFDC), and other state agencies (see Lin, 1990). Since interviewers determined that they had reached the correct respondent by verifying the name, court date, and number of children in the court case, mismatches are probably minimal (Miller & Groves, 1985). Interviewers were not provided and respondents were not prompted with other information from the CRD. Most questions in the interview asked about events in 1986.

For several reasons, this analysis included only mothers with physical custody. Since it is still by far the most common custody arrangement and since survey participation rates are often higher for women, most studies of separated families or about child support focus on families in which the mother is the resident parent (see, e.g., Seltzer, 1990; Teachman & Polonko, 1990; U.S. Bureau of the Census, 1990). Applying the results of the present study is straightforward if it focuses on this group. Since the PS questionnaire, like most questionnaires, was designed with this modal case in mind, information for families with more complex custody arrangements is less precise.<sup>2</sup>

### The Validation Comparison

The analysis compared the court's records about child support with answers obtained in the telephone survey, which used the calendar year 1986 as the reference period. The CRD provided most of the award and payment characteristics used in the analysis as well as the criterion to use in estimating response errors. The CRD operationalizes child support as cash transfers from the nonresident parent to the resident parent for the children under an order from a court.<sup>3</sup> This operationalization focuses on legal agreements because the court can only record and enforce agreements that it knows about. This court-based operationalization exploits Wisconsin's requirement that court-ordered cash transfers for child support be made to the county Clerk of Courts. The court records all kinds of

---

<sup>2</sup>Since the large majority of resident parents are mothers (approximately 80% in Wisconsin), it is not possible to distinguish errors associated with the parent's gender from those associated with the parent's resident status. Because the description of the child support system and the variables in the analysis apply regardless of the parent's gender, I refer to resident and nonresident "parents" in those descriptions. In presenting the analysis, I refer to resident "mothers" and nonresident "fathers" in order to take the definition of the analytic sample into account.

<sup>3</sup>The data analyzed here define child support as income for the resident parent, as is done in the Current Population Survey (CPS), rather than as expenditures by the nonresident parent, as is done in some publications using the Survey of Income and Program Participation (e.g., U.S. Bureau of the Census, 1988). "Court-ordered" refers to provisions ratified by, that is, ordered by, the court. The CPS refers to agreements that the court imposes on parents as court ordered and other agreements as "voluntary" even when they are incorporated in a court-approved agreement (U.S. Bureau of the Census, 1990, p. 21).

support owed and the amount of monthly support payments owed and paid. Amounts owed and paid are standardized on a monthly basis in the CRD.

Although the analysis treated the CRD data as though they provided an accurate criterion, no criterion is free from error. Details about the usefulness of the CRD data as a criterion for evaluating self-reports in the PS are given in Schaeffer et al. (1991). The analysis presented here attempted to include measures to capture some of the principal possible discrepancies between the CRD and the PS (see below).

## Event Characteristics and Errors

The principal factors affecting response errors in reports about events are knowledge, elapsed time, respondent motivation, and characteristics of the thing asked about: its complexity, salience, and social desirability (Cannell, Miller, & Oksenberg, 1981; Sudman & Bradburn, 1974). The importance of knowledge is obvious, since respondents cannot report what they never knew. The time elapsed since the event asked about can contribute to omissions or telescoping errors (Neter & Waksberg, 1964; Sudman & Bradburn, 1974). Respondents who are not committed to giving accurate responses may not take the time required to correct these memory errors (Cannell et al., 1981). Reporting accuracy decreases as events become more complex, less salient, and more socially desirable or less socially undesirable (Belson, 1986; Bradburn & Sudman, 1979; Duncan & Mathiowetz, 1985; L. C. Hamilton, 1981; Wyner, 1980); these factors may be more important than elapsed time (Mathiowetz & Duncan, 1988). Respondent characteristics, such as age and gender, may also be associated with accuracy (Borus, 1966), but this may be because the characteristics of events vary by social group (Mathiowetz & Duncan, 1988). Although I also consider some of the other factors that previous research has found to influence reporting errors, I focus on the characteristics of the event being asked about.

An event's salience is often described as affecting how memorable it is (e.g., Mathiowetz & Duncan, 1988; Sudman & Bradburn, 1974), but salience is a complex construct. In various contexts, saying that an event is salient can mean that whether or not it occurred is easy to remember accurately, that the event appears distinct from similar events, that details about the event can be remembered easily, and so forth. I distinguish among several characteristics of events that may affect what might be considered salience, broadly defined: how complex the events are (their frequency, regularity, and similarity), how distinct they are from similar events (their clarity), and how much reinforcement they receive from the social context in which they occur. When events are complex, are indistinct from each other, and receive little social reinforcement, respondents are less likely to retrieve correct values to use in estimating values that they cannot recall directly.

The effects of these different characteristics of events may be difficult to distinguish in an analysis. Factors constituting salience may be conceptually

distinct but highly correlated in practice. This collinearity may make it impossible to estimate accurately the independent effects of the dimensions of salience just described. For example, how many *similar* child support payments are received is limited by how many payments are received at all. Thus, it may be difficult to determine whether response errors are affected more by the number of payments received or by their similarity.

Table 9.1 presents a summary of the measures that I used to examine errors in reports about child support and indicates how I expected them to affect errors in self-reports. The table notes whether the measure was used to predict support owed, paid, or both; the source of the measure (CRD or PS); and its expected effect on reporting errors.

### Complexity: Event Frequency, Regularity, and Similarity

The complexity of patterns of unemployment has been shown to influence errors in reports about employment (Duncan & Mathiowetz, 1985), but I have found no similar investigation for cash transfers. The analysis distinguished among those who received no payments, those who received them for all 12 months of the reference period, and those who received them for some months. Since the absence of an award or payment during the entire reporting period was a simple “event,” I expected it to be the easiest pattern to report accurately. In comparison, reporting about awards or payments that were present for all 12 months of the reporting period should have been more difficult, but easier than reporting about a mixed experience.

This analysis could not completely disentangle the frequency and regularity of events. If payments were never received, the absence of the event was regular; if payments were received all 12 months, the event was both regular and frequent. Although respondents who received fewer than 12 payments may still have received those payments regularly (e.g., a payment every 4 months), it was difficult to identify such patterns, since there are many ways in which infrequent payments can be made regularly. Nevertheless, when payments were made for some months, they were *relatively* less regular than when they were never made or were made for all 12 months.

Changes in the amount owed or paid should also have increased the difficulty of reporting accurately. Respondents who used a decomposition strategy to estimate the yearly total must have recalled more than one value for the monthly payments; respondents who used a simple estimation strategy (e.g., multiplying the value of a single payment by 12) would have calculated an inaccurate answer.<sup>4</sup> Reporting accurately about payments was also expected to be more

---

<sup>4</sup>The frequency of changes in the amount of support owed and paid is highly correlated with overall frequency: Amounts cannot change frequently unless they are owed or paid frequently. The problems resulting from this confounding are more acute for payments, since they are less stable than the amount owed. In this analysis, changes in the amounts of payments referred only to changes in amounts other than zero.

TABLE 9.1. Characteristics of the Event, Method, and Respondent

Independent Variables	Model	Source	Description	Effect on Errors	Mean SD
<b>Complexity: frequency, regularity, and similarity</b>					
Owed some months	Owed	CRD	1 = support owed 1-11 months in 1986	++	0.24 0.43
Owed all months	Owed	CRD	1 = support owed 12 months in 1986	+	0.66 0.47
Change in amount owed	Owed	CRD	1 = amount of support owed changed	+	0.15 0.36
Paid some months	Paid	CRD	1 = support paid 1-11 months in 1986	++	0.49 0.50
Paid all months	Paid	CRD	1 = support paid 12 months in 1986	+	0.33 0.47
Change in amount paid	Paid	CRD	1 = amount of support paid changed	+	0.72 0.45
Some partial payments	Paid	CRD	1 = 1-12 partial payments in 1986	+	0.51 0.50
Retroactive order	Both	CRD	1 = a retroactive order was in effect in 1986	+	0.13 0.34
<b>Salience: reinforcement of the agreement and method of payment</b>					
Private payment agreement	Owed	CRD	1 = court approved private payment in 1986	+	0.03 0.16
Informal agreement	Owed	PS	1 = respondent reports an informal agreement in 1986	+	0.03 0.16
Direct payments	Paid	PS	1 = respondent reports nonresident parent paid some support directly to the resident parent in 1986	+	0.09 0.29
<b>Clarity: other transfers</b>					
Some alimony owed	Owed	CRD	1 = alimony owed some months in 1986	+	0.08 0.26
Some alimony paid	Paid	CRD	1 = alimony paid some months in 1986	+	0.06 0.24
Family support paid	Both	CRD	1 = family support paid 1-12 months in 1986	+	0.04 0.19

(Table 9.1 continued)

TABLE 9.1. Continued

Independent Variables	Model	Source	Description	Effect on Errors	Mean SD
Health insurance owed some months	Both	CRD	1 = nonresident parent was to provide children's health insurance 1-11 months in 1986	+	0.18 0.38
Health insurance owed all months	Both	CRD	1 = nonresident parent was to provide children's health insurance 12 months in 1986	+	0.58 0.50
Other transfers owed some months	Both	CRD	1 = nonresident parent was to provide other transfers 1-11 months in 1986	+	0.13 0.33
Other transfers owed all months	Both	CRD	1 = nonresident parent was to provide other transfers 12 months in 1986	+	0.26 0.44
Extra payments made	Paid	PS	1 = respondent reports nonresident parent made extra payments in 1986	+	0.32 0.47
Some payments made to social services	Paid	CRD	1 = payments forwarded to social services agency 1-12 months in 1986	+	0.26 0.44
Memory decay					
Days from end of reference period to interview	Both	PS	Number of days between end of reference period and date of interview	+	284.96 51.41
Social desirability					
Withholding due to delinquency	Both	CRD	1 = withholding due to delinquency was in effect in life of the case	+	0.14 0.35
Respondent motivation					
Ever refused	Both	PS	1 = respondent refused before interviewed	+	0.08 0.26
Income not reported	Both	PS	1 = respondent did not report income	+	0.29 0.46
Both parents interviewed	Both	PS	1 = both parents were interviewed	-	0.66 0.48

(Table 9.1 continued)

TABLE 9.1. Continued

Independent Variables	Model	Source	Description	Effect on Errors	Mean SD
<b>Method factors</b>					
Resident status in PS differs from that in CRD	Both	PS-CRD	1 = resident status in PS differs from that in CRD	+	0.05 0.21
Total paid imputed from amount owed	Paid	PS	1 = the amount paid was imputed from the award because respondent said all payments were made	-	0.52 0.50
None owed imputed	Owed	PS	1 = the amount owed was imputed as 0	-	0.11 0.31
None paid imputed	Paid	PS	1 = the amount paid was imputed as 0	+	0.11 0.32
Female interviewer	Both	PS	1 = female interviewer	?	0.84 0.37
<b>Respondent characteristics</b>					
Age	Both	PS	Age in years	+	33.36 6.84
High school	Both	PS	1 = high school education	-	0.46 0.50
More than high school	Both	PS	1 = more than high school education	-	0.47 0.50
Income	Both	PS	Household income in 1986 excluding child support and alimony	?	11,970.54 8,788.30
Number of children	Both	CRD	Number of children in the original court order	+	1.89 0.92

*Note:* Data are from the Wisconsin Parent Survey 1 and the Court Record Database. Respondents were 964 divorced mothers who had physical custody in most of 1986 according to the court record. The base for individual statistics deviates from this total because of missing information.

difficult when the nonresident parent made partial payments, since when the amount of the payment differed from the amount owed, respondents had to remember more information.

In some cases, for example, when parents have been separated for several months before their divorce becomes final, the court orders retroactive support. When retroactive support is ordered, the amount owed for previous months is entered in the monthly CRD data in the month in which it was ordered. However, the CRD records all payments, including retroactive support, as made in the month in which the support is actually paid. It is unclear whether respondents with retroactive support orders reported about support owed or paid for a

month or support owed or paid *in* a month. Whichever accounting system respondents used, on average, the presence of retroactive support orders should have increased errors in PS reports about support owed and paid when compared with the CRD.

## Salience

### *Reinforcement of the Agreement*

Occasionally the court authorizes a private payment agreement, in which parents are allowed to exchange support directly. Since private payment agreements require less contact with the courts, their terms may receive less institutional reinforcement, and respondents may then report the amount of support owed less accurately. Because the parents are allowed to exchange support directly, amounts recorded in the CRD should be lower than those in the PS. The effect may not be large, however, since some parents who have private payment agreements continue to report to the court the payments that they have made, and these payments are recorded by the court.

When private behavior and institutional requirements reinforce each other, awards and payments should be more salient and easier to remember and report accurately. In the PS, some parents reported that they had an informal agreement about support with the other parent, which is the agreement that they reported about. In almost all of these cases, the court record shows that the parent has a court-ordered award. These informal modifications may interfere with memories about the award made by the court, however, and thus may increase errors in reporting about awards; reports about support paid should not be affected.

### *Method of Payment*

Although all payments are supposed to be made through the courts, parents sometimes exchange child support directly. The PS asked whether or not the nonresident parent had made such direct payments. If parents' reports in the PS are accurate, the amount of support recorded in the CRD should be lower than that in the PS when direct payments occur.

## Clarity: Other Transfers

Although the PS attempted to describe child support in a way that would obtain answers comparable with the CRD, parents may be involved in other exchanges that reduce the distinctness of cash child support transfers.

Alimony is support for a former spouse, but some respondents may consider that alimony benefits children, if only because alimony may free up other resources, which the resident parent can then spend on the children. Thus, when alimony is owed or paid, reports about child support awards and payments may be less accurate. In some cases, the court awards "family support" rather than

child support.<sup>5</sup> The description of child support in the PS was intended to include family support, and interviewers were instructed to treat family support like child support; but the difference in terminology may have led some respondents to exclude family support without the interviewer's knowledge.

Although additional transfers required of the nonresident parent are not paid through the courts, the court records when such other transfers (e.g., health insurance, mortgage or rent payments, or school tuition) are owed. When the court requires additional transfers, errors in reports about awards and payments should increase, since some respondents may include these additional obligations.<sup>6</sup> It is unclear, however, whether errors should be larger when additional payments are owed for the full 12 months of the reference period (e.g., health insurance) or for a shorter time (e.g., school tuition), partly because these obligations are heterogeneous. Additional obligations that a respondent includes for all 12 months could substantially increase the difference between the CRD and PS. When such other obligations are constant, however, they may be perceived as more distinct from cash child support obligations.

Whether or not the court recorded the presence of other obligations, the PS asked if the nonresident father made payments other than for regular child support, and the amount recorded in the CRD is expected to differ from that reported in the PS when he did.

When the resident mother receives AFDC, the court forwards support payments to the social services agency, which then sends up to the first \$50 to the mother. When the resident mother received AFDC, the interviewer was instructed to ask for the amount of support that she actually received, since many mothers know only the amount that the court forwards to them. Thus, when the court has forwarded payments to the social services agency, PS reports are expected to be lower than the CRD.

## Reporting Errors versus Problems in the Criterion

Unlike the other factors considered above, private payment agreements, informal agreements, direct exchanges of support, and obligations other than regular support also potentially affect the accuracy of the CRD as a criterion (see Schaeffer et al., 1991, for a fuller discussion). Finding that reporting errors increase when these circumstances are present may suggest that the court is not able to monitor important aspects of nonresident parents' support of their children or

---

<sup>5</sup>Family support and child support differ with respect to which parent has tax liability for the support. Awards for family support are relatively rare (see Table 9.1). The CRD data used in this analysis only distinguishes family and child support for payments, not for awards. Therefore, information about family support paid is used to predict errors both in child support awards and in payments.

<sup>6</sup>Since the court record does not include information about payments for these other obligations, this analysis uses the CRD information about whether the obligations exist to predict errors in reports about both the amount owed and the amount paid.

that the questions operationalizing the CRD conceptualization of support need improvement.

## Other Independent Variables

In addition to indicators of the structure and characteristics of experience with child support, I included several other factors that previous research suggests may affect errors in self-reports (see Table 9.1). The impact of memory decay was examined by including the number of days between the end of the reference period and the date of the interview. Child support is a legal obligation, and in the aggregate, nonresident fathers' reports about child support appear to be influenced by social-desirability factors (Schaeffer et al., 1991). Resident mothers' reports may be indirectly subject to similar pressures: Mothers who wish to protect a father who has not paid support may underreport the amount owed or overreport the amount paid. Thus, whether or not support had been withheld from the nonresident parent's pay because of delinquent payments was included in the analysis. Respondents who initially refused to be interviewed or who were unwilling to report income may have been less motivated to report child support accurately. When both parents were interviewed, it was often because one parent was able to supply a telephone number for the other; such cooperative respondents may be more accurate reporters.

Several characteristics of the instrument may also affect the accuracy of self-reports compared with the CRD. The PS used the respondent's report about resident status to determine whether to ask about support received or paid. If a parent changed resident status without reporting the change to the court (or reported his or her resident status incorrectly in the PS), reports in the PS may differ from the information recorded by the court.<sup>7</sup> The structure of the PS questions assumed that the amount paid could be imputed from the amount owed if all payments were made. Since this structure ignores the possibility of partial payments, however, the procedure is expected to increase "overreporting" of support paid in the PS compared with the CRD. Some instrument design strategies that lead to imputed values, however, may increase accuracy. For example, if the respondent reports that there is no agreement for regular child support, it may be justifiable to impute zero as the amount owed and paid. Imputing that the value

---

<sup>7</sup>Whether respondents were asked about support that they were to pay or be paid depended on whether they reported in the PS that they were the nonresident or resident parent. If a respondent reported in the PS that she was the resident parent, she was asked about support that she received, not support that she paid out. If the CRD records that this respondent is the nonresident parent, there is no report of support paid out to match with the CRD and the respondent is treated as reporting no support paid. The analysis retained the small number of affected respondents (see Table 9.1) to estimate the effect of the CRD's and PS's different definitions of parents' resident status, but these "response" errors could be considered artifactual.

of awards or payments is zero based on such filter questions is expected to increase the accuracy of the PS when it is compared with the CRD.

The gender of the interviewer may affect accuracy if respondents ascribe different expectations about child support to male and female interviewers and consider these expectations when formulating an answer. Respondents' beliefs about these expectations could then lead to inaccuracies in reports about child support.<sup>8</sup>

Research cited above suggests that older and less educated respondents may be less accurate. In addition, the salience of child support owed or paid may depend on the respondent's income. Finally, as the number of children to be supported increases, the calculation of the amount owed may become more complex. This is because the amount of support that the court awards considers that there are "economies of scale": The amount of support ordered for the "first" child is typically larger than the amount of support ordered for additional children. Reports about awards are expected to be less accurate when the court order refers to more children; reports about support payments should not be affected.

## Dependent Variable

The analysis focused on the absolute value of errors, computed as the PS report minus the CRD value. Table 9.2 summarizes the distributions of the absolute errors and briefly describes the direction of errors.<sup>9</sup> Surprisingly, given the complexity of the construct, the length of the reference period, and the timing of the field period, the absolute errors were bunched at zero. Since the distributions were positively skewed, and since the process that governed reporting errors may have been proportional rather than absolute (e.g., an error of \$100 is larger when estimating a payment of \$200 than when estimating a payment of \$1,000), logs of the absolute errors were analyzed. Approximately one-third of the sample reported a value within \$100 of the CRD for both support owed and support paid. For support owed, overreports were more common than underreports, but the likelihood of positive and negative errors was approximately equal for reports of support paid. Overall, reporting about support paid appeared

---

<sup>8</sup> Respondents were not assigned randomly to interviewers, but the distribution of the workload was haphazard. Since the number of interviewers was smaller than the number of respondents, the significance tests reported in the tables are inflated, because the degrees of freedom for the tests in the tables are based on the number of respondents rather than on the number of interviewers.

<sup>9</sup> The CRD value is the sum of all child support and family support payments that the court forwarded to the resident parent plus any support payments forwarded to the social services agency when the resident parent was receiving AFDC. Since payments forwarded to the social services agency are retained, how they affect responses can be estimated.

TABLE 9.2. Descriptive Statistics for Errors in Reporting Dollars of Support Owed and Paid

Statistic	Support Owed	Support Paid
Absolute errors		
Mean	1,128.53	1,278.20
Standard deviation	2,206.34	2,812.84
Skewness	5.32	8.72
Logs—Mean	4.68	4.98
Logs—Standard deviation	3.08	3.00
Logs—Skewness	-0.46	-0.68
Direction of errors		
Percent less than -100	22	32
Percent -100 to 100	39	33
Percent greater than 100	39	35
<i>N</i>	(889)	(842)

*Note:* Data are from the Wisconsin Parent Survey 1. Respondents were 964 divorced mothers who had physical custody in most of 1986 according to the court record. The number of cases deviates from this total because of missing information. Because of 0s, 1 was added before taking natural logarithms.

to be slightly worse than reporting about support owed, whether one considers the average absolute error or the proportion reporting correctly (see also Schaefer et al., 1991).

## Results

Table 9.3 presents the results of tobit regressions predicting log absolute errors for support owed and paid. The tobit model was used because there were a large number of observations with the value 0; that is, the distribution of the dependent variable was truncated at 0. The tobit regression coefficients express the linear effects of the independent variables in the untruncated population (McDonald & Moffit, 1980). Since the dependent variable is the log of the absolute errors, the antilog of the coefficient expresses the ratio of the predicted value after the independent variable's effect has been applied to the value before the variable's effect (with both terms in the original metric). For example, the effect of owing support for some months compared with no months during the reference period was to increase the size of the absolute error, in the original metric, approximately 74 times.

The results suggest that the complexity of the respondent's support experience had a substantial impact on her accuracy. The errors made by those who were owed or paid support for only some months in 1986 were on the order of 74 and were 15 times larger than the errors of those with no support. In contrast,

TABLE 9.3. Tobit Regression of Log Absolute Errors on Characteristics of the Event, Method, and Respondent

Independent Variables	Support Owed			Support Paid		
	<i>b</i>	<i>t</i>	<i>e<sup>b</sup></i>	<i>b</i>	<i>t</i>	<i>e<sup>b</sup></i>
<b>Complexity</b>						
Owed/paid some months	4.30*	8.20	73.70	2.70*	5.86*	14.88
Owed/paid all months	1.61*	3.14	5.00	0.84	1.66	2.32
Partial payments	—	—	—	0.46	1.62	1.58
Change in amount owed/paid	1.68*	5.39	5.37	1.26*	3.39	3.53
Retroactive order	0.83*	2.48	2.29	0.76*	2.29	2.14
<b>Saliency</b>						
Private payment agreement	-1.07	-1.58	0.34	—	—	—
Informal agreement (PS)	-0.35	-0.52	0.70	—	—	—
Direct payments (PS)	—	—	—	3.09*	8.25	21.98
<b>Clarity</b>						
Some alimony owed/paid	0.71	1.69	2.03	1.80*	4.05	6.05
Family support paid	0.80	1.45	2.23	0.70	1.25	2.01
Health insurance owed some months	1.18*	3.26	3.25	0.76*	2.14	2.14
Health insurance owed all months	0.36	1.27	1.43	0.33	1.16	1.39
Other transfers owed some months	0.92*	2.80	2.51	0.19	0.55	1.21
Other transfers owed all months	0.40	1.54	1.49	0.73*	2.85	2.08
Extra payments made (PS)	—	—	—	0.15	0.65	1.16
Some payments made to social services	—	—	—	0.84*	2.69	2.32
<b>Memory decay</b>						
Days from end of reference period to interview (10s)	0.01	0.59	1.01	0.03	1.28	1.03
<b>Social desirability</b>						
Withholding due to delinquency	-0.13	-0.40	0.88	-0.23	-0.70	0.79
<b>Respondent motivation</b>						
Ever refused	-0.41	-1.00	0.66	0.22	0.54	1.25
Income not reported	0.14	0.60	1.15	0.04	0.14	1.04
Both parents interviewed	-0.06	-0.28	0.94	0.24	1.04	1.27
<b>Method factors</b>						
Resident status in PS differs from that in CRD	2.68*	4.03	14.59	2.94*	4.44	18.92
Total owed imputed as paid	—	—	—	0.42	1.54	1.52
None owed/paid imputed	-1.24*	-2.45	0.29	-2.08*	-4.01	0.12
Female interviewer	0.58	2.01	1.79	-0.16	-0.56	0.85

(Table 9.3 continued)

TABLE 9.3. Continued

Independent Variables	Support Owed			Support Paid		
	<i>b</i>	<i>t</i>	<i>e<sup>b</sup></i>	<i>b</i>	<i>t</i>	<i>e<sup>b</sup></i>
Respondent characteristics						
Age	0.01	0.55	1.01	0.04*	2.01	1.04
High school	0.38	0.88	1.46	-0.03	-0.07	0.97
More than high school	0.16	0.38	1.17	-0.25	-0.54	0.78
Income (\$1,000s)	-0.00	-0.34	0.99	-0.00	-0.23	1.00
Number of children	0.27*	20.20	1.31	0.05	0.36	1.05
Intercept	-0.65	-0.58		-1.54	-1.41	
Log likelihood		-1,973.0			-1,796.2	
<i>N</i>		(879)			(831)	

*Note:* Data are from the Wisconsin Parent Survey 1 and the Court Record Database. Respondents are divorced mothers who had physical custody during most of 1986 according to the court record. The dependent variable is the natural log of the absolute value of errors in support owed and paid. Because of 0s, 1 was added before taking logs. Dashes indicate that a variable was substantively relevant for one of the dependent variables but was omitted from the equation predicting the other. Coefficients (other than that for the interviewer's gender) for which *t* values are 2 or greater are indicated by "\*".

those who were paid support for all 12 months during 1986 did not make significantly larger reporting errors than did those who were paid none;<sup>10</sup> the error for those who were owed support all 12 months was approximately 5 times larger than it was for those who were owed no support. Errors also increased substantially when the amount of support owed or paid was variable: Errors were 5.4 times larger for support owed and 3.5 times larger for support paid. Partial payments do not appear to have had any additional effect on errors in reports about support paid once the similarity of payments was taken into account. As expected, the presence of a retroactive order also increased errors in both reports about child support: Errors approximately doubled when there was a retroactive order.

Results were more mixed for the indicators of the extent to which the amount of the court's award was reinforced in the respondent's experience. Respondents who reported that they had an informal agreement were not less accurate than others in reporting the amount of child support that they owed, supporting speculation that these "informal agreements" involve minor adjustments to the court's order. Resident mothers with private payment agreements also did not differ from other mothers in the accuracy of their reports. However, errors in reports of

<sup>10</sup>In this discussion, I treat coefficients for which the *t* value is 2 or greater as statistically significantly different from 0, except for the test of the effect of interviewer's gender. This test must be adjusted to consider that there were fewer interviewers than respondents.

support paid by resident mothers who said that they exchanged support directly were 22 times larger than were the errors of mothers without such direct exchanges.

Some transfers that respondents might be involved with that could be confused with child support increased reporting errors. If alimony was paid, errors in reporting about the amount of child support paid increased 6 times. Obligations other than regular cash child support were sometimes associated with increased errors. Errors in reporting the amount owed were larger if obligations for health insurance or other transfers were present for some months than they were if no other obligations existed. Errors in reporting the amount of support paid increased significantly if health insurance was owed for some months of 1986, if other transfers were owed for all months during 1986, or if the court forwarded part of the support paid to the social services agency because the mother was on AFDC.

This set of characteristics of respondents' experience with awards and payments taken as a whole offers six comparisons between how accurate respondents are when the structure of an event is relatively simple in their experience (something is never owed or paid or is owed or paid for all months) and when it is relatively complex (something was owed or paid for some months). In five of the six comparisons, errors appear to be larger when an obligation was owed or paid some months than when it is owed or paid all months, and in three of these comparisons, the coefficients differ significantly from each other. The most important such comparison is between the effect of being owed or paid support for some months and for all months. In both cases, the increase in reporting error was significantly greater for the complex than for the simple event structure (for support owed, the difference was 2.69,  $t = 8.92$ ; for support paid, the difference was 1.86,  $t = 7.03$ ).

The length of time between the end of the reference period and the interview does not appear to be associated with an increase in reporting error, probably because effects due purely to the passage of time occurred before interviewing began. Insofar as I have been able to measure the effects of respondent motivation and of social desirability for mothers, these factors do not appear to have affected reporting errors for resident mothers.

The method factors are more problematic. When a mother's resident status in the CRD differed from that in the PS, her reports in the PS necessarily differed from the CRD values. When I imputed that awards or payments were zero based on the skip patterns that implemented the CRD's operationalization of support, reporting errors decreased significantly. In contrast, using the respondent's report that "all" payments were made to impute the amount of the award as paid appeared to increase errors, although not significantly.

Respondents' characteristics appear to have had little effect on reporting errors. The exceptions are that increasing age was associated with increasing errors in reports of support paid and that as the number of children involved in the court order increased, errors in reporting the amount of the award increased, as expected.

## Improving Question Design

Many analytic models assume that response errors are similar for all respondents. The analysis reported here suggests that such an assumption may not be warranted, since respondents for whom the pattern of support payments owed and paid differed also differed in the size of their reporting errors. This pattern establishes conditions for an association, although not necessarily a linear association, between response errors and the amounts of child support “actually” owed or paid. The results in Table 9.4 suggest that those conditions may be present for reports about child support. The table compares the average amounts of support owed (or paid) with the average error for respondents owed (or paid) support no months, some months, or all months in 1986. Since positive and negative errors may cancel each other, the table presents errors rather than absolute errors. The table shows that the mean error is relatively small and negative for the group with the highest average support but is generally larger and positive for other parents. Investigators designing questions must, then, consider not only how to prevent self-reports from systematically underestimating or overestimating population means and variances but also how to make errors comparable for different groups of respondents, thus reducing relationships between errors and actual values that could bias results of multivariate analyses.

The analysis suggests that characteristics of the events themselves and method factors are, on the whole, more important predictors of the size of reporting errors than are the other classes of factors examined. The analysis reported here suggests three foci for experiments in improving the accuracy of self-reports about child support and similar variables: increasing the clarity of the events asked about, identifying respondents for whom the pattern of events differs, and examining the consequences of filter questions used to impute values.

Increasing the clarity of the events asked about requires careful work during instrument development to identify the kinds of events that respondents may confuse. Describing the categories that respondents use to label events and how those events are organized in respondents’ classification systems may enable question designers to address potential sources of confusion. For example, respondents may label either any transfers to the nonresident parent or cash transfers specifically for the children as “child support,” depending on the context. Respondents may confuse alimony with child support when they are asked about child support before being asked about alimony, but they may distinguish the two correctly when they are informed that the distinction is important and are asked to report about alimony first.

Identifying important patterns of events requires attention to how events of a given kind are socially and individually structured. In the case of estimating annual totals of child support payments, a filter question asking how many months during the preceding year the respondent received a child support payment could be used to divide the sample into those receiving no payments, payments during some months, and payments during all months. Respondents who

TABLE 9.4. Means of Errors and True Values, by Complexity (Frequency or Regularity) of Dollars of Support Owed or Paid

Variable and Statistic	Months of Support Owed or Paid (CRD)		
	None	Some	All
<b>Support owed</b>			
Mean of errors (PS - CRD)	1,036	1,154	-187
Standard deviation	1,763	2,556	2,393
Mean support owed (CRD)	0	2,457	4,260
Standard deviation	0	2,396	3,892
<i>N</i>	(69)	(217)	(603)
<b>Support paid</b>			
Mean of errors (PS - CRD)	1,413	122	-506
Standard deviation	2,628	2,569	3,706
Mean support paid (CRD)	0	2,488	5,375
Standard deviation	0	2,906	5,217
<i>N</i>	(162)	(393)	(287)

*Note:* Data are from the Wisconsin Parent Survey 1 and the Court Record Database. Respondents were divorced mothers who had physical custody during most of 1986. Cell entries were calculated in the original metric and statistics are rounded to the nearest dollar.

received payments for only some months could be asked a more intensive series of questions designed to help them recall the number of payments actually received (e.g., Means, Nigram, Zarrow, Loftus, & Donaldson, 1989). In addition, asking respondents in the latter two groups about whether all payments were in the same amount before proceeding to calculate a yearly total could also improve the accuracy of their reports.

Evaluations of such experiments must pay close attention to how filter questions increase or decrease both random and systematic errors. The analysis presented here provides examples of a filter that decreases "reporting" errors on average (imputing that the amount of support paid is zero based on the report that there is no agreement for support) and of a filter that increases systematic "overreports" (by imputing the total owed as paid when the respondent reports that all payments were made). In general, the effects of using filters on the error properties of the resulting composites that are used in analysis are not well understood (see, for example, Groves, 1989, pp. 464-468, and Krosnick & Berent, 1991).

## Conclusion

Although this analysis used the case of child support, similar considerations clearly apply to many common survey questions that implicitly ask for summaries of events. The level of detail that was necessary to explicate the case exam-

ined here suggests that when accurate reports are important, the patterning of events, the expected frequency of different patterns in the population, and events that may be confused with the event of interest must be carefully analyzed. Although the indicators of respondent motivation and of the extent to which child support events received additional social reinforcement were not important in predicting absolute errors, the operationalizations were somewhat limited, and these negative findings should be generalized only cautiously.

*Acknowledgments.* This research was supported in part by grants from the National Institute for Child Health and Human Development (HD-24571) and from the DHHS Office of the Assistant Secretary for Policy and Evaluation (ASPE 91ASPE236A) to Judith A. Seltzer and Nora Cate Schaeffer, and by the Graduate School Research Committee, University of Wisconsin, Madison. Computing was provided by the Center for Demography and Ecology, which receives core support from the Center for Population Research of the National Institute for Child Health and Human Development (HD-5876). Opinions expressed here are those of the author, not of the funding agencies. Helpful comments were provided by conference participants and by Judith A. Seltzer. I-Fen Lin provided exemplary research assistance, and Kay Tuschen made the bibliography program work.

# 10

## Judgments of Behavioral Frequencies: Memory Search and Retrieval Strategies

Geeta Menon

### Background

Surveys in marketing often employ questions that seek to determine the frequency with which respondents engage in different kinds of behavior. These behaviors range from very frequent ones (such as the number of times a day that one consumes coffee) to somewhat frequent ones (such as the number of times that one went shopping in the last month) to infrequent ones (such as the number of times that one has purchased a car in the last 5 years). A study by Blair and Burton (1987) indicated that the cognitive processes that respondents use vary depending on the relative frequency of the event. In other words, although it is easy to recall and count every instance for an infrequent behavior, it becomes more difficult to do so for a frequent behavior. Many researchers now maintain that in a survey situation in which respondents are asked a question relating to the frequency of a fairly frequent, nonsalient behavior, they do not do a straightforward recall and count of every occurrence of the target behavior. Instead, they provide an *estimate* based on various inference strategies (Blair & Burton, 1987; L. Ross, 1984; Schwarz, 1990a; Strube, 1987).

The use of estimation procedures by survey respondents is consistent with current theory that people are cognitive misers who use heuristics to arrive at decisions and judgments efficiently (Newell & Simon, 1972; Taylor, 1981; Wyer & Srull, 1986, 1989). To the extent that behavioral frequencies are reported based on inferential heuristics, they are *judgments* and are subjective. This led Schwarz (1990a) to conclude that the traditional distinction between “opinion questions” (presumably answered on the basis of somewhat unreliable judgmental processes) and “factual questions” (presumably answered on the basis of more reliable recall from memory) is misleading.

In this chapter, I posit that for different behaviors with the same high frequency, the strategies that respondents use to report frequencies could vary depending on how the requisite information is stored in memory and how it is

retrieved. To this end, I now examine the cognitive processing of autobiographical information about events in terms of their organization and structure in memory, and the consequent retrieval strategies used.

## Literature Review of Autobiographical Memory

Some definitions of autobiographical memory have included all events, whether experienced directly or through the mass media (e.g., Wagenaar, 1986; Whitten & Leonard, 1981), whereas others have included only personal experiences, excluding reported events or news stories (e.g., Larsen, 1985, 1987). For the purposes of this research, I have taken all personally experienced events to constitute autobiographical memory, with a focus on frequently experienced ones.

Autobiographical information about events can be stored in the form of individual episodes (i.e., in an *episodic* store) and/or in the form of general world knowledge about the event (i.e., in a *semantic* store). Tulving (1983) maintains that in recalling a personal episode (i.e., an event from autobiographical memory), the system under operation is the episodic memory. However, in recalling general knowledge about a certain class of episodes, the operating system is the semantic memory. He views episodic and semantic stores as being two discrete stores. For an autobiographical behavior, it is possible to have multiple stores accessible at the same time (see Brewer, chap. 1, this volume, for a classification of different kinds of autobiographical memory).

The three major "empirical" works in the area of autobiographical memory are based on studies that the researchers conducted using themselves as subjects (M. Linton, 1975, 1978, 1982; Wagenaar, 1986; R. T. White, 1982). Each of these researchers maintained a diary for different time periods (ranging from 6 months to 7 years), with the intention of determining the accessibility of episodic information over time. The common finding was that although all events were encoded episodically, as the number of experiences in any event class increased, semantic knowledge about the event increased and its context became increasingly confusable. Interference by similar events was identified as the factor that contributed most to the inability to recall specific occurrences of an event.

There is growing consensus that frequent experience of similar events fosters semantic memory, which lacks specific time or location indicators. Such decontextualization turns episodic information into semantic knowledge (Bahrick & Karis, 1982; Nelson, 1983; Strube, 1987; Tomkins, 1980). Generalized knowledge has its roots in episodic experiences. Over time, the identity of these roots is lost, together with the context in which such information was acquired (Bahrick & Karis, 1982).

This distinction between the episodic and semantic memory systems is an important one. Depending on the relative accessibility of the episodic or the semantic store, the strategy adopted by a respondent in arriving at a frequency

judgment should vary. For example, respondents should use the recall-and-count strategy if only the episodic store is accessible. As the episodic store becomes less accessible, respondents should resort to estimation strategies based on different heuristics available in the semantic store(s). If both stores are accessible, the respondent may use a combination of recall-and-count and estimation strategies.

In the next section, I use the literature reviewed thus far to develop a model of autobiographical memory. This model helps explain how information pertaining to different kinds of behavior may be stored and, consequently, the retrieval strategies that respondents will use in a survey situation to respond to a behavioral frequency question.

## Proposed Autobiographical Memory Model

Two dimensions, the regularity and the similarity of the behavior, are hypothesized to affect the storage and retrieval of the relevant information in the context of responding to a behavioral frequency question.

### Storage of Behavioral Information in Autobiographical Memory

#### *Impact of Regularity of the Behavior*

It has been suggested that for some behaviors a rate of occurrence may be stored in memory (Blair & Burton, 1987). Intuitively, because of the inherent nature of a regular behavior (R) and its consequent fixed periodicity of occurrence, it is likely that respondents store a general rate of occurrence of the behavior. In estimating the frequency of washing hair, therefore, the accessible information is more likely to be the general rate of occurrence of the behavior, which is applied to the reference time frame. This general rate-based estimation could also be used by respondents in modified forms such as some adjustment being made to the figure after it is arrived at, estimating and then deleting noninstances of the otherwise regular behavior, and so on. It is important, however, to understand the exact cognitive mechanisms that respondents use in arriving at the final figure.

On the other hand, in the case of irregular behavior (IRR), since occurrences are not periodic, information about a single, general rate of occurrence is less likely to be stored and used. Instead, respondents are more likely to rely on episodic recall or estimation procedures such as counting for a short period and extrapolating for the time frame of interest if some episodes are more accessible because of their salience and/or vividness.

*Proposition 1.* The likelihood of storing a general rate of occurrence depends on the regularity of the behavior. Regular behaviors facilitate the storage of a single rate of occurrence, whereas irregular behaviors do not. Therefore, in reporting a frequency judgment, respondents would be more likely to use estimation strategies for regular behaviors than for irregular behaviors.

### *Impact of Similarity of the Behavior*

The theory of the transition from episodic to semantic memory appears to be the most plausible explanation for the inability of respondents in a survey situation to recall and count every episode of a frequent behavior. This theory can also be extended to accommodate differential storage of dissimilar (DIS) and similar (S) behaviors. In general, if different occurrences of a behavior are almost identical (e.g., washing your hair), the behavior can be classified as a similar one. On the other hand, if the behavior is more heterogeneous (e.g., stopping to talk to friends), the behavior can be classified as a dissimilar one. It is more likely that respondents will tend to have a semantic store for similar behavior (R. T. White, 1982; Wagenaar, 1986) because of the homogeneity of the different occurrences and consequently lower accessibility to individual episodes. On the other hand, because dissimilar behaviors are highly heterogeneous, it is less likely that the respondents will have a single semantic store for these behaviors and more likely that the information will be maintained in episodic stores.

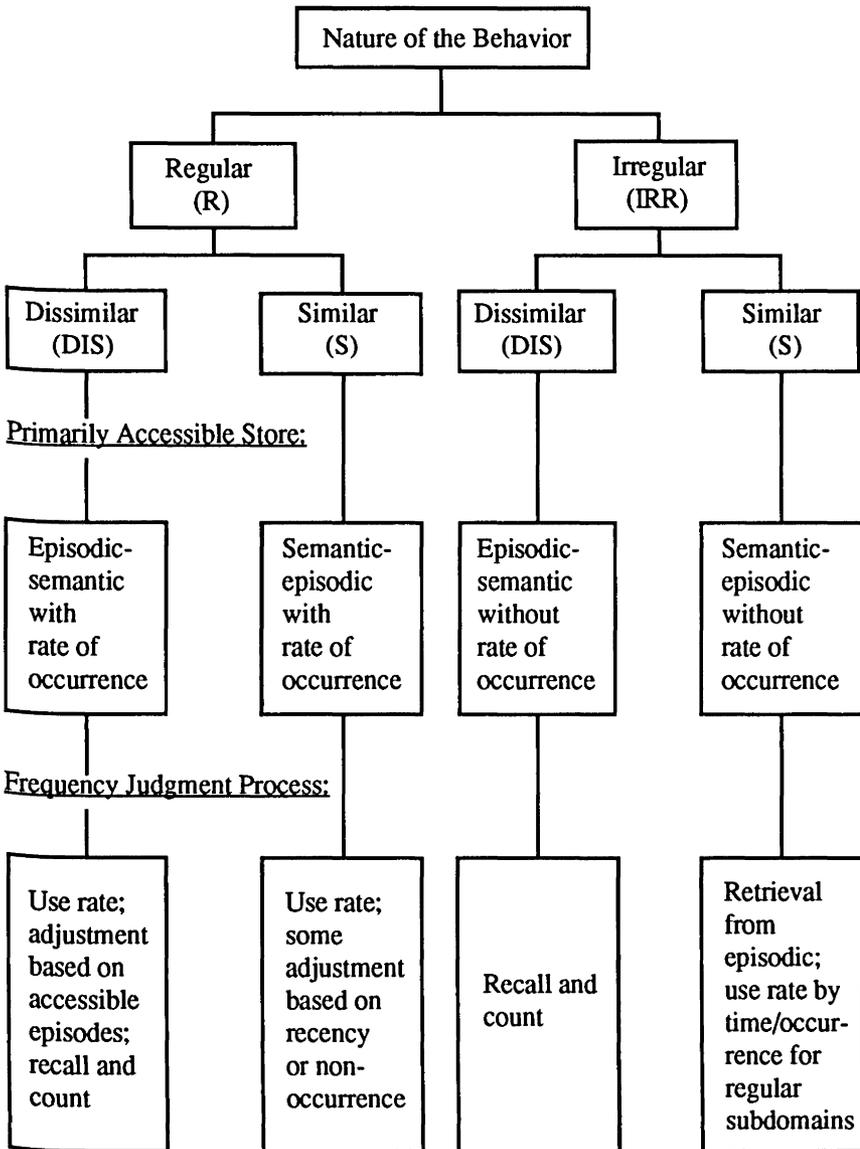
*Proposition 2.* The type of store in which information about a behavior is maintained depends on the similarity of the behavior. The greater the dissimilarity of a behavior, the more salient will be the individual occurrences of the behavior and the more likely that such information is accessed through episodic stores; if the behavior is similar, the more likely that individual occurrences cannot be discriminated easily and that the semantic store is more accessible. Therefore, in reporting a frequency judgment, respondents will be more likely to use counting strategies for dissimilar behaviors than for similar behaviors.

### Retrieval of Information and Reporting a Behavioral Frequency

The above discussion indicates that while the (ir)regularity of the behavior determines the (in)accessibility of a ready rate of occurrence as a basis on which to make a behavioral frequency estimate, the similarity of the behavior determines the location of the more accessible information as the semantic or episodic-semantic stores. Figure 10.1 illustrates the store that the proposed autobiographical memory model predicts would be tapped in order to arrive at a frequency estimate of the target behavior.

The figure illustrates that if the behavior is a regular-similar (R-S) one, the procedure that the respondent will use to report a behavioral frequency is straightforward rate-based estimation for the time period of interest to the researcher. In addition, a few minor adjustments *may* be made on the basis of a recent occurrence or nonoccurrence of the event. The primary accessible store, however, is the semantic store.

If the behavior is a regular-dissimilar (R-DIS) one, the accessible store is likely to be episodic in nature (because of the dissimilarity of the behavior) but with a rate of occurrence (because of the regularity of the behavior) that the respondent can easily resort to in arriving at a frequency. However, because the dissimilarity of the behavior may cause some episodes to be more salient than



*Note:* “Semantic-episodic implies a more accessible semantic store with few accessible episodes, whereas “episodic-semantic” implies more accessible episodes with/without accessible semantic stores within subdomains.

FIGURE 10.1. Proposed Model for the Storage and Retrieval of Information about Frequent Behavior.

others, there may be a tendency for respondents to use these available episodes to make adjustments to the frequency estimate. In addition, the dissimilarity of the behavior may prompt some respondents to use a recall-and-count strategy.

For an irregular-similar (IRR-S) behavior, the more accessible store is still going to be a semantic store because of the homogeneity of the behavior. However, there will be no readily accessible rate of occurrence stored in this semantic store because of the irregularity of the behavior, and the respondent will need to use more cognitive effort in order to estimate the behavioral frequency using some strategy other than a general rate-based one. For example, the respondent may estimate by decomposing the behavior into time and/or situational subdomains and using rates within each subdomain when this is possible (i.e., when there is regularity within subdomains); or the respondent may have to search memory much more to arrive at "available" (à la Tversky and Kahneman, 1973) episodes, which are less accessible for this kind of behavior, and use this as a basis on which to compute a frequency judgment. This process is going to be extremely arduous for the respondent, however, since the accessible store is a semantic one.

Lastly, for irregular-dissimilar (IRR-DIS) behavior, information is likely to be maintained in an episodic format with no general rate of occurrence. The heterogeneity of the behavior is likely to be high, and no general rate of occurrence will be available to which the respondents can resort. They will have to compute behavioral frequency judgments on the spot based on the episodes that are accessible to them, using a counting strategy.

The implications of the preceding discussion are that the cognitive effort required by the respondents for behavioral frequency questions depends on the nature of the behavior. Whereas R-S behavior requires the least cognitive effort because of the highly accessible rate of occurrence, R-DIS behavior requires a little more effort, given that some episodes are likely to be still accessible. IRR-DIS behavior requires still more effort, given that there is no ready single rate of occurrence that the respondent has available to compute a frequency judgment. The respondents will have to rely primarily on accessible episodes to compute a frequency judgment directly. Finally, the most complex processes are predicted to be for IRR-S behaviors, which have very low accessibility of episodes combined with an absence of the general rate of occurrence. The cognitive effort is likely to be the highest in this last case.

*Proposition 3.* The cognitive effort required by respondents to arrive at behavioral frequency judgments will have the following significant interaction between regularity and similarity:

$$R-S < R-DIS < IRR-DIS < IRR-S.$$

## Research Design

Two experiments were conducted in which subjects were asked for responses to frequency questions about 12 different behaviors (3 for each regularity-

similarity combination). The behaviors were selected after extensive pretesting among subjects from the target population of this research. While trying to ensure that the behaviors fell into each of the regularity × similarity cells, I also tried to ensure that the behaviors chosen were popular ones, engaged in by a large majority of the target population. In addition, I wanted to match behaviors on regularity while varying them on similarity and vice versa. The behaviors finally used in the experiments are presented in Table 10.1.

In the main experiments, subjects were presented each of the 12 behaviors and were asked how many times they had engaged in it during the last week. Responses were elicited in an open-ended format. The process used was tracked through concurrent and/or retrospective verbal protocols in Experiment 1, and the time taken was monitored through a response-time measure in Experiment 2. The order of presentation of the behaviors was perfectly counterbalanced in all experiments using Wagenaar’s (1969) diagram-balanced Latin square design. At the end of the interview, each respondent was asked to rate the behaviors on regularity and similarity dimensions. These were used as manipulations check variables. Both experiments were conducted among a sample of students from the subject pool of the Department of Business Administration of the University of Illinois at Urbana-Champaign.

### Independent Variables

Two independent variables were measured within subjects in both experiments:

1. *Regularity*: The two levels chosen were regular behavior (occurred with a fixed periodicity) and irregular behavior (did not have a fixed periodicity of occurrence).
2. *Similarity*: The two levels chosen were dissimilar behavior (one occurrence differed from another in terms of where, with whom, and what happened) and similar behavior (one occurrence was similar to the next in terms of where, with whom, and what happened).

TABLE 10.1. Behaviors Used in the Experiments

Regular-Similar (R-S)	Regular-Dissimilar (R-DIS)
Brushing your teeth	Going out in the evenings
Washing your hair	Having dinner
Attending class	Socializing with friends
Irregular-Similar (IRR-S)	Irregular-Dissimilar (IRR-DIS)
Leaving a message on someone’s answering machine	Eating at a fast-food place
Drinking water from a public fountain	Making unplanned stops to talk to friends during the day (more than a “Hi!”)
Buying soft drinks from a vending machine	Snacking

## Dependent Variables

The main dependent measure used in Experiment 1 was concurrent verbal protocols obtained while the respondent was arriving at frequency judgments for the target behaviors. This measure provided information on the strategies (counting vs. estimation) that subjects used in formulating frequency judgments. If the process used in computing frequency judgments was not clear through the protocols, or if subjects forgot to give concurrent protocols, they were asked to reflect back on the procedure that they had used and protocols were obtained in a retrospective fashion. These protocols were tape recorded and later coded. The coding scheme used was previously developed, tested, and used in a project on proxy reporting that was being conducted by Seymour Sudman and others at the Survey Research Laboratory of the University of Illinois.

In Experiment 2, the major dependent measure was the response time taken by subjects in reporting frequencies for the behaviors being investigated. The response time should indicate the amount of cognitive effort that is required to form a frequency judgment and thus is an indicator of the validity of the proposed model. Response times were elicited in milliseconds. Subjects were brought into a computer laboratory in groups of five or six and administered the questionnaire on computers programmed to record the time taken to answer the 12 different behavioral frequency questions. They were first allowed to go through a few practice questions before starting on the main questionnaire.

## Data Analyses

Analyses were conducted using individual-level data—that is, by picking out the best behaviors for each of the four cells at the individual level—and then aggregating across all subjects. The determination of the best behaviors at the individual level was based on the following considerations: (a) the behavior chosen had to be one that subjects engaged in at least once in the last week; (b) an attempt was made to choose behaviors such that those varying on similarity were matched on regularity and vice versa at the individual level; and (c) having fulfilled the above two conditions, if there still was a choice, the behavior with more extreme “regularity” and/or “similarity” ratings was chosen.

The advantage of looking at individual-level data in this manner is that one can aggregate across behaviors and thereby ignore any kind of effect that the specific behavior examined may have on the processing strategy and accuracy of frequency reports.

I shall now briefly examine the results of the two experiments (for a more detailed discussion of the data, see Menon, 1991).

## Manipulations Check

Ratings of the behaviors obtained from subjects on the “regularity” and “similarity” dimensions on 7-point semantic differential scales (1 = irregular/dissimilar

and 7 = regular/similar) were used for checking the manipulations. The manipulations worked identically in both experiments. The regularity and the similarity manipulations were effective, since the associated main effects were significant: the main effects of regularity on the “regularity” ratings were, for Experiment 1,  $F(1,28) = 428.39, p < .001$ , and for Experiment 2,  $F(1,27) = 401.54, p < .001$ ; the main effects of similarity on the “similarity” ratings were, for Experiment 1,  $F(1,28) = 279.62, p < .001$ , and for Experiment 2,  $F(1,27) = 230.20$ .

## Results

Table 10.2 presents the summary of the data for both experiments.

### Experiment 1

Since regular behaviors occur with a fixed periodicity, it was hypothesized that subjects would use an estimation strategy based on some rate of occurrence when reporting a frequency judgment for a regular behavior. On the other hand, since such a rate is not accessible to subjects in the case of an irregular behavior, they would need to resort to some other technique (e.g., a counting strategy) in order to arrive at a frequency judgment (Proposition 1).

In addition, since episodic stores are more accessible for dissimilar behaviors than for similar behaviors, a greater reliance on episodic recall was predicted for the former than for the latter case (Proposition 2).

As predicted (see top half of Table 10.2), the majority of the subjects used some kind of estimation strategy in arriving at a frequency report for regular behaviors (76%). On the other hand, the percentage of subjects using estimation strategies was significantly lower for irregular behaviors (34%). Subjects seemed to rely much more heavily on counting strategies in the latter case. Also, reliance on episodic recall, as is indicated by the use of counting strategies, was higher for dissimilar behavior (57%) than for similar behavior (33%).

As an additional dependent measure of the usage of rates in formulating frequency judgments, the protocols were examined for the presence of event cues (i.e., mentions of person, place, time, or other characteristics of the event). It was hypothesized that when subjects used a counting strategy, they would be prone to use more event cues. On the other hand, since an estimation strategy requires the use of a decontextualized rate, the usage of event cues was hypothesized to be much lower. As Table 10.2 indicates, the mean numbers of event cues used were exactly as predicted.

### Experiment 2

The prediction that subjects would use rates in reporting frequency judgments about regular behaviors was also tested using response time as a dependent measure in the second experiment. This prediction would translate into subjects

TABLE 10.2. Usage of Rates in Formulating Frequency Judgments, by Regularity and Similarity of the Behavior

Dependent Measure and Regularity of the Behavior	Similarity of the Behavior	
	Similar	Dissimilar
Experiment 1 ( <i>N</i> = 29)		
Percent using estimation strategies		
Regular	89.5	62.5
Irregular	44.7	24.1
Mean usage of event cues		
Regular	1.20	2.17
Irregular	2.45	2.90
Experiment 2 ( <i>N</i> = 28)		
Mean response time (seconds)		
Regular	4.1	6.5
Irregular	6.8	7.1
Mean effort ratings (1 = none, 7 = a lot)		
Regular	1.5	3.2
Irregular	3.5	4.0

taking less time to make such a judgment if it was based on a rate than if they were to enumerate each occurrence and count them up. Therefore, in general, it was predicted that the response time for a regular behavior would be less than for an irregular behavior. The results were examined by looking at (a) the actual average response time that subjects took to answer the frequency question and (b) the effort ratings from each experiment for regular versus irregular behaviors indicating the amount of cognitive effort that subjects perceived themselves as having expended in the task of formulating a frequency judgment.

This hypothesis received overwhelming convergent support from both measures (see bottom half of Table 10.2). The difference between the average response time for an irregular behavior and a regular behavior was 1.6 seconds. A one-way analysis of variance using response time as the dependent measure indicated a significant main effect of regularity,  $F(1,55) = 5.75$ ,  $p < .02$ .

In addition, subjects perceived the cognitive effort that they had to expend in arriving at a frequency judgment as being significantly greater for irregular behaviors than for regular behaviors,  $p < .01$ . Hence, we can safely conclude that the cognitive effort required and the time taken to arrive at a frequency judgment are less for a regular behavior than for an irregular one.

According to Proposition 3, since frequency judgments for regular behavior are based on rates, we can expect the process to be quicker for regular than for irregular behavior. In addition, since R-DIS would render some episodes accessible, some subjects may engage in a counting strategy. This would mean that,

on an average, a behavioral frequency judgment for R-S would take less time than for R-DIS.

For irregular behaviors, however, (a) subjects do not have access to a rate, and (b) they have to rely on episodic recall. The latter would be easier for dissimilar behavior but very difficult for similar behavior. Hence, we could expect the response time to be shorter for R-DIS than for IRR-S.

In order to examine the above pattern of results, I looked at the average response times for each of the four behavioral categories included in this analysis. A repeated-measures analysis of variance revealed *no* significant regularity  $\times$  similarity interaction,  $F(1,27) = 2.48$ ,  $p < .127$ . The main effects of both independent variables were, however, significant (regularity:  $F(1,27) = 5.52$ ,  $p < .026$ ; similarity:  $F(1,27) = 6.06$ ,  $p < .021$ ). This indicates that it takes less time to report about regular behaviors than about irregular behaviors and less time to report about similar behaviors than about dissimilar behaviors.

For convergent validity, I also examined the average effort that subjects said that they had put into responding to each behavioral frequency question. The effort ratings reflect exactly this same pattern, indicating that the effort that subjects *thought* they had expended and the amount that they had actually expended are consistent with each other. Repeated-measures analyses of variance using effort ratings as the dependent variables revealed a significant regularity  $\times$  similarity interaction, together with significant main effects. This interaction was significant at  $p < .09$ ,  $F(1,27) = 3.15$ , with the main effect of regularity significant at  $p < .004$ ,  $F(1,27) = 9.88$ , and that of similarity significant at  $p < .001$ ,  $F(1,27) = 16.21$ . However, this result was not in the predicted direction. The model had predicted that IRR-DIS behaviors would be less difficult to report about than would IRR-S behavior, since in the latter case subjects would not have access to individual episodes of the behavior nor would they have a rate. This does not seem to be reflected in the findings. A post hoc explanation is that in the case of IRR-S behavior, subjects had to search in fewer places in memory for potential occurrences of the behavior, given its similarity in terms of contextual factors such as person, place, and what happened. On the other hand, in the case of IRR-DIS behavior, although the accessibility to individual episodes was greater, a larger number of places had to be searched in memory owing to the dissimilar or heterogeneous nature of the behavior. Hence, the response time for IRR-DIS was more than for IRR-S. This factor was not considered while developing the model.

## Summary

The two experiments, using different dependent variables (verbal protocols and response time), were designed to test the model for the storage and retrieval of information in computing frequency judgments of different kinds of behaviors. The use of two different dependent measures was considered particularly attractive in its contribution to the convergent validity of the results.

The model for storage and retrieval of information about frequent behaviors

tested in this research (presented in Figure 10.1) would suggest that frequencies for regular behaviors, with a fixed periodicity of occurrence, would be based on a general rate of occurrence retrieved from memory. In other words, it was hypothesized that for regular behaviors, a rate would be stored in memory that would be retrieved and extrapolated for the reference time frame in order to arrive at a frequency judgment. Given this strategy, the associated time taken to report a frequency about a regular behavior would, conceivably, be less than that taken to report the frequency of an irregular behavior, since, in the latter case, subjects would have to rely more on episodic recall that would potentially take more effort and be more time consuming. The dependent variables used in Experiments 1 and 2 support this model well. Although the verbal protocols revealed that subjects did use rates as a basis for arriving at frequency judgments for regular behavior, the response-time data together with effort ratings obtained from subjects revealed that this task was perceived as taking less effort and being less time consuming than was reporting about irregular behaviors.

The second dimension that was examined was the similarity of the behavior, or the extent to which different occurrences of the same behavior varied in terms of where the behavior was engaged in, with whom, and what happened. It was expected that the more similar the behavior, the more accessible the semantic store of the behavior, and that the more dissimilar the behavior, the more accessible the episodic stores of the behavior. This dimension also seems to affect the strategy used by respondents in reporting a frequency judgment, as well as the accuracy of the report. The verbal protocols obtained in Experiment 1 indicated that subjects tended to use counting strategies more often when the behavior was dissimilar than when it was similar. More subjects relied on counting strategies even when a rate was available for an R-DIS behavior than when one was available for an R-S behavior. Hence, there was some support for the differential accessibility of semantic versus episodic stores.

A finding that was, however, not predicted by the model was that IRR-DIS behaviors took longer to report about than did IRR-S behaviors. The opposite was expected, since respondents would have access to neither a rate nor episodic stores in the case of IRR-S behaviors and would consequently have to spend more time arriving at a frequency judgment. A post hoc explanation for this reversal is that given the dissimilarity of an IRR-DIS behavior, a larger number of places would need to be searched in memory before arriving at a frequency judgment. This is not true for an IRR-S behavior.

In conclusion, the manner in which information about mundane, frequent behaviors is organized and stored in memory seems to have a significant impact on the manner in which such information is retrieved and ultimately used in formulating a judgment. Therefore, it is extremely important to consider this in designing questionnaires in which the purpose is to determine "How many times did you . . . ?"

# 11

## On Providing Population Data To Improve Respondents' Estimates of Autobiographical Frequencies

*Edward Blair and Kathleen Williamson*

Recent research has shown that survey respondents are not very accurate in estimating autobiographical frequencies (cf. Marquis, Marquis, & Polich, 1986; Burton & Blair, 1991). Autobiographical frequency questions, once considered simple to answer, are now known to present respondents with a formidable memory task that often results in large errors.

One possible way to reduce the burden that frequency questions place on respondents' memories, and possibly improve subsequent estimates, is to provide respondents with data about the distribution of frequencies in the population. Population data are always available, from survey pretests if from no other source. Such information may help respondents to scale their estimates and may facilitate a rapid "anchor-and-adjust" estimation process that is more accurate than the processes that it replaces.

However, the value of providing population data is by no means assured. There are reasons to believe that respondents may not use population data when estimating autobiographical frequencies, even if these data are provided. Also, if respondents use population data, there are reasons to believe that they will give *less* accurate frequency estimates as a result.

We undertook the present research to learn more about the conditions, if any, under which researchers might be able to use population data to improve estimates of autobiographical frequencies. We also hoped to gain some broader knowledge about the cognitive processes by which people make frequency estimates.

### Background

Our current knowledge regarding frequency estimation in surveys is as follows: We believe that respondents answer autobiographical frequency questions by using information stored in long-term memory plus information that may be

available from the questionnaire. Relevant information in memory includes information about specific episodes of the event being measured as well as non-episodic information such as rules (“I go to church every Sunday”), relationships (“I go to restaurants more than I go to church,” “I go to restaurants more often than most people”), totals (“I once added up how many times I had been to restaurants in a month when I was eating out a lot, and it came to 50, so that would be an upper boundary”), and other information. Information available from the questionnaire can include the context provided by previous questions, magnitude cues provided by response categories (Schwarz & Hippler, 1987), and population data if provided.

The cognitive process used by any given respondent to answer a frequency question is defined by both content and procedure, that is, by the information that is used to generate the frequency estimate and by the way in which the frequency is estimated from this information. Respondents appear to change processes as task conditions change (Blair & Burton, 1987), which implies the existence of some mechanism for selecting a process. This mechanism appears to be automatic—respondents cannot describe it or answer questions about it—but in Burton and Blair (1991), it is suggested that the mechanism balances criteria of effort and accuracy. Variables such as method of administration and respondent motivation determine how much effort the respondent will exert; then the respondent somehow chooses a process that will deliver acceptable accuracy within the constrained level of effort. The perceived effort and accuracy for any particular process depends on the availability, accessibility, and vividness of relevant information.

### Will Respondents Use Population Data?

When population information is not explicitly provided in frequency questions, respondents show little inclination to answer these questions through processes that use population norms (Blair & Burton, 1987; Burton & Blair, 1991). This is not surprising, because such processes require respondents to estimate norms and fit themselves into the population distribution. We usually would not expect respondents to have the norm data available in memory or to feel confident about estimating norms.

Explicitly providing population data in the question will make norm information available, and because the provision of information carries some demand characteristic to use it, respondents might be cued to make judgments about their relative standing in the population. These judgments, along with the population data provided, will furnish a “low-effort” basis for frequency estimation. Even so, it is not clear whether respondents will use this estimation method.

One possibility is that respondents will use population data whenever their confidence in judgments of relative standing is higher than their confidence in the absolute frequency estimates obtainable within their effort constraints. This implies that population data will more likely be used when events are frequent, irregular, pallid, or similar than when they are not, since respondents will have

weaker memory resources to guide absolute frequency estimates under these conditions. Population data are more likely to be used for public behaviors, because respondents will have more confidence in their judgments of relative standing for these behaviors.

A second possibility is that population information will only rarely be used, even when provided, since this information is pallid compared to autobiographical memories. For example, researchers who study inference and decision making have found that people tend to underutilize population data in estimating magnitudes (cf. Nisbett & Ross, 1980), especially when the population data compete with personal experience (e.g., Hansen & Donoghue, 1977). It seems likely that the vividness of information influences respondents' perceptions about the accuracy of estimates drawn from that information, and hence influences their likelihood of using it. Recollections of episodes, even if fragmentary, are likely to be more vivid than recollections of personal rules and relationships, which in turn are likely to be more vivid than population data. This logic would explain why respondents to frequency questions make more use of episode retrieval when given more time to respond, even if episode retrieval does not improve estimates (Burton & Blair, 1991); it also would imply low use of population data.

Other possibilities are that respondents will use the first process at hand rather than the best process to estimate autobiographical frequencies or that respondents somehow will use all of the information at hand in formulating estimates. These possibilities would imply high use of population data and would be consistent with studies showing that people are affected by response categories in estimating autobiographical frequencies (e.g., Schwarz, Hippler, Deutsch, & Strack, 1985).

### How Will Population Data Affect Estimates?

Even if respondents use population data in making frequency estimates, will the result be beneficial? For population data to help respondents, they will need to have a reasonable sense of how their personal frequencies compare with those of others. In fact, for population data to *improve* estimates, respondents will need to have a better sense of relative frequency than of absolute frequency. We think that this condition will not often be met.

A plausible scenario is that providing respondents with information about the population distribution will encourage them to move autobiographical frequency estimates toward the middle of that distribution. This might occur because of the leading effects of appeals to the norm, or simply because people think that others behave as they do. The move toward the middle of the distribution will reduce the variance of frequency reports. It will also tend to lower their mean, since most behavioral frequency distributions have longer tails on the high side and are likely to show more effects from compression on this side. There will be large gains in individual-level accuracy for some respondents who give wild answers without the stabilizing influence of the population data. On the other

hand, the movement toward the middle of the distribution will cause a loss in accuracy for many individuals and, by reducing discrimination among respondents, will make the frequency reports less useful for cross-tabular or correlational analyses. Overall, the effects of population information on the quality and usefulness of frequency estimates will depend very much on the quality of those estimates without population data, and on their intended uses.

## Methods and Results

To explore the possible effects of population data on frequency estimates under various conditions, we conducted two studies. Both used self-administered questionnaires, with junior-level students from Marketing classes at the University of Houston as respondents ( $N = 178$  in Study 1,  $N = 536$  in Study 2). These students mostly are in their early 20s, have independent living arrangements, and should average higher cognitive skills than the population at large.

### Study 1

#### *Methods*

Study 1 had two phases. In the first phase, respondents were asked a series of open-ended questions measuring absolute autobiographical frequencies. One-half of the respondents provided frequencies for a 1-week time frame, and one-half provided frequencies for a 6-week time frame. After answering these absolute frequency questions, respondents were asked how confident they were that each answer was correct ("very confident," "moderately confident," "slightly confident," or "not at all confident"). Next, respondents were asked how their absolute frequencies would place them within their Marketing class. One-fourth of the respondents provided these relative estimates in thirds (e.g., whether their absolute frequency of eating at restaurants would place them in the highest 33%, the middle 33%, or the lowest 33%); one-fourth answered in quartiles; one-fourth answered in quintiles; and one-fourth provided open-ended percentiles. (The four relative frequency measures were crossed with the two absolute frequency time frames in a  $4 \times 2$  between-subjects design, with random assignment of questionnaires to respondents.) Finally, respondents were asked how confident they were that each relative frequency estimate was correct.

Behaviors measured in Study 1 were the following: eating hamburgers, consuming soft drinks, washing clothes, eating at restaurants, seeing a movie at a movie theater, watching movies on TV or videotape, making long distance telephone calls, reading magazines, shopping for clothing, making a purchase from a vending machine, writing checks, withdrawing money from an automatic teller machine, and purchasing gasoline. These behaviors were intended to cover a range of vividness, frequency, regularity, and publicness. We had used several of the behaviors in previous studies of respondents' cognitive processes (see Blair & Burton, 1987; Burton & Blair, 1991).

In the second phase of Study 1, conducted 2 months later, the same respondents were asked for open-ended, absolute frequencies for five of the same behaviors (consuming soft drinks, eating at restaurants, phoning long distance, shopping for clothes, and writing checks). All questions covered a 6-week time frame and contained distributional information compiled from the first phase data. Half of the respondents received this information in the form of median splits from Phase 1; half received quartile splits. For example, for the soft drink frequency question, respondents in the median condition were told that, for those students who gave a 6-week frequency other than zero in the previous questionnaire, "50% gave a number in the range 1 to 30, and 50% gave a number in the range 30 or more." Respondents in the quartile condition were told, "25% gave a number in the range 1 to 10, 25% gave a number in the range 10 to 30, 25% gave a number in the range 30 to 50, and 25% gave a number in the range 50 or more."

After answering each frequency question, respondents were asked how confident they were that the answer was correct. Then, after answering all of the frequency and confidence questions, respondents were asked how helpful they found the distributional information in coming up with accurate answers for each frequency question ("very helpful," "moderately helpful," "slightly helpful," or "not at all helpful").

### *Phase 1 Results*

The first results that we considered were respondents' self-assessments of relative standing from Phase 1. We were looking for even distributions in these assessments; that is, if we asked respondents whether their frequency of consuming soft drinks would place them in the highest third, middle third, or lowest third of the class, we were looking for 33% to answer "highest third," 33% to answer "middle third," and 33% to answer "lowest third." An uneven distribution—for example, 9% putting themselves in the highest third, 37% in the middle third, and 53% in the lowest third—would indicate errors in relative placement. This would suggest that encouraging respondents to rely on relative assessments by providing information about the population distribution might harm rather than help the accuracy of frequency estimates.

We expected differences in evenness across the various measures of relative frequency. We thought that asking respondents to place themselves in thirds or fifths would produce distributions with a strong spike in the middle category, that quartiles would produce a better spread because they would not allow respondents to place themselves in a middle category, and that open percentiles would produce the best spread because they would avoid categories altogether.

We also expected the evenness of distributions to vary across the behaviors that we measured. We thought that on the more public behaviors, the distribution would be more even because respondents would have a better sense of their relative standing for these behaviors, and that on the less public behaviors, more people would consider themselves to be medium.

Table 11.1 shows results for respondents who were asked to place themselves in thirds. The respondents' ability to place themselves accurately in thirds varied across behavior and generally was not promising. On only 3 of the 13 behaviors did even 20% of the subjects place themselves in the top third. We did not see the middle spike that we had expected. We thought that most respondents would represent themselves as medium; instead, it appeared as if respondents with low frequencies *knew* that they were low, while other respondents distributed themselves around the middle category.

Table 11.2 shows results for respondents who were asked to place themselves in quartiles. The quartiles generally were more even than the thirds (and quintiles). We had expected this. Even so, the distributions were not as even as one would like. The behaviors with the smoothest distributions were soft drink consumption, watching TV or videotape movies, phoning long distance, and check writing. We thought that the most even distributions would occur for the most public behaviors, but this did not appear to be the case.

Table 11.3 summarizes the percentages of respondents who placed themselves in the top categories under each relative assessment procedure: open percentile estimates, quintiles, quartiles, and thirds. The top category was used most often in the open estimates, suggesting, as we expected, that the very use of categories may encourage people to represent themselves as medium. Overall, all procedures showed a tendency for respondents to rate themselves as medium or low. This implies that providing population data may lead to lower mean estimates if respondents use these data to guide their answers to frequency questions.

We next considered respondents' confidence ratings regarding the accuracy of various estimates. We wanted to know whether respondents had higher or lower confidence in their ability to assess relative frequency (compared with absolute frequency). Lower confidence in the relative estimates might imply that respondents would hesitate to use population data as a basis for generating autobiographical frequency estimates.

We expected confidence in the accuracy of absolute estimates to vary substantially across the behaviors that we measured. We thought that confidence would be high for infrequent and/or vivid behaviors such as clothes shopping and long distance phoning and low for behaviors such as soft drink consumption and check writing. We thought that respondents would show more confidence in absolute frequency estimates when the time frame was 1 week than when it was 6 weeks.

For relative frequency estimates, we expected respondents to show more confidence as the estimates became more coarse (highest confidence for thirds, lowest for open percentiles). We thought that confidence in the relative estimates generally would be lower than confidence in the absolute estimates but that we would find some crossovers that might be exploited in subsequent research.

Table 11.4 shows the percentages of respondents who said that they were "very confident" in the accuracy of their absolute and relative frequency esti-

TABLE 11.1. Percentage of Respondents Placing Themselves in Relative Thirds<sup>a</sup>

Behavior	Lowest Third	Middle Third	Highest Third
Eating hamburgers	53	37	9
Consuming soft drinks	41	41	18
Washing clothes	25	55	21
Eating at restaurants	27	42	31
Going to movies	73	18	9
Watching movies on TV or videotape	43	4	11
Phoning long distance	41	32	27
Reading magazines	27	55	18
Shopping for clothes	52	46	2
Making vending machine purchase	50	41	9
Writing checks	52	32	16
Withdrawing money from an ATM	72	21	7
Purchasing gasoline	50	39	11

<sup>a</sup>Based on Study 1 respondents who were asked which relative third they would fall in ( $n = 44$ ).

TABLE 11.2. Percentage of Respondents Placing Themselves in Quartiles<sup>a</sup>

Behavior	Lowest Quartile	Second Quartile	Third Quartile	Highest Quartile
Eating hamburgers	44	21	23	12
Consuming soft drinks	19	26	35	21
Washing clothes	12	19	44	26
Eating at restaurants	14	23	40	23
Going to movies	42	33	19	7
Watching movies on TV or videotape	28	28	23	21
Phoning long distance	35	19	26	21
Reading magazines	19	42	28	12
Shopping for clothes	33	35	21	12
Making vending machine purchase	40	35	16	9
Writing checks	30	16	28	26
Withdrawing money from an ATM	58	16	14	12
Purchasing gasoline	28	14	28	30

<sup>a</sup>Based on Study 1 respondents who were asked which relative quartiles they would fall in ( $n = 43$ ).

mates. As expected, confidence in the absolute frequency estimates varied substantially across behaviors, with the lower frequency behaviors showing higher confidence. Also as expected, confidence in absolute estimates dropped when the time frame went from 1 week to 6 weeks. Confidence in the relative estimates varied less across measurement forms and, as expected, generally was lower than confidence in the absolute estimates.

TABLE 11.3. Percentage of Respondents Placing Themselves in Highest Categories

Behavior	Highest 20% <sup>a</sup>	Highest Fifth <sup>b</sup>	Highest Fourth <sup>c</sup>	Highest Third <sup>d</sup>
Eating hamburgers	20	7	12	9
Consuming soft drinks	22	22	21	18
Washing clothes	24	4	26	21
Eating at restaurants	11	9	23	31
Going to movies	9	2	7	9
Watching movies on TV or videotape	9	9	21	11
Phoning long distance	17	18	21	27
Reading magazines	7	4	12	18
Shopping for clothes	1	7	12	2
Making vending machine purchase	20	4	9	9
Writing checks	9	11	26	16
Withdrawing money from an ATM	2	4	12	7
Purchasing gasoline	26	13	30	11

<sup>a</sup>Based on Rs who were asked what (open-ended) percentile they would fall in ( $n = 46$ ).

<sup>b</sup>Based on Rs placing themselves in quintiles ( $n = 45$ ).

<sup>c</sup>Based on Rs placing themselves in quartiles ( $n = 43$ ).

<sup>d</sup>Based on Rs placing themselves in thirds ( $n = 44$ ).

TABLE 11.4. Percentage of Respondents "Very Confident" in Frequency Estimates

Behavior	Absolute Estimates		Relative Estimates			
	6 weeks	1 week	Open	Fifths	Fourths	Thirds
Eating hamburgers	36	78	28	24	26	34
Consuming soft drinks	28	40	35	33	28	45
Washing clothes	64	83	22	36	26	32
Eating at restaurants	27	60	13	27	23	30
Going to movies	72	98	24	51	33	50
Watching movies on TV or videotape	30	54	20	24	28	43
Phoning long distance	58	83	20	33	35	36
Reading magazines	30	49	17	20	21	25
Shopping for clothes	58	90	28	36	21	32
Making vending machine purchase	28	52	35	36	35	34
Writing checks	32	61	30	31	47	36
Withdrawing money from an ATM	64	91	39	49	60	59
Purchasing gasoline	67	96	39	47	35	43

### Phase 2 Results

In the second phase of Study 1, the same respondents were asked to answer autobiographical frequency questions that contained information about either the population median or the population quartiles. They later were asked how con-

fidient they were that their answers were accurate and how helpful they found the population information in formulating their answers.

We were interested in learning whether the response distributions for these open-ended frequency questions would change (from Phase 1) in a manner suggesting that respondents were influenced by the population data. We also were interested in seeing whether the population information was regarded as helpful and whether respondents became more confident in the accuracy of their answers.

The five behaviors measured in Phase 2 were the following: consuming soft drinks, eating at restaurants, phoning long distance, shopping for clothes, and writing checks. These behaviors were chosen on the basis of confidence ratings for the 6-week frequency estimates provided in Phase 1 and on the basis of having few respondents who reported 6-week frequencies of zero. Long distance phoning and clothes shopping were used to represent behaviors with high confidence ratings for the 6-week absolute estimates (much higher than confidence ratings for the relative estimates); soft drink consumption, eating at restaurants, and check writing were used to represent behaviors with low confidence ratings. We were particularly interested in check writing, since we had found in Burton and Blair (1991) that many respondents said that they “just guessed” in estimating check-writing frequencies over a 6-week time frame.

We thought that the Phase 2 response distributions for clothes shopping and long distance phoning would change very little from the Phase 1 distributions, since respondents had high confidence in their abilities to estimate these frequencies without population data. Similarly, we thought that there would be very little change in respondents’ confidence in these estimates and that the population data would not be considered helpful.

We expected the population data to have more effect on the results for soft drink consumption, eating at restaurants, and check writing. We thought that respondents who received population data would show lower means and variances in their frequency estimates for these behaviors (for reasons discussed earlier) and would be more confident about the accuracy of their answers. We also thought that effects would differ between respondents who received population quartiles and respondents who received medians. We expected confidence and helpfulness ratings to be higher when quartiles were given, since the quartiles provided more information; but we thought that frequency estimates would change more when medians were given, since this condition emphasized the middle of the distribution.

Table 11.5 shows percentages of respondents who rated the population information as “helpful.” Roughly 30–40% rated this information at least “somewhat helpful,” but few rated it as “very helpful.” The ratings did not differ much across behaviors. More information—that is, quartiles—was not rated as any more helpful.

Table 11.6 offers a different view of the perceived value of population data. Respondents generally expressed no more confidence in their frequency estimates in Phase 2, where they received population data, than in Phase 1, where

TABLE 11.5. Percentage of Respondents Who Found Data Helpful

Behavior	At All Helpful		Very Helpful	
	Quartile	Median	Quartile	Median
	Splits Given	Splits Given	Splits Given	Splits Given
Consuming soft drinks	41	45	6	7
Eating at restaurants	41	43	4	7
Phoning long distance	29	37	0	8
Shopping for clothes	35	36	0	8
Writing checks	33	37	3	10

TABLE 11.6. Percentage of Respondents "Very Confident" in Absolute Estimates

Behavior	Phase 1:	Phase 2:
	No Population Information Given	Quartile or Median Splits Given
Consuming soft drinks	28	24
Eating at restaurants	27	22
Phoning long distance	58	54
Shopping for clothes	58	58
Writing checks	32	47

they did not. The one exception was check writing. Confidence ratings were also similar between respondents who received quartiles and respondents who received medians (results not shown).

Table 11.7 shows the means and standard deviations for respondents' frequency reports under the various measurement conditions. Means and standard deviations fell when respondents were given population information for soft drink consumption, long distance phoning, and check writing. However, these differences were not statistically significant.

TABLE 11.7. Mean Reported Frequencies

Behavior	Phase 1	Phase 2	
	No Population Information Given	Quartile Splits Given	Median Splits Given
Consuming soft drinks	41.2	31.7	32.0
Eating at restaurants	18.9	16.2	18.3
Phoning long distance	10.0	7.0	8.8
Shopping for clothes	3.6	3.7	3.6
Writing checks	18.7	18.1	14.7

We had expected the largest effects to occur for soft drink consumption, eating at restaurants, and check writing—the three behaviors for which respondents showed the least confidence in their absolute frequency estimates. Effects were largest, although not significant, for soft drink consumption, long distance phoning, and check writing.

We also had expected that providing only a population median would cause more compression in estimates than would providing population quartiles. These differences did not appear in any systematic or significant fashion. (Table 11.7 shows standard deviations; we also looked at percentiles and extreme values.)

## Study 2

### *Methods*

Study 2 measured frequencies for the same five behaviors as in Phase 2 of Study 1, again with a 6-week time frame, in a new group of 536 respondents. Respondents in Study 2 were assigned to one of five conditions: One group received no population data; one group received median splits as in the second phase of Study 1; one group was given the median for each behavior in the form “the average person gave an answer of . . .”; one group received quartile splits as in the second phase of Study 1; and one group received these quartile splits and was asked for closed responses using the quartiles as categories. As in the second phase of Study 1, respondents indicated how confident they were in the accuracy of each frequency estimate after providing it. They did not, however, rate the helpfulness of the distributional information.

Study 2 allowed us to test the effects of providing population data to a larger sample of respondents than in Study 1 and with some different presentation formats.

### *Results*

As in the second phase of Study 1, we expected the effects of population information to be larger for soft drink consumption, eating at restaurants, and check writing than for long distance phoning and clothes shopping. We also thought that if effects emerged, the four groups that received population information would all show lower means and standard deviations in their frequency reports than would the unaided group. Among the aided groups, (a) we expected lower figures (i.e., stronger effects) from the two groups that received medians than from the two groups that received quartiles, on the assumption that movement toward the middle of the distribution would be stronger when only the middle was presented; (b) between the two groups that received medians, we expected lower figures from the group that received this information in the “average person” format, because this format makes a more explicit appeal to the norm; and (c) between the two groups that received quartiles, we expected lower figures from the group that provided categorical answers, on the assumption that this format more explicitly asks respondents to consider the population data.

Table 11.8 shows the means and standard deviations for the various measurement conditions. There were no significant differences across the measurement conditions, except for isolated paired comparisons on individual behaviors, and there was no significant variation in effects across behaviors. As in Study 1, we considered percentiles and extreme values as well as standard deviations, with no difference in results.

The standard deviations that are reported in Table 11.8 tend to be lowest for the group that received quartiles and answered in categories. We believe that this is irrelevant to the effects of population data on frequency estimates. Instead, we believe that it occurred because we estimated this group's means and variances using a smoothing algorithm that fits continuous distributions to categorical data (LaMotte & Blair, 1988). The open-ended answers from the other groups are subject to "clumping" around numbers such as 5, 10, etc. This clumping is caused by response errors and increases the variance of reports. However, the smoothed distributions estimated for the categorical group are less subject to clumping.

Table 11.9 shows the percentages of respondents who were "very confident" in their answers under the various measurement conditions. Like the frequency reports themselves, these confidence ratings did not change when population information was provided, except for the group that provided the frequency estimates in categories. We regard the higher confidence ratings from this group as resulting from the categorical measure and not from the provision of population information.

## Conclusions

Overall, our results do not suggest that respondents used population data to guide autobiographical frequency estimates. The means and standard deviations of reported frequencies did not change significantly when population data were provided. Also, respondents' confidence in the accuracy of their frequency reports did not change. These findings, combined with the lack of differences in response to various formats for the population data, suggest to us that respondents did not use the information.

Since we did not measure response accuracy in this research, we cannot make any definite statements about the effects of population data on respondents' accuracy in estimating autobiographical frequencies. However, if respondents do not use population data, then there is no reason to believe that population information will help improve their answers.

From a viewpoint of learning more about cognitive processes, we find the results interesting because they imply that respondents do not use all available information in generating autobiographical frequency estimates. Instead, the results are consistent with an argument that cognitive processes compete for use and that information is not pooled across winning and losing processes. The

TABLE 11.8. Mean Reported Frequencies and Standard Deviations

Behavior	No Population Information Given	Quartiles Given, Open-ended Response	Quartiles Given, Closed Response	Median Given in 50/50 Format	Median Given in "Average" Person Format
Consuming soft drinks	39.2 (39.0)	33.5 (38.4)	31.0 (28.7)	34.1 (28.3)	28.7 (27.6)
Eating at restaurants	18.5 (15.4)	19.1 (14.8)	19.8 (14.0)	19.0 (13.7)	17.0 (12.6)
Phoning long distance	6.9 (9.1)	10.1 (26.5)	7.2 (13.5)	9.4 (15.2)	8.0 (13.2)
Shopping for clothes	2.7 (3.2)	2.9 (2.8)	3.2 (3.9)	3.0 (3.1)	2.1 (2.1)
Writing checks	20.4 (23.0)	16.8 (24.4)	14.0 (13.0)	20.1 (29.5)	18.1 (20.0)

Note: Standard deviations are in parentheses.

TABLE 11.9. Percentage of Respondents "Very Confident" in Frequency Estimates

Behavior	No Population Information Given	Quartiles Given, Open-ended Response	Quartiles Given, Closed Response	Median Given in 50/50 Format	Median Given in "Average" Person Format
Consuming soft drinks	26	30	26	26	54
Eating at restaurants	21	28	30	33	63
Phoning long distance	59	67	62	55	81
Shopping for clothes	70	60	59	62	63
Writing checks	47	47	48	37	84

results also are consistent with an argument that population information is disadvantaged in this competition, perhaps because it is pallid.

From a viewpoint of learning how to improve estimates of autobiographical frequencies, we find the results disappointing because they imply that providing population data will not accomplish this purpose. However, three caveats temper this conclusion. First, population data may be helpful in improving estimates for *threatening* frequency questions, which we did not study. The mechanism would not be based in cognitive processes but rather in legitimizing correct reporting (Blair, Sudman, Bradburn, & Stocking, 1977, make a parallel argument regarding their question-length manipulation). Second, population data seem likely to influence respondents' estimates of *nonautobiographical* frequencies, where

memory resources will play a lesser role. Third, different research procedures may unveil an influence of population data that is not seen in our results.

For example, Schwarz et al. (1985) showed that frequency estimates change as the boundaries of response categories are raised and lowered, and they argued that this change occurs because respondents use the categories as information about the population distribution. They may be wrong in this argument—the change may have some other cause—but they may also be right. Population information may have an effect, but it may be necessary to raise or lower the provided norms to see this effect or to look at individual-level data. Also, Schwarz has suggested (in private conversation) that the effect of population data may disappear when the information is labeled as such, because respondents somehow adjust for this information once they recognize it as an explicit influence.

We think that it would be interesting to see the effects of population data studied by methods that were different from what we used, and especially to see these effects studied where validating data were available. Our guess is that the results would echo our results; they would be interesting from the viewpoint of learning about cognitive processes but not from the viewpoint of improving autobiographical frequency estimates for mundane events.

# 12

## Retrospective Reports: The Impact of Response Formats

*Norbert Schwarz, Hans-J. Hippler,  
and Elisabeth Noelle-Neumann*

As the chapters in this volume indicate, basic psychological research into individuals' autobiographical memories has provided many fruitful insights into the cognitive processes that underlie autobiographical recollections. In only a few years, this research has had a pronounced impact on the strategies that applied researchers use to obtain veridical retrospective reports in sample surveys. Nevertheless, the methods that are typically used in psychological experiments and sample surveys differ in important ways. Most importantly, the majority of psychological investigations are based on retrospective reports that subjects provide in a free-response format. In contrast, respondents in a survey interview are typically asked to check one of several response alternatives presented to them by the survey interviewer.

This difference in procedures reflects the fact that cognitive psychologists are interested in obtaining information that is sufficiently detailed to allow analyses of the underlying cognitive processes. Moreover, the small number of subjects typically used in psychological experiments allows these researchers to employ detailed coding strategies to exploit the rich material that they collect. Survey researchers, on the other hand, are primarily interested in obtaining veridical factual information about some content domain, and how respondents arrive at their reports is not of particular interest in this applied research. Moreover, the potential richness of respondents' free reports would most likely be reduced by the interviewer's ability to make appropriately detailed notes, hence rendering the transcripts less informative than they are under the more elaborate recording conditions of the psychological laboratory. In addition, economic considerations require that the interview be kept as short as possible, and the large number of respondents sampled typically prohibits the use of data collection methods that require extensive coding of the obtained material. As a result, survey researchers prefer to assess retrospective reports in a closed-response format that requires respondents to check one of several precoded response alternatives presented to them.

In this chapter, we review the procedures used by survey researchers, emphasizing the impact of precoded response alternatives on respondents' reports. Whereas previous discussions of the use of response alternatives focused primarily on issues of opinion measurement (for reviews, see Schwarz & Hippler, 1991; Schwarz, Hippler, & Noelle-Neumann, 1992), we shall now apply the concepts developed in this domain to the assessment of retrospective factual reports. In doing so, we hope to address the dual audience of this volume. On the one hand, we want to draw cognitive psychologists' attention to factors that have so far been neglected in basic research on autobiographical memory, whereas on the other hand, we want to inform survey researchers about unintended "side effects" of the data collection strategies used.

## What Did You Do? Recollections of Activities

In the first part of this chapter, we explore the impact of response formats on respondents' recollections of the activities in which they engaged. We begin by comparing open-ended questions with closed questions, and then we address the impact of the order in which various response alternatives are presented.

### Open- and Closed-Response Formats: How Do They Differ?

Experimental studies that compare the answers provided in an open- or closed-response format are, to our knowledge, limited to the domain of opinion questions. We assume, however, that the obtained findings are also likely to hold for autobiographical reports, although this hypothesis deserves systematic investigation. Most importantly, comparisons of open- and closed-response formats converge on the finding that these formats may yield considerable differences in the marginal distribution as well as the ranking of items. Suppose, for example, that respondents are asked, in a free-response format, what they have done today. It seems likely that only a few respondents would spontaneously report that they took a shower. If "taking a shower" were provided as an item on a list of daily activities, on the other hand, we may assume that most respondents would endorse it. Conversely, if we presented a list of daily activities that omitted "taking a shower" as an item, we may safely assume that rarely would any respondent report this activity, even if a generic "other" category provided the opportunity to do so.

This thought experiment reflects a robust finding in the domain of opinion questions (e.g., Bishop, Hippler, Schwarz, & Strack, 1988; Schuman & Presser, 1977). Research in this domain indicated that any given opinion is less likely to be volunteered in an open-response format than to be endorsed in a closed-response format. On the other hand, opinions that are omitted from the set of response alternatives in a closed-response format are unlikely to be reported at all, even if an "other" category is explicitly offered (Bradburn, 1983; Molenaar, 1982). Several processes are likely to contribute to these findings.

First, precoded response alternatives may remind respondents of options that they may otherwise not consider. From a cognitive perspective, open-response formats present a free-recall task to respondents, whereas closed-response formats present a recognition or a cued-recall task. As in other domains of research, recognition and cued-recall tasks result in higher degrees of recall (cf. Smyth, Morris, Levy, & Ellis, 1987). Thus, open-response formats are more adequate when the investigator is interested in the salience that an activity or issue has for respondents, where the order in which a respondent retrieves different activities and the total number of respondents who retrieve a particular one are of primary interest (cf. Bodenhausen & Wyer, 1987). Closed formats, on the other hand, are more appropriate when the investigator is interested in a fairly complete report about a large set of specified activities to determine their relative frequencies.

Second, in an open-response format, respondents are unlikely to report spontaneously information that seems self-evident or irrelevant. In refraining from these responses, they follow conversational maxims (Grice, 1975; for applications to survey research, see Schwarz, Strack, & Mai, 1991; Strack & Martin, 1987; Strack & Schwarz, 1992) that require speakers to provide information that is informative and relevant to the topic of the conversation. This results in an underreporting of apparently irrelevant information, such as "taking a shower." This underreporting is eliminated by closed-response formats, where the explicit presentation of the proper response alternative indicates the investigator's interest in this information. For the same reason, respondents hesitate to endorse an "other" alternative in a closed-response format, even if explicitly provided. Such a generic response violates conversational maxims because it does not convey what the respondent actually did, thus providing little information. In fact, in most natural conversations, generic answers of the "other" type would be considered inappropriate.

In addition, respondents may frequently be uncertain if information that comes to mind does or does not belong to the domain of information in which the investigator is interested. Again, closed-response formats reduce this uncertainty, resulting in higher responses. In the domain of opinion questions, this differential complexity of open- and closed-response formats is reflected in higher nonresponse rates in the open-response format, particularly among less educated respondents (Schuman & Presser, 1981).

As a first conclusion, we note that open- and closed-response formats pose different cognitive tasks and result in different patterns of responses, thus undermining the comparability of data obtained under open- and closed-response format conditions. Next we consider the impact of the order in which different response alternatives are presented.

## Order Effects in Retrospective Reports

Suppose that respondents are given a list of daily activities and are asked to check each activity in which they engaged. Would the order in which these activities are presented affect respondents' performance on this cued-recall task?

The empirical answer is “yes.” Some studies from the archives of the Institut für Demoskopie Allensbach illustrate this point. These data are based on split-ballot experiments conducted in representative surveys based on quota samples of the adult population of the Federal Republic of Germany. Most of the studies were conducted in the 1950s, all under the direction of Elisabeth Noelle-Neumann.

In two of these studies (IfD 1008, May 1957, and IfD 1022, September 1958), respondents were asked:

Could you please tell me, with the help of this list, what you happened to do last Saturday?

Respondents were provided a list of 28 activities, and the order in which these activities were presented was reversed for half of the respondents. Table 12.1 shows the key findings.

Whereas 34% of the respondents to the 1957 survey (shown in the first two columns of Table 12.1) reported that they worked on their job “last Saturday” when this item was presented as the first one on the list, only 25% did so when this item was presented last. Conversely, 15% reported that they slept in when this item was presented first, whereas 10% reported doing so when this item was presented last. The 1958 data, shown in the second and third columns, nicely replicate this pattern. Thus, pronounced primacy effects of up to 9 percentage points emerged when respondents were asked to recall activities that they engaged in “last Saturday,” that is, a maximum of 6 days previously.

These order effects do show reality constraints, however. In June 1960 (IfD 1043), the Allensbach researchers asked respondents what they did Whitsunday, using a list of 30 activities similar to the one used in the previous studies. As shown in the last columns of Table 12.1, respondents were again more likely to report that they slept in when this item was presented first (46%) rather than last (35%). However, the likelihood that they reported working at their jobs during this major holiday was not affected by presentation order (10% and 11%, respectively).

On the other hand, response-order effects are likely to be more pronounced when a specific temporal reference is missing. For example, in February 1977 (IfD 3039), respondents were asked, “What do you generally do on weekends, that is, during your leisure time on Saturdays and Sundays?” and were provided a list similar to the one used in the previous studies. In this case, 37% of the respondents reported that they typically “sleep in” when this item was presented first, whereas only 21% reported doing so when the item was presented last. Thus, a primacy effect of 16 percentage points emerged on reports of one’s typical behavior, which were not constrained by a specific temporal reference. Although a temporal distance of two decades, and accompanying changes in the content of the list, do not permit a direct comparison of the reported studies, the findings suggest that response-order effects are more pronounced in reports of one’s usual behavior than in reports of one’s behavior on a specified recent day.

How are we to account for the emergence of response-order effects in retrospective behavioral reports? Although we cannot verify respondents’ reports,

TABLE 12.1. Reported Activities as a Function of Presentation Order (%)

Activity	"Last Saturday"				"Whitsunday"	
	1957		1958		1960	
	Presentation Order		Presentation Order		Presentation Order	
	First	Last	First	Last	First	Last
Worked	34	25	26	22	10	11
Slept	15	10	17	12	46	35
<i>N</i>	(347)	(317)	(318)	(309)	(701)	(672)

*Note:* Data based on quota samples of German adults (IfD 1008, May 1957; IfD 1022, September 1958; IfD 1043, June 1960). Respondents were randomly assigned to one of the two presentation-order conditions.

their random assignment to the order conditions eliminates the possibility that the obtained differences reflect differences in real behavior. Hence, the obtained primacy effects need to be explained on the basis of the impact of presentation order on respondents' performance on a cued-recall task. We suggest that several, not mutually exclusive, processes may contribute to the emergence of these order effects.

First, successful retrieval requires time. Recalling specific events may take up to several seconds (Reiser, Black, & Abelson, 1985), and repeated attempts to recall may result in the retrieval of additional material, even after a considerable number of previous trials (e.g., Means, Mingay, Nigram, & Zarrow, 1988; Williams & Hollan, 1981). It is conceivable that respondents' motivation to engage in this effort declines over the course of the list, resulting in more successful retrieval attempts at the beginning than at the end. If so, the responses to the first few items should be more veridical than responses to later items, a possibility that deserves systematic investigation in future research.

Moreover, numerous studies have indicated that the more recall increases, the better the retrieval cue matches the representation stored in memory (for a review, see Strube, 1987). Most notably, Wagenaar (1986) observed in a study of his own memory that he could retrieve any apparently forgotten event as more specific retrieval cues became available. The retrieval cues presented in standardized lists of daily activities, however, are not very specific. We assume that they are most likely to result in successful retrieval when respondents take their time to think about them, thus constructing more specific personal retrieval cues from the generic material presented in the list. However, not only may the likelihood that they do so decrease as a function of motivational deficits, but also respondents' idiosyncratic elaboration of the presented cues may be impaired by the elaborations that they generated in response to preceding cues.

Krosnick and Alwin (1987, see also Krosnick, 1992) noted in a discussion of response-order effects on opinion questions that "items presented early in a list

are likely to be subjected to deeper cognitive processing; by the time a respondent considers the later alternatives, his or her mind is likely to be cluttered with thoughts about previous alternatives that inhibit extensive consideration of later ones" (pp. 202–203). As a result, respondents may be most likely to generate detailed and effective recall cues in response to the generic cues presented at the beginning of the list, resulting in higher endorsements of the same item when presented early rather than late. Again, this hypothesis would suggest that responses to early items are more veridical than responses to later ones.

The assumption that recall may depend on respondents' elaboration of the retrieval cues presented to them has an interesting additional implication. It suggests that the direction of order effects may depend on the mode in which the retrieval cues are presented. In the above examples, the list of daily activities was presented in a *visual mode*; that is, respondents were given a list of activities to read. Sometimes, however, a short list of activities may be *read* to respondents by the interviewer—for example, as part of a telephone interview. How would such an *auditory presentation mode* affect respondents' elaboration of the retrieval cues presented to them? In this case, respondents have little opportunity to elaborate on the cues presented early in the list, because the time that is available for processing the early items is limited by the speed with which the interviewer moves on to read the next item. As Krosnick and Alwin (1987) suggested in their discussion of response-order effects in opinion questions, "Under these circumstances, respondents are able to devote most processing time to the *final* item(s) read, since interviewers usually pause most after reading them" (p. 203). In addition, respondents may find it difficult to keep all response alternatives in mind without visual cues. As a result, we may expect that items presented near the end of the list will receive more elaboration and may, therefore, be more likely to generate successful retrieval cues. If so, recency effects should be likely when the list is read to respondents by an interviewer.

In combination, these considerations predict an interaction effect of presentation order and presentation mode. When the list is presented in a visual mode, items presented early in the list should receive more processing, resulting in primacy effects. When the items are read to respondents, however, those at the end of the list should receive more processing, resulting in recency effects. To the extent that more elaborate retrieval cues increase the veridicality of recall, we may also expect that retrospective reports are more veridical for early than for late items if a visual presentation mode is used but are more veridical for late than for early items if an auditory presentation mode is used. To date, these hypotheses have not been systematically tested in the domain of autobiographical recall, although the general hypothesis of an interaction effect of serial position and presentation mode received some support in the domain of opinion questions (for a review, see Schwarz, Hippler, & Noelle-Neumann, 1992).

Finally, it seems likely that response-order effects will, to some degree, depend on the specific context created by preceding items. However, the specific effects are difficult to predict because the content of preceding items may facilitate as well as inhibit the recall of material related to subsequent items. More-

over, respondents may use their answers to preceding items to infer what their answers to subsequent items may plausibly be, without engaging in extended recall efforts. Thus, a respondent who recalled that he worked last Saturday may conclude, without spending too much effort on checking the veridicality of this conclusion, that this implies that he could hardly have "slept in." If so, the general effects of effort and elaboration addressed above may occasionally be diluted by the impact of the local context created by the specific items presented.

In summary, these speculations suggest that response-order effects raise a number of interesting theoretical issues for research on autobiographical recall. Moreover, the possibility that the direction of response-order effects may depend on the mode in which the list is presented points to possible limitations in the comparability of data collected under self-administered and face-to-face interview conditions on the one hand and under telephone interview conditions on the other (see Schwarz, Strack, Hippler, & Bishop, 1991, for a more detailed discussion of mode differences).

## How Often Did You Do It? Estimating the Frequency of Activities

Much as precoded response alternatives affect *which* activities respondents recall, they also influence respondents' estimates of how *frequently* they engaged in a given activity. This is particularly likely when the activities are frequent and mundane. As numerous studies have indicated, individual instances of mundane and frequent behaviors, such as watching TV or having a drink, are not separately represented in memory (for reviews, see Bradburn, Rips, & Shevell, 1987; Schwarz, 1990a; Strube, 1987). Instead, individual episodes tend to blend into one generic representation, thus making it difficult to determine their frequency on the basis of a "recall-and-count" procedure (see Blair & Burton, 1987; Sudman & Schwarz, 1989). Accordingly, respondents have to rely on estimation strategies to determine the frequency of mundane activities. Although a number of different estimation strategies have been documented in the literature (for reviews, see Bradburn et al., 1987; Schwarz, 1990a), a strategy that is particularly important in survey interviews involves the use of precoded response alternatives.

## The Impact of Frequency Scales

In survey research, respondents are often asked to report the frequency with which they engage in a certain behavior by checking the appropriate alternative from a set of response categories provided to them. Although the selected alternative is assumed to inform the researcher about the respondent's behavior, the possibility that a given set of response alternatives is far more than a simple "measurement device" is frequently overlooked. Instead, it may also constitute an important source of information for the respondent, which they may actively use in arriving at a judgment.

With regard to frequency scales, respondents assume that the range of the response alternatives reflects the researcher's knowledge of, or expectations about, the distribution of the behavior in the "real world." Thus, they assume that the behavior of the "average" person is represented by the values stated in the middle range of the response scale, whereas the extremes of the scale correspond to the extremes of the distribution (Schwarz, Hippler, Deutsch, & Strack, 1985; Schwarz & Hippler, 1987). Accordingly, respondents may use the range of the response alternatives as a frame of reference in estimating their own behavioral frequency. This results in higher frequency reports when the response scale provides high-frequency rather than low-frequency response alternatives.

A study on leisure time activities can serve as an illustration. In this study (Schwarz et al., 1985, Experiment 1), we asked German adults to report how many hours a day they spent watching TV. Previous research by Darschin and Frank (1982) indicated that the average daily TV viewing in West Germany was slightly more than 2 hours. To test the impact of different response alternatives, half of the sample received a scale ranging in half-hour steps from "up to  $\frac{1}{2}$  hour" to "more than  $2\frac{1}{2}$  hours," while the other half received a scale ranging from "up to  $2\frac{1}{2}$  hours" to "more than  $4\frac{1}{2}$  hours." As shown in Table 12.2, the range of the response alternatives had a pronounced impact on respondents' reports.

When the reports are coded to reflect estimates of either  $2\frac{1}{2}$  hours or less, or more than  $2\frac{1}{2}$  hours, only 16.2% of the respondents reported watching TV for more than  $2\frac{1}{2}$  hours when presented with the low-frequency scale. In contrast, 37.5% did so when presented with the high-frequency scale. Subsequent research indicated that the impact of the range of response alternatives on the obtained reports is very robust and replicates over a wide range of content domains, including media consumption, shopping behavior, sexual behavior, and medical complaints (e.g., Billiet, Loosveldt, & Waterplas, 1988; Schwarz & Bienias, 1990; Schwarz & Scheuring, 1988, 1992; see Schwarz, 1990a, for a review). In all domains, respondents reported higher frequencies on scales with high-frequency than with low-frequency response alternatives.

## The Role of Competing Information

These findings indicate that individuals are unlikely to have detailed knowledge about the *absolute* frequency and duration of mundane activities stored in memory. As a result, they have to rely on estimation strategies. Theoretically, the less episodic information is accessible in memory, the more likely they should be to have to rely on estimation strategies. Two studies bear directly on this implication.

In one of these studies (Chassein, Strack, & Schwarz, 1987), we increased the accessibility of episodic information by asking some respondents to list all TV programs that they could recall from the preceding week before they reported their TV viewing for that week. Other respondents were given a chance to refresh their memory by browsing through a TV program guide, and a final group

TABLE 12.2. Reported Daily TV Viewing as a Function of Response Alternatives

Response Alternatives	%
Low frequency	
Up to ½ hr	7.4
½ hr to 1 hr	17.7
1 hr to 1½ hr	26.5
1½ hr to 2 hr	14.7
2 hr to 2½ hr	17.7
More than 2½ hr	16.2
High frequency	
Up to 2½ hr	62.5
2½ hr to 3 hr	23.4
3 hr to 3½ hr	7.8
3½ hr to 4 hr	4.7
4 hr to 4½ hr	1.6
More than 4½ hr	0.0

Note: This table is adapted from Schwarz, Hippler, Deutsch, and Strack (1985, Table 1).  $N = 132$ .

of respondents had to report their TV viewing without a preceding recall task. As expected, respondents who had to recall last week's TV programs, or who had a chance to browse through last week's TV program guide, were not significantly affected by the range of the response alternatives provided to them. In contrast, a significant effect of the response alternatives emerged for respondents who had to report their TV viewing without a prior recall task, replicating the finding reported above.

In a similar vein, we observed in a related study (Schwarz & Bienias, 1990, Experiment 3) that individual differences in the cognitive accessibility of self-related information may moderate the impact of scale range. Previous research in personality psychology indicated that individuals who chronically focus their attention on the self provide more accurate self-reports, presumably because relevant self-knowledge is cognitively more accessible to them (see Wicklund, 1982, for a review). This suggests that these individuals should be less influenced by the range of the response scale provided to them, because they may have better access to relevant episodic information. In line with this assumption, individuals who scored low on the "private self-consciousness" scale developed by Fenigstein, Scheier, and Buss (1975) were significantly affected by the range of response alternatives provided to them, whereas individuals who scored high on this scale were not.

In combination, these findings indicate that the impact of response alternatives depends on the cognitive accessibility of relevant episodic information: The more episodic information is accessible in memory, the less respondents have to rely on a scale-based estimation strategy. From an applied point of view,

this suggests that response alternatives may influence some populations more strongly than others and that their impact should be more pronounced in proxy reports than in self-reports. We shall consider each of these implications in turn.

## Self and Proxy Reports

If the impact of scale range increases as the accessibility of episodic information decreases, response alternatives should be particularly likely to influence the obtained reports when respondents are asked to report about other members of their households, whose behavior may not be well represented in memory (see Sudman, Bickart, Blair, & Menon, chap. 16, this volume, for a more detailed discussion). At least two experiments support this assumption.

For example, in one of our studies (Schwarz & Bienias, 1990, Experiment 1), we asked American college students to report their own TV viewing, that of a close friend, and that of a "typical undergraduate" of their university along high- or low-frequency response scales. As predicted, the impact of scale range was most pronounced for reports on the behavior of a typical undergraduate and least pronounced for self-reports, indicating that respondents are less aware of the behavior of others than of their own behavior. Similarly, we observed in a related study (Schwarz & Bienias, 1990, Experiment 2) that proxy reports of beer consumption were more affected by the range of response alternatives than were self-reports of beer consumption.

In addition, these findings bear on an alternative explanation of scale-range effects that has been put forth by Bradburn and Danis (1984), who suggested that the impact of response alternatives may be due to respondents' self-presentation concerns. According to this account, respondents in the above studies may have correctly recalled their TV or beer consumption, but they may have hesitated to report it once they noticed that this would entail checking an extreme response category, which presumably would indicate an unusual behavior. If so, the observed effects would occur at the "editing" rather than at the "judgment" stage of the question-answering process (cf. Strack & Martin, 1987).

Note, however, that the self-presentation account offered by Bradburn and Danis (1984) implies that the range of the response alternatives should affect self-reports *more* than reports about the behavior of others, because respondents are presumably more concerned about the image that they present of themselves than the image of their peer undergraduates. The opposite was the case in both of the above studies. Accordingly, the obtained pattern of results suggests that the impact of scale range reflects respondents' estimation strategy rather than their concerns about self-presentation and social desirability. In line with this conclusion, we observed in another experiment (Schwarz & Bienias, 1990, Experiment 3) that respondents' personality disposition to worry about the image that they presented of themselves, as assessed by Fenigstein et al.'s (1975) "public self-consciousness scale," was unrelated to the impact of response alternatives, whereas individual differences in the cognitive accessibility of self-related information did moderate the impact of scale range, as discussed above.

## Differential Effects for Different Subpopulations

That the impact of response alternatives is particularly pronounced when relevant episodic information is not available has another important implication for survey practice. Specifically, it suggests that individuals for whom the behavior is important, and hence is given a lot of attention, should be less likely to be influenced by response alternatives than should individuals for whom that behavior is of less interest. If so, the use of response alternatives may sometimes obscure differences between subpopulations and may thus undermine one of the key uses of survey data. A 1991 study on the frequency of psychosomatic symptoms illustrates this problem.

In this study, we (Schwarz, Kupfer, and Scheuring) asked patients of different psychosomatic clinics to report the frequency with which they experienced a variety of symptoms, along scales that presented either high- or low-frequency response alternatives. According to the clinical diagnoses, about half of the patients suffered migraine headaches, whereas the remaining patients were treated for other complaints. As expected, patients who were *not* diagnosed as having migraine headaches reported higher frequencies of migraine-related symptoms on the high-frequency than on the low-frequency range scale, replicating our previous findings. Patients who were diagnosed as having migraine headaches, however, were *not* significantly influenced by the frequency range of the scale presented to them. Apparently, the symptoms that brought them into therapy in the first place were so well represented in memory that these patients did not need to rely on the scale to determine an answer.

Depending on the specific values provided on the response scale, this differential susceptibility of different groups of respondents may sometimes obscure subpopulation differences that would otherwise be observable. In the present example, reporting migraine-related symptoms on a high-frequency scale raised the frequency estimates of patients who were *not* diagnosed as having migraine headaches to a level that was not significantly lower than the level reported by migraine patients, obscuring the differences that would be expected on the basis of the clinical diagnosis.

As these preliminary findings illustrate, the use of precoded response alternatives may occasionally obscure differences between subpopulations. This reflects the fact that the range of the response alternatives is more likely to influence respondents for whom the respective behavior is less well represented in memory, resulting in differential scale effects for different groups of respondents.

## Understanding the Question

So far, we have seen that the range of the response alternatives presented to respondents as part of a behavioral frequency question may influence the obtained estimates. In addition, however, the range of the response alternatives may influence respondents' interpretation of what the question refers to in the

first place. This is particularly likely when the behavior under study is ill defined and open to interpretation, as is frequently the case when researchers are interested in subjective experiences.

Suppose, for example, that respondents are asked to indicate how frequently they were “really irritated” recently. Before the respondent can give an answer, he or she must decide what the researcher means by “really irritated.” Does this refer to major irritations, such as fights with one’s spouse, or does it refer to minor irritations, such as having to wait for service in a restaurant? If he or she has no opportunity to ask the interviewer for clarification, or if a well-trained interviewer responds, “Whatever you feel is really irritating,” the respondent might pick up some pertinent information from the questionnaire. One such piece of information might again be the frequency range provided by the scale.

For example, respondents who are asked to report how often they are irritated on a scale ranging from “several times daily” to “less than once a week” may relate the frequency range of the response alternatives to their general knowledge about the frequency of minor and major annoyances. Assuming that major annoyances are unlikely to occur “several times a day,” they may consider the target of the question to be instances of irritation that are less severe than are those considered by respondents who are presented a scale ranging from “several times a year” to “less than once every three months.” Experimental data support this assumption (Schwarz, Strack, Müller, & Chassein, 1988). Respondents who reported their experiences on the former scale subsequently reported less extreme examples of annoying experiences than did respondents who were given the latter scale. Thus, the types of annoying experiences that respondents reported were determined by the frequency range of the response alternatives in combination with the respondents’ general knowledge rather than by the wording of the question per se.

Accordingly, the range of the response alternatives may not only affect respondents’ frequency estimates but also influence their understanding of the question, thus determining which behaviors or experiences they attempt to recall.

## Conclusion

In combination, the findings reviewed here indicate that quantitative autobiographical reports of behavioral frequencies are strongly influenced by the range of the response alternatives provided to respondents. Most importantly, the range of the response alternatives may influence respondents’ understanding of the question and may provide a frame of reference that respondents use in computing a frequency estimate. The use of this frame of reference results in estimates of higher frequencies if the report is to be given on a scale with high-rather than low-frequency response alternatives, and its impact is more pronounced the less episodic information is accessible in memory.

From a theoretical perspective, the reported findings indicate that answers to quantitative “factual” questions are to a considerable extent *theory driven*. To

date, research on theory-driven reconstructions of autobiographical reports has primarily focused on individuals' subjective theories about stability and change in personal attributes (for a review, see M. Ross, 1989). In contrast, the subjective theory that is at the heart of the response-range effects discussed above pertains to respondents' assumptions about the ways in which researchers construct meaningful scales. Both lines of research indicate that a judgmental approach to retrospective reports that emphasizes the role of inference and estimation strategies may prove to be an important supplement to current research on autobiographical memory.

*Acknowledgments.* The reported research was supported by grant SWF0044-6 from the Bundesminister für Forschung und Technologie of the Federal Republic of Germany to Norbert Schwarz, by the Institut für Demoskopie Allensbach, and by the Cognition and Survey Research Program of the Zentrum für Umfragen, Methoden und Analysen.



# **Part III**

## **Event Dating and Time Estimation**



# 13

## Telescoping and Temporal Memory

*Norman M. Bradburn, Janelle Huttenlocher,  
and Larry Hedges*

Many survey questions ask respondents to report their behavior during some time period. These questions typically ask about the incidence or frequency of events that occurred during a particular period of time. Most often the question will specify a starting date (e.g., "Since January 1, 1990, . . .") or a time period (e.g., "During the past 12 months, . . .") that ends with the present. Sometimes, however, the reference period may begin and end before the actual time of reporting (e.g., "During the 2-week period beginning Sept. 2 and ending Sept. 16, how often did you . . .?"). Less often, respondents may be asked questions that call for specific dates (e.g., "When were you interviewed?").

More than a quarter of a century ago, Neter and Waksberg (1964) reported a puzzling phenomenon. When respondents were asked about events occurring in reference periods of differing lengths (1 month vs. 6 months), the frequency of events reported for the most recent month was greater in the 1-month reference period. Other studies have investigated reports of dates of events within particular time limits and found that more events are reported as occurring in the more recent portions of the reference periods (e.g., Bachman & O'Malley, 1981; Uhlenhuth, Haberman, Balter, & Lipman, 1977). These phenomena suggested that reports of events are moved forward in time—what is called "telescoping" in the survey literature.

Although telescoping has often been reported in the survey literature (Sudman & Bradburn, 1973), there is no generally agreed-on explanation for the phenomenon. Explanation would appear to lie in understanding the psychological processes involved in reporting the times of events. Therefore, we must turn to the psychological literature. In this chapter, we present a model that we have recently developed (Huttenlocher, Hedges, & Bradburn, 1990; Huttenlocher, Hedges, & Prohaska, 1988) and argue that it offers a good explanation for the telescoping phenomenon. An alternative, but similar, model has been developed independently by Rubin and Baddeley (1989).

In order to answer questions about the frequency of events over an interval, events must be associated in memory with some representation of time. If tele-

scoping reflects bias in memory, one might think that there is a lack of correspondence between the subjective experience of time and its objective measurement. Such a conclusion is not logically necessary, since the observed bias in reporting could occur through the reporting process itself rather than as the result of a biased representation in memory. Indeed, we shall argue that objectively measured time is preserved in memory but is only inexactly represented. Simply stated, it is the conditions of reporting in conjunction with the inexactness of the representation that produces the observed biases.

The time of an event may be represented by its place in a commonly shared calendar, such as an exact day, a month, or a year, or in calendar-like units that are defined in other ways, such as seasons (in winter), socially patterned time periods (during the academic year 1989–90), or idiosyncratic reference points (on my birthday). They may also be represented by elapsed time (e.g., days ago).

Everyday experience suggests that accuracy for temporal information declines with time. Baddeley, Lewis, and Nimmo-Smith (1978) have shown that there is a loss of exact information about dates over time. They asked subjects who came to their laboratory to report when they last participated in an experiment. The variance of the reports increased markedly with the length of time since the last visit (the standard deviation increased by 19 days for each 100 days of elapsed time). Increasing uncertainty about exact dates with greater elapsed time is an important element in our model.

## How Retrieval Produces Telescoping

The fundamental phenomenon that we are trying to explain is the tendency to overreport events in a reference period. How does our model explain this tendency? Consider events that occur near, but outside, the boundary of the reference period. If events are represented in memory without bias but with uncertainty, some events will be remembered as falling within the period and inaccurately reported and others will be remembered as falling outside the boundary and (correctly) not reported. Now consider events that occur near, but inside, the boundary. They will have an opposite fate. Some events will be remembered as falling within the period and correctly reported; others will be remembered as falling outside the boundary and incorrectly omitted from the report. Note that in both cases the boundary truncates the distribution of reports, sometimes incorrectly including them and sometimes incorrectly omitting them.

If the degree of uncertainty were the same for both cases, the effects of the temporal misclassifications would counterbalance each other and there would be no net overreporting. As we noted earlier, however, there is a relation between the degree of elapsed time and the degree of uncertainty such that the greater the elapsed time between the event and the report, the greater the uncertainty about the exact date of the event, although still the estimates are unbiased. This greater uncertainty manifests itself in a greater variance in reported dates for events that are more remote. Because the variance increases as one moves further back from

the boundary and declines as one moves forward from the boundary, an imbalance in reports will accumulate overreporting.

To see this effect more clearly, consider the following cases: Suppose that you were asking respondents about doctor visits in the past 30 days. Assume that visits are evenly distributed over days and that the memories for the dates of those visits are normally distributed around the actual date. Consider first those who visited a doctor exactly 30 days ago. In our model, one-half of those respondents would report a doctor visit in the period and one-half would not. Now take the case of a day on either side of the boundary. For those who visited the doctor 31 days ago, let us assume a mean report of 31 days and a standard deviation of 11 days; for those with visits 29 days ago, assume a mean report of 29 days and a standard deviation of 9 days. For those whose true value lies at 31 days, 46% ( $z = .09$ ) of the distribution will lie within the period, and this proportion will be incorrectly reported as having occurred within the period. This proportion of reports that is incorrectly put into the period is offset by 45% ( $z = .11$ ) of the distribution of those whose true date was 29 days ago but incorrectly reported as not having occurred within the period, thus leading to a 1% overreporting of visits.

The reference period is also bounded at the other end, usually by the present. There will also be some uncertainty about recent events, but reports of events that are remembered as having occurred more recently than they did will be truncated by the boundary of the present, since remembered events cannot have occurred in the future. Although the present puts a dramatic boundary on reports, it is not the impossibility of the past occurring in the future that is important but rather the fact that the uncertainty attached to more recent events is less than that attached to the older events. Because the variance of the estimates for dates of the recent events is so much less than that for the older events, there will be less backward telescoping than forward telescoping. Thus, the net effect of uncertainty on reports will be forward telescoping.

## Magnitude of Bias Effects

It is possible to compute the magnitude of forward-telescoping effects on the number of reports. It is convenient here to denote the dates as integers, for example, the number of days before the end of the reporting period or the number of days ago. Denote the reference period as the interval from date "a" to date "b." Therefore a subject reports an event if they recall it as having occurred on or between dates a and b. Let  $R_{(ab)}$  be the total number of reports and let  $R_i$  be the number of reports of events that occurred on actual date  $i$ . Thus,

$$R_{(ab)} = R_1 + R_2 + \dots = \sum_{i=1}^{\infty} R_i.$$

In practice, many of the terms of this sum will be zero, since events that

occur on dates very far outside the reporting period will have essentially zero probability of being reported in it.

The number  $R_i$  of reports on day  $i$  is the product of the number of occurrences on day  $i$  (denote  $O_i$ ) and the probability  $p_i$  that an event occurring on day  $i$  is reported; thus,

$$R_i = O_i p_i.$$

We want to know the *number* of occurrences in the reporting period. This is

$$O_{(ab)} = O_a + O_{a+1} + \dots + O_b = \sum_{i=a}^b O_i.$$

It is often reasonable to assume that the number of occurrences (or rate of occurrence) is the same across dates. Denoting the number of occurrences each day as  $O_\bullet$ , the number of occurrences in the reporting period becomes

$$O_{(ab)} = (a - b) O_\bullet,$$

and the number of reports  $R_i$  of events that occurred on day  $i$  becomes

$$R_i = O_\bullet p_i.$$

The ratio  $R_{(ab)}/O_{(ab)}$  of total number of reports to the actual number of occurrences in the reporting period gives an index of the amount of overreporting. Under our assumptions, this ratio is independent of  $O_\bullet$ :

$$\frac{R_{(ab)}}{O_{(ab)}} = \frac{\sum_{i=1}^{\infty} p_i}{(b - a)}.$$

Thus, the proportion of overreporting (or underreporting) depends only on the probabilities of reporting events that occurred on each day and on the length of the reporting period.

In our truncation model, the probability of reporting an event is determined by the probability that an unbiased, but normally distributed, memory for the event falls into the reporting period. Let  $X_i$  be a random variable corresponding to memories for events that actually occurred on day  $i$ ; then

$$X_i \approx N(\mu_i, \sigma_i^2).$$

The mean  $\mu_i$  is the actual day  $i$ , which corresponds to the fact that  $X_i$  is unbiased. The variance  $\sigma_i^2$  depends on the actual day. Several authors (e.g., Huttenlocher et al., 1988) have found that the standard deviation  $\sigma_i$  increases linearly with  $i$ :

$$\sigma_i = \alpha + \beta i.$$

Thus, the probability  $p_i$  of reporting an event that actually occurs on day  $i$  is

$$p_i = \Phi\left(\frac{b - \mu_i}{\alpha + \beta i}\right) - \Phi\left(\frac{a - \mu_i}{\alpha + \beta i}\right)$$

when the most recent date  $a$  in the reporting period is not the present and is

$$p_i = \Phi \frac{(b - \mu_i)}{(\alpha + \beta i)}$$

when  $a$  is the present and therefore it is impossible for  $X_i$  to be less than  $a$ . Here  $\Phi(x)$  is the standard normal cumulative distribution function.

An example with  $a = 1$ ,  $b = 5$ , and  $\sigma_i = 0 + .5i$  illustrates the magnitude of the effects. Let 0 to 1 be the present day, 1 to 2 be Day 1, 2 to 3 be Day 2, etc. Table 13.1 gives the details of the computations. Setting the mean for Day 1 at 1.5, for Day 2 at 2.5, etc., and computing the probability of reports for actual days during the month including the reference period,  $i = 1$  to 50, yields

$$\frac{R_{(ab)}}{O_{(ab)}} = \frac{5.27}{4.0} = 1.32,$$

and hence there would be a 32% overreporting of events in the reference period.

This argument applies whether or not the boundary is at the present. If the reference period both begins and ends in the past, it is possible for respondents to recall events as having occurred more recently than the near boundary and thus to not report them. In such a situation, events occurring in the reference period would be somewhat further back in time than if the boundary were the present and would have a somewhat higher degree of uncertainty attached to them. In such cases, the model would predict that there would be a greater degree of backward telescoping than when the boundary was the present, although there would still be net forward telescoping. We do not know of any studies that have tested this hypothesis.

What empirical evidence is there that respondents actually forward telescope reports of dates for remembered events? A study, reported in Huttenlocher, Hedges, and Prohaska (1988), provides confirmation of the model. In that study, a sample of 115 students at the University of Chicago was asked how many movies they had seen at the University. Approximately half (58) were asked using an 8-month (academic year) reference period; the remainder (57) were asked using a 2-month (Spring Quarter) reference period. Respondents were then asked to name the movies that they had seen and to report the dates on which they had seen them. A list of movies and their dates shown was available so that the reported dates could be checked for accuracy.

These data support findings from earlier studies. Forward telescoping was significant in the 2-month group and for the Winter and Spring Quarter in the 8-month group. (The data were collected at the end of the Spring Quarter; the Fall Quarter data also showed forward telescoping but were not statistically significant.) Respondents in the 2-month group reported more movies seen in the Spring Quarter than did respondents in the 8-month group. Contrary to Neter and Waksberg's (1964) assumption about similar findings in their study, examination of the actual dates indicated that the higher reporting for the 2-month group was not due to forward telescoping of movies from earlier periods but was due to reporting more movies actually seen in the shorter reference period.

TABLE 13.1. Probability of Reporting an Event That Actually Occurs on Day  $i$ 

Date $i$	$\Phi\left(\frac{b - \mu_1}{\alpha + \beta_1}\right)$	$\Phi\left(\frac{a - \mu_1}{\alpha + \beta_1}\right)$	$P_i$
1	1.000	0.841	0.159
2	0.999	0.309	0.691
3	0.952	0.159	0.792
4	0.773	0.106	0.668
5	0.579	0.081	0.499
6	0.434	0.067	0.367
7	0.334	0.058	0.276
8	0.266	0.052	0.214
9	0.218	0.048	0.171
10	0.184	0.045	0.140
11	0.159	0.042	0.117
12	0.139	0.040	0.099
13	0.124	0.038	0.086
14	0.112	0.037	0.075
15	0.103	0.036	0.067
16	0.095	0.035	0.060
17	0.088	0.034	0.054
18	0.082	0.033	0.049
19	0.078	0.033	0.045
20	0.074	0.032	0.041
21	0.070	0.032	0.038
22	0.067	0.031	0.036
23	0.064	0.031	0.033
24	0.062	0.030	0.031
25	0.060	0.030	0.029
26	0.056	0.030	0.028
27	0.057	0.030	0.026
28	0.054	0.029	0.025
29	0.053	0.029	0.024
30	0.051	0.029	0.023
31	0.050	0.029	0.022
32	0.049	0.028	0.021
33	0.048	0.028	0.020
34	0.047	0.028	0.019
35	0.046	0.028	0.018
36	0.045	0.028	0.018
37	0.044	0.028	0.017
38	0.044	0.027	0.016
39	0.043	0.027	0.016
40	0.043	0.027	0.015
41	0.042	0.027	0.015
42	0.041	0.027	0.014

(Table 13.1 continued)

TABLE 13.1. Continued

Date $i$	$\Phi\left(\frac{b - \mu_1}{\alpha + \beta_1}\right)$	$\Phi\left(\frac{a - \mu_1}{\alpha + \beta_1}\right)$	$P_i$
43	0.041	0.027	0.014
44	0.040	0.027	0.013
45	0.040	0.027	0.013
46	0.039	0.027	0.013
47	0.039	0.026	0.012
48	0.038	0.026	0.012
49	0.038	0.026	0.012
50	0.038	0.026	0.011
Total			5.27

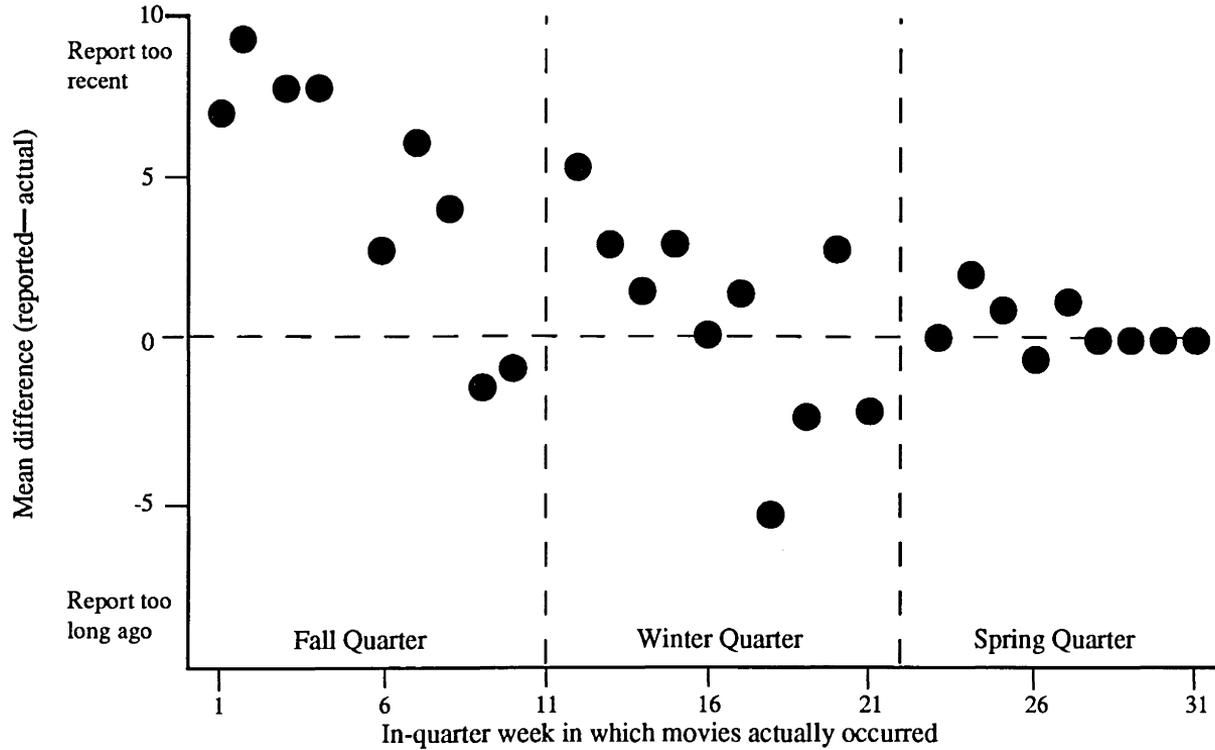
Examination of the data for each quarter (see Figure 13.1) confirms the model's prediction of forward telescoping at the furthest boundary and backward telescoping at the nearest boundary.

C. P. Thompson, Skowronski, and Lee (1988b) have directly tested the hypothesis that uncertainty of memory for events leads to forward telescoping. In a series of experiments in which undergraduates kept daily diaries of unique events, the authors found that the magnitude of telescoping was affected by the length of time since the event happened: More remote events were forward telescoped, more recent events were backward telescoped, although this tendency was not reliable. Forward telescoping was found both for personally experienced events and for events that the respondents had only heard about.

Thompson et al. (1988b) also measured respondents' degree of confidence in their memories for the dates of the events. They found both forward and backward telescoping for events about which respondents reported being uncertain about when they occurred but no significant telescoping for events about which respondents felt that they could confidently recall the dates. These findings are consistent with our model.

## "Bounded Recall" and the Reduction of Telescoping

Neter and Waksberg (1964) used a technique, which they called "bounded recall," that reduced forward telescoping. The technique requires asking the same questions on successive interviews with the same individual. Respondents are asked about their behavior in a particular reference period. In the second and subsequent interviews, they are asked to report their behavior during the preceding reference period, but they are reminded of the events that they gave in the previous interview. Although the first interview ("bounding interview") suffers from considerable forward telescoping, subsequent interviews do not.



Adapted from Huttenlocher, Hedges, and Prohaska (1988).

FIGURE 13.1. Plot of the Mean Differences (Reported In-Quarter Weeks Minus Actual In-Quarter Weeks) by In-Quarter Week in Which the Movies Actually Occurred

How does our model account for this effect? In the second interview, respondents are reminded of their responses in the first interview. The reminder acts to reduce the uncertainty about when some events happened. If respondents remember an event that took place in the previous period but think that it might have happened in the current period (and thus, absent the reminder, would tend to forward telescope it), they will have a check on their accuracy and will learn that they have already reported it. If they remember events that took place in the present period but recall them as having occurred in the previous period, again they will be reminded that they have not reported them and will check the accuracy of the date. Thus, the misplacing of events either forward or backward is discouraged by reducing the uncertainty.

## Reports of Specific Events

So far we have discussed questions that deal with reporting the number of events within a reference period. What happens when the question is one of recall of a specific event without any specified time period? Consider the question, "When did you last visit a doctor?" The task here is similar to reporting all events within a reference period but requires only finding the most recent of the events and reporting its date. In fact, one might view answering this question as merely one step in the process undertaken to answer the question on total events during a reference period.

The same elements obtain here as we previously discussed. We assume that the date is represented unbiased, but inexactly, in memory and that the greater the elapsed time, the greater the uncertainty. However, in this case there is no stated beginning boundary, although there is an ending boundary created by the present. In such cases, we might expect that there would be some backward telescoping caused by the ending boundary but no forward telescoping because there is no stated beginning boundary to truncate the reports.

There are two considerations, however, that contravene this expectation. The first is that there may be an implicit beginning boundary even though no boundary was explicitly stated. Such implicit boundaries arise from the reasonableness or unreasonableness of the question if it is interpreted without any boundaries. Thus, in answering the question in a validation interview, "When were you interviewed?" respondents might reasonably put a limit on the time period believing that no one would be asking about such a minor event if it occurred a very long time ago. If respondents do put such a self-imposed boundary on the question and the event actually occurred beyond that boundary, they will report it as being at the boundary rather than beyond.

The second factor at work is more subtle and involves the way in which respondents handle the increasing uncertainty that comes with greater elapsed time. Respondents seldom give an exact time (to the day) in response to a question about when something happened. Instead, they give an approximation that becomes less fine grained the greater the elapsed time. When things are recent,

they respond in terms of “days ago”; when they are somewhat longer ago, they respond in terms of “weeks ago” or some prototypical number of days such as “10 days ago” or “45 days ago.” At even longer time periods, they move to “months ago” or “years ago.” As memory becomes less certain, respondents round their answers to prototypical values, based on either the calendar (days, weeks, months) or the decimal system (5, 10, 15, 20, etc.). Sometimes the two principles coincide, such as “30 days” for 1 month or “60 days” for 2 months.

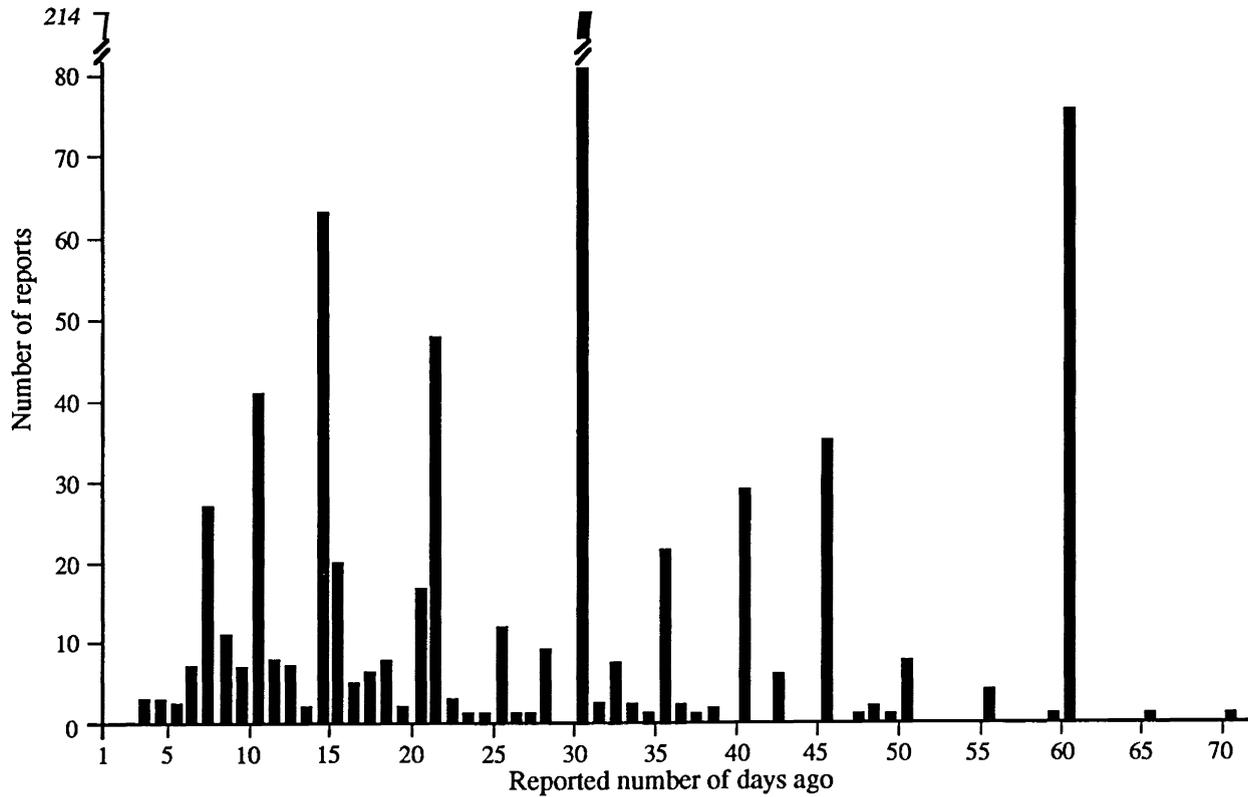
If the distance between adjacent rounded values increases with elapsed time and if reports are framed in terms of days, the distance in reported time given in days will also increase. In this case, forward telescoping occurs even if respondents generate their reports from an underlying distribution with uncertainty that is distributed symmetrically around the true date of the event, that is, in cases where there is no boundary to truncate the distribution.

To see the reason for this effect, consider the prototypic values 7, 10, 14, 21, 30, and 60 days. Suppose respondents choose the prototypic value nearest to their remembered value, and the reporting interval for each prototypic value is bounded by the midpoints between adjacent rounded values. Assignment of the rounded response 14 begins at remembered values of 12 (midway between 10 and 14) and ends at remembered values of 17.5 (midway between 14 and 21). Note that the distance between 12 and 14 is smaller than that between 14 and 17.5. Because the range of remembered values to which a particular rounded value is assigned is smaller in a forward than in a backward direction, a smaller range of values is rounded up (e.g., to 14) than rounded down. This asymmetry in the size of the rounding distance produces a net forward bias in reporting. By the same argument, forward bias due to rounding in terms of the decimal system also might occur with increase in elapsed time if people first round by 5s, then by 10s, and so on.

Our recent study (Huttenlocher et al., 1990) supports this model as an explanation for forward telescoping of reports for single events. That study used the validation procedures for the General Social Survey (GSS) to reinterview respondents about when they had been originally interviewed. In order to extend the range of elapsed time, the sample was enlarged and respondents were reinterviewed from 4 to more than 60 days after the original interview. A total of 629 respondents were reinterviewed.

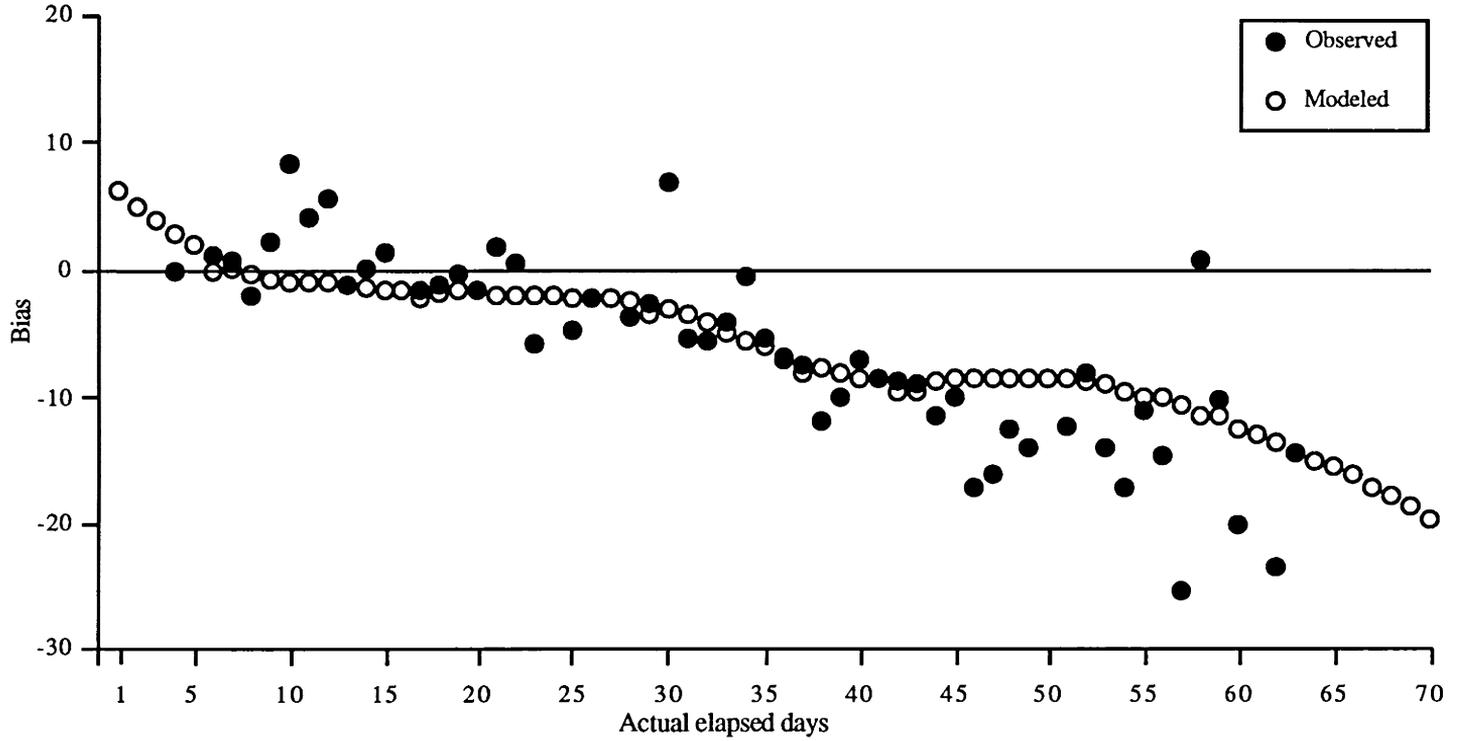
The data indicate that respondents were putting an upper bound of 60 days on their reports. Only 3 reports were more than 60 days even though there were 37 cases where the true number of elapsed days was greater than 60 days. All of these were reported as occurring less than 60 days ago.

The data also clearly show that respondents are rounding to prototypic values (see Figure 13.2). As with previous studies, there is increasing uncertainty with greater elapsed time, which manifests itself not only in the increasing variance of reports but also in increased use of rounded values. As a result of increasing distance between the adjacent rounded values, there is pronounced forward telescoping in the reported values (see Figure 13.3). Thus, it is apparent that



Adapted from Huttenlocher, Hedges, and Bradburn (1990).

FIGURE 13.2. Frequency of Each Reported Number of Days



Adapted from Huttenlocher, Hedges, and Bradburn (1990).

FIGURE 13.3. Observed and Modeled Bias in Reporting for Each Actual Number of Elapsed Days

even in the absence of explicit boundaries, forward telescoping will occur as a result of rounding procedures.

## Conclusions

In this chapter, we have presented an explanation for overreporting due to telescoping based on three hypotheses about the representation of events in temporal memory: (a) memory for the time of events is unbiased but inexact; (b) with greater elapsed time, there is greater uncertainty about dates; and (c) with greater uncertainty, temporal reports will be rounded to prototypical units. We have argued that these three elements of temporal representation cause telescoping, both forward and backward, when respondents are asked to report the occurrence of events within a bounded reference period.

Our explanation hypothesizes that the representation of the time of events is unbiased in memory and that the observed biasing (telescoping) comes from the recall process—owing partly to uncertainty and partly to rounding effects. It is possible, however, that there are other factors involved, such as motivational effects. Sudman and Bradburn (1973) hypothesized that telescoping might be, in part, due to the respondents' desire to perform the task required of them. When in doubt, respondents prefer to give too much information rather than too little. Although this hypothesis cannot be ruled out, the fit of the data in this experiment to our model is close enough that we do not feel that one needs an added assumption of motivational bias.

The explanation put forth in this chapter gives theoretical support to the two practices that are commonly suggested to reduce telescoping: making the reference period as short as possible and using bounded-recall techniques to reduce uncertainty about the dates of events. Further, the explanation focuses attention on one of the principal causes of telescoping: the increasing uncertainty about the date of events as time increases. Thus, techniques in addition to bounded recall would help reduce telescoping. An example of such a technique would be eliciting landmark dates well known to the respondent, such as a birthday, significant personal event, or public holiday, and then getting the respondent to relate the events being reported on to those landmark dates.

The techniques for reducing telescoping tend to be expensive, and many times researchers will not be able to use them. If one can establish empirically the parameters of the model described above, however, it may be possible to estimate the magnitude of the telescoping and then adjust the data for telescoping. Further work in refining this model and developing ways to estimate the parameters is needed.



# The Impact of Differing Memory Domains on Event-Dating Processes in Self and Proxy Reports

*John J. Skowronski, Andrew L. Betz, Charles P. Thompson,  
W. Richard Walker, and Laura Shannon*

The subjective impression that many of us have when we attempt to provide accurate dates for autobiographical events is that it is a very difficult task. This personal subjective experience is corroborated by our observations of respondents in event-dating studies. These respondents often spontaneously express a great amount of dismay (sometimes quite vehemently) when they are asked to provide exact dates for their autobiographical events.

Given that the respondents in studies of event dating often have many potential advantages over individuals who date events in a truly naturalistic context, this negative response to the event-dating task might be a bit surprising; that is, since the methodologies used in event-dating studies provide advantages to respondents that they might not possess when trying to date events in the “real world,” event dating in these studies should be relatively easy. For example, in many dating studies (e.g., Mingay, Shevell, Bradburn, & Ramirez, chap. 15, this volume), respondents are asked to record autobiographical events daily and submit these records, which are later used for dating. An event’s date can obviously become more salient in the course of recording the event, potentially leading to better date recall. Respondents are sometimes asked to turn in their records of events weekly instead of daily (C. P. Thompson, Skowronski, & Lee, 1988b). If subjects review the week’s events (as we have seen them do), this weekly submission technique could facilitate recall of event orders. These recalled-event orders could serve to enhance dating accuracy. Further, subjects are sometimes asked to date only “unique” events instead of the “truly commonplace” events (C. P. Thompson, 1985a). That the recorded events are even slightly out of the ordinary suggests the possibility that subjects might pay more attention to the unique events than to nonunique events and, hence, would provide more accurate dates for the unique events.

Even in the face of all these advantages, the objective accuracy data from these event-dating studies provide an indication of why people often express great dismay when they are asked to date events: They just do not seem to be very good at it. For example, in an event-dating study in which people tried to

date autobiographical events from the most recent 4 months of their lives, people occasionally misdated an event by as much as 3 months (C. P. Thompson, 1982). Further, substantial uncertainty concerning the exact date of an event is a relatively frequent occurrence: Respondents in a recent study reported that over 20% of their dating attempts for autobiographical events were pure guesses (C. P. Thompson, Skowronski, & Lee, 1988a).

Much of the recent work looking at the accuracy of event dating investigates the impact of this considerable uncertainty surrounding event dates. For example, the magnitudes of forward-telescoping and time-expansion biases (the tendencies to date events as younger or older than they really are, respectively) have been directly linked to the degree of uncertainty surrounding an event's exact date (Bradburn, Huttenlocher, & Hedges, chap. 13, this volume; Huttenlocher, Hedges, & Prohaska, 1988; Thompson et al., 1988b). However, these studies have usually not investigated the underlying psychological causes of this uncertainty or the situational and psychological factors that might lead to enhanced dating accuracy.

A few studies suggest that the uncertainty accompanying event dating, and the frequent poor dating performances exhibited by event daters, are a partial function of poor temporal memory. Respondents' self-reports of how they date events indicate that direct recall of an event's exact date is relatively infrequent. For example, in a recent study on retrospective dating for a 10-week period, respondents reported direct recall of an event's date for only 9% of the events in which they had participated and for only 4% of the events that they had encountered secondhand (Thompson et al. 1988a; also see Brewer, 1988).

## Memory and Event Dating

### The Possible Roles of Multiple Memory Domains

If people cannot usually recall directly the date of an event, then how are most event dates produced? Event-dating research suggests that various memory domains may play a significant role in the dating process. For example, data from one of our studies (Thompson et al., 1988a) indicate that event daters will typically access a number of different memory content domains when trying to estimate an event's date.

One of these domains is the ease with which a memory is retrieved. Bradburn, Rips, and Shevell (1987) have argued that the accessibility of a memory contributes to date estimates, with easily remembered events being dated as more recent than the exact dates and with difficult-to-recall events being dated as older than the exact dates. Data from Bradburn et al.'s (1987) studies, which involve the use of major world events as the dating targets, indicate that this accessibility bias will, indeed, sometimes occur.

Although a dater might not be able to recall an exact event date, recall of other, less precise forms of temporal information could allow a narrowing of possible event dates (e.g., "I remember that registration for school was right at

the beginning of April this year, instead of its usual spot at the end of March”).

Another useful memory domain might be the recalled details of an event. A dater might use such details to place the event in a general time period either by directly relating that information to another recalled event (e.g., “I remember that there was snow on the ground, and I remember that we only got snow this year at the beginning of the term”) or by using the recalled-event details in conjunction with general world knowledge (“I remember that there was snow on the ground, and it usually snows only in winter”).

An event dater may sometimes access prototypic or abstracted information, such as information about an event type or event class, in the reconstruction process. For example, in the process of dating an argument with a Child Psychology class instructor, a dater may recall that his or her Child Psychology class met only on Mondays. Hence, the dater may conclude that the argument must have happened on a Monday. The difference between this process and the direct-recall process is a subtle, but theoretically important, one. The idea that partial temporal information is directly recalled implies that temporal information is encoded in the event memory trace or takes the form of a temporal tag that is attached to that trace. The idea that temporal information comes from prototypic or abstracted memories suggests that the temporal information used to date an event is not stored with the event but is stored elsewhere in memory and is applied to the event during judgment.

Finally, a dater might sometimes use recalled-event relations to date events, narrowing the possible range of event dates by recalling whether an event occurred before or after another event with a known date. This strategy can be used either in conjunction with direct recall, such as when respondents can remember the exact event order (“It happened right after that Psychology midterm that I flunked”), or in conjunction with respondents’ general world knowledge, such as when the logical relations between the events strongly suggest an event order (“The review session must have come before the exam”).

## Empirical Implications

The fact that people can use many different memory domains in the process of date production suggests several interesting hypotheses about event dating and about the nature of the dating errors that people should make. Two of us (Skowronski and Thompson), along with a number of associates, have recently conducted a number of studies that illustrate the involvement of many different memory domains in the event-dating process and that differentiate the roles that these domains play in event dating.

### *Event Memory and Dating Accuracy: Evidence for the Impact of Multiple Memory Domains*

As we indicated earlier, one of the obvious factors that should affect dating accuracy is event recall. Assuming that one cannot directly recall the date of an event, the more details that one can recall about the event, the more accurately

one should be able to estimate the event's date. Several studies support this hypothesis. One of us (Thompson, 1985b) found that events that respondents thought they would recall well were dated more accurately than events that the respondents did not think they would remember well. In a more direct test (Thompson et al., 1988b), we found that telescoping bias was reduced (i.e., dating was more accurate) for well-remembered than for poorly remembered events. Further, for well-remembered events, the telescoping bias was nearly eliminated when the dates that subjects said were "pure guesses" (i.e., no temporal information was available anywhere in memory) were removed from the analyzed data set. Similar data (Thompson et al., 1988a) showed a link between respondents' dating strategies and dating accuracy; that is, when respondents said that they recalled the details of an event, they were substantially more accurate in their date estimate than when they said that their date estimate was just a pure guess. Finally, data provided by Brewer (1988) suggest that event memory is sometimes all that people have: People are much more likely to recall non-temporal than temporal details about that event.

In our recent study (Skowronski, Betz, Thompson, & Shannon, 1991), we attempted to substantiate further this link between event memory and event dating in several ways. In this study, we asked respondents to enter events in a diary. The diary sheets were collected weekly. At a later date, we read the events back to subjects (with some slight editing and shortening, as required) in a random order and asked them to date each event after it was read. As a dating aid, subjects were provided with a calendar that was blank except for the usual calendar information (day of week, month, date).

Central to our attempts to verify the memory-dating link in this study were several ratings that respondents provided for the events. One of these ratings assessed event memory. During the diary test, respondents were asked to rate how well they thought that they remembered each event. This metamemory measure is not a standard laboratory assessment of recall, but it does often produce comparable results (see Brewer, 1988). Using regression techniques, we treated these memory ratings as a predictor of dating accuracy. Obviously, a potential memory-dating link would be indicated if the memory ratings significantly predicted dating accuracy. It was indeed the case that subjects' memory ratings were a strong predictor of dating accuracy, semipartial  $r = -.187$ ,  $F(1,6005) = 260.90$ ,  $p < .001$ .

A second way in which the memory-dating link was assessed was to look at the impact that variables known to affect event recall have on the accuracy of event dating. If the recalled details of events are used to enhance dating accuracy, then one would expect the same factors that affect event recall also to affect dating accuracy.

To investigate this idea, two additional event ratings were collected. Respondents were asked to rate the pleasantness of each event (whether the event was pleasant or unpleasant for the subject) and the person-typicality of each event (whether the event was typical or atypical for the subject). Based on the results of past research, we expected subjects to remember (a) positive events better

than negative events; (b) pleasantness-extreme events (i.e., events that were thought to be either extremely pleasant or extremely unpleasant) better than pleasantness-neutral events (neither pleasant nor unpleasant); (c) person-atypical events better than person-typical events; and (d) both person-atypical and person-typical events better than typicality-neutral events. We obtained exactly these findings when we examined subjects' ratings of how well they remembered their autobiographical events. (However, there was also a higher-order interaction—see the subsection on Dating the Events of Others, below).

If event dating is partially dependent on event memory, one would also expect to see these same effects in the dating-accuracy scores. Multiple regression analyses indicated that these effects did occur: Pleasant events were dated more accurately than unpleasant events, semipartial  $r = -.053$ ,  $F(1,6006) = 20.51$ ,  $p < .001$ ; pleasantness-extreme events were dated more accurately than pleasantness-neutral events, semipartial  $r = -.028$ , quadratic  $F(1,6023) = 5.78$ ,  $p < .02$ ; person-atypical events were dated more accurately than person-typical events, semipartial  $r = .067$ ,  $F(1,6006) = 32.20$ ,  $p < .001$ ; and both person-typical and person-atypical events were dated more accurately than typicality-neutral events, semipartial  $r = -.031$ , quadratic  $F(1,6010) = 6.92$ ,  $p < .01$ .

The results of these analyses indicate that the recall of an event's details may play a significant role in accurately dating that event. However, we wondered if there was any possibility that some of these effects were mediated primarily (or solely) by event memory, while other effects were mediated by other memory domains. To investigate this idea, we reran the regression equations with the event-memory rating as a covariate in the models. Presumably, the effects that depend highly on event memory would be eliminated by the addition of the memory rating to the model; the effects that involve other memory domains would remain.

The results of the further analyses indicated that both the quadratic pleasantness and the quadratic typicality effects were eliminated by the inclusion of the memory term in the regression models. The elimination of the quadratic pleasantness effect is particularly surprising. We would have suspected both extremely pleasant and extremely unpleasant events to have a higher probability than non-extreme events of being temporally tagged or exactly dated in memory: It seems more likely that one would recall the date of a car wreck than the date on which one dropped a glass of milk on the kitchen floor. Hence, we would have thought that this quadratic pleasantness effect should have been somewhat resistant to the inclusion of memory in the regression models. Instead, the elimination of this effect by inclusion of the memory term in the models suggests that the heightened dating accuracy for pleasantness-extreme events might be solely a function of enhanced estimation accuracy due to heightened event recall and might not at all involve enhanced temporal tagging of extreme events relative to nonextreme events.

In contrast to the quadratic effects, the person-typicality and pleasantness variables were resistant to the inclusion of the memorability ratings in the model, semipartial  $r_s = .041$  and  $-.041$ ,  $F_s(1,6005) = 12.94$  and  $12.92$ ,  $p_s < .001$ ,

respectively. Even when event memory was accounted for, pleasant events were dated more accurately than unpleasant events, and person-atypical events were dated more accurately than person-typical events. Obviously, these outcomes suggest that respondents are relying on memory domains other than event memory in their event dating, and it seems likely that temporal tagging is one of these memory domains.

This certainly makes sense with respect to the effects of person-typicality and person-atypicality on event dating; that is, it is intuitively reasonable that person-atypical events are more likely to be temporally tagged than person-typical events: It seems more likely that one will have a temporal tag for flunking an exam (an unusual occurrence, one hopes, for the target) than for routinely passing an exam.

That pleasant events are dated more accurately than unpleasant events might seem a bit less intuitively obvious. As with person-typicality, one might argue that surprising events are more likely than unsurprising events to receive temporal tags. Because negative events tend to be more surprising than positive events, the surprisingness argument might lead one to expect greater dating accuracy for negative than for positive events.

There are two replies to this argument, one data based and one statistical. The data-based argument is that a positivity effect in event dating has appeared before (Thompson, 1985a). The statistical argument is that since both pleasantness and person-typicality were included in the same multiple regression model, the effect of pleasantness was independent of the effect of person-typicality. If dating accuracy for both factors was partly driven by event surprisingness, increases in dating accuracy due to the surprisingness of events (whether due to temporal tagging or to other reasons) might not be reflected in the results of the multiple regression analysis. The common variance due to event "surprisingness" might have been partialled out, a procedure that could have had a greater statistical impact on negative than on positive events, producing a positivity effect in dating.

Although this statistical argument is reasonable for one of our studies (Skowronski et al., 1991), it cannot account for the Thompson (1985a) data. Since typicality ratings were not collected in the latter study and the data were analyzed using ANOVA techniques, the selective partialing argument cannot be applied to the results. Thus, it seems reasonable to consider other explanations for the positivity effect in event dating.

One might argue that a positivity effect in event dating might be produced by temporal tagging processes: Positive events might be more likely than negative events to receive temporal tags. This is sensible when one considers the types of events that are frequently dated and are frequently positive, events such as birthdays, special events (e.g., concerts), anniversaries, and holidays (Thanksgiving, Christmas). These types of events are sometimes listed in respondents' diaries. Thus, it may be that the positivity effect in event dating simply reflects the tendency for many of our personal landmark events to be both positive and associated with a definite calendar date.

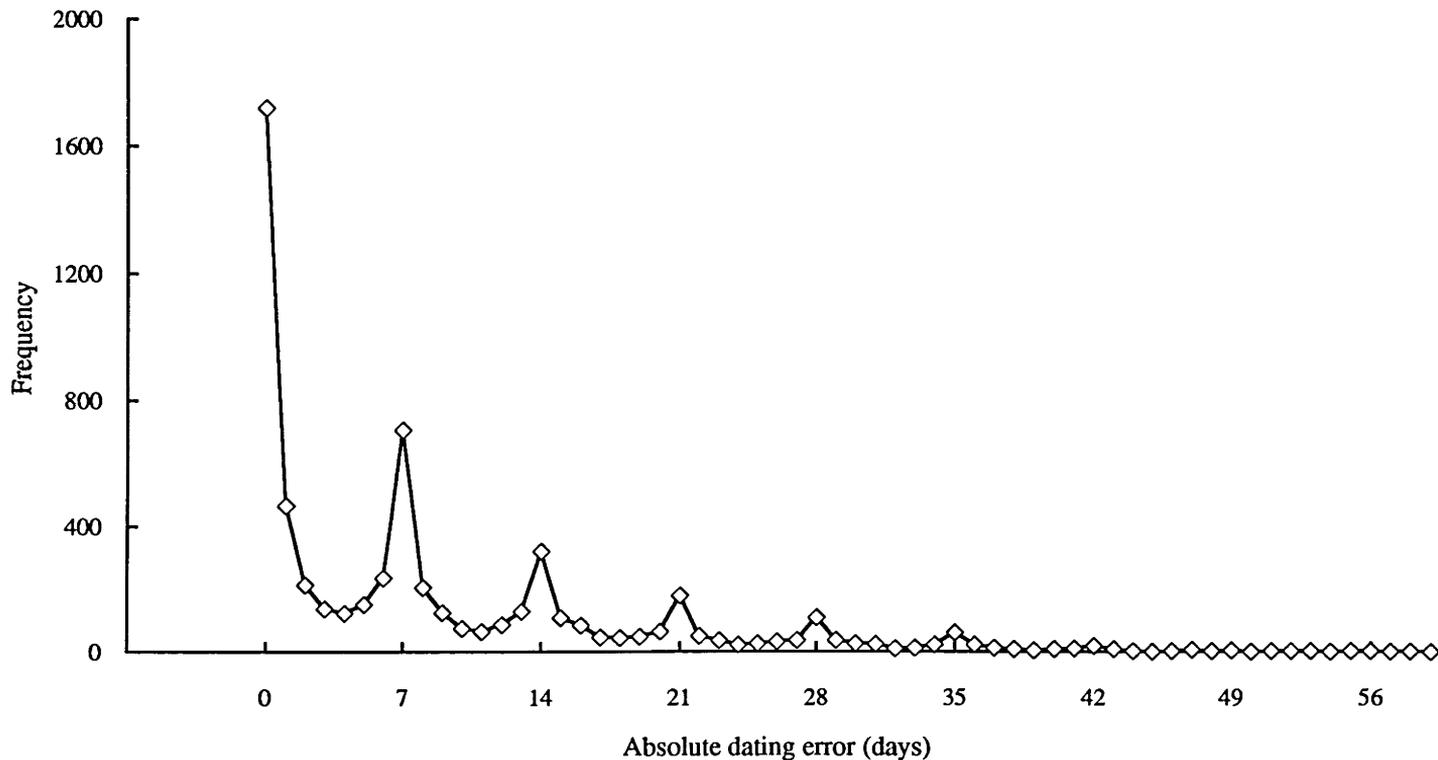
Although this “positive landmarks” hypothesis may contribute to the positivity effect in dating, it cannot fully account for that effect. The Thompson (1985a) data indicated that the effect of pleasantness persisted even when the identifiable positive landmark events were removed from the data set. In a follow-up analysis (Skowronski et al., 1991), we reached a similar conclusion. In that analysis, an independent coder judged the “exact date likelihood” of an event: the probability (high, medium, low) that an event would be stored with an exact date. This date-likelihood rating was then entered as an additional predictor in the regression model. Judged likelihood was significantly related to dating error. The coder was able to predict which events would be accurately dated—mean days dating error by judgment category were as follows: high probability = 2.81, moderate probability = 5.72, low probability = 9.45,  $F(2,6004) = 48.46$ ,  $p < .001$ . If the pleasantness effect were solely due to the fact that positive events were more often exactly dated landmark events than were negative events, the addition of this likelihood factor to the regression model should have eliminated the pleasantness effect. It did not: semipartial  $r = -.037$ ,  $F(1,6004) = 9.96$ ,  $p < .01$ . Hence, although the “positive landmark” explanation for the positivity effect in dating seems sensible, two studies now suggest that some other memory domain may also be partially responsible for this positivity effect.

Collectively, these memory and dating results suggest two conclusions. First, event memory does play a role in event dating. Well-remembered events are also well-dated events. Second, memory domains other than event memory also play a role in event dating. Although we have speculated that two of these other memory domains are the tendency to assign exact dates to events and the tendency to tag events with partial temporal information, the analyses performed above do not directly implicate these domains. Additional data from our study (Skowronski et al., 1991), as well as data from other studies, provide more direct evidence for the impact that these additional memory domains have on event dating.

### *Error Patterning: Seven Is a Magic Number*

Consider the proposition that people will often recall or reconstruct only partial temporal information about an event. One form of temporal information that we suspect is frequently recalled or that may be relatively easy to reconstruct is the day of the week on which an event occurred. If this is the case, then daters should make a relatively large number of “multiple-of-seven” errors (7 days, 14 days, 21 days, etc.) relative to “non-multiple-of-seven” errors; that is, subjects may be able to recall or reconstruct the day of the week on which an event occurred, but they may sometimes misremember the week in which that event occurred.

This is exactly what we observed (Skowronski et al., 1991). As Figure 14.1 clearly shows, respondents exhibited a very strong tendency to make multiple-of-seven errors: Subjects often were correct about the day on which an event occurred but misremembered the exact week. In addition to this basic effect, it is also interesting to note that error frequency declined with error magnitude, even for



Based on data from Skowronski, Betz, Thompson, and Shannon (1991).

FIGURE 14.1. Seven Is a Magic Number: The Frequency with Which Subjects Made Dating Errors at Differing Error Values

the multiple-of-seven errors. This systematic decrease in error frequency with increasing error magnitude for the multiple-of-seven errors suggests that information other than the day of the week must be involved in the date estimation process—otherwise, one would be just as likely to make a 7-day error as a 28-day error (after all, a Monday is a Monday . . .). Hence, even while pointing to the fact that one sometimes uses recall or reconstruction of the day of the week on which an event occurred to help date an event, these data also strongly reinforce the notion that multiple memory domains are used in event dating.

One other interesting aspect of the data presented in Figure 14.1 is that they show no evidence of the prototypic responses (or rounding errors) that have been described by Huttenlocher, Hedges, and Bradburn (1990). If these prototypic responses had been provided by our respondents, then the frequency plot should have exhibited peaks at non-multiple-of-seven error values. In particular, often-used estimation values, such as 5-day or 10-day errors, should have emerged with heightened frequency.

The absence of such rounding errors in our data might not be very mysterious but instead might be explained by one important methodological difference between our study (Skowronski et al., 1991) and the Huttenlocher et al. (1990) study. In the Huttenlocher et al. research paradigm, respondents were asked to estimate *how long ago* an event occurred. In our research paradigm, subjects were presented with a calendar and were asked to attempt to *provide the exact date* for an event.

These two task types might cause respondents to engage in memory searches of differing depth. In the Huttenlocher et al. paradigm, subjects might search memory just enough to provide a reasonably accurate dating value—a value that might often be expressed in terms of a rounded prototypic response. The task in our study may implicitly demand more accuracy from respondents, so that the respondents might engage in a more thorough search of memory before providing a response.

This line of reasoning suggests some interesting studies designed to investigate further these differing psychological processes. First, because the exact-date methodology is presumed to involve a more detailed memory search than the age-estimation methodology, respondents should take longer to provide date estimates when they are attempting to estimate an exact date for an event than when they are estimating the age of an event. Note that this prediction would not hold for event dates that are directly recalled; it should apply only for dates that are estimated—reconstructed from imperfect temporal information. It is only under these estimation conditions that respondents should be induced into a detailed search of many different memory domains.

A second study suggested by this reasoning might involve measuring the number of different strategies that respondents use. This should not be very difficult. In the Thompson et al. (1988a) study, we asked respondents to list all of the different strategies that they used to provide an event date, and respondents frequently reported using more than one strategy. The prediction is straightforward: Respondents who are attempting to provide an exact date for an event

should report that they use more strategies than should respondents who estimate the event's age (again, this should hold only for estimated dates, not for directly recalled dates). In addition, the strategies themselves might change. Respondents in age-estimation conditions might be more likely than respondents in exact-date conditions to resort to phenomenologically easier strategies, such as an event's accessibility.

There is also one important possible practical implication of this methodological difference for survey researchers who are interested in increasing the accuracy of date estimates obtained from respondents. If respondents engage in a more thorough memory search when looking for an exact date as opposed to an event's age, then they should generally be more accurate in exact-dating conditions than in age-estimation conditions.

### *Individual Differences in Estimation Accuracy: The Possible Role of Differential Encoding of Temporal Information*

One of the salient facts that we noted early in our studies of dating accuracy was that there were relatively large individual differences in average dating accuracy among individuals. For example, when dating events from the most recent 10 to 15 weeks of their lives, it has not been unusual in our studies for some respondents to exhibit an average dating error of only 2 or 3 days; whereas other respondents exhibit an average dating error of 12 or 13 days (e.g., see Thompson, 1982).

Given such a large spread in average dating accuracy, it is reasonable to wonder about the factors that differentiate good daters from poor daters. Quite by chance, when two of us (Thompson and Skowronski) were manually sorting daters from one of our data sets into "good" and "poor" groups for a median split analysis, we noticed that females were disproportionately represented in the "good" group and that males were disproportionately represented in the "poor" group. Statistical analysis of this difference indicated that it was significant, and reanalyses of the results of earlier dating studies indicated that this between-gender differential accuracy effect persisted across studies.

In reporting these findings (Skowronski & Thompson, 1990), we speculated that this between-gender difference is caused by male/female differences in the degree to which people attend to and encode temporal information. More specifically, we speculated that in our culture, females are expected to be the "keepers" of event dates. As the stereotype goes, it is the insensitive husband/father who always forgets the dates of the kids' birthdays, whereas the sensitive and caring mother/wife never forgets.

In our recent study (Skowronski et al., 1991), we attempted to test this general explanation and also to provide more specific information about the exact memory mechanisms underlying this between-gender effect. At least four psychological possibilities exist. First, females' superiority in event dating could be due to the fact that females are more likely than males to encode the exact date of an event in memory. This suggests that when forced to estimate event dates

(i.e., no direct date recall), females' estimates should be just as inaccurate as males' estimates.

A second possibility is that females have a more developed temporal schema than males; that is, females' heightened attention to and use of event dates might allow them to develop abstracted date-related knowledge that is useful in event dating. This idea suggests that even when forced to estimate the date on which an event occurred, females should be more accurate than males.

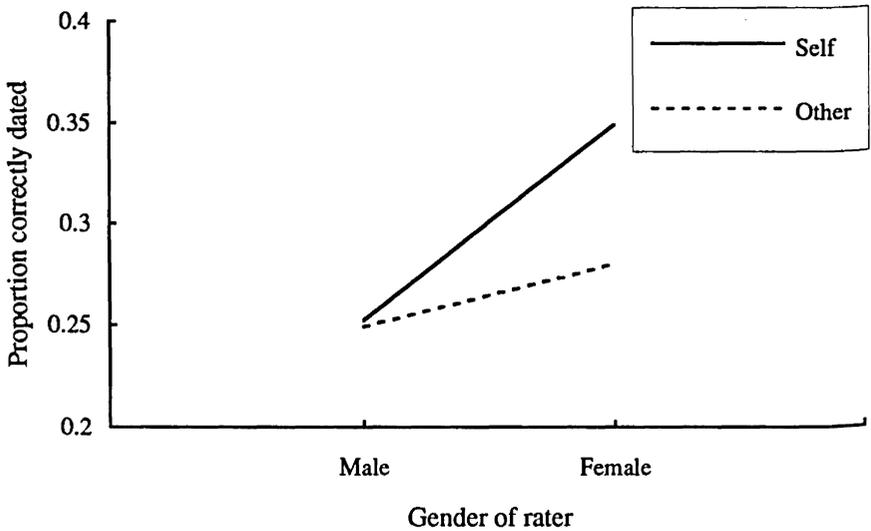
A third possibility is that females are more likely to encode partial temporal information for events. In other words, females might be more likely to attach temporal tags, such as the day of the week on which an event occurred, to events. These temporal tags would later be useful in reconstructing event dates, contributing to heightened estimation accuracy, but would also cause a heightened proportion of multiple-of-seven errors for females relative to the dating estimates provided by males.

A fourth possibility focuses on event memory as the mechanism causing the between-gender effect. Females could simply recall event details better than could males. Because the details of events can give clues to the events' dates, those who recall events well (i.e., females) should also date them well. This again suggests that when forced to estimate the date on which an event occurred, females should provide more accurate dates than should males.

There was one additional twist to this study that is worthy of note. In this study, each respondent kept a diary for both his or herself and for one other person. We reasoned that the dates provided for events from this other diary could also help to shed light on the memory processes involved in the between-gender accuracy effect. If females generally possessed a more well developed temporal schema that enabled them to estimate event dates better than did males, then this schema should be helpful in dating both self-events and events of others. However, if females' heightened dating accuracy was restricted only to self-events, the implication is that it is information derived from memory for the specific personal events, not a general abstracted schema, that is responsible for females' heightened accuracy.

The data from this study replicated the results of prior research: Females (mean days error = 8.43) tended to provide more accurate date estimates than did males (mean days error = 9.65),  $F(1,60) = 3.53$ ,  $p < .07$ . In addition, the data suggested that females' dating superiority was due to direct recall of event dates. This is, perhaps, most strongly supported by the finding that females more frequently reported the exact dates of events than did males, particularly when those events were self-relevant (see Figure 14.2),  $F(1,60) = 4.14$ ,  $p < .05$ .

The other three explanations for the between-gender accuracy difference received little support. When the exactly dated events were removed from the data, so that the accuracy scores were calculated only from erroneously dated events, females' estimated dates were no more accurate than males'. In addition, females did not make significantly more multiple-of-seven errors than did males, in either absolute or proportional terms. Finally, males' and females' reported memories for the events did not differ. These three null effects suggest



Based on data from Skowronski, Betz, Thompson, and Shannon (1991).

FIGURE 14.2. Females Are More Frequently Exact for Self-Events: The Diary Type (Self vs. Other)  $\times$  Gender Interaction for the Proportion of Dates Provided

that females' estimation strategies, and the memories used as input into those strategies, were no different from the males' estimation strategies and produced estimates of accuracy equivalent to those of the males.

The fact that females were more often exactly correct than males, but primarily for self-relevant events, is an important limitation on the female accuracy effect. This limitation led us to consider other conditions under which the female dating-accuracy advantage could be nullified or perhaps even reversed. In particular, we were reminded of the early gender research in social psychology that indicated that females were more persuadable than males, a finding that was modified by later results indicating that persuadability was much more dependent on females' and males' domains of expertise than on any inherent persuadability differences between the genders (for a review, see Deaux, 1984).

We conducted an exploratory content analysis on the events listed in the diaries with the idea of looking for similar expertise-like effects in event dating; that is, we looked for evidence suggesting that heightened female accuracy was limited to only some event domains and for evidence suggesting that males could be superior daters in other event domains.

To investigate these possibilities, we first classified the events into 73 separate categories that were developed from an initial pass through the events listed in the diaries. We then separately calculated for males and females the average dating error of the events in each category. This was purely an exploratory analysis: The mean dating error for some of these categories must be viewed as highly suspect because of the relatively small number of entries in these categories.

However, despite this limitation, the analysis yielded interesting and informative results. Female superiority in dating accuracy was fairly general: Females provided more accurate estimates than did males for 42 out of the 66 categories (63.6%) for which both genders provided estimates, a proportion that is significantly greater than chance ( $z = 2.11, p < .05$ ). Further, the reversals (categories in which males provided more accurate estimates than females) were generally not dramatic or large in magnitude. Yet, some of these reversals seem intuitively sensible: Males were more accurate than females in estimating the dates of activities involving cars, home improvements, and rest/recreation. The analysis suggests the possibility that males might sometimes be more accurate daters than females and that an investigation of dating accuracy that more precisely controlled the event content domain might be a worthwhile endeavor.

We must confess that many people are more fascinated by the gender aspect of these data than we are. We think that these between-gender effects are interesting because of the possibilities that are revealed for the causes of individual differences in event dating and because we think that these data provide enlightenment concerning the role of memory in event dating. However, we are also interested in gender effects for another reason: We wonder whether other researchers would obtain similar gender effects in other autobiographical memory paradigms. We are concerned that the gender effect might somehow be bound to the particular diary methodology that we employ.

We raised this concern in Skowronski and Thompson (1990), where we noted that females tended to write longer diary entries than did males and found that an ANCOVA on dating error using average entry length as the covariate eliminated the between-gender effect. Although our later study (Skowronski et al., 1991) found that a similar analysis did not account for the between-gender difference in exact-dating frequency, we would be greatly comforted if a similar between-gender effect emerged in other studies of the dating of personal autobiographical events.

### *Dating the Events of Others: Assessing the Utility of Proxy Reports of Event Dates*

As indicated in our discussion of gender effects in event dating, we (like other researchers—for example, see Mingay, Shevell, Bradburn, & Ramirez, chap. 15, this volume) have periodically attempted to assess the ability of our respondents to date the events of others who have a reasonably close relationship to the dater. This type of proxy event dating would seem to be intermediate between the dating of “public” events (such as the explosion of the space shuttle *Challenger*) and the dating of personal autobiographical events (such as when one last had chicken cordon bleu).

This question of proxy-report validity is important from both a practical and a theoretical perspective. From a practical perspective, survey researchers must often rely on proxy reports as the only available source of event information. For example, a mother might be the only reliable source of information on when a

child was last vaccinated. From a theoretical perspective, proxy reporting provides an interesting arena in which to test the relationship between the various memory domains and dating accuracy.

Although it is obvious that self-event dating will generally be more accurate than other-event dating, the reason is not at all clear. The most straightforward hypothesis is that other-memory will be generally worse than self-memory along all important dimensions; that is, for others' events, event memory will be worse, temporal tagging will be less frequent, and exact-date tagging will be less frequent than for self-events. This reasoning implies that there should be no interactions between the self–other variable and other variables of interest in a study.

However, there might be reason to suspect self–other differences in memory. For example, we may recall positive better than negative information when that information is relevant to the self, whereas we may recall negative better than positive information when that information is relevant to another (for a detailed discussion of affect–memory relationship, see Banaji & Hardin, chap. 5, this volume). Similarly, we may recall person-typical information better when that information is other-relevant but recall person-atypical information better when that information is self-relevant. If event memory is used to date events, then the presence of these interactions in the memory data would suggest the presence of similar interactions in the dating-accuracy data.

In the Skowronski et al. (1991) study, we investigated these possibilities. Multiple regression analyses on the memory data indicated that, unsurprisingly, self-event recall ( $M=4.76$ ) was superior to other-event recall ( $M=4.15$ ),  $F(1,6006)=234.94$ ,  $p<.001$ , and that the recall of others' events was less affected by extremely pleasant or unpleasant events (i.e., produced a flatter curve) than recall of self-events, semipartial  $r=-.023$ , diary target  $\times$  quadratic pleasantness  $F(1,6021)=6.95$ ,  $p<.01$ . There was also a three-way interaction between the self–other diary variable, person-typicality, and event pleasantness, semipartial  $r=-.024$ ,  $F(1,6002)=5.10$ ,  $p<.03$ . Interpretation of this interaction is difficult, but it seems to reflect the fact that positive events were recalled better than negative ones only when the events were self-events and neutral in person-typicality. Otherwise (i.e., for others' events and for very typical and atypical self-events), pleasantness had no effect on recall.

Again, if event-dating accuracy were largely determined by event memory, these same patterns should have emerged in the dating-accuracy data. They did not. Dating accuracy for self-events (mean days error = 8.45) was better than dating accuracy for others' events (mean days error = 9.62),  $F(1,6006)=22.21$ ,  $p<.001$ ; but the self–other variable did not interact with any other variables in the regression analyses. Instead, the same patterns that characterized the dating accuracy for self-events (extreme items more accurately dated, atypical items more accurately dated, positive items more accurately dated) similarly characterized the dating accuracy for others' events. This suggests that the self–other effect in event recall did not differentially affect the basic memory domains that are used in event dating. In other words, proxy daters do not seem to be particu-

larly biased in ways that are any different from self-event daters; the proxy daters just seem to have less information to use in recalling or reconstructing dates.

## Concluding Thoughts

The major thrusts of this chapter were to discuss the hypothesis that many different memory domains are involved in event dating, to understand the impact of these memory domains on dating accuracy, and to note conditions that affect the applicability of these memory domains to event dating. Certainly, a greater understanding of memory issues is a key to understanding these effects, and significant work on autobiographical memory and autobiographical memory biases continues (see Brewer, chap. 1, Ross & Buehler, chap. 4, and Banaji & Hardin, chap. 5, this volume).

However, it is obvious that memory is only one of the components that we must understand in order to comprehend dating error. Respondents may use many different estimation strategies in the course of concocting event dates; perhaps our understanding of event dating would benefit from analyses of estimation techniques in a manner similar to those used by Menon (chap. 10, this volume) to study frequency judgments. More generally, we need a greater general understanding of the cognitive processes involved in date estimation.

Regardless of the psychological causes, the data presented in several of the chapters in this book suggest that event dating is a domain where error is the norm rather the exception. This suggests that we should look with great skepticism on the event dates that often emerge in conversations with others, especially when those events are not particularly special. This also suggests that if we want to remember event dates ourselves, we should rely on an age-old technique: Record those dates so that we can go back and look at them later.



# **Part IV**

## **Comparisons of Self and Proxy Reports**



## Self and Proxy Reports of Everyday Events

*David J. Mingay, Steven K. Shevell, Norman M. Bradburn, and Carl Ramirez*

Respondents to surveys are often asked to report not only information that they acquired through direct personal experience but also information that they acquired indirectly. Much of this information pertains to other household members, with the respondent serving as a proxy reporter. For example, in the National Health Interview Survey (NHIS), a married woman in the sample household will often be asked detailed health-related questions about her husband and children.

The primary advantage of proxy reporting is that since one person can report for absent household members, fewer callbacks are needed, thus reducing costs and the length of field work. Although numerous studies have attempted to determine whether this advantage is offset by greater reporting error by proxies (e.g., Mathiowetz & Groves, 1985; Mosely & Wolinsky, 1986), the evidence is equivocal.

In a review article, Jeffrey C. Moore (1988) points out that the majority of studies comparing self and proxy reporting have conceptual and methodological shortcomings that make their findings questionable. In particular, validation data have rarely been available. The common assumption that a higher level of reporting indicates greater accuracy may not be valid. For example, reporting the date on which an event occurred as being more recent than it was (forward telescoping) is common and constitutes a bias toward overreporting (Sudman & Bradburn, 1974). In addition, higher reports for themselves than for other household members may reflect true incidence differences between self-respondents and proxies, caused by biases in the selection of the self-respondent (Mathiowetz & Groves, 1985). For example, the self-respondent may have been at home and thus available for interview during the day owing to illness; such a respondent may well have more health-related events than those members of the household who are not at home (Mosely & Wolinsky, 1986).

One of the few studies that obtained validation of reports (Cobb, Thompson, Rosenbaum, Warren, & Merchant, 1956) examined the prevalence of arthritis

and rheumatism based on self-reports, proxy reports, and a medical examination and found that similar point estimates were provided by self and proxies but that proxies showed higher response error variance. A similar result pertaining to heart disease was reported by D. J. Thompson and Tauber (1957). However, as J. C. Moore (1988) points out, the findings from both studies need to be treated with caution because numerous persons refused the medical examination and some of those who had the examination and were diagnosed with one of the health conditions may have been unaware that they had the condition.

The experiment reported in this chapter investigated the relative accuracy of self and proxy respondents who were college roommates. Given the frequency with which information about others is acquired indirectly, proxy reporting should be of interest to psychologists who study learning and memory in an everyday setting.

Despite this, little psychological research has been conducted on self/proxy reporting. Charles P. Thompson and his colleagues have conducted two studies using college students who were roommates (C. P. Thompson, 1982; Skowronski, Betz, Thompson, & Shannon, 1991; Skowronski, Betz, Thompson, Walker, & Shannon, chap. 14, this volume). However, several aspects of the experiments reduce their relevance to survey research. For example, subjects were asked about any nonrecurring autobiographical events that the diary keeper had chosen to record. In addition, subjects were not tested on their recall of the events. Instead, they were read their descriptions of the events; they then rated how well they remembered the events (a type of recognition task) and they provided dates.

The primary purpose of the study described in this chapter was to compare the relative accuracy of self and proxy reporting. Thus, it was crucial that the records used to validate subjects' reports were of high validity and reliability. Most studies of everyday memory have used diaries that subjects only mailed in once a week. Debriefings indicate that subjects often do not complete the diary until just prior to mailing it (C. P. Thompson, personal communication). Consequently, it is likely that certain events—especially those that are low in salience—will be forgotten and thus not recorded in the diary.

To reduce the likelihood of such forgetting, respondents were required to return the completed diaries on a daily basis. In addition, to ensure that they did not just glance over the list of activities and neglect to report certain ones that they had done, respondents were required to respond to every activity by circling a word, either "yes" or "no." Several other procedures intended to improve the quality of the data collected in the daily diaries are discussed below.

For each activity that respondents reported doing, they indicated the number of times that they had done the activity that day. Diaries were kept for 5 weeks, after which the respondent and roommate returned to the laboratory and were asked to report how often the respondent had engaged in each of these 20 activities over the 5-week period. There are a number of reasons for expecting that roommate proxies would perform less well than respondents on this task:

1. Proxies may not have learned about one or more of the occasions on

which their roommate had done the activity. In the absence of this information, proxies may assume that the respondent has only engaged in the activity on those occasions that they know about and therefore they may underreport the frequency. Alternatively, proxies may judge that they will not have heard of every occasion on which the respondent engaged in the activity and may use some heuristic to make an estimate of frequency that is greater than the number of occasions that they know about.

2. The proxy may have received incomplete information about the activity. For example, in the morning the respondent may have told the roommate that he or she was planning to go to a bookstore that day but subsequently failed to indicate whether the visit was actually made.
3. There are a number of reasons why the memory representation of the activity is likely to be more impoverished for the roommate proxy than for the respondent. The activities will typically be less consequential for the proxy than for the respondent. Whether observing the activity or hearing about it, the proxy is likely to pay less attention to it than is the respondent and is less likely to process the activity in a way that will lead to good recall (e.g., semantically or “deeply”; Craik & Lockhart, 1972). Typically, the respondent is likely to encode the activity itself better than the roommate is likely to encode the description of that activity—for example, because the activity itself is likely to last longer than the description of it and be presented in several modalities (e.g., visual, tactile, and auditory). In addition, the daily act of noting on a diary sheet the frequency of engaging in the activity may increase the memorability of the activity for the respondent. As a result of these and other factors, the respondent is likely to form a more detailed representation of each occasion on which the activity is performed than is the roommate proxy, and thus the respondent will be more likely to recall the occasion subsequently.

A recent study by Schwarz and Bienias (1990) provided some evidence that information about others is relatively impoverished. Students reported on TV consumption using either a high-frequency or a low-frequency response scale. The effects of the response scale on reporting were greater when students answered for others (a close friend and a typical undergraduate at their university) than for themselves. The authors suggest that subjects are more influenced by the scale range when reporting for others because there is a smaller amount of relevant information easily accessible in memory.

The chapters by Skowronski et al. and by Sudman, Bickart, Blair, and Menon (chaps. 14 and 16, this volume) also discuss the differences between the task of reporting for oneself and for someone else and present pertinent data. The Sudman et al. chapter adopts a social-cognitive perspective and provides evidence that some of the principles of social memory apply to reporting of proxies' events while others do not. The different factors associated with reporting by respondents and by proxies somewhat parallel the distinction between autobiographical events and events reported in the mass media (see Larsen & Plunkett,

1987, for an interesting discussion of the differences between these types of events).

A possible exception to the superiority of self to proxy reporting may occur with activities that are quite strongly socially desirable or socially undesirable (but see Mathiowetz & Groves, 1985). This is because respondents may distort their answers to present themselves in a favorable light. None of the activities asked about in this study appears to be as strongly socially desirable or undesirable for this population as the topics usually studied in research on this subject.

The frequency of an activity might also influence the relative accuracy of self and proxy reporting. As the frequency of an item increases, subjects change from using a strategy of recalling and counting the individual episodes of an event to using a heuristic-based frequency estimation strategy (e.g., Blair & Burton, 1986; Bradburn, Rips, & Shevell, 1987). It seems likely that proxies will often have to resort to using estimation strategies even for low-frequency activities, since they will have neither observed the activities nor been told about them (cf. Sudman et al., chap. 16, this volume). Thus, there may be an interaction between self/proxy performance and activity frequency, with proxies' reports being considerably poorer than respondents' reports only for low-frequency activities.

A greater use of estimation strategies might also be demonstrated by more use of numerically rounded numbers, in a manner analogous to the rounding of reports of elapsed time (Huttenlocher, Hedges, & Bradburn, 1990).

## Overview of Experiment

To investigate the reporting accuracy of self and proxy respondents, 20 everyday activities were selected for study. Over a 5-week period, 30 respondents kept daily diaries in which they recorded the frequency of doing these activities. They were tested the day following the last diary-keeping day. Respondents indicated the frequency of engaging in these activities during the 5 weeks and answered other questions about these activities. Roommates, who had not been told in advance that they would be asked to participate in this study, answered the same questions about the respondents.

## Method

### Subjects

Respondents consisted of 30 first-year undergraduate students at the University of Chicago who lived in a dormitory and had a roommate. They replied to a flyer inviting students to volunteer for a study of everyday activities. Proxies consisted of the roommates of the respondents.

## Materials and Procedure

Twenty quite diverse activities were selected. Three criteria were used to select the activities: (a) they could be unambiguously described in a few words, (b) they were not highly personal, and (c) they were unlikely to be engaged in several times in a single day. Candidate events were administered to 18 subjects in a pilot study. After reporting the frequencies with which they had engaged in each activity, subjects were debriefed to determine whether they had misunderstood the instructions or any of the activity descriptions. The 20 activities that best met the above-mentioned criteria were selected for use in the study.

At the start of the study, both respondents and their roommates came to the laboratory and were taken to different rooms. Respondents were instructed in the diary-keeping task and given practice in completing the daily diary. The importance of completing each diary and mailing it in no later than the following day was stressed. They were given 35 daily diary sheets, each with a different date starting with the following day and finishing 35 days later, and 35 stamped envelopes addressed to the laboratory. Each diary sheet listed the 20 activities in a different random order. Respondents were paid \$10 for their initial laboratory visit and were told that after the diary-keeping period they would be paid \$1 for each daily diary sheet that was received, a \$2 bonus for each week that they sent all of the diary sheets in, a \$15 bonus for returning all 35 diary sheets, and \$10 for a final laboratory session after the 5 weeks. Respondents were told only that this was a study of everyday activities and were not informed that their memories for the diary activities would be tested.

At the initial session, respondents' roommates completed a number of questionnaires that were unrelated to this study and were paid \$10. They were told that they had been invited to come to the laboratory along with their roommates merely for convenience and that they were participating in a different study. They were not told that they would be asked to return to the laboratory. The primary purpose of asking roommates to come to the laboratory was to screen out those likely to refuse to return 5 weeks later. If the respondent or roommate failed to attend this initial session at the scheduled time, neither was used in the study.

Respondents kept the daily diaries, starting with the day that followed their initial visit to the laboratory, for 35 consecutive days. At the end of each day, or early the following day, they indicated on the appropriate diary sheet whether they had done each activity that day by circling either the word "yes" or the word "no." For each activity done, they wrote a number to indicate how many times they had done the activity that day and wrote a letter from A to E to indicate how likely they would be to remember having done the activity after a week, with A indicating that they definitely would remember and E indicating that they definitely would not remember (respondents doing the activity more than once on a particular day gave an average memorability rating). Finally, respondents wrote down the date on which they had completed the sheet, and

then they mailed it back to the experimenter in one of the envelopes provided. On receiving each diary sheet, the experimenter noted the postmarked date and the date on which it was received. If a diary was late, the experimenter telephoned the respondent to check whether it had been mailed back.

A few days before the end of the 5-week diary-keeping period, the experimenter telephoned respondents to schedule a time for each respondent and roommate to attend the laboratory at the same time on the day following the last day of the diary-keeping period. Respondents and their roommates were told only that they would be administered several questionnaires. At the laboratory, respondents completed a series of self-completion questionnaires concerning the activities. Roommates were administered a series of self-completion questionnaires concerning both the diary-keeper's and their own engagement in these activities. This chapter presents the data that are relevant to a comparison of self and proxy reporting.

The first questionnaire completed by both the respondents and proxies listed the 20 activities, with each person getting a different random order. They were instructed to give a number indicating how often the respondent had done each of the 20 activities during the previous 5 weeks and were reminded that this was since the time that they first came to the laboratory.

Several other questionnaires were administered, including a debriefing questionnaire designed to assist in the interpretation of the results. Finally, the experimenter asked several other debriefing questions.

## Results

Three respondents ceased keeping diary sheets during the 35 days and were dropped from the study. The other 27 respondents came to the final test session, and all but one of their roommates also attended. Thus, the analyses are based on 26 respondent/proxy pairs, with 15 of the pairs being female and 11 male. All 35 daily diary sheets were received from 25 respondents, and 34 daily diary sheets were received from the other two. Two respondents were unable to come to the laboratory on the day following the last diary day. One attended a day later; the other came 2 days later. They were instructed to exclude activities done since the end of the diary-keeping period when answering the questions.

### Reporting of Individual Activities

For each activity, the sum of the frequencies reported by the respondent in the 35 daily diaries was taken as the true frequency against which the respondent's and proxy's frequency reports at the second laboratory session were scored. Table 15.1 shows the activities, ordered by the mean frequency reported in the daily diaries. Two activities have been excluded from the analyses: talking to someone other than parents on the telephone and seeing a movie at the University of Chicago's major film society (DOC). Talking on the telephone was done

TABLE 15.1. Mean Frequencies of Activities as Recorded in Daily Diaries and Reported by Respondents and Proxies (Excluding Activities Never Done)

Frequency Group	Mean 5-Week Frequencies		
	Diary Records	Respondents' Reports	Proxies' Reports
<b>Low-frequency group</b>			
Go to a music store	1.5	2.0	0.8
Go to a grocery store	2.8	2.2	1.7
Make a deposit in bank account	2.9	2.0	2.3
Go on a date	3.3	2.5	1.9
Do laundry	3.4	2.8	2.6
Doze in class	3.8	3.4	3.9
<b>Medium-frequency group</b>			
Write a check	4.1	3.2	2.0
Go to a bookstore	4.5	3.9	2.6
Leave Hyde Park	5.9	4.5	2.6
Go to the Regenstein Library <sup>a</sup>	8.5	9.0	10.0
Write a letter	8.5	4.7	5.4
Drink alcohol	8.9	4.8	4.6
<b>High-frequency group</b>			
Talk to parent on phone	10.7	8.9	6.8
Buy food at restaurant, fast food place or coffee shop	11.8	8.1	7.7
Nap, other than in class	12.6	10.0	16.9
Do a sport or physical activity	15.3	12.7	11.0
Watch TV	19.7	12.3	14.8
Eat evening meal in dorm	23.1	25.1	27.1

<sup>a</sup>The main University of Chicago library.

considerably more often than any of the other activities, and the debriefing of the respondents suggested that some calls may have been forgotten when the daily diary was completed. The diaries indicated that 15 respondents did not see any films at DOC and others saw many films. Since this bimodal distribution causes both conceptual and analytical difficulties, the activity was excluded from the analyses.

The analyses were based on 26 respondents reporting on 18 different activities. Reflecting predicted self/proxy performance differences as a function of frequency, the activities were divided into three frequency groups of equal size.

There were a total of 468 possible data points (18 activities engaged in by 26 respondents). For 47 of these, the diary records indicated that the respondent had not done the activity at all during the 5-week period. The type of error that the respondent can make on such an item, namely, reporting doing the activity in the reference period (a false positive), is quite different from the type of errors that

respondent can make on activities that have been done, namely, reporting not having done the activity (a false negative) or reporting doing the activity but giving an incorrect frequency. Thus, errors on activities not done during the 5 weeks will be reported separately.

Table 15.1 also shows the mean frequencies reported by respondents and proxies. For 14 of the 18 items, respondents' reports were closer to the correct answer than were proxies' reports (binomial probability  $p < .05$ ). The four activities that were reported more accurately by proxies than by respondents were "making a deposit in a bank account," "dozing in class," "writing a letter," and "watching TV." Taken together, these four items do not appear to have particular characteristics that would lead one to expect superior performance by proxies.

Underreporting was more common than overreporting for both respondents and roommate proxies: Respondents' and proxies' mean reported frequencies were lower than the diary records' frequency for 15 ( $p < .01$ ) and 14 activities ( $p < .05$ ), respectively. It is tempting to speculate that the overreporting of visits to the library by respondents might be due to social-desirability bias. However, the fact that proxies overreported by a larger amount than respondents does not support this explanation. In addition, another activity that one might expect to be overreported because of social-desirability bias, namely, that of going on a date, was in fact underreported by respondents.

The overreporting by both respondents and proxies of eating an evening meal in the dormitory may be due to the use of the heuristic of multiplying the number of weeks by 7. This would result in an overestimate, since the dorms do not serve an evening meal on Sunday. However, no ready explanation is apparent for why respondents overreported going to a music store. It would be misleading to place too much reliance on these numbers, however, since the fairly small number of subjects means that one or more outliers can make a dramatic difference.

To identify patterns better, the activities within each of the three frequency groups were collapsed. Not surprisingly, the mean reporting error increased as frequency increased. For the six lowest, middle, and highest frequency activities, the mean reporting errors of frequency were  $-0.50$ ,  $-1.69$  and  $-2.64$ , respectively. The corresponding standard errors were  $0.19$ ,  $0.37$ , and  $0.49$ .

To allow a better comparison of reporting accuracy pertaining to activities of very different frequencies, mean proportional errors were calculated. Table 15.2 shows these data for each of the activities.

Frequency reports by both respondents and proxies were generally quite poor. Occasions of drinking alcohol, writing a letter, and watching television were the activities that were most underreported. Alcohol consumption is underreported in general population surveys, and this appears to be at least partly due to an unwillingness to report excessive alcohol consumption (Bradburn & Sudman, 1979). This may, however, not be the cause of the underreporting in this study: The same social-desirability bias may not operate among undergraduate students, especially males, and number of drinking occasions may well not be as

TABLE 15.2. Mean Proportional Error of Respondents and Proxies (Excluding Activities Never Done)

Frequency Group	Proportional Error		Mean Memorability Rating <sup>a</sup>
	Respondents	Proxies	
Low-frequency group			
Go to a music store	0.33	-0.47	2.38
Go to a grocery store	-0.21	-0.39	2.89
Make a deposit in bank account	-0.31	-0.21	2.70
Go on a date	-0.24	-0.42	1.29
Do laundry	-0.18	-0.24	2.99
Doze in class	-0.11	0.03	3.62
Medium-frequency group			
Write a check	-0.22	-0.51	2.94
Go to a bookstore	-0.13	-0.42	2.94
Leave Hyde Park	-0.24	-0.56	1.30
Go to the Regenstein Library	0.06	0.18	2.99
Write a letter	-0.45	-0.37	2.28
Drink alcohol	-0.46	-0.48	2.35
High-frequency group			
Talk to parent on phone	-0.17	-0.36	2.51
Buy food at restaurant, fast food place or coffee shop	-0.31	-0.35	2.72
Nap, other than in class	-0.21	0.34	3.38
Do a sport or physical activity	-0.17	-0.28	2.76
Watch TV	-0.38	-0.25	3.19
Eat evening meal in dorm	0.09	0.17	3.30

<sup>a</sup>1 = Definitely remember 1 week from now; 2 = Probably remember; 3 = Might or might not remember; 4 = Probably not remember; 5 = Definitely not remember.

sensitive a question as amount drunk. In addition, although it is unlikely that the respondents' roommates would find the question particularly sensitive, they show a similar level of underreporting.

Turning to proxy reporting, activities are likely to differ in terms of the likelihood that students would know that their roommates had engaged in them, although this is complicated by the fact that they could learn of this in one or more of the following ways: by observing the activity, by being told about it by the roommate, or by being told about it by another person. Scrutiny of proxy reporting reveals some counterintuitive results, however. For example, the highest proportional error concerns an activity that the roommate seems likely to get to know about on most occasions, namely, leaving the University neighborhood of Hyde Park. Less surprisingly, the activity showing the second highest proportional error by proxies is writing a check.

## Reporting across Frequency Groups

To examine the relationship between reporting accuracy and activity frequency, and to examine data that are more robust than those for the individual items, the activities were divided into three equally sized groups, labeled “high frequency,” “medium frequency,” and “low frequency.” For respondents, absolute reporting errors were 0.36, 1.47, and 2.62 for the low-, medium-, and high-frequency groups, respectively. The corresponding numbers for proxies were 0.67, 1.79, and 1.66. Figure 15.1 shows the mean frequency reports of respondents and proxies as a proportion of mean actual frequencies for the three frequency groups. As indicated by numerical values of less than one, both respondents and proxies underreported the frequency of the activities in each of the three frequency groups.

Comparing reporting across frequency groups, and between self and proxy reporting, the only statistically significant effects were that proxy reports showed less proportional underreporting for the high-frequency activities than for the medium- and low-frequency activities (Duncan’s Multiple Range Test,  $p < .05$ ). The prediction that the relative performance of proxies compared with respondents would be worse for low-frequency activities was not supported, although there was a nonsignificant trend in this direction. Collapsing across the frequency groups, there was no significant difference in the frequency reports of respondents and proxies as a proportion of mean actual frequencies (means of 0.812 and 0.773, respectively;  $F < 1$ ).

All of the error data presented thus far have been signed error. Consider an activity that was done 10 times by all 26 subjects and suppose that 2 subjects reported doing the activity 10 times, 8 subjects reported doing it 20 times, and 16 subjects reported doing the activity 5 times. The errors would cancel out, and subjects’ performance on this activity would be scored as being accurate. In a survey, this would be appropriate if the statistic of interest was the mean frequency that the activity was engaged in. However, survey data are also used to explore the association between variables. For example, correlations may be computed or multivariate analyses performed. For these more complex analyses, absolute error is important, that is, the mean error with the sign ignored.

Figure 15.2 shows the mean absolute proportional error for the high-, medium-, and low-frequency activities, with a lower number indicating a lower error rate. Performance of respondents was considerably better than that of proxies at all three frequency levels. Collapsing across the frequency levels gives a similar result, with mean absolute proportional error of 0.46 for respondents, whereas it was larger by more than 50% at 0.71 for roommate proxies,  $F(1,25) = 14.72$ ,  $p < .001$ . The greater variability of proxies’ reports is also demonstrated by the considerably larger standard deviation for proxy reports than for self-reports (1.52 and 0.63, respectively). There was no significant difference between the mean absolute proportional errors of male and female respondents or between male and female proxies, although there was a tendency for females to perform better. The mean absolute proportional error of the 11 male respond-

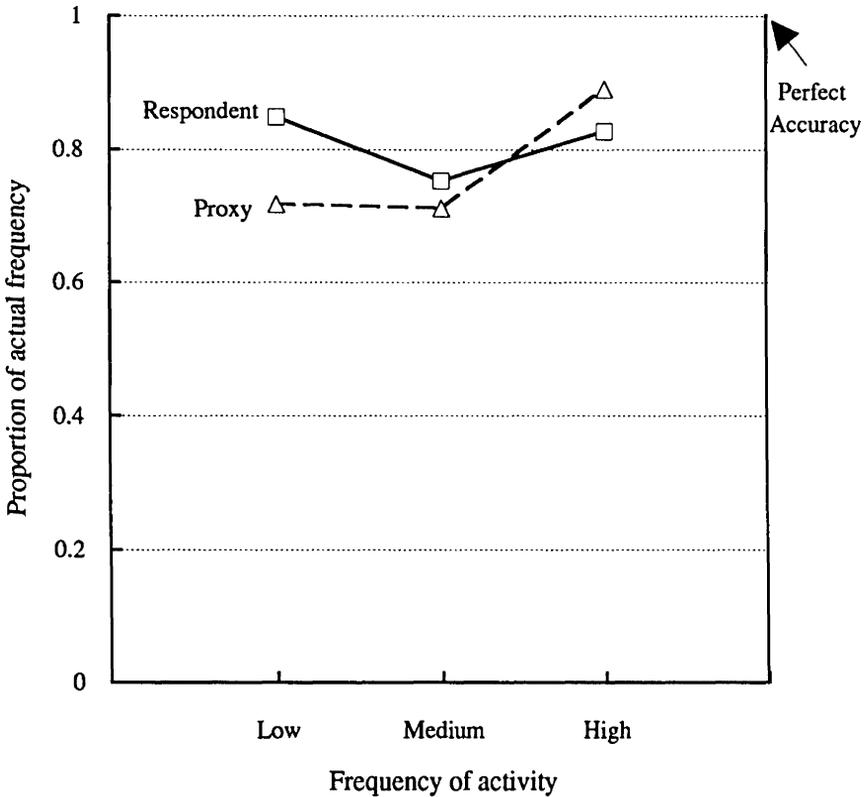


FIGURE 15.1. Mean Frequency Reports of Respondents and Proxies as a Proportion of Mean Actual Frequencies

ents was 0.55, and of the 15 female respondents, 0.41. The equivalent figures for male and female proxies were 0.77 and 0.65. The sample size was too small to assess reliably a difference between men and women.

To explore the consistency of the reporting patterns, we looked at self and proxy performance at the individual activity level. For 7 of the 18 activities, respondents had a significantly lower proportional error than did proxies ( $p < .05$  on one-tailed  $t$  tests). Proxy reporting was not significantly superior to respondents' reports for any of the activities.

The relative performance of respondents and roommate proxies can also be illustrated by comparing accuracy in the respondent/proxy pair at the individual activity level. Excluding the 22% of cases in which the respondent and proxy roommate gave the same answer, in 16% of cases the respondent was correct and the proxy wrong, and in 51% of cases the respondent had a smaller error than the proxy. Thus, the respondent was more accurate than the proxy in a total of 67% of cases. The 34% of cases in which the proxy was more accurate than

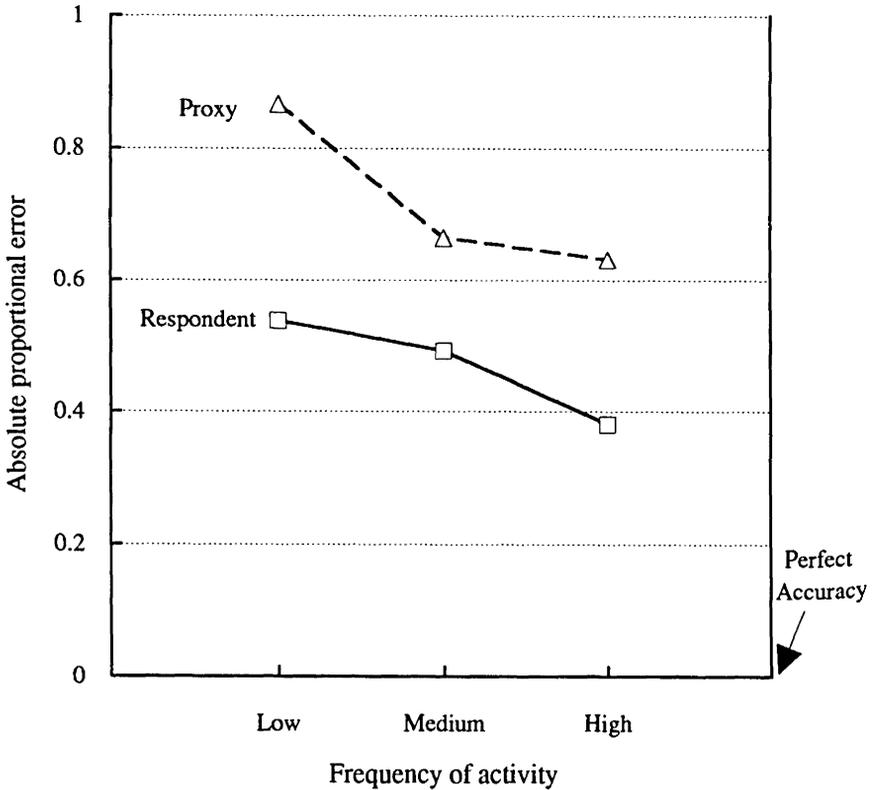


FIGURE 15.2. Mean Absolute Error as a Proportion of Mean Actual Frequencies

the respondent consisted of 8% of cases in which the proxy was correct and the respondent wrong and 26% of cases in which the proxy's error was smaller than that of the respondent.

The relative accuracy of respondents and roommate proxies was very similar across all three frequency groups. For example, proxies were more accurate than respondents in 31%, 34%, and 36% of cases in the high-, medium-, and low-frequency groups, respectively.

Although for the majority of activities the frequencies were fairly low, an analysis was conducted to determine whether there was a bias toward responding with rounded numbers, reflecting uncertainty in the answers (cf. Huttenlocher et al., 1990). Rounded numbers were defined as numbers divisible by 5. The response of 0 was reported separately, since it seemed unlikely that respondents who thought that the answer was, say, 0, 1, or 2, would wish to round to 0 unless they believed that this was the most likely answer. The percentage of actual frequencies that were nonrounded, rounded, and 0 were 77%, 13%, and 10%, respectively. For respondents, the equivalent percentages were 67%, 21%,

and 13%, and for proxies, 61%, 23%, and 18% (percentages total more than 100 owing to arithmetical rounding). There was no significant difference between the frequency of rounded numbers on the diary sheets and the number of rounded answers given by respondents, between the frequency of rounded numbers in the records and those in proxy reports, or between the amount of rounding by respondents and proxies ( $p > .15$  by chi-square tests). The failure to observe the rounding commonly found in reporting in surveys (e.g., Bradburn, Huttenlocher, & Hedges, chap. 13, this volume) may be due to the relatively low frequency of the activities.

Another analysis concerned activities that the respondent had not done during the 5-week reference period. Roommate proxies may experience particular difficulty with these items: They must decide whether their failure to retrieve instances of the activity is due to the fact that the respondent has not done the activity, has done the activity without informing the roommate, or has done the activity and informed the roommate, who no longer remembers being told. In contrast, respondents will find it relatively easy to answer the question. If they do not remember doing the activity in the last few months, there is a strong probability that they have not done it at all during the reference period.

Collapsing across all subjects and activities, the respondents correctly reported 79% of the activities never done. The equivalent figure for the proxies was 60%. Although the pattern of data was in the predicted direction, the difference between self and proxy reports was not significant (chi-square = 2.02, 1 *df*,  $p = .16$ ) because of the relatively small number of never-done observations (47).

A forced-choice debriefing question was administered to respondents and proxies asking whether they had expected to be tested on the diary items. The phrases, with the percentages of respondents choosing them, were as follows: "Yes, thought it was very likely," 54%; "Yes, thought it was possible," 15%; and "No, did not think this," 30%. The equivalent figures for proxies were 4%, 38%, and 59%. Expecting to be asked to recall the diary activities did not improve respondents' performance (mean absolute proportional errors for these response categories were 0.46, 0.52, and 0.45, respectively;  $F < 1$ ). Informal questioning suggested that the memorability ratings task was an important reason why the majority of respondents expected to be asked to recall the activities recorded in the daily diaries.

## Discussion

This experiment involved the investigation of accuracy in reporting on the frequency that 18 different everyday activities were engaged in over a 5-week period. Subjects consisted of pairs of students who were college roommates, with one student reporting for him or herself (respondent) and the other reporting for his or her roommate (proxy). Reports were validated using daily diary sheets completed by the respondent and mailed to the experimenter.

The major findings of this study can be stated quite succinctly. Underreport-

ing of frequency was more prevalent than overreporting for both respondents and proxies. Two measures used to compare the accuracy of self and proxy reports gave quite different results. When performance was measured by taking into account the direction of error as well as its magnitude (signed error), the roommate proxy achieved, on average, about the same level of accuracy concerning the respondent's activities as did the respondent. In contrast, using absolute error as the measure of performance, in which the direction of error is ignored (unsigned error), respondents were considerably more accurate than proxies. This absence of bias in proxies' reports means that first-order point estimates from proxies' reports are as valid as those derived from respondents' reports. In contrast, higher-order analyses (e.g., multivariate relationships) would be more valid using data obtained from respondents than from proxies. Most major government and academic surveys have both multiple users and multiple uses, with at least some higher-order analyses being conducted. Thus, if these results were generalized to those surveys, a strong case could be made for using respondent rules that minimize the use of proxies.<sup>1</sup> We shall first evaluate the internal validity of these findings and then discuss their relevance to household surveys.

### Validity of the Findings

The validity of these findings rests on the suitability of the diary-keeping methodology to validate the frequency reports of respondents and proxies.<sup>2</sup> We evaluate below the methodology as it pertains to this study. A. F. Smith, Jobe, and Mingay (1992) give a more comprehensive evaluation of the diary-keeping methodology.

To look for evidence that the quality of diary reports may have declined over time, we examined the number of activities recorded each week. Encouragingly, there were no statistically significant changes in reported events over the diary-keeping period. We also found no evidence that the diary keeping enhanced the memorability of the activities. Such an effect is likely to be more pronounced for activities that are mundane and routine—for example, watching TV and going to the grocery store. However, there was no association between memorability of the activities, as rated on the daily diary sheets, and the relative accuracy of respondents' and proxies' reports.

---

<sup>1</sup>Such a decision would also rest on the effects of different levels of proxy reporting on item nonresponse, person nonresponse, and household nonresponse. In general, response completeness appears similar: Maximizing self-response rates tends to result in slightly higher household and person noninterview rates, but this is offset by lower item nonresponse (J. C. Moore, 1988).

<sup>2</sup>Diaries are used in surveys in two other ways: to serve as memory aids to improve the recall of events at the time of the interview and as a primary data source (see Verbrugge, 1980, for a comprehensive review of the advantages and disadvantages of using diaries in health studies).

Several aspects of the experimental procedure should reduce the likelihood that the diary keeping enhanced memory for the activities. Since respondents only needed to indicate the number of times that they had engaged in each activity during the day, the diary sheets could be completed quickly and with less thought than the diaries commonly used in autobiographical memory studies, in which respondents write a brief description of each activity (e.g., Thompson, 1982; Skowronski et al., 1991, and chap. 14, this volume).<sup>3</sup> The ease of completing the diaries should minimize the amount of rehearsal engaged in when the activity is recorded in the diary, thereby minimizing the likelihood that the recording of the activity will enhance memory for the item. The likelihood that the respondent will answer the subsequent frequency questions by recalling the diary record rather than the activity itself should also be minimized. In addition, respondents were required to mail in the diaries at the end of each day or early the following day to reduce the likelihood that they would review the diary records. Finally, respondents completed a total of 35 diary sheets, a number sufficiently large that the recording task is likely to have become quite habitual over the course of the 5 weeks.

Smith and Jobe (chap. 8, this volume; A. F. Smith et al., 1992) and Thompson (1982) also found no evidence that diary keeping enhanced memorability of the events recorded. Based on a review of the literature, Verbrugge (1980) suggested that diaries are an appropriate procedure for validating the health-related reports of respondents in household surveys. Diaries also have proved to be a suitable methodology for collecting consumer expenditure data, especially when the diary keepers receive compensation (Sudman & Ferber, 1974; Sudman, Wilson, & Ferber, 1976).

## Relevance to Household Surveys

In most surveys, few, if any, respondents are students living at a college or university. Most proxy respondents are family members living in the same household rather than roommates, and thus they have different types of contact with the respondent. In addition, although some of the activities that were asked about in this study are similar to those queried in some surveys on health or other topics—for example, drinking alcohol and engaging in a sport—other activities are not. That these findings do apply under very different circumstances is suggested by their consistency with the Sudman et al. study (chap. 16, this volume) and the Cobb et al. (1956) and D. J. Thompson and Tauber (1957) papers cited earlier.

From a methodological perspective, the use of college students has a number of advantages, including their likely greater compliance with the diary keeping over the 5 weeks and their greater homogeneity in terms of characteristics likely to influence reporting accuracy.

The results of our research strongly argue for the further investigation of

---

<sup>3</sup>The majority of respondents reported taking less than 2 minutes on average.

self/proxy reporting patterns in household surveys, using diaries or other types of records to validate reports. Careful attention should be paid to avoiding the methodological shortcomings that bedevil the literature, well described by J. C. Moore (1988). Familiarity with several social science literatures offers further help in the research design, including the cognitive-psychological literature on autobiographical memory, the social-psychological literature on social desirability, and the survey literature on questionnaire design.

*Acknowledgments.* This research was funded by a grant SES-8520979 from the National Science Foundation. We are grateful to Lance Rips, Norbert Schwarz, and Seymour Sudman for their comments on an earlier draft of the chapter.

# The Effect of Participation Level on Reports of Behavior and Attitudes by Proxy Reporters

*Seymour Sudman, Barbara Bickart, Johnny Blair,  
and Geeta Menon*

The use of proxy reporters is very common for a wide range of surveys conducted by government agencies, universities, and the private sector. As examples, the Current Population Survey has a single household member report about the labor force participation of all household members; in the National Health Interview Survey and the National Crime Survey, proxy respondents are permitted to report about other household members if they are not available. In consumer research, information about multiple family members' expenditures is often obtained using proxies. The major benefit of using proxies is the reduced cost of data collection.

The use of proxy reporters raises important issues of data quality and of the processes used by respondents to answer about others. Earlier work has usually found that proxy reports are less complete than self-reports, although in many cases the differences are small and in a few cases proxy reporting is actually better (J. C. Moore, 1988).

Many of the comparisons reported in the literature, however, are difficult to interpret because proxies have been used only when the respondent is unavailable. Thus, reporting and sample biases are intermixed. In addition, the earlier studies have simply compared self and proxy reporting and have not provided information on how the cognitive processes used for self and proxy reporting might differ.

Our study addressed these limitations by interviewing pairs of respondents in a household and asking them to report both about themselves and about the other household partner. We used a range of research strategies, including face-to-face think-aloud interviews in a cognitive laboratory setting, telephone interviews, and laboratory experiments. The questionnaire covered a broad range of topics, including behavior, attitudes, and demographic information.

The study focused on several major issues related to proxy reporting. In earlier papers (Bickart, Blair, Menon, & Sudman, 1989; Menon, Sudman, Bickart, Blair, & Schwarz, 1990), we have explored the different cognitive processes

used by self and proxy respondents to answer behavioral frequency and attitude questions, with specific attention to anchoring and adjustment strategies.

This chapter summarizes some of these earlier results, discusses the convergence between self and proxy reports over the entire range of behavior and attitude questions in our study, and looks at a possible explanatory factor derived from our theoretical perspective: the level of participation and discussion between self and proxy.

## Theoretical Overview

Most researchers who use cognitive methods to explain survey responses have adopted a general model of the process that involves interpreting the question, retrieving appropriate information or a prior judgment, making the judgment, and reporting the response (Strack & Martin, 1987; Tourangeau, 1987). Self and proxy processes would not differ in terms of understanding the question but could differ on the other dimensions. In Table 16.1, we present a general overview of how self and proxy reports would differ during stages of encoding, storage, retrieval, and response generation of information.

The ability of self and proxy respondents to retrieve information from memory differs along four dimensions:

1. Whether the information is available in memory
2. The context in which the information is stored
3. The accessibility of the information
4. The extent to which information has been integrated into summary judgments

Information needed to answer questions about one's *own* behavior and attitudes should be more available in memory and more completely encoded. Our own behaviors provide a rich set of experiences, including information about what we wanted to do, what we actually did, how we felt while doing it, and so on (Tulving, 1972, 1983). Information needed to answer questions about *others* may not be available or may be accessed in the context in which respondents learned about the event (Larsen & Plunkett, 1987).

Information about others will only be available if the target person's behavior has been observed or if it is learned through discussion. The amount of information available should be related to the amount of time that the partners spend talking about the topic or the degree to which they participate together in the behavior. It should also be related to the importance of the behavior.

In many cases, respondents will be unable or unwilling to remember specific events but will use estimation procedures (Blair & Burton, 1987). Even here, we expect that joint participation in events will lead to similarity in reporting strategies between self and proxy reporters.

In addition to encoding differences, the storage of information about oneself versus others might differ for several reasons. First, self-relevant information has been shown to receive increased elaboration at encoding (Kuijer & Rogers,

TABLE 16.1. Information about Self and Other at Stages of Information Processing

Encoding	Storage	Retrieval	Response Generation
Self			
1. Actually experience event: a. Chronological b. Experienced in entirety	1. If episodic, related to actual	1. Cues related to experience	1. Self-presentation effects
2. Increased elaboration	2. If semantic, related to self	2. Search likely to be chronological  3. Event likely to be recently activated	
Other			
1. Learn about event through a. Observation b. Word-of-mouth/direct (from target)	1. If episodic, related to reception situation unless joint participation in event	1. Cues related to encoding situation	1. Less sensitive to demand
2. Not necessarily chronological	2. If semantic, related to "other" knowledge structure, presumably not as well organized	2. Chronological retrieval is difficult	2. May not be aware of sensitive information
3. Amount of elaboration depends on social distance			3. Social distance will influence both (1) & (2)

1979; Rogers, Kuiper, & Kirker, 1977). This is thought to create a more elaborate memory trace for the event, resulting in enhanced recall. However, several studies have shown that information relevant to important others, such as one's spouse, also receives increased elaboration (Bower & Gilligan, 1979; Kuiper & Rogers, 1979).

Information about self-behavior is better organized in memory than is information about others (Klein & Kihlstrom, 1986). In addition, respondents are more likely to have a summary judgment, such as a rate at which the behavior takes place, available about their own behavior than they are about the behavior of others. Rate-based estimates should, therefore, be more accurate when made about oneself than about others.

Finally, the impact of social-desirability and self-presentation concerns should be different for self and proxy respondents. Proxy respondents may, in general, be less likely to overreport socially desirable behavior and more willing to report about socially undesirable behavior, but they may lack the knowledge to do so. These effects may counteract each other and may be difficult to detect, but they cannot be ignored.

## Summary of Earlier Findings

Our earlier papers investigated the strategies used by proxy respondents, and the results that we obtained were consistent with the theoretical perspective in Table 16.1. Although the use of chronological sequences was fairly low for self as well as for proxy reports, self-reporters were about 60% more likely to use chronological sequences and event cues than were proxy respondents. Conversely, proxy respondents were more likely to use estimation and less likely to use counting methods.

This was true not only in total but also for 10 of 11 specific topic areas (see Bickart et al., 1989). Proxy respondents were significantly more likely than self-reporters to rely on "anchor-and-adjust" strategies. It is possible to increase the use of these strategies by questionnaire design. When the question about self was followed by a priming question on how much the respondent and the other person agreed or disagreed and then by a question about the other person's attitudes or behavior, the use of an "anchor-and-adjust" strategy was double that in other conditions where there was no priming or where the self and proxy questions were separated (see Menon et al., 1990).

## Hypotheses

The overview of the literature suggests the following hypotheses discussed in this chapter:

- H1. The correlation between self and proxy reports should increase significantly as the level of participation increases.

- H2. The correlation between self and proxy reports should increase significantly as the level of discussion increases.
- H3. The correlation between self and proxy reports should increase significantly as the perceived importance increases.

## Methods

The results in this chapter are based on face-to-face interviews in a cognitive laboratory with 50 pairs of partners in the same household and on telephone interviews with 200 pairs of partners in the same household in Champaign County, Illinois. In addition to the substantive questions shown in the tables presented below, respondents in the two samples were asked the following:

- How often do you and your partner . . . together?
- How often do you and your partner discuss . . .
- How important is . . . to you? (behavioral items) or
- How much do you care about the following issues? (attitudinal items)
- How confident are you that your answers about partner's . . . were accurate?

Based on their answers, respondents were dichotomized into higher or lower participation, more or less discussion, higher or lower importance, and more or less confidence.

The face-to-face interviews and the telephone interviews obtained slightly different information. Some behavioral questions asked face-to-face were not asked on the telephone, whereas some political attitude questions asked on the telephone were not asked face-to-face. The face-to-face interviews were much longer, since respondents were asked to think aloud about responses to every question, whereas think-alouds were asked for only a few questions in the telephone interview. The topics covered in the interviews are shown in Exhibit 16.1. The coding scheme used for the verbal protocols is shown in Exhibit 16.2.

In addition to direct measures of participation, it is possible to observe partners' behavior better in the home for activities such as reading books for pleasure, whereas activities such as reading books for work or school are conducted outside the home and are less easy to observe. Activities that occur in the home should generally lead to higher convergence in self and proxy reports.

For continuous variables, the measure of convergence is simply the correlation between self and proxy reports on the same question. For noncontinuous items, the percent agreement is taken as the measure of accuracy. In this chapter, we do not concern ourselves with directions of inaccuracy that result in bias, nor do we address, except in passing, the possibility that proxy reporters may be better than self-reporters for threatening questions.

## Convergence Results

Each table below presents comparisons separately for the telephone and face-to-face samples. In general, the results of the two methods were in close agree-

## EXHIBIT 16.1. Topics of Questions

---

Sources of news information  
 Newspapers read  
 Books read  
 Amount of TV watched  
 Drinking behavior  
 Eating habits  
 Work or school days missed  
 Date of last doctor visit  
 Reason for last doctor visit  
 Health  
 Crime victimization  
 Employment status  
 Education  
 Year of birth  
 Income  
 Voter registration  
 Voting  
 General interest in politics  
 Attitudes toward  
   Labor unions  
   Women's rights  
   Environmental groups  
   Ku Klux Klan  
   Various politicians  
   Family planning  
   Abortion

---

ment, lending additional support to the conclusions. There were a few puzzling differences that we shall discuss. It is not clear whether these were simply due to sampling variability or to methods differences.

The sample sizes were small by traditional survey standards, although fairly large by the standards of typical cognitive laboratory studies. For this reason, we did not conduct significance tests for each individual question but established statistical significance by observing the patterns of responses and using a binomial signs test.

## Participation

Table 16.2 presents the correlations between self and proxy reports for the two samples split by partners who reported participating together at higher and lower levels. On 14 of 15 comparisons, the correlations were higher for partners who participated more.

It may be seen that, in general, the hypothesis that greater participation should lead to greater convergence was strongly confirmed in both the telephone and face-to-face samples. The effects were consistent, but the magnitude of the

## EXHIBIT 16.2. Verbal Protocol Coding Scheme

Code	Description
<b>10s: Automatic Response</b>	
11	Automatic
12	Event did not occur/Nonevent
13	Retrieval of prior judgment
14	No probe, therefore no protocol
15	Don't know
16	Not applicable (for skips)
<b>20s: Counting Strategies</b>	
21	General recall and count
23	Counting by observation
24	Counting with adjustment for uncertainty
25	Counting with expression of uncertainty (no adjustment)
22	Counting by domain
26	Counting by domain with adjustment for uncertainty
27	Counting by domain with expression of uncertainty (no adjustment)
<b>30s: Rate-Based Estimates</b>	
31	General rate-based estimation
33	Rate-based estimation with adjustment based on specific incident (addition/s to estimate)
34	Rate-based estimation with adjustment based on general knowledge
35	Rate-based estimation with adjustment based on nonoccurrence (subtraction/s from estimate)
36	Rate-based estimation with adjustment for uncertainty
37	Rate-based estimation with expression of uncertainty (no adjustment)
32	Rate-based estimation by domain
38	Rate-based estimation by domain with adjustment for uncertainty
39	Rate-based estimation by domain with expression of uncertainty (no adjustment)
<b>40s: Enumeration-Based Estimates</b>	
41	General enumeration-based estimation
43	Enumeration-based estimation with adjustment based on specific incident (addition/s to estimate)
44	Enumeration-based estimation with adjustment based on general knowledge
45	Enumeration-based estimation with adjustment based on nonoccurrence (subtraction/s from estimate)
46	Enumeration-based estimation with adjustment for uncertainty
47	Enumeration-based estimation with expression of uncertainty (no adjustment)
42	Enumeration-based estimation by domain

(Exhibit 16.2 continued)

## EXHIBIT 16.2. Continued

Code	Description
50s: Anchoring Strategies	
51	Same as self
52	Based on prior answer
53	Anchor on self and adjust
54	Anchor on norm and adjust
55	Anchor on another specific person and adjust
56	Anchor on proxy and adjust
60s: Miscellaneous Codes for Attitude Questions	
61	Based on specific behavior/event
62	Based on discussions with other
63	Based on general knowledge
64	Based on attitude toward issue
70s: Search Strategies	
71	No order/search
72	Forward search
73	Backward search
74	Anchor on date and forward search
75	Anchor on date and backward search
76	Search by domain
77	Based on another event
78	Based on regularity of behavior
80s: Event Cues (for counting strategies)	
81	Person mentioned
82	Place mentioned
83	Emotional reaction to event mentioned
84	Time of event occurrence mentioned
85	Characteristic of event mentioned
86	Based on prior response (one answer triggers off another)
87	Based on cues used from question
90s: Reference Period	
91	Anchor date on public event
92	Anchor date on personal event
93	Anchor date on season of the year
	General characteristic of event/person (for estimation strategies):
94	"Always . . ."
95	"Never . . ."
96	"Nowadays . . ." / "Usually . . ."

TABLE 16.2. Correlations between Self and Proxy Reports, by Level of Participation and Interview Method

Behavior	Telephone		Face-to-Face	
	Higher Participation	Lower Participation	Higher Participation	Lower Participation
Number of hours watched				
TV				
Weekdays	.52	.42	.48	.64
Weekends	.66	.46	.60	.43
Number of times beer drunk in month	.72	.58	.78	.36
Number of times liquor drunk in month	.93	.69	.99	.45
Number of times drunk in year	.28	.24	.96	-.01
Number of times ate out per month	— <sup>a</sup>	— <sup>a</sup>		
Breakfast			.73	.04
Lunch			.87	.03
Dinner			.97	.66
Number of snacks per day	— <sup>a</sup>	— <sup>a</sup>		
Weekdays			.72	.26
Weekends			.67	.05
Average	.62	.48	.78	.29

<sup>a</sup>Not asked of telephone sample.

differences in correlations was higher for the face-to-face than for the telephone interviews. We had no prior reason to expect methods differences, and this may well be simply a function of sampling error or of the different questions. Nevertheless, these differences, along with some that we discuss later, suggest that some more careful methods comparisons be made.

Additional support for the participation hypothesis comes from Table 16.3, in which the overall correlations between self and proxy reports was sharply higher for books read for fun (presumably at home) than for books read for work or school and mainly away from home regardless of interview method or reported level of discussion.

## Effects of Discussion

Table 16.3 compares the correlations on behavior and attitude variables for partners who reported more or less discussion. On 13 of 16 behavior comparisons and 16 of 23 attitude comparisons (with one tie), correlations were higher for partners who discussed more. The hypothesis that more discussion leads to higher correlations between self and proxy reports was also supported, although

TABLE 16.3. Correlations between Self and Proxy Reports, by Level of Discussion and Interview Method

Variable	Telephone		Face-to-Face	
	More Discussion	Less Discussion	More Discussion	Less Discussion
<b>Behavior</b>				
Number of newspapers read	.36	.68	.47	.33
Number of books read				
for work/school	.17	.08	.26	.05
Number of books read for fun	.54	.42	.50	.08
Health rating	.61	.44	.31	.20
Days missed of work/school				
in past year	.48	.40	-.05	.82
Number of times seen doctor				
in past year	.50	.30	.65	.06
Seriousness of condition	.39	.20	.14	-.25
Income	.88	.79	.24	.33
Average	.49	.41	.32	.20
<b>Attitudes</b>				
Favorability to				
Ku Klux Klan	.26	.06	— <sup>a</sup>	— <sup>a</sup>
Labor unions	.76	.44	.56	.48
National Rifle Assn	.84	.58	.58	.31
Women's groups	.48	.28	.67	.51
Environment groups	.12	.31	.59	.67
Job that President is doing for				
Inflation	.57	.33		
Unemployment	.58	.38		
Trade deficit	.48	.24		
Economy	.54	.32		
Foreign competition	.50	.28		
Honesty of				
Politicians	.28	.46	— <sup>a</sup>	— <sup>a</sup>
President	.51	.50		
Senator Simon	.50	.41		
Gov. Thompson	.40	.36		
Senator Helms	.39	.40		
Gov. Dukakis	.33	.30		
Senator Kennedy	.40	.40		
Gary Hart	.76	.80		
Ronald Reagan	.49	.58		
Effectiveness of government				
drug abuse program	.29	.27	— <sup>a</sup>	— <sup>a</sup>
Average	.47	.38	.60	.49

<sup>a</sup>Not asked of face-to-face sample.

not quite as strongly as the participation hypothesis. The differences in the correlations between those who discussed more versus those who discussed less averaged around 10%, which is much lower than the differences observed in the face-to-face interviews between those who participated more or less and a little lower than those correlations for telephone respondents.

For noncontinuous items, given in Table 16.4, partners who discussed or participated more were more accurate in 14 of 17 comparisons. On average, partners on the telephone survey who discussed more were right 91% of the time, compared with 86% of the time for partners who discussed less. For the face-to-face sample, partners who discussed or participated more were right 85% of the time, compared with 78% of the time for partners who discussed less.

To summarize these three tables, in 80% of the 70 comparisons the hypothesis that greater participation or discussion leads to higher correlations between self and proxy reports was supported. These are clearly real differences and not artifacts, but some people may wonder why they are not still larger. Two reasons suggest themselves. First, it must be recognized that the questions dealing with participation and discussion between partners were themselves subject to the same cognitive processes as the substantive questions. Measurement error in these variables would reduce the estimated effects.

Second, since the level of agreement on the noncontinuous items, at least for these samples and these questions, was very high, there was little variability to explain. We would expect even larger effects for sample pairs other than husbands and wives where the social distance was greater.

## Importance

We hypothesized that agreement would be higher for topics important to respondents, but the results did not support this. After controlling for level of discussion or participation, there were no differences between respondents who rated topics as more or less important. It may be that the measure of importance was difficult for respondents to answer and was subject to substantial measurement error. This could explain the lack of observed differences. It is also possible, as we mentioned earlier, that another reason for the lack of clear results was that self-presentation issues arose because important topics may also have been threatening.

## Confidence

Confidence in answers is also a subjective variable that should reflect respondents' evaluations of how much they know about their partners. It would be expected that, to some extent, respondent confidence is related to the degree of participation and discussion between partners. The correlation between participation and confidence was usually less than .2 for behavior items; the correlation between discussion and confidence was higher, ranging from about .2 to .4 for attitudinal items. Others have reported, however, that many respondents

TABLE 16.4. Agreement between Self and Proxy Reports, by Level of Discussion or Participation and Interview Method (%)

Variable	Telephone		Face-to-Face	
	More Discussion/ Participation	Less Discussion/ Participation	More Discussion/ Participation	Less Discussion/ Participation
<b>Behavior</b>				
Read a book in month	94.3	82.9	84.2	70.0
Saw a doctor in past year	90.0	91.9	78.9	60.0
Registered voter	97.2	91.9	82.6	90.0
Voted in				
Presidential election	95.9	96.1	94.7	92.6
Voted in primary	85.4	77.1	82.9	76.2
Exercised in month	— <sup>a</sup>	— <sup>a</sup>	74.3	56.5
Employment status	92.0	100.0	85.0	84.6
<b>Attitudes</b>				
Government support for family planning	81.3	72.6	90.2	80.5
Support abortions for family planning	91.0	77.0	95.1	86.7
Average	90.9	86.2	85.3	77.5

<sup>a</sup>Not asked of telephone sample.

appear to be overconfident about their ability to recall events (see Ross & Buehler, chap. 4, this volume).

Since reports of confidence present issues of self-presentation and may be a function of overall self-confidence, we had no strong advance hypotheses about how well this variable would work as a predictor of convergence. The results in Tables 16.5 and 16.6 suggest that confidence in the answer was correlated with higher levels of agreement but that this effect was weaker than that observed in the participation and discussion hypotheses. In about 60% of comparisons between self and proxy reports, correlations were higher for those with higher confidence. The average correlations for telephone sample respondents with higher confidence were .60 on behavior items and .47 on attitude items, compared with .40 and .41, respectively, for those with lower confidence. For the face-to-face respondents, there was no clear trend. The average correlations for

TABLE 16.5. Correlations between Self and Proxy Reports, by Level of Confidence and Interview Method

Variable	Telephone		Face-to-Face	
	High Confidence	Low Confidence	High Confidence	Low Confidence
<b>Behavior</b>				
Number of books read for work/school	.16	.25	-.12	.36
Number of books read for fun	.84	.52	-.14	.41
Number of hours watched TV				
Weekdays	.59	.24	.58	.64
Weekends	.62	.15	.56	.45
Health rating	.62	.40	.32	.32
Days missed of work/school in past year	.49	.55	-.06	.34
Number of times seen doctor in past year	.38	.60	.11	-.17
Number of times beer drunk in month	.77	.57	.39	.47
Number of times liquor drunk in month	.92	.61	.98	.64
Number of times drunk in year	.60	.14	.63	.54
Number of newspapers read	— <sup>a</sup>	— <sup>a</sup>	.43	.47
Number of magazines read	— <sup>a</sup>	— <sup>a</sup>	.40	.26
Number of snacks per day	— <sup>a</sup>	— <sup>a</sup>		
Weekdays			.00	.22
Weekends			.04	.38
Average	.60	.40	.29	.38
<b>Attitudes</b>				
Favorability to				
Labor unions	.53	.59	.62	.44
National Rifle Assn	.72	.66	.40	.58
Women's groups	.46	.25	.68	.52
Environmental groups	.36	.06	.66	.64
Job that President is doing for			— <sup>b</sup>	— <sup>b</sup>
Inflation	.50	.38		
Unemployment	.52	.41		
Trade deficit	.38	-.03		
Economy	.48	.10		
Foreign competition	.40	.15		

(Table 16.5 continued)

TABLE 16.5. Continued

Variable	Telephone		Face-to-Face	
	High Confidence	Low Confidence	High Confidence	Low Confidence
Honesty of			— <sup>b</sup>	— <sup>b</sup>
Politicians	.37	.53		
President	.55	.60		
Senator Simon	.50	.46		
Gov. Thompson	.38	.71		
Senator Helms	.46	.38		
Gov. Dukakis	.36	.40		
Senator Kennedy	.38	.24		
Gary Hart	.79	.79		
Ronald Reagan	.52	.72		
Effectiveness of government				
drug abuse program	.23	.44	— <sup>b</sup>	— <sup>b</sup>
Average	.47	.41	.59	.54

<sup>a</sup>Not asked of telephone sample.

<sup>b</sup>Not asked of face-to-face sample.

TABLE 16.6. Agreement between Self and Proxy Reports, by Level of Confidence and Interview Method (%)

Variable	Telephone		Face-to-Face	
	High Confidence	Low Confidence	High Confidence	Low Confidence
<b>Behavior</b>				
Read a book	89.7	80.4	61.7	85.0
Saw a doctor in past year	83.9	86.7	75.0	73.0
Drank beer in last month	90.3	93.3	66.1	89.7
Drank liquor in last month	84.4	66.7	79.7	79.3
Registered voter	93.9	90.0	91.7	80.6
Voted in Presidential election	96.7	100.0	97.5	90.7
Voted in primary	83.8	71.4	85.3	72.7
<b>Attitudes</b>				
Government support for family planning	84.0	75.0	94.7	76.2
Support abortions for family planning	91.4	73.0	92.5	90.9
Average	88.7	81.8	82.7	82.0

those with high confidence were slightly larger on the attitude items and actually lower on the behavior items.

Initially, we looked at the telephone sample data first and concluded that confidence was a good predictor of agreement between self and proxy reports. The face-to-face results weakened that conclusion substantially. There may have been possible methods effects, but again in the absence of a theory to explain them, we treated the differences as having occurred by chance. Thus, we concluded that the confidence measure mainly reflected participation and discussion and did not contribute much new information by itself.

## Summary and Conclusions

The hypotheses that are related to participation and discussion improving the convergence of self and proxy reports were strongly confirmed, although the magnitude of the effects may be smaller than some might have anticipated. The hypothesis that perceived importance of the topic would also increase convergence was not confirmed for the items used in this study. As we mentioned earlier, issues of self-presentation may have interfered with the effects of importance.

Confidence in answers appeared to predict convergence weakly and not as well as the direct participation and discussion questions. These questions appear to be the ones for survey researchers to use to measure level of convergence in real-world surveys, as well as for possible data adjustment.

The size of the effects that we found was moderate. Obviously, the variables that we looked at were related to proxy reporting, but they clearly were not the only ones. We need to do more work to control for other factors, especially self-presentation. If we know that there are significant response errors in self-reports of threatening behavior, it is not obvious that high convergence between self and proxy reporting will be possible or will even indicate valid information.

We recognize that our current methodology limits the reliability and validity of reports of participation and discussion. One way to remedy this is to conduct laboratory experiments where the level of participation and discussion are controlled and then to conduct surveys asking about the process that occurred.

As a final remark, we have learned that by trying to understand differences between self and proxy reporting, we now have a little better idea of how people report about themselves.

*Acknowledgments.* Research reported in this chapter was supported by grant SES-8821362 from the National Science Foundation.



# Reconstruction of Relationship Memories: A Mental Models Approach

*Diane Holmberg and John G. Holmes*

As survey researchers, we must often rely on the introspection of strangers. If we wish to know a person's opinion on some complex political issue, we have little recourse but to ask him or her. We can take pains to present the questions in as simple, comprehensible, and unbiased a manner as possible; in the end, however, the difficult task of remembering, organizing, and synthesizing disparate information into a single coherent attitude, and then mapping that attitude onto the available response alternatives, must fall to the respondent alone. Research by Converse (1970) and others suggests that this task may often be too much for respondents: They may end up responding with transitory and inconsistent "nonattitudes" that provide little information about their true beliefs.

Of course, it is not too difficult to accept the fact that respondents may be constructing attitudes toward abstruse social policy questions on the spot. The formation of stable opinions on such issues requires information, time, and effort that only the most interested respondents may be willing to invest. We would no doubt feel on much firmer ground if we were to ask respondents only about matters directly pertaining to their own experiences: "How long have you felt the way you do now?" "What factors led you to hold your current opinion?" Such "simple" questions would seem to be readily open to direct introspection. Unfortunately, research by Nisbett and Wilson (1977) suggests that people may have very little direct access to their own mental processes. When people are asked to justify their actions or their attitudes, they show very little direct awareness of important causal factors leading up to their behaviors or beliefs. Objectively influential factors are not noticed or are regarded as unimportant; objectively unimportant factors are viewed as having had a major impact. People appear to be no more accurate in identifying important influences on their own behaviors or attitudes than are informed observers. In short, we cannot always depend on respondents to give us consistent, accurate, reliable information about what their attitudes toward complex issues are, or why they hold these attitudes.

We may gain small comfort in the thought that questions of *why* people acted the way that they did are still quite complex. Surely we would be safe if we were asking for “just the facts.” People must at least be able to tell us what they did, or how they felt, at some point in the past. No new integration of information is necessary: All that is required is a simple memory for events as they transpired at that time.

Again, social-psychological research reveals some disheartening news. From the time of Bartlett’s (1932) pioneering work on the nature of schema-driven memory, we have known that a “simple” memory task is not as simple as it might appear. We do not seem to store complete records of events that can be accessed intact, in spite of the fact that it often feels that way. Instead, our memories seem to be largely reconstructed; and our thoughts, feelings, and opinions at the time of recall can have a large impact on the nature of that reconstruction.

For example, Bem and McConnell (1970) showed that when subjects were asked to write counter-attitudinal essays that changed their opinions toward student control over the university curriculum, they tended to recall their previous attitudes as being similar to the ones that they now held. Although their opinions had changed, they believed that they had always felt the way that they did now. Similarly, Goethals and Reckman (1973) had students who were initially pro- and anti-busing take part in discussion groups that changed their attitudes. When the students were later asked to recall their initial attitudes, they showed a crossover such that those who had initially been pro-busing now recalled themselves as having been even more anti-busing than those who had initially been anti-busing now recalled themselves as having been. Again, subjects’ memories were changed in line with their current attitudes.

More recently, research by Michael Ross and his colleagues showed that these reconstructive effects are not limited to attitudes. Similar biasing effects are seen in memories for evaluations of a romantic partner (McFarland & Ross, 1987), for assessments of study skills (M. Conway & Ross, 1984), and for ratings of menstrual pain (McFarland, Ross, & DeCourville, 1989).

These and other studies have clearly demonstrated that remembering is not a passive accessing of stored information but instead is an active, integrative process that makes extensive use of currently available information to fill in gaps in existing memories. It is important for researchers, including survey researchers, to understand how this memory reconstruction process operates. An understanding of the process of memory reconstruction might, in the future, allow researchers to “correct” for memory biases if they so desire, arriving at a closer estimate of subjects’ actual past scores.

## Previous Theories of Memory Reconstruction

Two major theories of the reconstruction process in memory have emerged in recent years. Both theories emphasize the importance of the state of the individual at the time of recall. The first, which we shall call “mood-congruence

theory" (e.g., Bower, 1987; Isen, 1987), emphasizes the importance of mood at the time of recall in determining the nature of the eventual memory. Current mood serves as a kind of retrieval cue. Thus, when subjects are in a negative mood at the time of recall, the bad mood primes their recollections of negative events. For example, although *formerly* depressed people recall their parents no differently than do never-depressed people, *currently* depressed individuals recall their parents as having been rejecting and punitive (Lewinsohn & Rosenbaum, 1987).

The second theory, a theory of autobiographical memory proposed by Michael Ross (1989), also stresses the importance of the subjects' current state in determining the nature of the memory. According to Ross, when someone asks us what we were like in the past, and we are unable to recall, we answer by taking our current state as a benchmark and then invoking personal theories of consistency or change to determine whether there is any reason why we would have been different in the past. If we cannot think of any reason for having changed (as in the Bem and McConnell, 1970, and Goethals and Reckman, 1973, studies, where subjects did not realize that their attitudes had been changed), then we will simply give the same answer that we would give today. Often this heuristic will lead to an exaggeration of consistency. However, if people have reason to think that they might have changed when they actually have not (e.g., they have taken an unsuccessful study skills course), the tactic leads to an exaggeration of change.

These two theories have certainly proved productive and have stimulated very interesting research; however, some questions still remain unanswered. For instance, most of these studies have asked people to recall overall summary ratings of one type or another. In fact, laboratory researchers such as D. L. Hamilton (1981) have argued that social events are largely remembered in terms of such summative impressions rather than in terms of the episodic details of interactions. However, memories for events that could have important emotional consequences might conceivably show a very different pattern of results from memories for short-term laboratory interactions. In such real-world events, memory for specific details might prove more important. Will reconstruction still be seen if we ask individuals to focus on the details of particular events?

Also, what will be the effect if the events are particularly salient or related to an area of life in which the individual has a great deal of knowledge? We suspect that in memories relating to particularly important events or people, minor fluctuations in mood or attitude at the time of recall will be unlikely to have large effects on the nature of the memory. Instead, one's overall impression or attitude, formed over a long period of time, will be recruited to help fill in gaps in one's memory.

As an example, let us pretend that you have been having a particularly unpleasant and stressful day. At this point, a student arrives for a meeting and announces that (a) he has not completed the paper that he promised would be ready by today and (b) he needs you to write a brief letter of recommendation right away for him to include with a scholarship application. The first item in the

letter asks you to describe your initial impression of the student. What sort of response do you give?

According to mood-congruence theory, the student should brace himself for a relatively unfavorable review. Given that you are in a bad mood now, your memories will be skewed toward the negative aspects of your first impressions.

Ross's theory, too, suggests that the student may be in trouble. According to this theory, you would assess your current (relatively negative) feelings toward the student and then decide whether there is any reason why those feelings should have changed in the time that you have known him. Of course, if you realized that your overall bad mood was unfairly coloring your current judgment of the student, you might attempt to adjust your assessment to take this factor into account. Unfortunately, previous research (e.g., Schwarz & Clore, 1983, Experiment 2) has shown that people usually fail to adjust for the impact of their overall mood on judgments, unless their attention is somehow directed to the possible effect of extraneous events on that mood. Thus, unless the potential impact of your bad mood on the assessment of the student is pointed out, you might not make adjustments to your judgment. Even if you try to adjust, you might not make sufficient allowances for the impact of your current mood, and you might still rate the student more negatively than you normally would.

We suspect such a scenario might indeed take place when the student is one whom you do not know very well. Your first meeting may not have made much of an impression on you. When trying to reconstruct what you probably thought back then, your current impression is as good a starting place as any. But what would happen if the student were one with whom you had been working closely for a number of years and with whom you had had many important interactions? You still might not remember your initial impressions all that well (first meetings often do not seem significant at the time); however, you have at hand a much better heuristic for memory reconstruction than your current feelings. In this case, you are able to draw on the general picture that you have built up over the years of the student's strengths and weaknesses. You still cannot remember your past impressions, but your well-elaborated *mental model* of the student allows you to fill in the gaps in your memory with relative ease and confidence.

## A Mental Models Approach

We would argue that such elaborated mental models are very important components of memory reconstruction. They may take the form of abstract *scripts*, which contain the individual's expectations or theories about events in the social world. These scripts contain information on how sequences of events "should" proceed. Individuals can use this scripted information to aid them in memory recall or reconstruction. For example, Bower, Black, and Turner (1979) gave subjects elements of a "getting-up-and-going-to-work" script out of its usual order (e.g., she started the car, drank coffee, got up, and put on her coat). How-

ever, subjects tended to recall the scripted events in their typical order, suggesting that they were using the script as part of their recall strategy.

Mental models may also take the form of *stereotypes*. In one well-known study (M. Snyder & Uranowitz, 1978), subjects read a short life history of a woman named Betty K. Immediately after reading the story, subjects were told that Betty was a lesbian or were told that she was heterosexual or were given no information about her sexual orientation. In a later recognition-memory test, subjects presented with the lesbian label made more false alarms on items that fit the lesbian stereotype than did those who were not told about Betty's sexual preferences. It seems that when subjects' memories were unclear, they used the information contained in their stereotypes to make educated (or biased) guesses about which statements probably appeared.

Similar results were found in a study exploring the concept of personality *prototypes* (Cantor & Mischel, 1977). Subjects were presented with a list of traits that were consistent with either introversion or extroversion. In a later recognition-memory test, subjects were quite confident that they had seen many traits that were consistent with the prototype that they had been exposed to but that had not, in fact, been shown in the original session. Again, subjects used the knowledge contained in their broader mental model to help them fill in gaps in their memories with plausible, but inaccurate, details.

We believe that mental models can take on an even more specific and contextualized form. They can contain information not only about generalized groups or types of people but also about particular individuals. These individualized mental models contain complex, integrated sets of impressions and information built up about other people over time. The term is similar to that used in the attachment or interdependence theory literature (Bowlby, 1980; Hazan & Shaver, 1987; Kelley & Thibaut, 1978), where individuals are said to build up working models of self and relationship partners. These mental models consist of cognitive (e.g., "Sue is always on time"), affective (e.g., "I love Sue"), and evaluative (e.g., "Sue is a reliable, trustworthy individual") components and thus are structurally similar to the concept of attitudes in the social-psychological literature.

Our mental model of a given individual can become more and more complex over time as we interact with that individual and come to know and understand more about him or her. These mental models are functional in the sense that they help us to *understand* a person's behavior in the present and to *predict* that person's likely behavior in the future. Importantly for our purposes, they may also help us to *recall* what that person was like in the past. When asked to recall a specific event, there will likely be areas where our memories are vague or incomplete. Our mental model of another individual can help us fill in the gaps in our memories of interactions with that individual, in a top-down fashion. Just like the subjects in the stereotype and prototype studies described above, we may not remember each event that happened, but we can use our broader knowledge to help us make reasonable inferences about what *probably* happened.

“Memory” may actually consist of backwards reasoning: “Given all that I know about this individual, it must have been the case that he or she acted in the following manner . . . .”

Often these mental models will help us remember (or reconstruct our memories) more accurately. For example, if we know that a friend is an avid athlete, it may help us to recall that she took part in a 10-km run 3 years ago. Our knowledge of her athletic habits (e.g., details of the training regimen that she follows in the weeks before a big race) can provide us with a wealth of interrelated elements, any one of which might provide a retrieval cue for the larger set of memories. In a sense, our mental model of our friend provides us with a set of working hypotheses that can help us engage in guided scanning of our memories for particular pieces of information.

Such guided scanning will aid memory when our friend’s actions were indeed consistent with our larger mental model. However, if circumstances were such that she behaved in a way that was inconsistent or irrelevant to our mental model, such guided scanning might prove counterproductive. Our biased search strategy might prevent us from locating the inconsistent memories. Faced with the resulting “gap” in our memories, we might then invoke the currently inappropriate mental model to help us reason out what “probably” happened, leading to erroneous recall. For instance, we might find ourselves “recalling” that our friend did quite well in that race, when it was actually the one time that she was ill prepared and performed very poorly. In some circumstances, these reconstructed memories may feel quite real, since our detailed knowledge about our friend can allow us to flesh out the false memories with many authentic-sounding details.

## Research Overview

Our goal was to conduct studies that measured individuals’ mental models of important others and explored the effects of these individuals’ mental models on their memories for specific events. Such studies can be difficult and time consuming, because they require a prospective design. One must first obtain information on the nature of a subject’s mental model, collect information about an event at the time that it occurs, and then return to the subject at a later date to collect memories of the target event. The time and effort involved in such designs may explain why little research has been done in this area.

We decided that married couples would be an ideal population for our initial studies. Over years of daily, intensive interactions, spouses no doubt build up highly complex and stable mental models of their partners. Spouses also interact frequently enough to give us a wide variety of target events to choose from in any given time period. In this chapter, we report results from two separate studies on memory processes in married couples. The first study measured the subjects’ mental models at a single point in time and then examined the biasing

effects of these models on specific memories over a relatively short period. The second study examined changes in mental models over time and how these changes in mental models affect more general memories over longer periods.

## Study 1: Short-Term Reconstruction

In the first study, 67 couples from Waterloo, Ontario, were recruited for a study of “how couples think and feel about their everyday experiences together.” Respondents had to have been married or living together for at least 2 years prior to the study. Couples responded to newspaper and poster advertisements. The sample was heterogeneous in terms of social class, age, and education. On average, the couples were very satisfied with their relationships (average satisfaction score = 6.4 on a 7-point scale).

### Method

Couples first came to an introductory session in which the procedures of the study were explained to them. They also filled out a number of questionnaires assessing perceptions of their relationship on a number of dimensions. One of these questionnaires, a 34-item trust scale (Boon & Holmes, 1990), was used as our measure of subjects’ mental models of their partner in the context of their relationship. This trust scale is based on conceptions of mental models from attachment and interdependence theories. It measures people’s perceptions of their partner’s responsiveness and dependability, their confidence in their ability (as a couple) to deal with any conflicts that may develop, and their faith or confidence in the future of the relationship. True to our conception of mental models, it appraises subjects’ assessments of their partner in the context of the relationship in cognitive (e.g., “Our marriage could easily be explained in terms of he contributes this and I contribute that”), affective (e.g., “At times it feels as though my partner and I share a special bond”), and evaluative terms (e.g., “My partner is a thoroughly dependable person”).

Previous research on trust in close relationships (see J. G. Holmes & Rempel, 1989, for a review) shows that trust provides a broad organizational framework that guides subjects’ reactions to their partners in a number of domains, including attributions, rules/scripts, and affective responses. Based on these and other studies (e.g., Bradbury & Fincham, 1990; Holtzworth-Munroe & Jacobson, 1985), we predicted that lower trust individuals, compared with higher trust individuals, would display a more negative, “distress-maintaining” pattern of attributions when they were asked to explain why their partner behaved the way that he or she did in a given interaction. Most importantly for the current chapter, however, we predicted that these initial differences in attributional patterns would be accentuated over time. When recalling the earlier events, subjects would draw on their mental models of their partner (operationalized as trust) to reconstruct the details of those events. Over time, as initial memories faded and

were replaced with reconstructions driven by mental models, attributions would become more polarized in line with one's overall view of one's partner. These attributional differences were also expected to be associated with differences in feelings toward the partner at the time of the event: The same action would lead to more negative emotions if it were attributed to a negative motivation on the part of the partner than if it were attributed to uncontrollable external circumstances.

To obtain attributions and emotions in response to a number of different events, we had the couples in our sample fill out daily diary forms 4 days a week for a period of 3 weeks (Sharp & Holmes, 1990). On these "time-line" forms, subjects recorded any interaction that they experienced within a given 3-hour period that had an impact on their feelings for their partner. They then rated how each event made them feel (++ , + , - , --) and briefly explained why it made them feel that way.

In Step 2 of the diaries, participants focused on the one event during the recording period that had the most impact on their feelings for their partner. They described the event in more detail and elaborated on why they thought that the interaction had an impact on their feelings. They then responded to five scaled attribution questions, designed to tap their interpretation of the event or the meaning that it had for them at the time. These questions asked about the frequency of occurrence of similar events, the motives behind the partners' actions in this event, and the globality of the cause of the event (the extent to which the partner's actions in this event were informative about the state of the relationship in general). These attributional domains have proven to be the most diagnostic of relationship satisfaction in past research (Bradbury & Fincham, 1990).

Finally, subjects filled out a Daily Summary Sheet, which asked questions pertaining to their overall mood and their satisfaction with their relationship and their work for that day. All forms were filled out privately; the importance of maintaining confidentiality was strongly emphasized during the introductory session.

From the pool of 12 response days, we selected two events (one positive and one negative, as rated by the respondent) as our target events. We first asked subjects to complete a Daily Summary Sheet for the day of recall in order to assess their mood and feelings about the relationship at that time. We then gave subjects the brief description, in their own words, of each event, as recorded on the first page of the diaries. We asked them to recall, as accurately as possible, the detailed descriptions and attributions that they had provided in Step 2 of the diaries. Half of the respondents recalled the positive event first; half recalled the negative event first. We also asked them to recall the answer that they had given to the scaled question, "How did you feel about your interactions with your partner today?" for each target day in order to explore their memories for their summative evaluations of feeling states. The memory questionnaire was mailed to the subjects approximately 1.5 weeks after the diary portion of the study was completed. This practice gave us an average time span from target day to recall day of 34 days.

## Results

In the first set of analyses, we sought to predict subjects' recall of their ratings on the attributional questions filled out after describing each event in detail. These ratings are critical because they reflect subjects' coding of the meaning of the event in their lives. The five questions were averaged to create a single attributional index. Where necessary, scales were reversed before averaging so that higher numbers always represented more positive attributions (e.g., positive events seen as frequent and global, negative events seen as infrequent and specific). Separate regressions were run for the positive and negative events. In each case, Time 1 attributions were forced into the equation first, followed by blocks of variables representing subjects' mental models, their Time 2 ratings of current feelings and mood, and Time 1 ratings of event severity and mood. The order of entry of these blocks was varied so that the structural equations could be compared.

As can be seen in Table 17.1, for both the positive and negative events the single strongest predictor of the recalled attributions was the subject's actual Time 1 responses to the attributions. This finding might suggest that subjects were actually recalling their initial responses quite well; however, it should be noted that good reconstruction is very hard to distinguish from actual recall. If the subjects' mental models of their partners lead them to engage in particular attributional patterns at Time 1, and those mental models remain stable from Time 1 to Time 2, then subjects could achieve what looked like excellent recall simply by reinvoking their mental models. The correlations between trust and attributions at Time 1 were, in fact, quite high (.58 for the negative event and .37 for the positive one), suggesting that mental models may indeed guide interpretation at any given point in time. Thus, by responding to the target event at the time of recall as they "normally" would, subjects could achieve responses that were very similar to their true Time 1 responses without actually doing any remembering at all.

It is, therefore, somewhat difficult to decide how much the apparent accuracy represents actual accuracy in memory and how much it represents the effects of stable mental models. However, once the Time 1 attributions are entered into the model and covaried out, the focus shifts to the factors that are predictive of the *discrepancy* between actual and recalled attributions. What factors are predictive of the biases in the subjects' memories? We were chiefly interested in comparing the predictive power of trust, representing our mental models theory of memory reconstruction, with the predictive power of mood or feelings toward the partner at the time of recall, representing mood congruence and Ross's theory of memory reconstruction, respectively.

As can be seen in Table 17.1, trust turns out to be the single best predictor of these biases in recall for the positive event. As trust increases, so does the positive bias or distortion in one's recall. Feelings toward the partner at the time of recall (as assessed in the Daily Summary Sheet that they filled out on the day of recall) are also predictive of positive memory biases, as Ross's theory of mem-

TABLE 17.1. Predictors of Recall of Attributions

Event	B	F	p
Positive event			
Time 1 attributions	0.704	107.78	.0001
Trust	0.210	12.30	.001
Time 2 feelings	0.125	7.32	.01
Time 1 mood	-0.145	5.16	.025
Negative event			
Time 1 attributions	0.687	44.91	.0001
Time 1 feelings	0.104	8.71	.005
Trust	0.158	6.25	.015 <sup>a</sup>

<sup>a</sup>Effect not significant for globality component.

ory would predict, although the effects are not quite as strong as those for trust. Both effects remain significant when they are entered last into the equation, controlling for all other variables. Although there is some support in the data for the importance of current feelings toward partner in memory reconstruction, as Ross's theory would predict, there is no support for the importance of overall Time 2 mood, as mood-congruence theory would predict. Correlations between Time 2 mood and recalled variables never exceed .10.

However, Time 1 mood does enter as a significant predictor of biased recall for the positive event. Note that the beta has a negative value here, indicating that the *more* positive the mood of the subject on the day of the event, the *less* positive the bias displayed in recalling the initial attributional ratings. Apparently, very positive emotional states are difficult to recapture. At Time 1, fresh from the glow of a very positive interaction, partners may make very generous attributions. By Time 2, subjects may not be able to recapture the affect that sustained their very positive attributions, resulting in recall that underestimates their initial generosity in interpreting the event.

Turning to the negative event, we see that feelings toward partner at the time of recall are not at all predictive here, in contrast to the positive event results. Feelings toward partner at the time of the event *are* predictive, however. The worse the action made subjects feel at Time 1, the more negative distortion they exhibit. This result suggests a type of event-driven memory. Subjects may not be able to recall their exact attributions, but they probably remember that they were particularly unhappy with their partner at the time. The more unhappy they were, the more they "blow up" the event in retrospect. There is an interesting asymmetry here. Extreme reactions to the positive event become lost over time, whereas extreme reactions to the negative event are further accentuated with time. The good becomes average, and the bad becomes worse. This asymmetry may be due, in part, to the fact that people engage in more spontaneous causal reasoning when faced with events of a negative valence (see Schwarz, 1990b,

for a review). The negative material, having received more cognitive elaboration, may remain more accessible for recall; in contrast, the positive events are blithely accepted without further cognitive work, and thus they easily slip away.

Trust is again a significant predictor for the negative event, although its effects are not as strong as for the positive event. We broke down the attributional index to explore this weaker result. We discovered that trust remained a strong predictor of biased attributional recall when we looked at the frequency and motive components of the index separately. The problem lay in the globality component; here, trust was not a significant predictor at all. Furthermore, the globality component was not showing the correlations with the other components that the past literature would lead us to expect (i.e., for negative events, one would expect that more negative motive evaluations would be correlated with high globality).

Further investigation revealed that the correlation between motives and globality was indeed in the predicted direction for the lower trust couples in our sample ( $r = .30$ ); however, for the higher trust couples, the correlation showed a startling reversal ( $r = -.70$ ). For these higher trust couples, the worst events, the ones that revealed very negative motives on the part of their partner, were the most circumscribed by attributions, telling them the least about their partner or their relationship in general. In retrospect, this defensive (or perhaps simply realistic) refusal to interpret the occasional negative act on the part of one's partner as being very significant in the larger scheme of things fits well with past research on very trusting individuals (Holmes & Rempel, 1989).

Nevertheless, this pattern of responses makes the relationship between the attributional index and trust quite complex. Certain subcomponents are positively correlated for some subjects and negatively correlated for others. Despite these complications, however, trust is still a stronger predictor than Time 2 mood or feelings, its theoretical competitors, neither of which is a significant predictor at all.

Overall, then, trust seems to be a more important predictor of individuals' biased recall of attributions than their mood or feelings toward their partner at the time of recall. A second set of analyses revealed parallel results for the recall of feelings toward partner on the particular day of the event. Whereas the trust scores represent participants' assessments of *overall* attitudes toward their partner, these daily assessments of feelings capture the ups and downs that are bound to happen in any relationship. As one would expect, trust and daily feelings are somewhat correlated, indicating that higher trust individuals tend to rate themselves as feeling more positive about their interactions with their partner on a given day ( $r_s = .36$  for the negative-event day and  $.27$  for the positive-event day). These correlations, although significant, are still low enough to make it clear that trust and daily feelings are distinct concepts. It is certainly the case that even very high trust individuals, who are quite positive about their relationships in general, can have isolated days when they feel extremely negative toward their partner. Our questions here are parallel to those asked for the attributional items. Can individuals accurately recall their feelings toward their partner

on the day that the target event occurred? If not, what are the best predictors of the biases in their recall?

As Table 17.2 shows, actual Time 1 feelings are again the most important predictors of recalled feelings, suggesting accurate recall or at least good reconstruction. When we move on to look at predictors of the residual distortion or bias in the recall of feelings, we again see that trust is the single most important predictor of that bias. It is more critical than Time 2 feelings, which are marginally significant if they are entered into the model before trust but have no residual power if trust is entered first. Thus, trust captures the variance explained by Time 2 feelings and more, because trust, an exogenous variable, was capable of forecasting people's feelings toward partners on the day of recall (they are correlated at  $r = .34$ ). Note, however, that if one did not have trust as an index of mental models, one might assume that Time 2 feelings played a direct role in reconstruction. Again, subjects' general mental models seem to be more important predictors of biases in their recall than do specific feelings toward their partner on the day of recall.

Indeed, the strength of the trust effects is quite impressive. Within the feelings analyses, trust explained 30.2% of the variance in memory distortions for the positive event and 12.4% of the variance for the negative event. Trust may be a more important predictor for the positive events than for the negative events because negative events receive more attention and cognitive elaboration at the time of encoding (e.g., Berscheid, 1983), making the memory for those events clearer. Positive events, which were not so well elaborated at the time of encoding, may be more likely to be lost over time, necessitating greater use of mental-model reconstructive strategies.

The regression analyses demonstrated that trust is indeed a very important predictor of memory biases. To look at its effects more closely, we ran analyses of variance examining changes in both feelings toward partner and attributions as a function of trust. Valence of event and time of rating were treated as

TABLE 17.2. Predictors of Recall of Feelings toward Partner on Event Day

Event	$\beta$	$F$	$p$
Positive event			
Actual Time 1 feelings	0.382	29.72	.0001
Trust	0.378	21.34	.0001
Time 2 feelings	0.109	3.16	.08 <sup>a</sup>
Negative event			
Actual Time 1 feelings	0.543	45.42	.0001
Trust	0.529	12.54	.001
Event severity	-0.690	10.01	.002
Time 2 feelings	0.230	3.62	.06 <sup>a</sup>

Note: Correlation of trust and Time 2 feelings is .34 for both events.

<sup>a</sup>Effect removed if entered last in model.

repeated-measures variables. Trust was cut using a median rather than the usual tertile split because previous research (Holmes & Rempel, 1989) suggested that the lower trust group in the present sample was better seen as medium rather than low trust. In fact, the average marital satisfaction rating of the medium-trust couples was 6.1 on a 7-point scale. Thus, these were *not* unhappy couples; they were simply somewhat insecure or uncertain about some aspects of their relationship. The results of the analyses of variance are shown in Table 17.3.

The top panel of Table 17.3 shows the analysis of feelings toward one's partner. Feelings toward partner form the affective core of one's mental model. We see, not surprisingly, that high-trust individuals felt more positively toward their partner than did medium-trust individuals on both the positive- and negative-event days. This difference in feelings at Time 1, as we predicted, becomes accentuated at Time 2. High-trust couples are able to hold on to their relatively positive feelings about their partners over time, despite the tendency for extreme scores to regress toward the mean. Lower trust couples, in contrast, may be using their more negative, less confident mental models to reconstruct their earlier feelings and therefore become even more negative over time.

The middle and bottom panels of Table 17.3 show the results of the attributional analyses. Again, highly significant effects were found for trust. Higher trust subjects gave much more generous attributions than lower trust couples, for both positive and negative events. We also see differential changes in these attributions over time in the two groups. Looking first at the positive events, we see that high-trust individuals maintained their generous attributions in their recall, again despite the tendency for regression toward the mean to pull down extreme scores. Medium-trust individuals, in contrast, were lower at Time 1 and dropped even more when recalling their attributions, increasing the distance between themselves and high-trust individuals. For the negative events, the medium-trust respondents actually maintained their very negative evaluations over time, against the pull of regression to the mean. High-trust individuals, in contrast, became more positive in their motive ratings for the negative event over time.

Overall, we see evidence of lower trust individuals displaying a pattern of "distress-maintaining" attributions at Time 1, which only becomes accentuated at Time 2. Sadly, the insecurity that colors their mental models is then perpetuated and reinforced by their biased negative memories. High-trust individuals, on the other hand, tend to exhibit and maintain a pattern of "relationship-enhancing" attributions (Holtzworth-Munroe & Jacobson, 1985).

The only result that does not fit this overall pattern is the drop in high-trust respondents' globality ratings, indicating that they became less generous in recall. This result can perhaps be explained by recalling the unexpected strong negative correlation between their motive and globality ratings at Time 1. If subjects really were being defensive in asserting that their partners' most negative acts were the least informative about the relationship in general, then it is possible that they were losing some of that defensiveness in recall. When they had achieved some distance from the event, they might have been able to see in

TABLE 17.3. Memory for Feelings and Attributions

Trust Level	Positive Event		Negative Event	
	Time 1	Time 2	Time 1	Time 2
	Feelings <sup>a</sup>			
High	6.39	6.27	4.83	4.76
Medium	6.08	5.73	4.50	3.94
	Attributions: Motive Ratings <sup>b</sup>			
High	5.54	5.44	3.21	3.47
Medium	5.30	4.99	2.52	2.49
	Attributions: Globality Ratings <sup>c</sup>			
High	4.08	4.21	3.41	3.14
Medium	3.68	3.41	2.81	2.93

<sup>a</sup>High scores indicate more positive feelings. Scale 1–7, midpoint 4. Effects: Trust  $p < .005$ ,  $T \times Tm = .05$ .

<sup>b</sup>Scale 1–6, midpoint 3.5. Effects: Trust  $p < .0001$ ,  $T \times Tm = .03$ .

<sup>c</sup>High scores indicate more positive conclusions. Scale 1–5, midpoint 3.0. Effects: Trust  $p < .0001$ ,  $V \times Tm \times T p < .001$ .

retrospect that it did have at least some meaning for the relationship as a whole.

A result that deserves further comment is the asymmetry in results for positive versus negative events for the medium-trust couples. Perhaps contrary to intuitions, given their relatively high satisfaction levels, these “uncertain” respondents have a model that makes it difficult for them to “hold on” to positive conclusions but easy to maintain very negative conclusions. These results are quite reflective of previous research on medium-trust individuals (Holmes & Rempel, 1989). Faced with feelings of insecurity and uncertainty, these individuals engage in *hypothesis testing* concerning their partner’s caring and responsiveness. Events are continually examined to determine their meaning with respect to core relationship issues. In past studies, such couples have exhibited a *risk-averse* bias in their hypothesis testing. They are particularly conservative in reaching broad generalizations on the basis of specific positive events. The risk of being hurt by drawing an unwarranted positive conclusion is too great. The current results show that even the hesitant positive attributions drawn at Time 1 tend to be lost over time. On the other hand, insecure individuals tend to be vigilant for signs of negative behavior and are prepared to consider the possibility that such behavior is indicative of their partner’s lack of concern for them. They make relatively extreme negative attributions, and the current results show that they are able to maintain these extreme attributions over time, even given the countervailing pressure of regression toward the mean. This differential response to positive and negative events suggests that mental models cannot be thought of as simple unidimensional structures, where the partner is seen as

either good or bad, and memories are shaped accordingly. Instead, we see evidence of specific, context-sensitive information—that is, “My partner has let me down in the past; therefore, I can’t allow myself to get my hopes up over one little positive action.”

Overall, then, the results of Study 1 suggest that (a) mental models are important predictors of biased recall, (b) they seem to be more important predictors than concurrent mood or feelings for partner, and (c) their biasing effects are systematic and context sensitive.

## Study 2: Long-Term Reconstruction

Study 2 had three major goals. First, we wanted to see if the biasing effects of mental models could be observed in open-ended responses. (We did collect open-ended data in Study 1, but it has not been fully analyzed). Second, we wanted to see if mental models could have an impact on memory for major events in the course of the relationship, as well as on the more focused daily interactions recorded in Study 1. Third, we wanted to explore the effects of mental models as they change over more extended periods of time.

Study 1 measured mental models at only one point in time and assumed that they would remain stable over the course of the study. In fact, the correlation of trust scores over a 2-year period in a large sample of married couples was .72, suggesting that this assumption was a reasonable one. Mental models are conceptualized as complex cognitive structures built up over the course of many interactions, and they are unlikely to be drastically affected by any single event or set of events. In fact, the results of Study 1 show that people’s memories change in ways that reinforce and perpetuate their existing models. However, models will not necessarily stay the same indefinitely. A once-trusted partner may engage in an extended series of actions that gradually erode his or her credibility. Given a sufficient time frame and the appropriate sample, we are quite likely to find a number of individuals whose mental models have undergone a major shift. The question that then arises is whether these individuals rewrite their entire past, including the time when their relationships were actually very positive, to bring them more into line with their revised models.

## Method

Data from the Early Years of Marriage project, a 4-year longitudinal study of newlyweds conducted at the University of Michigan (Joseph Veroff, Elizabeth Douvan, and Shirley Hatchett, principal investigators), allowed us to address these questions. Huston, McHale, and Crouter (1986) have shown that in the first few years of marriage, “after the honeymoon is over,” a significant drop in marital satisfaction occurs on average. A newlywed sample is therefore particularly likely to contain individuals whose mental models of their partner will change over the first few years. In the Michigan study, 373 couples (199 black,

174 white) who had applied for marriage licenses in Wayne County, Michigan, took part in a longitudinal study of couples' adjustment to the early years of marriage. To be eligible for the study, it had to be the first marriage for both partners, they had to be of the same race, and the wife had to be less than 35 years old (to allow for the possibility of childbearing during the course of the study).

The first wave of data collection took place 3 to 7 months after the couple was married. Couples were interviewed in their own homes. They first responded to extensive standardized interviews, which asked the partners separately about many aspects of their married life. Couples also provided data in the form of a narrative interview. Here, couples were asked together to tell the story of their relationship in their own words, from the time when they first met up until the present, and on into the future. The couples were encouraged to tell the story in any way that seemed most comfortable to them; they were both encouraged to participate, and it was stressed that they could agree or disagree about any point in the story. Two years later, in the third year of their marriage, couples were recontacted and asked to respond to the same standardized interview. Narrative data were also collected, again asking couples to tell the story of their relationship from the very beginning. Thus, for the early stages of their relationship (courtship, wedding, and honeymoon), we have their memories for the same events, collected 2 years apart.

One of us (Holmberg), along with Joseph Veroff, conducted a study to explore changes in these memories for early relationship events over the 2-year period. The sample for the study consisted of 26 couples (16 black, 10 white) drawn from the larger population. Half of these couples were selected because they had experienced a sizable decline in their overall marital well-being over the 2-year period (at least half a standard deviation for each partner, average drop = 1.4 *SDs*). The other couples were matched with this "decline" group on Year 1 well-being; however, these couples remained stable in their marital well-being scores over the 2-year period (i.e., did not vary up or down more than 0.15 *SDs*).

This well-being score provides an approximation of individuals' mental models, albeit at a rather general summary level. It contains information regarding the person's overall view of the relationship, focusing on both satisfaction and trust. Again, it contains both affective (i.e., "How happy is your marriage?" "How satisfied are you with your marriage?" "When you think of what you put into and get out of your marriage, how happy are you?") and cognitive/evaluative components (i.e., "How stable is your marriage?" "How certain are you you'll still be married in 5 years?" "How often in the past month have you considered leaving your spouse?").

Coders blind to experimental hypotheses listened to tapes of the couples' narratives from Year 1 and Year 3 and coded them for their affective content. Each time that one of the partners made an affective statement, it was recorded and then coded on a number of dimensions, including *valence* (Is it a positive, negative, or neutral emotion?), *speaker* (Who is mentioning the emotion?),

*source* (Who is experiencing the emotion?), *object* (Whom or what is the emotion directed toward?), *context* (Where in the story does the emotion take place?), and *motivation* (Is the motivation underlying the emotion agentic, communal, both, or neither?). In this study, only statements describing the courtship, wedding, or honeymoon were analyzed. This portion of the story is common to both the Year 1 and Year 3 narratives.

We chose to focus on these affective statements in our initial coding because we believed that they might be particularly sensitive in picking up memory changes. Memories could undergo changes at several levels. First, one's memories for the actual details of events might change. Someone with a very negative mental model of his or her spouse could potentially forget a positive act by the partner completely or could "recall" a negative act that never really happened. Alternatively (or perhaps additionally), partners could recall the actual events more or less as they really occurred, yet their interpretations of those events could change dramatically. A negative act that was initially construed as simply somewhat thoughtless or careless might later be remembered as having reflected malicious intent (i.e., an attributional shift, similar to those seen in Study 1). These different interpretations would certainly arouse different emotions in the partner. We therefore believed that the affective statements would be particularly sensitive indicators of subtle changes in the tone or meaning of the memories for the speakers.

We had a number of specific hypotheses. Among them, we predicted that compared with the stable well-being group, couples who showed a decline in marital well-being over the 2-year period (i.e., whose mental models became more negative over time) would show an increase in negative affective statements when describing the early stages of their relationship. This increase in negative affect would be particularly apparent in statements where the partner was the object, indicating that the negative bias in affect does not extend to all memories but only to those tied in to the now more negative mental model of the partner. We expected no differences between the groups in Year 1, when they had equally positive mental models.

We also expected to see changes in the motivational underpinnings of the emotions expressed. According to Bakan's (1966) agency/communion distinction, an agentic motivation is a self-expressive motivation. It represents a desire to be your own person, to achieve your own goals, to be independent. A communal motivation, in contrast, represents a desire to be connected or related to other people. Liking, loving, and depending on others all reflect a communal motivation. Statements may also reflect elements of both motivations simultaneously. This usually occurs when there is a conflict between a person's need for self-expression and his or her communal needs. Emotions aroused by fighting or arguing would be one example. When a couple argues, both members want to express their own point of view, yet still keep the larger goal of maintaining and improving the relationship in mind. These "both" statements essentially express ambivalence—trying to fight out a compromise between own needs and relationship needs.

We expected to see a rise in memories for agentic emotions and a decrease in communal emotions over the 2 years for couples in the decline group, suggesting more of a concern with their own needs and goals and less of a concern with the closeness of the relationship. Recalled through the filter of a more negative mental model, their past paths might now seem more separate and less intertwined. We also predicted a rise in the number of ambivalent statements, reflecting both agentic and communal motivations. These statements reflect a conflict between own needs and relationship needs—a state of increased psychological distance that one would expect to be more evident in the decline group.

## Results

To test these hypotheses, we ran a series of  $2 \times 2 \times 2$ —group  $\times$  time  $\times$  gender—ANOVAs. Group was the “decline in well-being” group versus the “stable well-being” group; time was a repeated-measures factor, with two levels, Year 1 and Year 3; and gender was also treated as a repeated-measures, or within-subject, variable. This technique is common in studies of married couples, where one expects husbands’ and wives’ responses to be correlated. Race was initially included in all analyses. However, since it never entered into the interactions of interest, it was removed from the final version of the analyses.

The results of our first analysis are shown in Table 17.4. The dependent variable for the analysis was the percentage of all of the speaker’s affective statements that were negative. The total number of negative affective statements was not used as the index because preliminary analyses uncovered differences between the groups in how much affect they revealed overall. To control for these discrepancies, we divided the number of statements in the relevant category by the total number of affective statements for that speaker. As Table 17.4 shows, results were as predicted for the males: The two groups were approximately equal in the percentage of negative statements in Year 1. By Year 3, however, there was a large difference between the groups. The decline group had become more negative in their memory over time, while the stable group had become less negative. By Year 3, the difference in percentage of affect that was negative was almost 2:1. In contrast, there were no significant differences for the females.

We also broke our analysis down by the object of the emotion—whom or what the emotion was directed toward. As predicted, there were no significant effects when the object was self or some outside person or situation. Instead, the effects were confined to negative affect directed toward the partner. The pattern of results was identical to that shown in Table 17.4. For males, the two groups were the same in Year 1 but very different in Year 3, while there were no significant differences for the females. Males, it seems, fit in very well with our theory. As their mental models became more negative, the negativity colored their memories of the early stages of their relationship. Their stories were tinged with more negative affect, and they especially remembered their partner as having inspired more negative affect, even at a point in time when they were happiest.

TABLE 17.4. Proportion of Emotions That Were Negative

Group	Males		Females	
	Year 1	Year 3	Year 1	Year 3
Decline in well-being	.27 <sup>ac</sup>	.35 <sup>b</sup>	.32 <sup>a</sup>	.32 <sup>a</sup>
Stable well-being	.28 <sup>a</sup>	.18 <sup>c</sup>	.25 <sup>ac</sup>	.32 <sup>a</sup>

$F = 7.49, p < .02$ ; means with common subscripts do not differ at  $p < .05$  using Tukey's HSD test.

Why did men show differences here, while women did not? When interpreting these differences, one should keep in mind the public nature of people's narratives. Their memories were recounted together in the presence of a stranger, the interviewer. Deaux (1976), Parsons and Bales (1955), and others have suggested that women in our society are socialized to be more concerned with interpersonal relationships and are more aware of the feelings of others than are men. If women are indeed more aware of others' feelings, they may be less likely to accentuate the negative affect in their stories for fear of hurting their spouses. And if women are seen as relationship "experts," they may be less inclined to accentuate the negative aspects of the past when facing current problems, in order to avoid looking like a failure in front of the interviewer. This self-presentational account receives some support from the fact that we did not observe male-female differences in the amount of memory bias in Study 1, where partners' responses were private.

In addition, if self-presentational concerns account for the male-female differences in recall bias, then one would predict that no such differences would emerge when the measure is more subtle, one that is less easy to control. That is indeed what we found. The results of the motivation coding showed differences in the number of "ambivalent" statements, reflecting both agentic and communal motivations at the same time. Again, these statements generally indicate problems in reconciling one's own agentic needs and communal or relationship needs. As predicted, we saw an increase in these ambivalent statements for the decline group, while the stable group actually decreased slightly (see Table 17.5).

## Summary

In summary, in our first study there was evidence that mental models lead to changes in recall that reinforce and perpetuate preexisting mental models. In the short term, then, these models are likely to be relatively stable and impervious to sudden changes (unless the act is of such import that it calls into question one's entire belief structure about the partner—infidelity might be one such event.) Over the course of the 2 years covered in Study 2, however, we saw substantial changes in the nature of some subjects' mental models, presumably reflecting

TABLE 17.5. Proportion of Emotions That Reflect Both Agentic and Communal Motivations

Group	Year 1	Year 2
Decline in well-being	.09	.20
Stable well-being	.12	.06

$F = 6.44, p < .05$ ; all means differ at  $p < .05$  using Tukey's HSD test.

changes in their actual interactions. The frightening aspect of these data is that subjects then showed corresponding changes in their memories for the earliest stages of their relationship, when they had once reported that they were happy. Their new, more negative mental models were reinforced again by memory biases: Not only are things bad now, but they have always been bad. Such biases can lead to a dangerous downward spiral. The worse your current view of your partner is, the worse your memories are, which only further confirms your negative attitudes.

Furthermore, as data from Study 1 show us, it may be difficult to escape from this spiral once it has begun. Those individuals with more negative mental models of their partners were reluctant to draw inferences from positive events, a trend only accentuated in the recall data. Such an asymmetrical pattern of attributions can make it almost impossible to return to a stable, positive view of a partner. If a partner is blamed for failures but is not given any credit for good deeds, how can he or she ever prove worthy of trust again?

## Some Concluding Thoughts

There are a great many questions that still remain to be answered about the effects of mental models on memory. One important question lies in the area of the exact nature of the memory change. As noted previously, people's memories for details of the actual events could be changing, or they could be remembering the same events and only changing their interpretations. Direct comparison of protocols for open-ended depictions of specific events at two points in time is needed to address these questions. Fortunately, these data are available in the Study 1 data set. Unfortunately, the coding is extremely time consuming, and only very preliminary analyses have been completed to date. These analyses suggest that subjects are in fact changing the actual details of their memories for the events. The proportion of "hits" (correct recall of thought units) is low, whereas the proportion of misses and false alarms is much higher.

Another pattern that has emerged from preliminary analyses is the powerful effect of event valence on the valence of details that are recalled and on the valence of details that are reconstructed. When the event is a predominantly negative one, for instance, people are better able to recall correctly the negative

details than the positive or neutral details. Their false alarms are also predominantly negative. Thus over time, shades of grey are lost and events are painted more in black and white. These effects may be due largely to the priming of elements consistent with the retrieval cue valence. However, the issue is probably more complex than that: For example, not only is an argument a category of event in terms of affect, but the prototypical script that it elicits will largely consist of negative features that are then available as guides to reconstruction.

These effects are tempered somewhat for the high-trust individuals. For example, only 51% of the details that they recalled correctly about the negative event were negative; the comparable figure for medium-trust individuals was 72%. Thus, trust does have a strong impact, but not even a high-trust individual will recall a negative action as being essentially positive, at least not within the time frame that we considered. Their mental models are more specific and contextualized than that: They are sensitive to the nature of the original event.

The open-ended coding will allow us to explore the issue of event valence more thoroughly and also to shed light on many other questions about the exact nature of the memory change. Unfortunately, interpretations of even these data are somewhat ambiguous. We know only the details of the target interaction that the respondent initially recorded. Thus, although our hit data are "clean," it is possible that some of our so-called false alarms are actually true memories of details of the event that did occur but that were not recorded at Time 1. Laboratory studies, in which we can record and/or control all details of the Time 1 interactions, may be needed to answer some of our questions with confidence.

Further study and theoretical work are also needed to understand the nature of the processes underlying the workings of these mental models. How exactly do the mental models work their effects on memory? One possibility is suggested by Higgins' chronic accessibility model (Higgins, 1990). This model posits that mental structures that are accessed very frequently (and one's views of spouse would be a prime example) become "chronically accessible." They are always near the top of the mental stack, ready to be accessed in many different situations. Of course, in the present studies, subjects were specifically asked to recall situations that they viewed as relevant to their feelings for their partners, a prime that should have further increased the salience of their mental models. When one tries to fill in gaps in memories, one's current mood and feelings toward the person in question are also salient cues that might be used in recall, as mood-congruence theory and Ross's theory suggest. However, if one has a chronically accessible mental model of the person in question, the present evidence suggests that it will indeed be used as an aid in reconstruction.

## Implications

Our mental models theory of memory reconstruction may prove useful in its own right, in terms of understanding memory processes. Its effects on memory for events in marriage also have important clinical applications. Marital thera-

pists may well be advised to avoid recapitulating past events as a lever for change until central issues of trust and attachment between the partners have been worked out; otherwise, recalling negatively biased past events may only make matters worse.

Survey researchers could potentially reap benefits from the mental models approach as well. If a person's current mental model has a stable biasing effect on his or her recall, then one could, in theory, use measures of the person's current mental model to correct for these biases, thus arriving at a recalled score that is closer to the person's actual original response.

More importantly, however, survey researchers might learn to look at these memory biases not as "errors" to be corrected but instead as interesting data in their own right. Many large surveys are panel studies, asking respondents similar questions at multiple points in time. If a few questions that ask respondents to recall previous responses were included, one could get interesting information on memory biases in large samples with very little additional effort. One could then work backward, inferring the nature and strength of subjects' current mental models from the degree of memory bias exhibited. This method could provide an efficient way of indirectly assessing attitudes that subjects may be unable or unwilling to articulate directly. In short, the concept of mental models, once it is more fully understood, may prove to be a powerful tool in survey research.

*Acknowledgments.* This chapter was prepared with the support of a Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowship to Holmberg and a SSHRC Research Grant to Holmes. Research reported in Study 2 was supported by a grant from the National Institute of Mental Health. Many thanks to members of the Early Years of Marriage research project at the University of Michigan for their contributions to this study. Special thanks to Joe Veroff for his valuable collaboration on this research.

**Part V**  
**Memories of the Past and Judgment**  
**of Personal and Social Change**



# 18

## Biasing Effects of Retrospective Reports on Current Self-Assessments

*Leslie F. Clark, James E. Collins II, and Susan M. Henry*

Our memories of past life experiences may alter our mood and influence our sense of happiness or satisfaction with our current lives (Strack, Schwarz, & Gschneidinger, 1985; Tversky & Griffin, 1991). The effects of recounting autobiographical memories on judgments are worthy of investigation because these judgmental biases are inherent in many survey and interview contexts (e.g., Salovey, Sieber, Jobe, & Willis, chap. 6, this volume). Retrospective reporting often requires the written descriptions of a specific past life experience. After such reporting, individuals may also be prompted to assess their life satisfaction, happiness, or well-being. This chapter focuses on retrospective reporting tasks and the consequences that such tasks have on respondents' subsequent assessments of *current* personal well-being.

We first describe research examining differential consequences of two types of description tasks. Next, we outline theories used to explain the occurrence of these effects. Finally, we discuss the implications that these different tasks have for survey response.

### Episodic versus Abstract Recounting

An autobiographical event can be described in terms of vivid and concrete details depicting how the event occurred. However, this same event can also be described in terms of what the respondent believes to be the underlying causes and reasons for the event. We shall refer to these two modes of responding as an "episodic" focus and an "abstract" focus, respectively.

*Episodic recounting tasks* use instructions that focus respondents on the details of the past event. Respondents are encouraged to recount exactly what happened to the best of their ability. Details may include the time, the location, dialogue, and the respondents' personal reactions to the event as it occurred. If recounting requires the respondent to note the temporal sequence of actions or events, it too is an episodic task.

*Abstract recounting tasks* are those that tap a respondent's causal analysis of a given past event. Respondents may be asked to enumerate the perceived causes and reasons for a past life event's occurrence, but they are *not* asked to focus on episodic details.

## Related Processes

Conceptually, the task of episodic recounting may share features with the process of ruminative thinking, which involves the experience of thoughts that involuntarily come to consciousness. These thoughts have been described as a mental "reliving" of the past event. Crime victims, war veterans, and parents bereaving the death of a child freely report experiencing ruminative thoughts (Horowitz, 1976, 1982; Meyer & Taylor, 1986; Silver, Boon, & Stones, 1983). This rumination about such negative life events is associated with increased stress and emotional distress (Horowitz, 1976, 1982; Silver et al., 1983). Rumination also appears to interfere with attention and concentration (Mandler, 1982; Nolan-Hoeksema, 1987) and to enhance the occurrence of other negative memories (Nolan-Hoeksema, 1987).

Tasks that encourage abstract recounting are conceptually related to the causal analysis that may be experienced by individuals who attempt to adjust to a traumatic life event. This process is referred to as a "search for meaning" and has been noted in cancer patients and incest victims (Silver et al., 1983; Taylor, 1983). Interestingly, such an interpretive process has been considered as a coping strategy (Lazarus & Folkman, 1984; Silver & Wortman, 1980; Taylor, 1983; Taylor & Schneider, 1989).

Contrary to rumination, a causal analysis of negative life events has been related to lower psychological distress (Silver et al., 1983) and higher self-esteem (Silver et al., 1983; Weber, Harvey, & Stanley, 1987). Individuals may also perform a causal analysis of past positive life events, which probably happens less frequently than analysis of negative events. Such an abstract focus in questioning positive life events has been found to be associated with depression (Berman, 1988; Nolan-Hoeksema, 1987). It is important to note that survey wording may invoke processes similar to rumination and search for meaning.

## An Example of Recounting

In the studies reviewed in this chapter, we and other researchers investigating the effects of these recounting tasks instructed respondents to (a) recount either a pleasant or unpleasant event and (b) use some manipulation of the episodic versus abstract distinction. Respondents then wrote relatively brief descriptions of a personally experienced life event.

Table 18.1 presents example protocols from such a survey in which we examined the effects of recalling past love relationships by college students (Clark & Collins, in press; Collins & Clark, 1988). In this survey, the respondents were instructed specifically to recount either falling in love or falling out of love with

TABLE 18.1. Example Protocols from College Students Recounting Past Love Relationships

---

Condition:	How Focus: Falling in Love
Response:	A friend of mine introduced me to her best friend. She was from California and attended L.S.U. and was just visiting in Memphis. We were at a bar and I asked her to dance. We hit it off real well and she spent the rest of her visit with me. She then came in town every weekend on the train to see me from L.S.U. I wrote her a lot of letters and she wrote me too. I guess that is how we fell in love.
Condition:	Why Focus: Falling in Love
Response:	I think I fell in love with this person because he was first of all physically attractive. He also was extremely fun to be around, always a good time. There was just something about him that I was attracted to. He definitely wasn't my type but he wasn't too serious, he had a sense of humor, and he made me feel happy. He was also a Christian and I was looking for a Christian to date.
Condition:	How Focus: Falling out of Love
Response:	The girl I was dating invited me to take her to see her father. When we got there it was late and I only got to meet half the family. She showed me off and then wouldn't talk to me, touch me (hold me or my hand). I felt so alone. I got the chance to see the whole thing (weekend) as if it weren't me. I realized that the way I was treated was because she wasn't in "love" at all. It was just a casual/sexual relationship. This crushing awareness made me feel "the fool" and that I had not been sensitive toward what she was or what she wanted from me.
Condition:	Why Focus: Falling out of Love
Response:	The reason I fell out of love was because the trust, the understanding, and the security I felt with him had begun to deteriorate. Not only were we drifting apart mentally and emotionally, but sexually as well. The once fulfillment of becoming one was no longer enjoyable.

---

their past partner. The reader will notice differences between the two protocols that describe falling in love. It may be apparent that neither of these protocols is a pure prototype of episodic or abstract recounting; that is, both include specific details (a product of episodic recall) and both include conclusions generated by the respondents (indications of an abstract recounting). However, the episodic account (how) more clearly describes a specific event, and the abstract account (why) focuses on the reasons for falling in love with that particular person.

## Consequences of Episodic versus Abstract Recounting

The episodic-abstract dimension has been manipulated in several studies addressing the impact that remembering autobiographical events has on self-

assessments of current well-being. These manipulations can be categorized in terms of (a) description length, (b) temporal distance of the event, and (c) descriptions of how the event occurred versus why it occurred. We discuss, in turn, research using each of these manipulations.

### Indirect Manipulations

It was reasoned that encouraging respondents to write longer, more vivid accounts of the past events would lead them to include vivid and concrete details of what happened. Shorter (pallid) description tasks, although not specifically encouraging an abstract focus, might lead respondents to use simple labels of the event, thereby blocking a more episodic recounting of what happened. The temporal distance of the event has also been manipulated, such as comparing present versus past events or comparing recent past versus distant past events, the assumptions being that respondents should have a clearer, more episodic memory of more recent events. Memories of events further in the past are more likely to consist of fewer episodic details and to contain more abstract features such that the respondent's stored conclusions and the reasons underlying the event's occurrence.

### Vivid versus Pallid Descriptions

The vividness of respondents' descriptions were manipulated in a study by Strack et al. (1985, Experiment 2). In a vivid-recall condition, respondents were instructed to recall a past event and to re-experience the event as vividly as possible, including how they felt at the time and what led to these feelings. In the pallid-recall condition, respondents were asked to recall three separate events and to describe them briefly in only a few sentences.

The vivid-recall instructions led respondents to report more positive moods after describing a pleasant event than after describing an unpleasant event. Pallid-recall instructions produced no mood differences based on the pleasantness of the past life event. Furthermore, respondents given vivid-recall instructions reported current life satisfaction that was affected by the hedonic valence of the past event; that is, remembering positive life events led respondents to give higher current satisfaction ratings than did remembering negative life events. Pallid-recall instructions, on the other hand, created contrast effects on this same measure of current life satisfaction. In other words, recounting positive past life events led to lower current life satisfaction ratings than did recounting negative past life events.

### The Temporal Distance of the Event

How long ago the event occurred determines, in part, how remembering will affect judgments of one's current life. Although Strack et al. (1985, Experiment 2) found that remembering a *present* event influenced judgments of current satisfaction in the direction of the remembered event's valence (i.e., positive events

led to positive satisfaction ratings), remembering a *distant* past event produced a contrast effect in judgments of life satisfaction. Therefore, distant positive-event recounting lowered satisfaction ratings compared with distant negative-event recounting. This effect was replicated by Tversky and Griffin (1991) using recent past versus distant past events.

These findings have been explained in terms of how representative the past event is perceived to be to one's current life (Strack et al., 1985; Tversky & Griffin, 1991). Present events or past events that have only recently occurred are seen as being more representative of one's current life and hence are more likely to be used as the basis for judging one's current satisfaction. Past events, on the other hand, are not seen as a sufficiently representative basis for judgment. Instead, memories of distant past events appear to be used as a standard against which individuals compare their current life.

We found this same crossover interaction pattern for life satisfaction in an elderly sample of retired individuals over age 60 (Clark, Henry, & Collins, 1991). In this case, respondents were asked to describe a pleasant or unpleasant life event from their recent past (5 to 7 years ago) or their distant past (25 to 30 years ago). Individuals describing a relatively recent past event assimilated their self-assessments of current life satisfaction to the valence of the past event. On the other hand, individuals describing an event occurring more than 25 years ago showed life satisfaction ratings that contrasted to the valence of the past event that they had described. In a more direct manipulation of episodic versus abstract recounting, researchers have instructed respondents to focus on how the event occurred (episodic focus) versus why the event occurred (abstract focus).

### Recounting Details versus Conclusions: Answering How versus Why

The distinction between abstract and episodic recounting has been examined directly in a study by Strack et al. (1985, Experiment 3). They defined thinking in episodic terms as remembering the details of the event and *how* it unfolded. Thinking about the past in abstract terms, on the other hand, involved considering the causes and reasons for *why* the event happened.

The instructions for the how-focus conditions were as follows:

Describe in detail how the event occurred. We are not interested in why this happened, but rather how it came about.

The instructions for the why-focus conditions were:

Describe in detail why the event occurred. We are not interested in how the event came about, but rather why it happened.

The how-focus instructions were intended to engage participants in a "mental reliving" of the experience, hence serving to elicit vivid descriptions. The why-focus instructions, on the other hand, were intended to elicit an *explanation* of the experience, including the relevant causes and reasons for what happened.

Strack et al. (1985) found that thinking about *how* a past event occurred led respondents to experience mood changes congruent with the valence of the

event. In addition, focusing on the details of a negative past event lowered reports of current life satisfaction, whereas focusing on the details of a positive past event was associated with higher current life satisfaction.

Thinking about *why* a past event occurred, however, led individuals who described negative events to report higher life satisfaction than did subjects who described positive events. Strack et al. (1985) suggested that individuals using the abstract mode tended to draw a comparison between that past event and their current life situation. Therefore, if they remembered a negative event in an abstract fashion, their life seemed better by comparison, whereas if they remembered a positive experience, their life appeared worse by comparison.

In a study of remembering personal events among the elderly, we again manipulated episodic versus abstract recounting with the use of “how” versus “why” instructions (Clark, Henry, & Collins, 1991; see also Clark, Collins, & Henry, 1991). We found that a how focus in recounting led to judgments of how respondents felt in the present that matched the affect associated with the past event. On the other hand, a why focus led elderly respondents to judge their current feelings of good or bad quite differently: They contrasted their present life circumstances with the valence of a recalled past event. Consequently, recounting past negative events led to judgments of more positive current feelings than did recounting past positive events.

In a separate survey sample of retired individuals, respondents reported life satisfaction using a standardized measure of life satisfaction in which they answered “true” or “false” to statements such as “As I grow older, things seem better than I thought they would be” (V. Wood, Wylie, & Sheafor, 1969). These results showed the same effects of episodic versus abstract recounting. Specifically, how-focus recounting led to assimilation effects on this standardized measure of satisfaction, whereas why-focus recounting led to contrast effects on this same measure. The means for these two surveys using elderly samples are presented in the top half of Table 18.2.

## Explanations of the Assimilation and Contrast Effects

In this section, we discuss explanations for these assimilation and contrast effects that go beyond the mere issue of whether or not an event is seen as representative of one’s life situation. Specifically, we address three explanations that may be applied to the assimilation and contrast effects described in this chapter.

### The Use of Mood as the Basis of Judgment

Besides being seen as representative of one’s current life, past events can have assimilation effects on judgments in another way. Remembering a past event may reinvolve feelings based on the pleasantness (or unpleasantness) of the prior event, thus leaving one with a positive (or negative) mood state. When asked to

TABLE 18.2. Means for How and Why Focuses, by Event Valence Interaction

Self-Assessment	How Focus		Why Focus	
	Positive	Negative	Positive	Negative
Elderly Samples				
Elderly Survey 1				
Feel at this moment <sup>a</sup>	2.50	2.04	2.03	2.39
Elderly Survey 2				
Life Satisfaction Index— Z form for the elderly <sup>a</sup>	10.39	9.00	9.63	10.54
College Sample				
Love relationships				
Current relationship satisfaction <sup>a</sup>	6.81	5.44	6.63	7.56
Current life satisfaction:				
Understood relationships <sup>a</sup>	7.75	6.50	6.25	7.88
Not understood relationships	6.68	6.63	7.38	6.00

<sup>a</sup>Indicates that the planned comparisons showed that both assimilation and contrast effects in the crossover interactions were significant ( $p < .05$ ).

assess current life satisfaction, individuals may mistakenly assume that this mood is reflective of their current life situation and not due to the process of remembering a past positive or negative event.

Schwarz and Clore (1983) demonstrated this assimilation effect in a telephone survey with moods affected by the weather. Individuals reported more positive moods and higher life satisfaction on sunny days than they did on rainy days. However, first asking individuals about the weather made them aware that their mood might be elicited by weather conditions and hence mood would be an irrelevant source of information on which to base their judgments of life satisfaction. Once individuals were made aware of the weather, mood ceased to have an impact on their judgments of life satisfaction (see also Clore & Parrot, 1991).

Schwarz (1990b) talks about such effects in terms of individuals basing judgments about their life on currently available information in memory. Mood is one such source of information. Although one's mood may be positive or negative owing to factors irrelevant to the judgment at hand (e.g., weather, memories of past events), individuals may mistakenly perceive their mood or feelings as providing relevant information for the current judgment. Schwarz (1990b) suggests that "feelings-as-information" effects are due to individuals using a simple heuristic, namely, "How did I feel about it?" to make a relatively complex judgment (e.g., life satisfaction).

### The Use of a Past Life Experience as a Standard for Judgment

The "contrast" function of past events for current judgments is said to occur because of a comparison process (Tversky & Griffin, 1991). Specifically, indi-

viduals use information regarding the pleasantness or unpleasantness of a past event as an anchor against which to compare their current life situation. In Strack et al. (1985), individuals were asked to describe past events using valence (either pleasant or unpleasant) as the only retrieval cue. Such a memory task probably produced events that were markedly pleasant (or unpleasant). This extremity of valence led to the use of this past event as an anchor of pleasantness (or unpleasantness) against which individuals then compared the relative pleasantness of current life events. After invoking a positive standard (recalling an extremely positive past event), individuals found their current life comparatively less satisfying. Conversely, individuals who recalled an extremely unpleasant past event set up a negative standard against which their current life seemed more positive by comparison.

### Autobiographical Memories as Primes

A relevant line of work used to explain both assimilation and contrast effects is the set/reset hypothesis of Martin (1985; Martin, Seta, & Crelia, 1990). The set/reset hypothesis maintains that previous tasks can serve to prime or “set” a particular category or concept that will be used in making later judgments. In our case, pleasant past events may prime (“set”) positivity, which then biases further responses in what appears to us to be assimilation effects. Priming or setting a positive valence leads to more positive responses, and priming a negative valence leads to more negative responses.

Under certain conditions, however, respondents may have the impression that the previous task will affect their subsequent responses. In an attempt to partial out the effect of the previous task, respondents mentally “reset” to a different category (e.g., *not* positive). Contrast effects then are due to an overcompensation of respondents attempting to avoid the influence of the prior task. This awareness of being primed need not be explicit for “resetting” to occur. It has been shown, however, that individuals must (a) be able to detect the inappropriateness of the prior task (e.g., remembering a life event) and (b) be both willing and able to expend the cognitive effort necessary for resetting to occur (Martin et al., 1990).

This interpretation would lead us to assume that abstract recounting makes respondents more aware of the inappropriateness of the recounting task as a basis for responses about their current life. Perhaps episodic recounting is so self-involving that respondents are unable to see that the threat of the task is coloring their subsequent judgments. It is also possible that episodic recounting is so involving that respondents do not have the capacity required for resetting. Abstract recounting may be less cognitively effortful, so that resetting takes place and contrast effects result. A set/reset explanation of the assimilation and contrast effects brought on by retrospective reporting is highly speculative. Further research that tests specific predictions based on this model is needed. However, this model does account for both assimilation and contrast, as well as specifying the conditions under which each should occur.

## Boundary Conditions for the Assimilation and Contrast Effects

Research on two factors shown to be prerequisites for these assimilation and contrast effects is described below.

### A Sense of Understanding about a Past Experience

One reason why more distant past events may invoke a standard of comparison is that we create stories to explain why certain social experiences turned out as they did (successfully or unsuccessfully). It is possible that such stories become more simplified with time; that is, more distant events can be seen more easily in abstract terms.

Depending on personal insight, cognitive effort, time elapsed, and a host of other moderating variables, people may be at very different stages of meaningfully integrating information about past events into a sense of understanding about what happened and why. In addition, for some events, a story of what happened and why it happened may not have been easily achieved. The social and personal meaning of some experiences may be complex and ambiguous. An example of such social events might be experiences that individuals have had with past love partners with whom they are no longer involved.

In an attempt to assess the impact of such prior understanding, we conducted a survey instructing college students to provide retrospective reports of past love relationships (Clark & Collins, *in press*). Love relationships, especially their initiation and dissolution, may often provide confusing social experiences. In general, love relationships have been found to require cognitive processes involved in understanding them (Harvey, Flanary, & Morgan, 1986; Perlman & Duck, 1987). Such processes might include attributional processes, multiple interpretations, and the integration of information into an ongoing cognitive impression of the relationship and its participants. When relationships end, individuals often attempt to understand by creating a story or "account" of what happened and why (Harvey, 1987).

We examined prior understanding as a prerequisite for the assimilation and contrast effects (Clark & Collins, *in press*). Respondents' current level of understanding was manipulated by having them describe a past love relationship that they either understood or did not. Individuals were defined as having prior understanding of their relationship if they could (a) predict becoming involved with their previous partner and (b) understand why their relationship had developed as it did.

This study sought to replicate Strack et al. (1985) by using judgments of one's current relationship as well as life in general. Respondents' satisfaction ratings of their current relationship showed the assimilation and contrast pattern. Specifically, there was a two-way interaction between thought focus and valence of experience. As shown in the bottom half of Table 18.2, those who described

the actual details of falling *out* of love with a past partner reported lower satisfaction with a current partner than did subjects who gave concrete descriptions of falling *in* love with a past partner. Furthermore, those who explained falling *out* of love in the past reported being more satisfied with their *current* partner than did those who described falling *in* love with someone.

Respondents' assessments of general life satisfaction was affected by their sense of understanding of the past love relationship. For understood relationships, perceptions of current general life satisfaction were a function of the assimilation and contrast effects described earlier. Satisfaction ratings were higher for how-focus than for why-focus conditions after participants discussed falling in love. Conversely, after describing falling out of love, why-focus participants reported higher satisfaction than did how-focus participants. This pattern was lacking for relationships that participants did not understand prior to the experiment; that is, these effects were disrupted when individuals lacked a sense of understanding about the past event. Prior understanding appears to be a prerequisite for the assimilation and contrast effects described by Strack et al. (1985).

Contrast effects on general satisfaction also disappeared when individuals lacked prior understanding. As in Tversky and Griffin's work (1991), individuals may have seen this past event as less representative (and hence less appropriate) to use as a standard against which to compare their current life satisfaction.

### Discontinuity: How "Old Times" Become a Standard

Strack, Schwarz, and Nebel (1987) have reported data that are relevant to the above finding. Specifically, they manipulated the discontinuity that students perceived between a future event (one that they anticipated occurring after graduation) and their present life. In the discontinuity condition, students were told that the event would be part of a new phase of their lives and that the future might be quite different from their lives now. Strack et al. (1987) found that this discontinuity instruction was necessary for the contrast effect to occur under a why-focus condition. Without the perception of discontinuity, individuals failed to use the anticipated event as a standard against which they could compare their current life satisfaction. It is possible that in the studies reported above, prior understanding of a relationship functioned in a way similar to Strack et al.'s (1987) discontinuity condition; namely, understanding allows one to put the event in the past by categorizing the event as being "behind them."

### Applications for Survey Measurement and Future Research

The format of a survey could invoke such categorization and the contrast effects that follow. However, if the respondent does not yet understand the event or still considers it a current concern, then a survey of this topic may prove frustrating to respondents and can even elicit assimilation effects. Similar to this interpreta-

tion, much of the research outlined in this chapter has direct implications for the survey researcher. We now turn to several issues of survey measurement.

Past research on survey measurement has attended to such issues as how a response scale and its verbal anchors can eliminate acquiescence and nay-saying (Alreck & Settle, 1985). The findings reported above raise new concerns. What may seem a subtle difference in the wording to the survey researcher can, in fact, greatly determine the memory and judgment processes brought to bear by a survey respondent.

In general, effects such as those reported in this chapter can enter survey data in two ways. First, survey questions can lead a respondent to sample from a domain of events that were experienced and encoded as vivid or emotionally intense. Such events may lead to strong episodic memories. Describing such strong memories would be likely to lead respondents to show assimilation effects on subsequent self-assessments. Memories of certain types of life events may be more likely to be stored in abstract terms, leading to subsequent contrast effects. Second, the format of a survey itself can determine whether an individual retrieves memories that were encoded episodically or stored in more abstract terms. In either case, the way in which retrospective reports are acquired may create rather than document the respondents' subsequent self-assessments.

### Phrasing the Initial Questions

Assessments of current mood and subjective well-being have been shown to be the result of whether a survey begins by asking an individual to recall past episodic details or to draw conclusions about the event. If first asked "how," then episodic memories will be tapped. If first asked "why," the nature of the recounting will be abstract. Self-assessments on different measures following how-type questions will assimilate to the valence of initially probed memories. Responses following why-type questions will show contrast effects, with later responses being contrasted away from the valence of the remembered event. Thus, as the research reported in this chapter indicates, the phrasing of initial questions could produce the pattern of responses to the remainder of a survey.

### Predetermining a Negativity Effect

The topic domain of events that are surveyed are subject to differential retrievability. It may be that certain recounting tasks are susceptible to oversampling salient past events. Salient events tend to be those that were negative events (although see Banaji & Hardin, chap. 5, this volume). The question for the survey researcher becomes: What expectations might we have about the way in which negative life events are stored and thus likely to be recounted?

#### *Tragic Events*

Unpleasant experiences often make salient memories. If negatively valenced events are overrepresented in response to free-recall survey items, and these are

recounted in episodic terms, then assimilation effects may dominate subsequent survey results. As was mentioned earlier, individuals experiencing negative events may attempt to understand why the event happened (see Schwarz, 1987), thus storing conclusions about the event in memory. Certain negative events then may be more likely to invoke such a “search for meaning” and, upon recounting, may lead to subsequent contrast effects.

### *A Robust Effect*

Even if participants are asked to sample or describe positive and negative events equally, the greater weighting of negative memories may still prevail. Survey researchers should thus note that instructions may appear neutral but may still set recounting in motion wherein the negative events form the basis for subsequent effects. Even if the overall pattern of subsequent responses is one of positive endorsement, it could still reflect a contrast effect from negative-event recounting. Future research is needed to delineate the separate and combined effects of surveying positive versus negative events.

## Begging a Contrast

### *Explained Events*

As mentioned above, certain events may lend themselves to causal analysis. If someone can understand why an event happened, then it may be stored in abstract, nonepisodic terms as well as in episodic terms. The result is a contrast effect: The responses following a negative recounting will be positive. For instance, surveys of bereavement may produce opposite effects depending on whether the pain is fresh or the loss has been understood or accepted.

If we assume that individuals problem solve more about negative events and do reach conclusions, then survey judgments concerning negative and understood events will not be based on mood. Instead, surveys that ask individuals to explain negative events may force mostly contrast effects.

Methodologically, a further complication deserves mention. The impact of techniques such as panel designs is unclear. Consider again bereavement or any process of adjustment. If someone is asked over and over in subsequent surveys to answer “why” a tragedy befell them, then such surveys may actually serve as interventions that hasten adjustment. If respondents are prompted to answer “how” it happened, then this particular survey design could potentially block adjustment. In this case, subsequent self-assessments of current unhappiness or low life satisfaction would occur through the repeated guidance of respondents to focus on the specific details of a negative event (answering “how”). Such a survey format might actually lead to high amounts of artificially induced rumination. Future research must still contend with whether autobiographical recording as a method of measurement could, in fact, change the process under study.

### *The Reverse of Mood Congruence*

Much research has shown that the mood state of an individual can influence the subsequent valence of recalled information (Bower, 1981). However, this chapter shows that the reverse is also true. The valence of a recalled event can lead to a similarly toned mood. Such an assimilation effect appears straightforward enough; that is, recall affects the reporting of mood and subjective well-being when mood is seen as informative. Positive-event recounting produces positive moods; negative-event recounting leads to negative moods.

However, if survey items subsequent to the initial recounting are viewed as “reset” tasks, or if the valenced event is viewed as currently apart from one’s life, then the remainder of the survey’s responses may simply reflect a contrast effect; that is, instead of judgments being based on one’s current moods, leading to assimilation effects, judgments will be made using the past event as a standard, leading to contrast effects.

### Other Methodological Issues

#### *Unitizing*

The survey researcher may also want to consider these possible effects. How does unitizing affect survey responses? That is, by drawing comparisons between past and present life or between present life and the future, does the survey build in a higher likelihood of contrast effects than would be expected?

#### *Issues of Generalizability*

*Response formats.* One may be inclined to think that these assimilation and contrast effects are limited to Likert response scales. Evidence from one study reported here suggests that this is not the case. Assimilation and contrast effects were obtained with a standardized index of life satisfaction that required respondents to indicate “true” or “false” to a number of statements. Further research is still needed to assess what effect recounting has on other standardized measures. It is also possible that types of standardized measures regarding personal life events may themselves focus individuals into an episodic versus abstract “set” that carries over into subsequent self-assessments.

*Proxy reporting.* The effects reported here may not be completely tied to autobiographical memory alone. The reporting of how an event occurred to someone else may still lead respondents to experience empathy and thus assimilate their subsequent responses to the mood that they are experiencing. Describing why an event occurred to one’s mother, spouse, or co-worker may lead one to draw contrasting parallels in one’s own life. Future research is needed here to document whether use of another’s experience is the basis for these assimilation and contrast effects.

*Societal events.* Research reported in this chapter has focused largely on life events that are highly personal to the individual. One may question whether *societal* events such as war or other events of political, economic, or technological advancement (see Schuman, Rieger, & Gaidys, chap. 20, this volume) may produce patterns of judgmental effects similar to those described here. It is possible that description tasks that encourage respondents to focus on societal events rather than on personally impactful events may produce contrast instead of assimilation effects; that is, after describing a societal event, that event may create a standard that leads to contrast effects for the respondent's subsequent personal assessments.

## Conclusions

This chapter has described research that indicates the impact of recounting retrospective autobiographical events on respondents' subsequent judgments. Specifically, respondents who give detailed, vivid accounts of past life events later produce judgments of current life that assimilate to the valence of the event retrospectively reported. Respondents who give descriptions of the causes and reasons for the past event's occurrence, on the other hand, tend to contrast subsequent judgments away from the valence of the past event.

These assimilation and contrast effects of the retrospective reporting of autobiographical events have been shown to occur both for elderly and for college samples. In addition, these effects hold for the reporting of many different types of autobiographical events, ranging from love relationships to changes in jobs and residence to death of a family member. This distinction of recounting in episodic versus abstract terms has been manipulated through (a) length of retrospective reports, (b) temporal distance of the past event, (c) vivid versus pallid instructions, and (d) asking respondents to describe "how" versus "why" the event occurred.

We have concluded that it is important for survey researchers to be aware of these consequences of retrospective recounting. For the reasons noted above, these robust assimilation and contrast effects are potentially occurring on a large scale in many surveys. We have further discussed the possibility that survey researchers may unknowingly provoke such effects through the ways in which surveys are constructed. Finally, the work in this chapter provides an example of how research on the consequences of autobiographical recounting may add to our understanding of response effects in surveys.

# 19

## Reconstructing Social Change through Retrospective Questions: Methodological Problems and Prospects

*Karl-Heinz Reuband*

Retrospective questions are usually employed in social and psychological research in order to trace individual change over time or to explain the present by past events or behavior. Within this tradition, several studies have, for instance, asked questions about patterns of upbringing and child socialization. The descriptions have then, in turn, been linked with present outlooks and personality variables (see, e.g., Filipp, 1981). The usefulness of this strategy—if considered as a measurement instrument for objective conditions—rests on the individual validity and reliability of the questions.

Instead of being used on the individual level, retrospective questions can, however, also be employed on the aggregate level. If a representative sample of individuals is asked about past events and these questions have a distinct time reference, the responses can be grouped according to time period. Under these circumstances, it is possible to describe social and cultural change across time—even including those periods for which no survey data have been collected on the respective topic. Survey research might not even have existed at that time.

When used for macro-level description, aggregate-level validity and reliability rather than individual-level validity and reliability count. People might err in their time locations, but as long as these errors cancel each other out on the aggregate level, it does not matter. In order to assess the usefulness of retrospective questions on the aggregate level, it might be sufficient to have aggregate-level data as long as these can be grouped according to cohort membership for different time periods. Such an analysis is made possible when questions from earlier surveys are replicated in later time periods. These have been my aims in analyzing childhood upbringing in West Germany within this century.

### Methodology

The baseline study for my analysis was Almond and Verba's 1959 study, "Civic Culture," which included not only questions on political matters but also a few

questions on upbringing in family and school. The study involved nationwide representative surveys from various countries, including West Germany (Almond & Verba, 1963). The original data are available for secondary analysis through the Zentralarchiv für empirische Sozialforschung, University of Cologne (Study No. 28). Thirty years later—in 1989—I replicated the questions on family influence in order to test the usefulness of retrospective questions for measuring change. Again, a nationwide survey was taken as the empirical basis, providing comparable conditions of data collection. This time the survey was done by the GFM-Getas institute with a sample of 1,891 respondents. As in the earlier survey, a random sample was used.

The original question in the 1959 survey was the following:

When you grew up, let us say, when you were about 16 years of age, what kind of influence did you have in family decisions that involved you? Did you have much, little, or no influence?

The second subsequent question ran as follows:

How was it at this time, when a decision was made that you did not like? Did you have the feeling that you could protest without inhibitions, did you have some inhibitions, or was it better not to protest?

In the replication, the same wordings were used. However, the age reference was slightly altered by the survey institute,<sup>1</sup> by using the reference “at this time” instead of “when you were about 16 years of age” and by establishing the age reference by means of prior questions about the employment of one’s mother when the respondent was between about “14 and 16 years” of age. Using a lower age reference is not without problems. One cannot rule out that under these circumstances respondents shift the reference period to earlier years and thus provide a rather conservative estimate of their influence when 16 years old.

One could argue, however, that memory is too vague anyhow to describe exactly what went on when between 14 and 16 years old or when 16 years old. All that one can get through memory might be relatively rough descriptions of past events and behavior. Moreover, once we are interested in relative trends over time rather than agreement on an absolute level, the differential wordings lose importance.

## Assimilation or Contrast Effects?

When grouping the retrospective questions on family decision making, cross-sectional survey data from various years have uniformly shown that influence in

---

<sup>1</sup>The replication was done in cooperation with a commercial survey research institute when consulting on other questions. As some kind of compensation, the institute incorporated my questions free of charge into one of its ongoing surveys. Unfortunately, the questions were slightly altered without notifying me, so that the present analysis can only be used to assess trends over time and a less absolute level of agreement. I plan to replicate the exact original wording in future studies.

the family has increased for children and youth over time. Families around the turn of the century are described as being rather authoritarian, leaving little influence to the child and youth, whereas in the second half of this century, the pattern has partially diminished (see, e.g., Reuband, 1988, 1992). What one does not know as long as one relies on cross-sectional surveys is whether the remembered part is reflective of the socialization patterns of the respective time, a product of selective memory, or both.

Memory is selective, as various studies have shown (see, e.g., Moss & Goldstein, 1979; Reuband, 1980; M. Ross, 1989). Past experiences are often seen in the light of present experiences and attitudes. Taking the present prevalent climate of behavior into consideration, people might see less change than there has been in reality, providing some form of assimilation effect. Under these circumstances, the authoritarian pattern in earlier times should be underestimated. However, it is equally plausible for people to accentuate change by basing the recall on implicit theories according to which social change has occurred (see Ross, 1989). This tendency would provide some kind of contrast effect.

The scanty data that exist for Germany on the retrospective recall of socialization patterns are mixed. Based on a comparison of a single cohort across a 30-year time period (1954–1984), a study by Zinnecker (1985) concluded that people tend to shift their perspective of influence in retrospect: They tend to view their past life as being much more regulated and rigid than they perceived it 30 years earlier. They increasingly think of themselves as having been brought up “harshly.” This finding would lend support to a contrast-effect interpretation.

However, the question that was used in that study simply referred to whether the parents dealt with the respondent “harshly” or not. Other indicators, with more precise references, might be less influenced by retrospective distortion. Moreover, these are data documenting that for the periods for which trend data are available—the 50s, 60s, and 70s—change has, in fact, occurred in the direction specified by recall. Thus, in 1955, 45% of the 15–24 years olds said that they were dealt with harshly by their parents; in the replication in 1984, only 35% in the same age bracket said so (Zinnecker, 1985). In another survey in 1962, 45% of the respondents aged 16–18 years replied that way, and in 1983, 19% did so (Allerback & Hoag, 1985). In the same series, a question similar to the one used here was also asked on how much influence respondents had in family decisions. In that study as well, a trend toward greater participation emerged in the time period of observation (see Allerback & Hoag, 1985); and what is especially noteworthy, the trend was roughly parallel to the retrospective series (Reuband, 1992). Although recall might somehow be impaired, these data for the postwar period suggest that the basic trends might still be reproduced.

## Findings

In the following, I use both the 1959 Almond and Verba data through secondary analysis and my new data for the period from 1900 onward. When the marginals

for both questions are compared across the 30-year time span, remarkable similarities evolve. As can be seen in Tables 19.1 and 19.2, there are only slight variations in the depiction of one's past. If the family has become less authoritarian, as cross-sectional data tend to point out when grouped according to cohort membership, far greater striking differences should have been observed in the data across time.

In a next step, summarized in Figures 19.1 and 19.2, the responses on "no family influences" and "being inhibited in protesting" are grouped according to cohort membership. Once we know the age of the respondent, we can locate him or her in this continuum, given the knowledge about the years in which he or she was 16 years old. The figure seems to indicate a change in memory *within* each of the cohorts. Among the 18–25 year olds in 1959, 16% declared that they had no influence, but 30 years later, the rate within this cohort had increased to 45%. Parallel drops in ascribed family power can be observed in the other cohorts over time. If the question of inhibitions in protesting against family decisions is taken for comparison, a similar pattern of diminishing influence is evident. People apparently see themselves as being less powerful than they really were. They seem to enhance the contrast to present-day situations.

Why and how the contrast effect become effective in this case is not clear. Leaving apart the slightly different question wordings, there are basically two

TABLE 19.1. Influence in Family Decisions at Young Age Over Time (%)

Influence	1959	1989
Much	15	17
Little	45	47
No	40	35
Total	100	100
<i>N</i>	(867)	(1,893)

Note: "Don't know," other responses, and "no answers" excluded. "Don't know/do not remember" in 1959 and in 1989 = 6% of the total sample.

TABLE 19.2. Inhibitions in Protesting against Family Decisions at Young Age Over Time (%)

Protest	1959	1989
Without any problem	48	45
With some inhibitions	22	28
Better not do	31	27
Total	100	100
<i>N</i>	(835)	(1,891)

Note: "Don't know" and "no answers" excluded; in 1959 = 8%, in 1989 = 6% of the total sample.

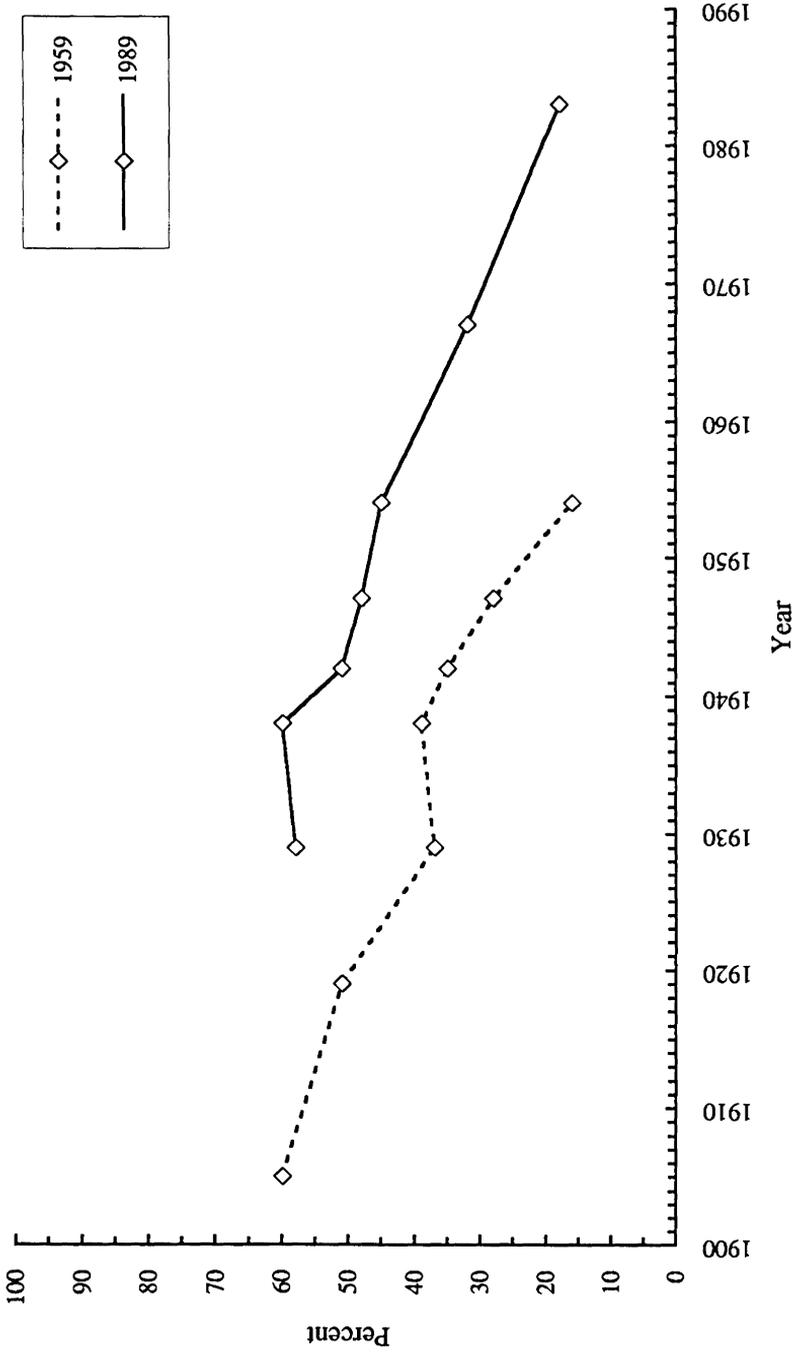


FIGURE 19.1.1. Having No Influence in Family Decisions Over Time

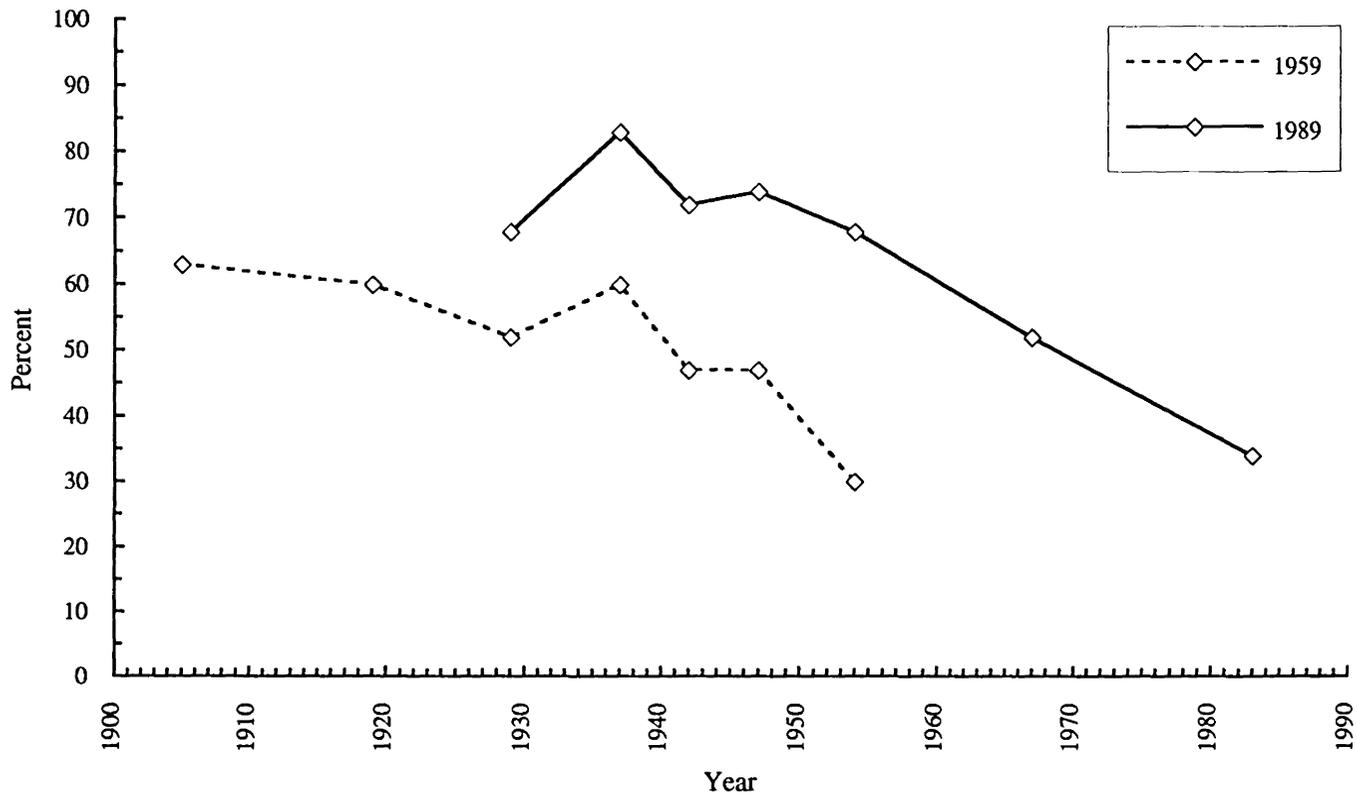


FIGURE 19.2. Feeling Inhibited To Protest against Family Decisions Over Time

possibilities that seem suggestive. One is that people might ascribe a strong change to the period under observation and might structure their recall of their past according to it. They might see much change because they think that much change should have happened. They impute their theories about change on the reconstruction of their personal biography.

The other possibility emanates from the standards of perception per se, and in the standards of linguistically labeling behavior. Categories such as “much,” “little,” or “no” influence, as used in the question wording, are naturally vague. They could easily change when different normative standards become prevalent in society: Influences that used to be “much” influence compared with the usual experiences of the time might have become relatively speaking “little” influence later. If this is true, shifting normative conceptions will promote shifting definitions of one’s biography. Under these circumstances, the perception of the past will undergo change not when time passes per se but only when normative conceptions of family influence change in the population.

Given the ambiguity of the question-wording differences between the 1959 and 1989 surveys, the parallelity of the trend as assessed by the retrospective questions might be more intriguing, however, than the absolute level of agreement. If both indicators tap the same dimensions, both should undergo the same kind of change. If this comparison is done, the retrospective trends based on the two surveys come out as rather parallel. This even applies to short-term fluctuations such as in the late 1930s, when authoritarian family patterns underwent a slight revival. The data, whatever their absolute level, document at least *trends* correctly.<sup>2</sup> A distinct tendency for the older or the younger ones to shift their recollection more than other cohorts cannot be observed—except, perhaps, for the very young group, who reveal slightly more marked change in perception.

## Summary

My research has shown that retrospective questions might be useful for delineating *aggregate* change with regard to general trends. Absolute levels, however, might differ; under which conditions this is the case deserves further study. There is reason to believe that people use present standards for viewing the past. I do not know whether this is due to selective perception per se or to normative standards that give rise to the use of specific linguistic categories. If the latter are the main reason, the problem of distortion should become less when standardized, concretely specified situations are made the basis of comparison. Further research into the process of constructing the past and remembering concrete details of events is warranted.

---

<sup>2</sup>One problem, however, remains unresolved that deserves future research: What cannot be controlled for is the fact that in *both* surveys members of different cohorts have a differential time lag regarding the period that they describe. This differential is preserved in the replication, since each cohort undergoes the same shift in time.



## Collective Memories in the United States and Lithuania

*Howard Schuman, Cheryl Rieger, and Vladas Gaidys*

Although most research on autobiographical memory has been on individual memory for events of everyday life, some attention has been paid to memories of public events, both political and nonpolitical. In research on memory for public events, two links between individual memory and the larger social group have been suggested and are much debated. One is the “now print” mechanism posited for the recording of surprising and important information, which connects the individual’s memory to the social group by the biological necessity “to survive and leave progeny” (R. Brown & Kulik, 1982, p. 38). The other is the “reconstructive” approach, which links the individual’s memory to the social group through the meanings that individuals assign to memories of historical events, especially through the “alignment” of the narrative of one’s private life with the narrative of public history (Neisser, 1982b). Largely missing from the studies of memory for public events is consideration of how the social and historical locations of individuals contribute to the formation of memory for public events.<sup>1</sup>

In contrast, research on collective memory emphasizes the importance of these locations. *Collective memories* are memories of their own historical past that are shared by groups of people. When individuals live through public events, their autobiographical memories may be linked with collective memories (N. R. Brown, Shevell, & Rips, 1986), whether the events are directly experienced or learned through secondhand reports (Larsen, 1988). However, when individuals learn of public events that took place before their birth, they neces-

---

<sup>1</sup>There are some qualified exceptions to this statement. R. Brown and Kulik (1982) note the higher frequency of reports by blacks of events concerning leaders importantly related to the cause of black advancement. Neisser (1986) refers to the importance of his own social and historical location—an immigrant desiring a secure American identity—in the symbolic meaning of his memory of hearing about the bombing of Pearl Harbor. Aside from these comments, however, social and historical location have been of little interest to psychologists working on these issues.

sarily receive such “memories” secondhand and symbolically—by cultural artifacts (Schwartz, 1991), social practices and daily routines (Connerton, 1989), or an oral tradition (Petrov, 1989). We argue that regardless of how an individual comes to learn of a public event of the recent past, social variables—mainly age, but also education, sex, and group identification—as well as the sequence of historical events in which individual lives are embedded, strongly affect individual and group memories.

In this chapter, we use findings from surveys in the United States and Lithuania to illustrate connections among social variables, historical location, and collective memories. First, we review briefly some basic findings on collective memory in the United States. Next, we describe in more detail our results from a replication in Lithuania. In the conclusion, we compare the findings from the two countries and speculate on the functions of collective memory. Throughout, our emphasis is on the meaning that individuals find in events, rather than on the validity of their reports.

## The Basic Survey Question

The initial hypothesis on collective memory, drawn from Karl Mannheim’s (1928/1952) classic essay on “The Problem of Generations,” suggested a life-course approach to understanding the nature of collective memory. The first assumption is that people tend not to remember events or changes that preceded their own consciousness of the larger political and social world; hence, events that occurred before birth, or indeed before adolescence, tend to be less meaningful to people than do events that they themselves lived through (Gallatin, 1980; Sigel & Hoskin, 1977). Furthermore, it seems plausible that the first event that people experience as important—as “personally gained in real situations,” as Mannheim (1928/1952, p. 296) put it—benefits from a kind of primacy effect, so that later events are usually given less weight in memory. Thus, this hypothesis would predict that people who grew to adulthood during World War II would be less apt to recall the Vietnam War as important than would those who grew to adulthood during the Vietnam era.

The national survey that provided the test of the initial hypothesis was conducted in 1985 using this wording:

The next questions concern how people think about the past. There have been a lot of national and world events and changes over the past 50 years—say, from about 1930 right up until today. Would you mention one or two such events or changes that seem to you to have been *especially* important?

There aren’t any right or wrong answers to the question—just whatever national or world events or changes that come to mind as especially important to you.

If only one event or change was mentioned spontaneously, respondents were probed for a second event. In addition, a follow-up question asked why the event was important:

What was it about [the event] that makes it seem especially important to you?

The questions were open-ended, so that the respondent supplied his or her own frame of reference.

Four findings of interest here emerged from the 1985 study, the results of which are reported in detail in Schuman and Scott (1989). The main finding supports the initial hypothesis and thus a generational interpretation of American memories: There is a strong tendency for people to recall as important those events and changes that occurred during their own adolescence and early adulthood. Two examples are found in the relation of age to mentions of World War II and the Vietnam War. In each case, it is especially people who were between the ages of about 15 and 25 at the time of the particular war who mention it as an important event from the past half century. A number of other events and changes show similar relations to adolescence and young adulthood.<sup>2</sup> These results are consistent with life-course perspectives on political socialization (Braungart, 1984) and are also consistent with the findings of Fitzgerald (1988) and Rubin, Wetzler, and Nebes (1986) on the large proportion of autobiographical memories from the period of young adulthood.

Second, people tended to think about public events in quite personal terms. This was brought out clearly when respondents were asked to explain why they thought an event was important. In the case of World War II, for example, older respondents often spoke not of the larger meaning of the war but of their own particular involvement in it—for example, injuries sustained in combat or the hardships of rationing on the homefront.

Third, when reporting an event that occurred before or after their late adolescence and young adulthood, people sometimes saw that event in terms that seem to have been affected indirectly by the experiences of their own youth. In particular, those who were in their formative years during the war in Vietnam but who nevertheless mentioned World War II described it as a “good war” significantly more often than did those who had been young adults in the early 1940s when World War II was occurring. Furthermore, the answers of respondents from the pre-Sputnik era conveyed more emotional awe at space exploration than did the answers of younger respondents, for whom astronauts and satellites were part of the natural world of their own childhood.

Finally, there was some evidence of the importance of social group identification for the events that individuals reported. Virtually all of the mentions of changes in the role of women were made by women. Similarly, about 44% of the blacks who were interviewed mentioned civil rights, whereas only 5% of the whites did so.

---

<sup>2</sup>Three puzzling results were also identified: the absence of relations of age to mentions of space explorations, the development of the computer, or civil rights (the last category lacked an age relation for whites, although there was a relation for blacks). However, a later experiment showed mentions of the computer to be related inversely to age when a closed, rather than an open-ended, question was used (Schuman & Scott, 1987).

## Evidence from Lithuania

In several ways, then, the results of the original study indicate that people tend to recall and define the “importance” of recent historical events in personal terms that are a function of experience in the formative years, and there was some demonstration of the importance of group membership in the selection of important events. However, there are several aspects of collective memory that are best explored in cross-national comparisons.

First, data gathered in Lithuania in 1989 provide a test of the generational hypothesis under historical conditions strikingly different from those surrounding the United States in 1985, which was generally a time of internal stability and even complacency. In contrast, the Lithuanian survey was conducted during a period of rapid movement toward rejection of Soviet rule and increasingly explicit claims of Lithuanian independence. For example, in August 1989 (a month before the survey), there had been widespread observances of the 50th anniversary of the secret Ribbentrop-Molotov pact that consigned the Baltic countries to the Soviet Union, and in December (3 months after the survey reported here) the Lithuanian Communist Party defined itself as distinct from the Communist Party of the USSR. The movement culminated on March 11, 1990, when a newly elected Lithuanian legislature declared Lithuania to be independent of the USSR. Figure 20.1 provides a time line of important events in Lithuanian history over the past 60 years.<sup>3</sup> Data gathered in such a period of highly charged political conflict allow us to test whether recent events, such as the momentum of the Sajudis independence movement, will replace memories of earlier events even in older individuals.

Second, American responses to the basic history question almost inevitably confound what is significant to the United States with what is important to the larger world. For example, mentions of World War II, the most frequently given event, refer to both an American and a world event. Likewise, man’s first landing on the moon, the third most frequently mentioned event, was a triumph for the United States as well as a “giant step for mankind.” Only by considering evidence from a population outside the United States, and perhaps less directly involved in world events, can we separate the national or even more local levels from the world level at which events are remembered and judged important.

Lithuania provides a particularly interesting case, since it is not only outside the United States but has been at once part of another great nation on the world stage and yet has strong claims to having a separate identity that rejects such a larger association. As a result, collective memories in Lithuania could include events at several levels:

1. Lithuanians may think of events of obvious importance to the world as a

---

<sup>3</sup>For an account of major events over the past half century in Lithuania and the other Baltic Republics, see Nahaylo and Swoboda (1989). Events leading up to the declaration of independence by Lithuania are described in Olcott (1990) and Senn (1990).

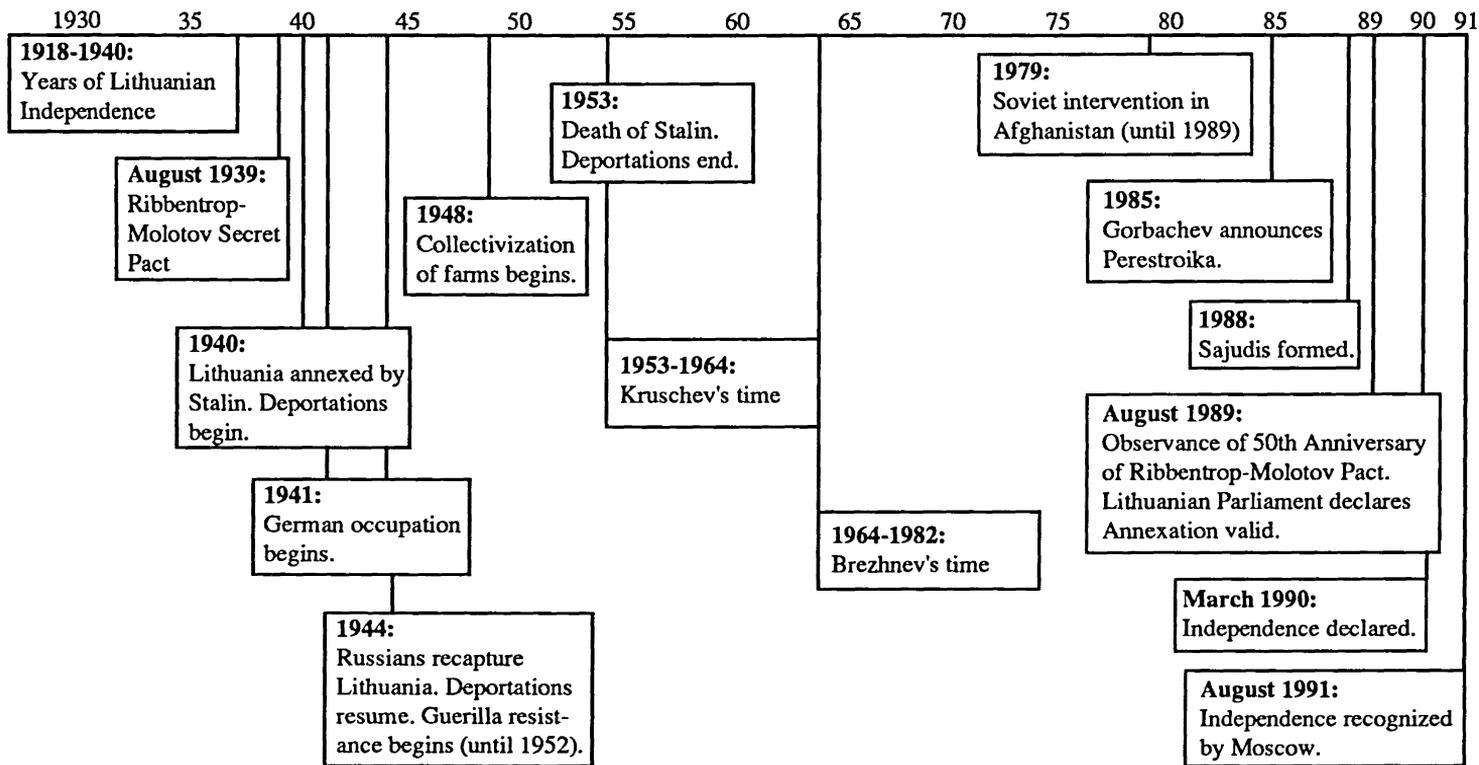


FIGURE 20.1. Time Line of Relevant Events in Lithuania

- whole and therefore may provide answers similar to many of those given by Americans; or
2. As a part of the Soviet Union and subject to the news as it came out of Moscow for most of the last 50 years, many Lithuanians may substitute for events mentioned by Americans some parallel events important to the recent history of the USSR; or
  3. Lithuanians may focus on events that were formative in recent Lithuanian history: the occupation and annexation of Lithuania by Stalin at the time of World War II, or the move toward independence over the past several years.

Thus, data from Lithuania provide a way to identify the political and social level most crucial for collective memories that was blurred in the data from the United States.

Finally, evidence from Lithuania allows us to look more closely at a social variable hypothesized to be important for collective memories—group identification. Specifically, we can examine ethnic differences in collective memories, since Lithuania includes a sizable proportion of Russians—about 9% of the population—as well as the majority population (80%) of ethnic Lithuanians. In the United States, aside from the difference between blacks and whites in the importance given to the civil rights movement, there are few apparent racial differences in the content of collective memories (Schuman & Scott, 1989). Lithuania provides another setting in which the effects of highly politicized group differences can be explored.

In sum, the Lithuanian data allow us to investigate more fully the formation of collective memories but to do so with an entirely different population that experienced a quite different series of events locally and nationally.

## Method

Our data were gathered in Lithuania in the fall of 1989, using translations of the two open-ended questions previously used in the United States in 1985. In addition, demographic data on ethnic identity, age, education, and gender were obtained. The questions were asked between September 30 and October 6, 1989, as part of a larger survey carried out by the Public Opinion Research Center of the Institute of Philosophy, Sociology, and Law, Academy of Sciences of Lithuania. The population was defined as all registered voters, 18 years and older, living in Lithuania, with the exception of those in hospitals or on military bases. Respondents whose mother tongue was Lithuanian were given the questionnaire in the Lithuanian language; others (mainly Russians) were given the questionnaire in Russian. The final sample size was 1,454; we have confined our analysis to Lithuanians ( $n = 1,172$ ) and Russians ( $n = 130$ ), omitting the smaller number of Poles ( $n = 91$ ) and the scattering of other nationalities ( $n = 61$ ).<sup>4</sup> Answers to the

---

<sup>4</sup>Following a common procedure in surveys in the Soviet Union, most respondents filled out the questionnaire themselves, including writing their responses to the open-ended

collective memory questions were coded by a Lithuanian historian, with help from the survey director and other workers at the Public Opinion Research Center. No measure of coding reliability is available, but the codes that form the main basis for this chapter are relatively straightforward.<sup>5</sup> We were able to use the follow-up “why” question to clarify certain points, but the answers were not extensive enough to allow the kind of detailed analysis carried out with the 1985 American data (Schuman & Scott, 1989).

## Results

### *Events Mentioned as Important*

The events mentioned most often in response to the collective memory question—by at least 4% of either the Lithuanian or the Russian respondents—are presented in Table 20.1.<sup>6</sup> The figures show the percentage by nationality of those giving each response as one of their two mentions. For example, 57.7% of the Lithuanians who mentioned at least one event gave a response coded under the heading of “the rebirth of Lithuania” as either their first or second answer to the event/change question.<sup>7</sup> Table 20.1 shows percentages based on the combined first and second mentions because the combination correlates .99 with percentages of first mentions only and therefore can be considered essentially identical to them; for later analytic purposes, it is desirable to use the larger percentages based on both mentions.

Overall, there were striking differences between the responses of the Lithuanian and Russian samples. More than half of the Lithuanian sample mentioned

---

questions, which were on separate sheets from the main questionnaire. Frequently the interviewer left the questionnaire for a brief period (an hour or less) and then returned to pick it up, placing it within a sealed envelope to demonstrate its confidentiality. However, if a respondent desired, the interviewer read the questions.

<sup>5</sup>In the 1985 study in the United States, the coding reliability for events and changes averaged 95% agreement between coders (Schuman & Scott, 1989).

<sup>6</sup>Although the 4% threshold is arbitrary, it provides a fairly natural breaking point for the large Lithuanian sample and is already small for the tinier Russian sample. Omitted from the table entirely are the 14% of the total Lithuanian sample and 16% of the total Russian sample who were unable, or conceivably unwilling, to mention any event or change from the past 60 years. The figures are only slightly higher than was true in 1985 for Americans, where 11% of the total sample could not mention any event or change (Schuman & Scott, 1989). As in the United States, education is the major determinant of giving or not giving a response, ranging for the total sample from 18% “no answer” for those with only primary education to 6% of those with a university education. Younger age is also an independent predictor of not giving any mentions, whereas gender is not; the opposite occurred in the United States.

<sup>7</sup>Most of the responses coded here actually used words very close to the name of the category. Others referred to *Sajudis* (the independence movement) or to the return of national culture and were therefore coded into the “rebirth” category.

TABLE 20.1. Highly Mentioned Events by Lithuanians and Russians (% and Ranking)

Event	Mentions by Lithuanians <sup>a</sup>	Mentions by Russians <sup>a</sup>	<i>t</i> <sup>b</sup>
Rebirth of Lithuania	57.7 (1)	15.0 (3)	-7.41***
Annexation/occupation	17.9 (2)	1.9 (8.5)	-3.76***
World War II	16.5 (3)	39.3 (2)	5.67***
Deportations to Siberia	15.7 (4)	3.7 (6.5)	-2.95**
Perestroika	14.9 (5)	51.4 (1)	8.11***
Ribbentrop-Molotov Pact	9.7 (6)	1.9 (8.5)	-2.52*
Collectivization of private farming	4.9 (7)	0.0 (10)	-1.12
Postwar guerilla resistance	4.2 (8)	0.9 (9)	-1.24
Afghanistan war	2.4 (9)	4.7 (5)	1.36
Krushchev's time	0.8 (10)	7.3 (4)	3.59***
Man in space	0.2 (11)	3.7 (6.5)	3.42***
<i>N</i>	(1,011)	(109)	

<sup>a</sup>Within columns, each row represents a dichotomy of those mentioning the event at all divided by the total (1,011 for Lithuanians, 109 for Russians) mentioning any event.

<sup>b</sup>These are *t* ratios for Lithuanians versus Russians, based on logistic regressions of each event on nationality, education, age, and gender, with \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

the "rebirth" of Lithuania, whereas only 15% of the Russian sample gave such a response. Almost exactly the opposite occurred for responses centering on perestroika: This was the single most popular answer for Russians (51%), whereas only 15% of the Lithuanians gave such an answer. In much the same vein, the second most frequent response given by Lithuanians, the annexation of Lithuania in 1940, was barely present in the list of events mentioned by Russians.<sup>8</sup>

These differences of upwards of 35% point to the gulf between the two nationality groups in how the past half century is remembered. Although many of the Russians (unfortunately not identifiable in the sample) moved to Lithuania after World War II and thus did not directly experience such early events as annexation, respondents of all ages and both nationalities experienced more recent events. Yet, perestroika was mentioned much less often by Lithuanians, who nominated the "rebirth" of Lithuania as often as Russians mentioned perestroika. Individuals in the two ethnic groups were then experiencing the same stream of events but were selecting different events as important according to the perspectives of their ethnic groups.

Moreover, even where the events mentioned sounded similar, there may have been sizable differences in how the events were conceptualized. A telling

<sup>8</sup>Lithuanians and Russians do not differ significantly in gender or age, but Russians are significantly more educated than Lithuanians. However, educational differences do not account for the differences discussed here, since the *t* ratios in Table 20.1 are based on logistic regressions with controls for education, age, and gender.

example is “World War II,” which is almost the only event in Table 20.1 that seems on its face to transcend the issue of Lithuanian independence, providing a common bond between Lithuanians and Russians. In their nominations of war-related events, both Russians and Lithuanians referred to “the years of war” or the German occupation, but Russians were significantly more likely ( $p < .001$ ) than Lithuanians to refer to the “Great Patriotic War,” whereas Lithuanians were more likely ( $p < .01$ ) to refer to the event as “the Second World War.” In addition, once we consider the further elaboration provided by the follow-up question on *why* the event was mentioned, there are also large differences in the way World War II was remembered by the two nationalities. As Table 20.2 shows, Russians were quite likely to think about the war in terms of “Soviet patriotism” (29%), whereas hardly any Lithuanians offered such a reason, noting instead either a personal experience during the war or the larger effect of the war on the division of Europe and the imposition of the Soviet system. There were some similarities to be sure (e.g., mention of “lives lost, cruelty, fear”), but if the Russian subsample had been somewhat larger, even such relatively small differences between Lithuanians and Russians might well have reached statistical significance rather than turning out to be due simply to sampling error.

In sum, Lithuanians and Russians in Lithuania tend to live within different psychological worlds. Moreover, it is also clear that for Lithuanians today memories of the past overwhelmingly concern not world events or events of the Soviet Union (except insofar as they have involved Lithuania), nor at the other extreme events at the regional or city level, but events that concern Lithuanian nationality.<sup>9</sup> Lithuanian identity provided the matrix within which events were remembered as important in answer to a question about world and national history over the past 60 years.

### *Generational Differences in What Is Remembered*

The major theoretical hypothesis of this research has been that people will tend to remember events and changes from the period in which they themselves grew up, that is, their adolescence and early adulthood. For Americans, much of the history of the past half century has been subdivided into a set of separate and seemingly unrelated events, such as World War II, the assassination of John F. Kennedy, man’s first landing on the moon, and Watergate. Cohorts can also be subdivided according to their relation to such specific events. However, almost all of the events mentioned by Lithuanians came from two widely separated periods of time, yet in an important sense they all focused on the same subject: Lithuanian independence. The years 1939 to 1953 saw the annexation of Lithuania, occupation by the Germans during World War II, Stalin’s deportations of

---

<sup>9</sup>We cannot be as certain of this latter point as we would like, since the questionnaire in which the memory questions were embedded included other questions focused on the national level and we were unable to control for possible context effects.

TABLE 20.2. Reasons for Mentioning World War II<sup>a</sup>

Reason	Lithuanians	Russians	<i>t</i> <sup>b</sup>
Lives lost, cruelty, fear	37.7	45.2	1.27
Personal experience during the war	26.1	12.9	-1.07
Divided Europe, created "socialist system," system," created fate of Lithuania	20.3	6.5	-2.14*
Destroyed economy	6.5	0.0	-0.72
Soviet patriotism	0.7	29.0	3.40**
Other	9.4	6.4	
Total	100.0	100.0	
<i>N</i>	(138)	(31)	

<sup>a</sup>Based on respondents mentioning World War II as 1<sup>st</sup> or 2<sup>nd</sup> event who also gave a reason.

<sup>b</sup>These are *t* ratios for nationality (Lithuanians vs. Russians) based on logistic regressions of each reason on nationality, education, age, and gender.

\*  $p < .05$ .

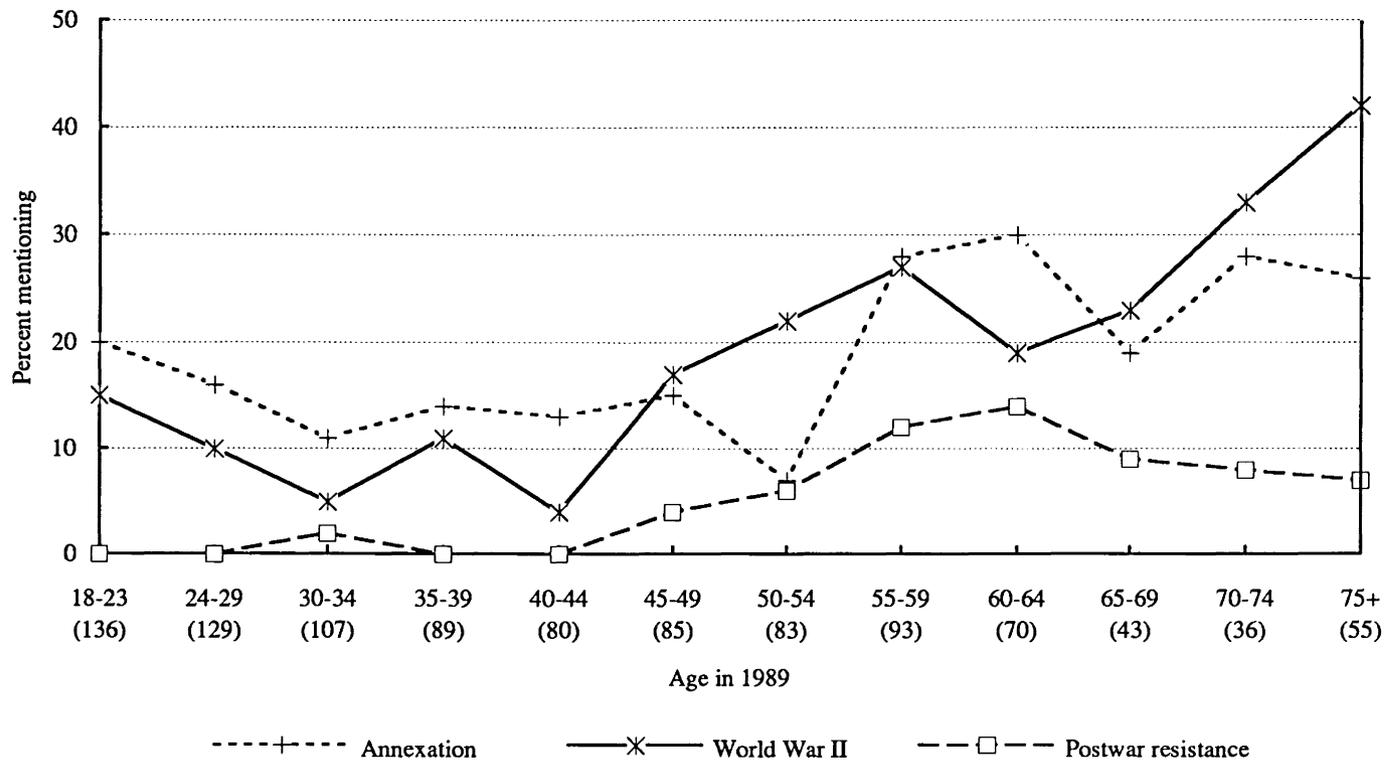
\*\*  $p < .001$ .

thousands of Lithuanians to Siberia, the forced collectivization of the countryside, and finally an unsuccessful guerilla struggle against Soviet control. At the other end of the time line, Gorbachev's perestroika and the "rebirth" of Lithuania were from the several years preceding the survey, five at the very most. There is one seemingly ambiguous event: the Ribbentrop-Molotov Pact occurred in the earlier period, but it first became a focus of public attention only late in 1987.

Thus, there appear to be in the collective memory of Lithuanians just two time periods that are of importance: the 1940s' and early 1950s' loss of independence and the late 1980s' renewed spirit of independence. The one exception is the occasional mention of the war in Afghanistan, closer to the present than to the past, yet distinct in both time and type from the other events in Table 20.1.

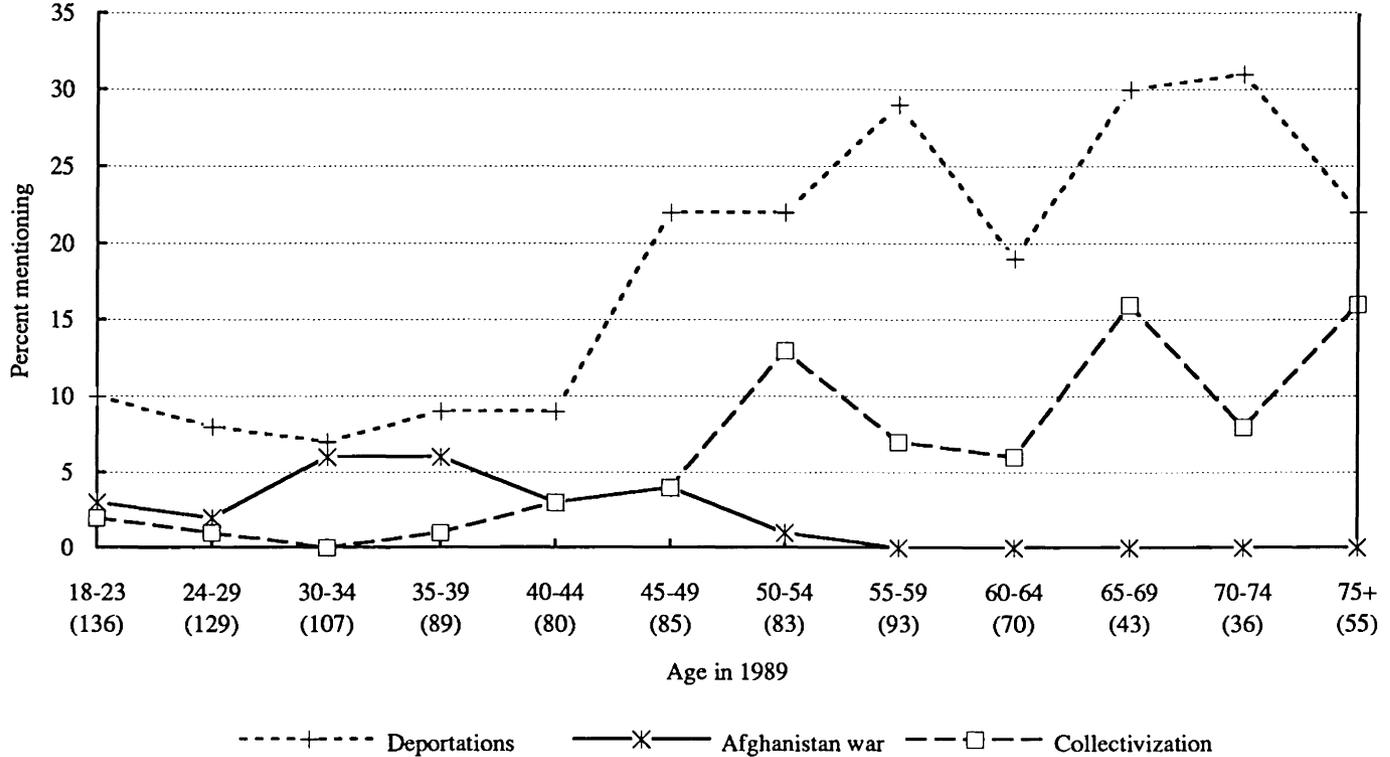
This bifurcation of history shows up quite clearly in Figures 20.2-20.4 and in Tables 20.3 and 20.4, which portray the relation of age to collective memories, although there are also interesting deviations from an entirely simple pattern. The three figures present the bivariate relations of age to mentions of the nine most highly mentioned events. The nine rows in Table 20.3 provide the exact percentages that are graphed in these figures. Finally, Table 20.4 uses logistic regression to test the age relations to the nine events statistically, controlling for both education and gender, and with an additional test for curvilinearity.

Age shows a significant linear relation to each of the nine events, with the relation being positive (older Lithuanians more likely to have high mentions) for the five events that occurred between 1940 and the early 1950s and negative for the four events that occurred much more recently. Thus, those who are young now are the ones more likely to mention Lithuania's rebirth or other events of



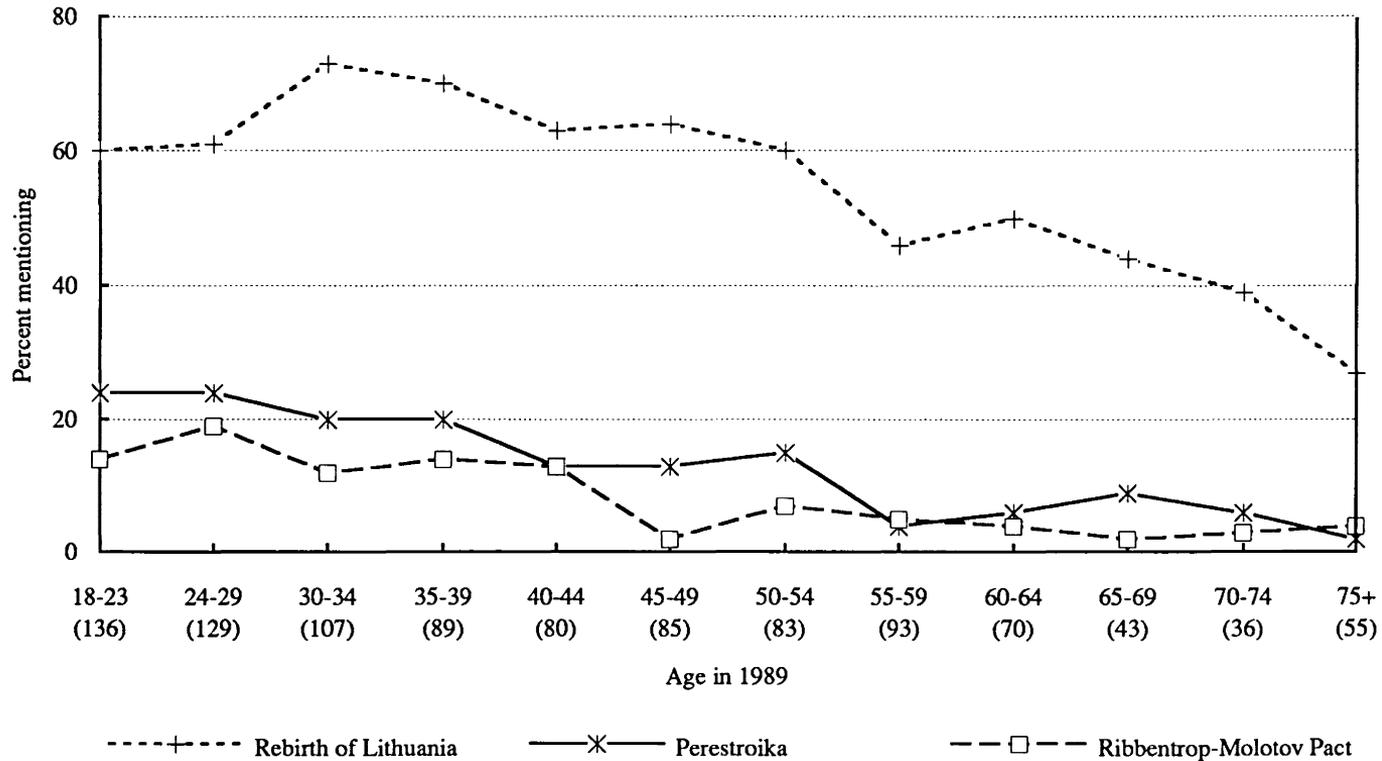
Figures in parentheses are base Ns.

FIGURE 20.2. Datable Events by Age: Annexation, World War II, and Postwar Resistance (Lithuanians only)



Figures in parentheses are base Ns.

FIGURE 20.3. Datable Events by Age: Deportations, Afghanistan War, and Collectivization (Lithuanians only)



Figures in parentheses are base Ns.

FIGURE 20.4. Datable Events by Age: Rebirth of Lithuania, Perestroika, and Ribbentrop-Molotov Pact (Lithuanians only)

TABLE 20.3. Percentage of Lithuanians Nominating Nine Most Mentioned Events, by Age in 1989<sup>a</sup>

Event	18-23	24-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75+
Rebirth of Lithuania	60	61	73	70	63	64	60	46	50	44	39	27
Annexation/Occupation	20	16	11	14	13	15	7	28	30	19	28	26
World War II	15	10	5	11	4	17	22	27	19	23	33	42
Deportations to Siberia	10	8	7	9	9	22	22	29	19	30	31	22
Perestroika	24	24	20	20	13	13	15	4	6	9	6	2
Ribbentrop-Molotov Pact	14	19	12	14	13	2	7	5	4	2	3	4
Collectivization of private farming	2	1	0	1	3	4	13	7	6	16	8	16
Postwar guerilla resistance	0	0	2	0	0	4	6	12	14	9	8	7
Afghanistan war	3	2	6	6	3	4	1	0	0	0	0	0
<i>N</i>	(136)	(129)	(107)	(89)	(80)	(85)	(83)	(93)	(70)	(43)	(36)	(55)

<sup>a</sup>Each row item is the percentage of those in the age group mentioning at least one event (*N*) who nominate the event. Since some of these respondents mentioned two events, the column percentages do not sum to 100.

TABLE 20.4. Relations of Major Event/Change Categories and Age, Education, and Gender<sup>a</sup>

Variable	Annexation of Lithuania (1940)	World War II (1941–45)	Postwar Guerilla Resistance (1944–52)	Deportations (1940, 1944–53)	Collectivi- zation of Farming (1948–53)	Afghanistan War (1979–89)	Perestroika (1985)	Rebirth of Lithuania (1988–89)	Ribbentrop- Molotov Pact (1939 & 1989)
Education	4.88	—	—	—	-3.61	—	2.76	—	3.48
Gender	-4.27	—	—	2.16	—	-2.51	(-1.83)	(1.80)	—
Age (linear)	5.02	4.67	3.52	3.94	2.05	-2.39	-3.83	-4.07	-2.61
Age	2.71	2.29	-3.06	—	—	-2.58	—	-2.65	—

<sup>a</sup>Based on logistic analysis of each major event or change using four predictors: age (6 categories), education (5 categories), and gender (1 = Men, 2 = Women). The cell figures are statistically significant ( $p < .05$ )  $t$  ratios (coefficient/standard error), with those in parentheses of borderline significance ( $.10 > p > .05$ ). Each analysis was done with and without an additional term for age squared to test for curvilinearity; if the age-squared term was not significant, results are shown only for the model omitting it. The sample size for these analyses was 989, a number smaller than that shown in Table 20.1 because cases with missing age, education, or sex were omitted from the logistic analyses. Nominal two-tailed statistical significance levels for this table are  $t = 1.64, p < .10$ ;  $t = 1.96, p < .05$ ;  $t = 2.58, p < .01$ ;  $t = 3.29, p < .001$ .

the recent past, whereas those who were in their youth at the time of the forced incorporation of Lithuania into the Soviet Union are particularly likely to remember events from that period.

In mentions of the earlier events, the structure of history becomes important. Rintala (1968) identifies three factors that increase the size of a political generation: the degree of political participation, the length of events, and the spatial limits of events. The events of 1940–1952 were unusual in extent or duration for each of these factors. Virtually all Lithuanians were affected by the changes of this period, which were probably experienced in peoples' lives not as discrete, categorical events (as coded here) but as a stream of interrelated changes that took place over the same time period. As a result, most of the earlier events are highly mentioned by all older cohorts, without a peak in mention associated with a single cohort.

One must also distinguish personal experience from chronology in considering the relation of age to the events mentioned. Particularly interesting in this regard is the emphasis by young people today on the Ribbentrop-Molotov Pact, which dates from half a century ago but became public knowledge only in the 2 years preceding the survey. The past and the present in this case are really interchanged from the standpoint of personal experience.

In addition to clear linear relations for age, five of the events also show significant nonlinear trends that are captured by adding a quadratic term for age. Two of these relations involve the oldest events in Table 20.3, and in both instances the nonlinear feature seems to be the result of some increase in mentions by the youngest members of our sample, especially those 18 to 25. We interpret the increase in mentions of Lithuania's annexation to be due to the renewed focus today on that event 50 years ago, thus basically similar to an emphasis on the "rebirth" of Lithuania. In fact, we note that the category of "rebirth" itself produces a nonlinearity that is the mirror image of the one for annexation, with 18 to 29 year olds mentioning "rebirth" less than those 10 years older. What seems to happen is that some of the youngest people speak of the issue today more in terms of the original annexation rather than by using phrases referring to "rebirth." It is likely that this is more a matter of intellectual sophistication and a concern for history than a substantive distinction of importance, since there is a significant interaction between age and education such that, among young people, a disproportionate number of highly educated respondents mentioned annexation.

The similar higher mention among the young of World War II is more puzzling, but if the reasons shown in Table 20.2 are considered, the explanation becomes clearer. When the World War II answer was probed, about a quarter of the respondents referred to personal experience during the war, and almost all of these are people now over 50 years of age. However, a second reason offered for mentioning World War II was that it led to the division of Europe and the imposition of a "socialist" system on Lithuania, and this reason was the one given mainly by the youngest age category. Again it appears that this current cohort of people in their early 20s includes some who refer back to the original occupation

of Lithuania as their way of thinking about the events of today involving rebirth. This result provides some qualification to the more general hypothesis about generations and collective memory, but at the same time we should emphasize that the much stronger trend in Table 20.1 and Figure 20.2 is linear.

The nonlinearity for the postwar guerilla struggles is different, for it represents an unexpected lowering of mentions of that event by the oldest respondents. The finding helps support the more exact hypothesis about age, namely, that it is particularly those who were in their adolescence and early adulthood who recall a national event. The annexation of Lithuania and World War II occurred in the early 1940s, and it is the oldest people in our sample (e.g., those in their 70s) who were in the critical age range at that point. For example, someone age 70 when the survey was done was 21 years of age in 1940 when Lithuania was occupied by the Russians and 22 when the Germans attacked the Soviet Union. However, the postwar guerilla struggles occurred mainly in the late 1940s and early 1950s. Thus it is people now around 60 (age 21 in 1950) who were adolescents and young adults at that time. Someone 60 in 1989 was 21 in 1950, and therefore this should be the peak age, according to the generational theory developed from Mannheim, for remembering the postwar struggles.

The nonlinearity for the Afghanistan war calls for a still different, but equally meaningful, interpretation. The war is the one event in Table 20.3 that is neither very old nor quite recent: It began in 1979 and continued through much of the 1980s. It was those in their 30s now who were in their 20s at the beginning of this period and who especially mentioned the war; note that nobody age 55 or beyond mentioned Afghanistan. Moreover, the negative relation to gender indicates that mentions of the war came disproportionately from males, who were the ones subject to the military draft.

We can summarize more briefly the findings for education. The annexation of Lithuania, the key historical event in Table 20.3, was most often mentioned by the best educated respondents, as was the Ribbentrop-Molotov pact with which annexation was directly associated. We take these relations to reflect a kind of intellectual sophistication that leads to a concern for political and legal history.

Collectivization of farming, on the other hand, was most often mentioned by the least educated respondents, regardless of age. We do not know the urban-rural origin of these respondents; but in Lithuania, as in almost all countries, education is lowest in rural areas, and the association of low education with mention of the forced collectivization of agriculture quite likely derives from personal connection to the land, perhaps through family members who were directly affected by Stalin's decision to dispossess individual farmers.<sup>10</sup>

---

<sup>10</sup>We also added terms for age-by-education interactions to the logistic equations. Four events showed significant ( $p < .05$ ) interactions. These were negative for the annexation of Lithuania and for World War II and positive for the postwar guerilla struggles and the rebirth of Lithuania. However, the positive correlation between age and education in the

## Conclusion

The United States in 1985 and Lithuania in 1989 appeared almost as different as two Western countries could be. The United States was in a particularly stable period in its 200-year history, with no challenges to its identity as a nation in present consciousness or in the memory of its citizens. Lithuania, however, was hurtling in a space of a few months from seemingly submissive incorporation within an apparently stable superpower to bolder and bolder steps toward complete independence. The contrast creates striking differences in the content of the collective memories of the two peoples, yet it supports rather than eliminates equally striking similarities in the relation of life course to the temporal nature of what is remembered.

The differences in content are obvious enough by comparing Table 20.1 in this chapter with the parallel Table 1 in Schuman and Scott (1989). Only a single event, World War II, occupies roughly similar positions in the two tables, appearing as the most mentioned event by Americans in 1985 and as the event mentioned third most often by Lithuanians. Some of this difference is doubtless due to time as well as geography, for more recent data gathered for Americans show a strong awareness of the changes in eastern Europe, although not specifically about Lithuania.<sup>11</sup> The importance of national identity is evident for both countries, however, as indeed the contrast between ethnic Lithuanians and Russians within Lithuania already makes clear.

More striking than the differences in content, however, are the similarities between Americans and Lithuanians in the relation of age to events remembered. For both peoples, adolescent and early adulthood experiences play an important role in shaping later collective memory, so that different generations or cohorts are distinctive in what they mention as most important about the past half century. The similarity is brought out well by graphing together the 1985 American and 1989 Lithuanian mentions of the one common event, World War II, by birth cohorts, as shown in Figure 20.5. In both cases, mention of the war has a predominantly monotonic relation to age, although mentions by Lithuanians have a curvilinear component that reflects more mentions of World War II by the youngest respondents than by those slightly older.

The graph also reflects indirectly a dip in the memories of World War II by Lithuanians born in the 1920–1929 period. These respondents reached age 21 between 1941 and 1950, a period that saw not only World War II but also the annexation of Lithuania and Stalin's repression of resistance by means of deportations and in other ways; hence, they mentioned several other events from

---

sample is strong enough to create very small cells for certain combinations, which makes interpretation difficult, so that only the interaction for mention of annexation lends itself to a clear interpretation.

<sup>11</sup>In data collected in the spring of 1990 from the metropolitan Chicago area, 34% of all respondents who gave at least one event mentioned the reunification of Germany, new freedom in the Soviet Union, or the breakdown of communism in Eastern Europe.

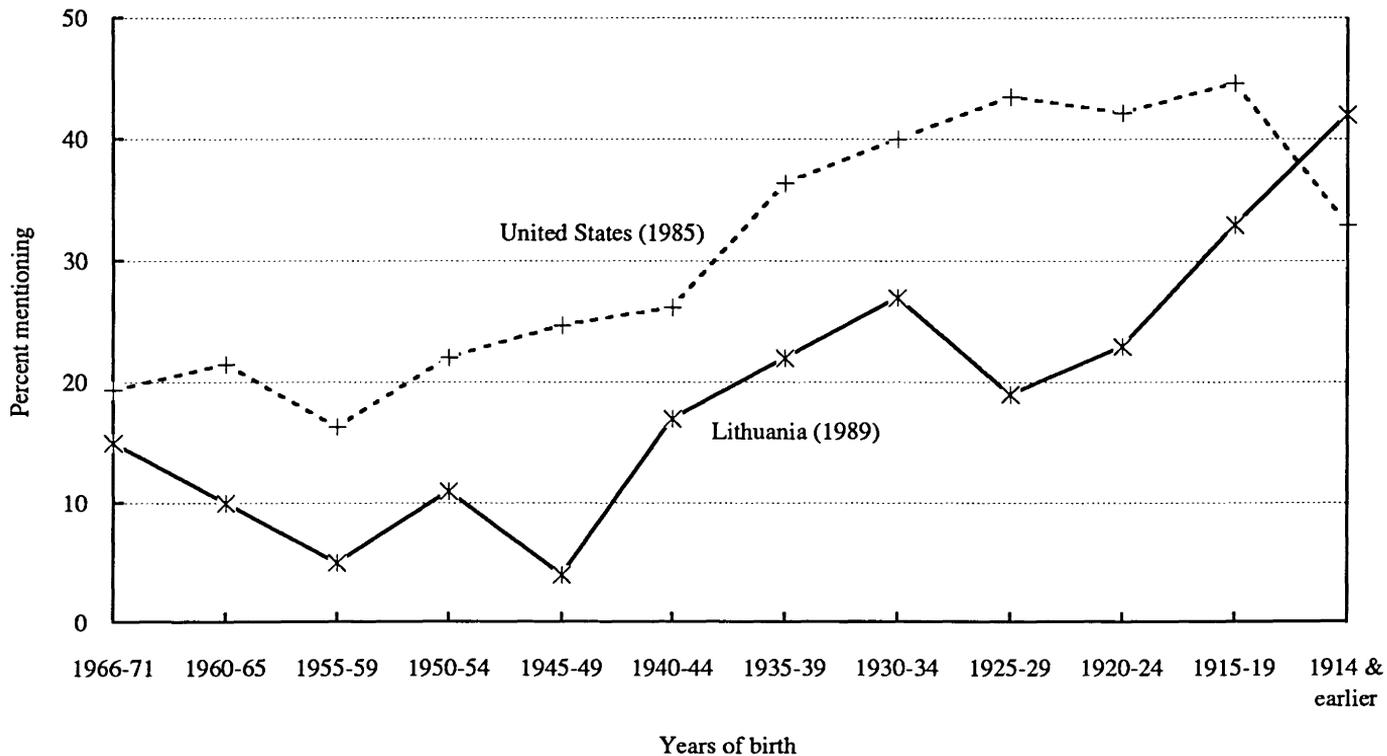


FIGURE 20.5. World War II Mentions by Cohort, United States and Lithuania

that same period, which accounts for the drop in World War II mentions. No such major political and social events competed with World War II for the attention and later memories of Americans during the 1941–1950 years, and thus there is no dip for the U.S. sample. Similarly, prior to 1941 the Great Depression attracted mentions of the oldest U.S. cohorts, whereas the Depression had less of an impact on heavily rural Lithuania during the 1930s.

Differences in historical circumstances between the United States and Lithuania are also reflected in the high level of agreement for the most frequently mentioned group of events in Lithuania: the process of “rebirth.” It may be that the high mention of “rebirth” by Lithuanians of all ages represents a period effect, reflecting the impact of objectively important recent events. However, the recent events included under the category of “rebirth” have made a larger impression on younger respondents than on older ones, even though the “rebirth” movement at the time of the survey had the potential to reverse the effects of earlier events that older respondents experienced in their youth—the annexation and Stalinization of Lithuania.

Overall, then, the relations of age to remembered events in Lithuania substantially replicate the previous findings in the United States and the generational hypothesis that explains them. At the same time, the results of this study allow us to extend the theory by identifying the extent to which youths in the same historical and geographical location (such as Lithuanians and Russians in present-day Lithuania) may identify as important *different* events from a commonly experienced flow of history. Moreover, the data presented here indicate that a single objective historical event may be experienced so differently by members of different groups that the subjective events are essentially distinct from each other.

In addition, the Lithuanian data emphasize the part that the structure of history plays in the formation of collective memories. Not only the rate of social or political change but also the length, uniqueness, and objective impact of events are important in creating distinct generational perspectives.

Moreover, the findings from the United States and Lithuania enable us to speculate on the functions of collective memories. Clearly, group identity may contribute to the selection of events seen as important; this may be especially true when group differences are highly politicized or groups see themselves as oppressed. In turn, it is plausible that, as theorists such as Halbwachs (1950/1980) and Strauss (1959) have suggested, a shared understanding of history supports group identity. Thus, Edward Shils (1981, p. 168) writes of a “consensus through time”—that is, “the persistence of the past into the present, through the maintenance of the partial identity of a society between its past and its present”—that provides a group with a stable identity even as its members are continually being replaced. If Hosking (1989) is correct in arguing that unexpurgated memory of the historical past provides necessary guidance for a society’s future course, particularly in totalitarian societies, the preservation of memories that have been repressed from official historical discourse, such as the Lithua-

nian memories of Stalinist deportations and executions, may become especially important when the course that a society is to take is contested.

Our findings indicate ways in which the content of individual memories of public events are shaped by life course, social structures, and patterns of historical events—variables previously neglected in much of the work on autobiographical memory. The complex ways in which these variables contribute to individual recollections of recent history demonstrate links between an individual's memory and the larger collectivity not adequately captured by either the "now print" mechanism or the "reconstructive" approach. Consideration of these variables in future work on autobiographical memory can assist in a richer understanding of memory and cognition in social contexts.

*Acknowledgments.* The authors wish to thank Katherine Bischooping, Victor Nakas, Jacqueline Scott, and V. Stanley Vardys for their comments on an earlier version of this chapter. The research was supported by National Science Foundation grant SES-8911588 and by a grant from the National Institute on Aging (1 RO1 AG08951).



# References

- Aborn, M. (1953). The influence of experimentally induced failure on the retention of material acquired through set and incidental learning. *Journal of Experimental Psychology*, *45*, 225–231.
- Abrams, D. B., Follick, M. J., Biener, L., Carey, K. B., & Hitti, J. (1987). Saliva cotinine as a measure of smoking status in field settings. *American Journal of Public Health*, *77*, 846–848.
- Adler, A. K. (1984). *Developmental effects on event memory*. Unpublished manuscript, Hamilton College, Clinton, NY.
- Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203–231.
- Alexander, I. (1990). *Personology: Method and content in personality assessment and psychobiography*. Durham, NC: Duke University Press.
- Allerback, K., & Hoag, W. (1985). *Jugend ohne Zukunft? Einstellungen, Umwelt, Lebensperspektiven*. Munich: Piper.
- Almond, S., & Verba, A. (1963). *The civic culture: Political attitudes and democracy in five nations*. Princeton: Princeton University Press.
- Alreck, P. L., & Settle, R. B. (1985). *The survey research handbook*. Homewood, IL: Irwin.
- Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, *77*, 171–179.
- Aristotle. (1905). De memoria et reminiscencia [On memory and reminiscence]. In J. M. Robertson (Ed.), *The works of Aristotle*. London: George Routledge and Sons.
- Bachman, J. G., & O'Malley, P. M. (1981). When four months equals a year: Inconsistencies in student reports of drug use. *Public Opinion Quarterly*, *45*, 536–548.
- Back, K. W. (1956). The well-informed informant. *Human Organization*, *14*, 30–33.
- Back, K. W. (1988). Myths in the lives of leaders of social movements: The case of the family planning movement. *Biography*, *11*, 95–107.
- Baddeley, A. (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.
- Baddeley, A. D., Lewis, V., & Nimmo-Smith, I. (1978). When did you last . . . ? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 77–83). London: Academic Press.

- Bahrack, H. P., & D. Karis (1982). Long-term ecological memory. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 427–465). London: Academic Press.
- Bakan, D. (1966). *The duality of human existence*. Boston: Beacon Press.
- Banaji, M. R. (1986). *Affect and memory: An experimental investigation*. Unpublished doctoral dissertation, Department of Psychology, Ohio State University.
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44*, 1185–1193.
- Banaji, M. R., & Greenwald, A. G. (1991). *Measuring implicit attitudes*. Paper presented at the meeting of the American Psychological Society, Washington, DC.
- Barclay, C. R. (1986). Schematization of autobiographical memory. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 82–99). Cambridge, England: Cambridge University Press.
- Barret, D. M. (1938). Memory in relation to hedonic tone. *Archives of Psychology*, *131* (No. 223).
- Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 193–243). Cambridge, England: Cambridge University Press.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.
- Bauer, R. A. (1964). The obstinate audience: The influence process from the point of view of social communication. *American Psychologist*, *19*, 319–328.
- Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology*, *56*, 669–679.
- Belson, W. A. (1986). *Validity in survey research*. Aldershot, Hants, England: Gower.
- Bem, D. J., & McConnell, H. K. (1970). Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, *14*, 23–31.
- Benowitz, N. L., Kuyt, F., Jacob, P., Jones, R. T., & Osman, A. (1983). Cotinine disposition and effects. *Clinical Pharmacological Therapy*, *34*, 604–611.
- Berman, W. (1988). The role of attachment in the post-divorce experience. *Journal of Personality and Social Psychology*, *54*, 496–503.
- Berscheid, E. (1983). Emotion. In H. H. Kelley, E. Berscheid, A. Christensen, J. H. Harvey, T. L. Huston, G. Levinger, E. McClintock, L. A. Peplau, & D. R. Peterson, *Close relationships* (pp. 110–168). New York: Freeman.
- Bickart, B. A., Blair, J., Menon, G., & Sudman, S. (1989). Cognitive aspects of proxy reporting of behavior. *Advances in Consumer Research*, *17*, 198–206.
- Billiet, J., Loosveldt, G., & Waterplas, L. (1988). *Response-effecten bij survey-vragen in het Nederlands taalgebied*. Leuven, Belgium: Katholieke Universiteit, Sociologisch Onderzoeksinstituut.
- Bishop, G. F., Hippler, H.-J., Schwarz, N., & Strack, F. (1988). A comparison of response effects in self-administered and telephone surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 321–340). New York: Wiley.
- Blair, E., & Burton, S. (1986). Processes used in the formulation of behavioral frequency reports in surveys. *Proceedings of the American Statistical Association Section on Survey Research Methods* (pp. 481–487). Washington, DC: American Statistical Association.

- Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research*, *14*, 280–288.
- Blair, E., Sudman, S., Bradburn, N., & Stocking, C. (1977). How to ask questions about drinking and sex. *Journal of Marketing Research*, *14*, 316–321.
- Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin*, *99*, 229–246.
- Block, G. (1982). A review of validations of dietary assessment methods. *American Journal of Epidemiology*, *115*, 492–505.
- Bodenhausen, G. V., & Wyer, R. S., Jr. (1987). Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 6–41). New York: Springer-Verlag.
- Bohmstedt, G. W. (1983). Measurement. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 70–121). New York: Academic Press.
- Boon, S. D., & Holmes, J. G. (1990). *Interpersonal trust, attachment and emotion*. Paper presented at the meeting of the International Society for the Study of Personal Relationships, Oxford.
- Borus, M. E. (1966). Response error in survey reports of earnings information. *Journal of the American Statistical Association*, *61*, 729–738.
- Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research theory* (Vol. 1, pp. 230–325). London: Academic Press.
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*, 129–148.
- Bower, G. H. (1987). Commentary on mood and memory. *Behavioral Research and Therapy*, *25*, 443–455.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in comprehension and memory. *Cognitive Psychology*, *11*, 177–220.
- Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality*, *13*, 420–432.
- Bower, G. H., & Mayer, J. D. (1985). Failure to replicate mood-dependent retrieval. *Bulletin of the Psychonomic Society*, *23*, 39–42.
- Bower, G. H., & Mayer, J. D. (1989). In search of mood-dependent retrieval. *Journal of Social Behavior and Personality*, *4*, 121–156.
- Bower, G. H., Monteiro, K. P., & Gilligan, S. G. (1978). Emotional mood as a context of learning and recall. *Journal of Verbal Learning and Verbal Behavior*, *17*, 573–585.
- Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss*. New York: Basic Books.
- Boyle, N. (1991). *Goethe: The poet and his time* (Vol. I). New York: Oxford University Press.
- Bradburn, N. M. (1983). Response effects. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 289–328). New York: Academic Press.
- Bradburn, N. M., & Danis, C. (1984). Potential contributions of cognitive research to survey questionnaire design. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 101–129). Washington, DC: National Academy Press.
- Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, *236*, 157–161.
- Bradburn, N. M., & Sudman, S. (1979). *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. San Francisco: Jossey-Bass.

- Bradbury, T. N., & Fincham, F. D. (1990). Attributions in marriage: Review and critique. *Psychological Bulletin*, *107*, 3–33.
- Braungart, M. M. (1984). Aging and politics. *Journal of Political and Military Sociology*, *12*, 79–98.
- Brewer, W. F. (1986). What is autobiographical memory? In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 25–49). Cambridge, England: Cambridge University Press.
- Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 21–90). Cambridge, England: Cambridge University Press.
- Brewer, W. F. (1992). The theoretical and empirical status of the flashbulb memory hypothesis. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: The problem of "flashbulb" memory* (pp. 274–305). Cambridge, England: Cambridge University Press.
- Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, *6*, 473–486.
- Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.
- Brewer, W. F., & Pani, J. R. (1983). The structure of human memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 17, pp. 1–38). New York: Academic Press.
- Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230.
- Brigham, J. C. (1988). Is witness confidence helpful in judging eyewitness accuracy? In M. M. Grunenberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 77–82). Chichester: Wiley.
- Brody, N. (Ed.). (1987). The unconscious [Special issue]. *Personality and Social Psychology Bulletin*, *13*(3), 291–429.
- Brooks, J. O., & Watkins, M. J. (1989). Recognition memory and the mere exposure effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 968–976.
- Brown, N. R., Shevell, S. K., & Rips, L. J. (1986). Public memories and their personal context. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 137–158). Cambridge, England: Cambridge University Press.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–99.
- Brown, R., & Kulik, J. (1982). Flashbulb memories. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 23–40). San Francisco: Freeman.
- Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337.
- Bruce, D., & Read, J. D. (1988). The how and why of memory for frequency. In M. M. Grunenberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 317–322). Chichester: Wiley.
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge: Harvard University Press.
- Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly*, *55*, 50–79.

- Butler, R. N. (1968). The life review. In B. Neugarten (Ed.), *Middle age and aging* (pp. 486–496). Chicago: University of Chicago Press.
- Byrne, M., Troy, A., Bradley, L. A., Marchisello, P. J., Geisinger, K. F., Van der Heide, L. H., & Prieto, E. J. (1982). Cross-validation of the factor structure of the McGill Pain Questionnaire. *Pain, 13*, 193–201.
- Campell, J. W. (1981). *The phenomenology of recall: Introspective reports of remembering stories*. Unpublished manuscript, Trenton State College, Trenton, NJ.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology 1981* (pp. 389–437). San Francisco: Jossey-Bass.
- Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology, 32*, 38–48.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York: Springer-Verlag.
- Chassein, B., Strack, F., & Schwarz, N. (1987, April). *Erinnerungsstrategie und Häufigkeitsskala: Zum unterschiedlichen Einfluss von relationaler versus episodischer Erinnerung auf Häufigkeitsurteile*. Paper presented at the 29th Tagung experimentell arbeitender Psychologen, Aachen, Germany.
- Cinciripini, P.M., & Floreen, A. (1983). An assessment of chronic pain behavior in a structured interview. *Journal of Psychosomatic Research, 27*, 117–124.
- Clark, L. F., & Collins, J. E. (in press). Remembering old flames: How the past affects assessments of the present. *Personality and Social Psychology Bulletin*.
- Clark, L. F., Collins, J. E., & Henry, S. M. (1991). *The tendency to review autobiographical memories: The impact of recollections on present well-being*. Unpublished manuscript.
- Clark, L. F., Henry, S. M., & Collins, J. E. (1991). *The effect of autobiographical remembrances on current well-being of the elderly*. Unpublished manuscript.
- Clark, M. S. (1982). A role for arousal in the link between feeling states, judgments, and behavior. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition* (pp. 263–289). Hillsdale, NJ: Erlbaum.
- Clark, M. S., Milberg, S., & Erber, R. (1984). Effects of arousal on judgments of others' emotions. *Journal of Personality and Social Psychology, 46*, 551–560.
- Clark, M. S., Milberg, S., & Erber, R. (1988). Arousal-state-dependent memory: Evidence and implications for understanding social judgments and social behavior. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition and social behavior: New evidence and integrative attempts* (pp. 63–83). Toronto: Hogrefe.
- Clark, M. S., Milberg, S., & Ross, J. (1983). Arousal cues arousal-related material in memory: Implications for understanding effects of mood on memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 633–649.
- Clore, G. L., & Parrot, W. G. (1991). Moods and their vicissitudes: Thoughts and feelings as information. In J. P. Forgas (Ed.), *Emotion and social judgment* (pp. 107–123). Oxford: Pergamon Press.
- Cobb, S., Thompson, D. J., Rosenbaum, J., Warren, J. E., & Merchant, W. R. (1956). On the measurement of prevalence of arthritis and rheumatism from interview data. *Journal of Chronic Diseases, 3*, 134–139.
- Cockshut, A. O. J. (1974). *Truth to life: The art of biography in the nineteenth century*. London: Collins.
- Cofer, C. N. (1967). Does conceptual organization influence the amount retained in

- immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory* (pp. 181–214). New York: Wiley.
- Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology, 71*, 858–866.
- Cohen, G. (1986). Everyday memory. In G. Cohen, M. W. Eysenck, & M. E. LeVoi (Eds.), *Memory: A cognitive approach*. Milton Keynes, England: Open University Press.
- Collins, J. E., & Clark, L. F. (1988). *Person perception of past lovers*. Paper presented at the annual meeting of the American Psychological Association, Atlanta.
- Connerton, P. (1989). *How societies remember*. Cambridge, England: Cambridge University Press.
- Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. R. Tufté (Ed.), *The quantitative analysis of social problems* (pp. 168–189). Reading, PA: Addison-Wesley.
- Conway, M. A. (1990). *Autobiographical memory: An introduction*. Milton Keynes, England: Open University Press.
- Conway, M. A. (1991). In defense of everyday memory. *American Psychologist, 46*, 19–26.
- Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology, 47*, 738–748.
- 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.
- Crovitz, H. F., & Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society, 4*, 517–518.
- Croyle, R. T., & Uretsky, M. B. (1987). Effects of mood on self-appraisal of health status. *Health Psychology, 6*, 239–253.
- Darschin, W., & Frank, B. (1982). Tendenzen im Zuschauerverhalten. Telekopie-Ergebnisse zur Fernsehnutzung im Jahre 1981. *Media Perspektiven, 4*, 276–284.
- Deaux, K. (1976). *The behavior of women and men*. Monterey, CA: Brook/Cole.
- Deaux, K. (1984). From individual differences to social categories: Analysis of a decade's research on gender. *American Psychologist, 39*, 105–116.
- Denzin, N. K. (1989). *Interpretive biography*. Newbury Park, CA: Sage.
- DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Deceiving and detecting deceit. In B. R. Schlenker (Ed.), *The self and social life* (pp. 323–370). New York: McGraw-Hill.
- Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York: Russell Sage Foundation.
- Duncan, G. J., & Hill, D. H. (1985). An investigation of the extent and consequences of measurement error in labor-economics survey data. *Journal of Labor Economics, 3*, 508–532.
- Duncan, G. J., & Mathiowetz, N. A. (1985). *A validation study of economic data*. Ann Arbor: University of Michigan, Institute for Social Research, Survey Research Center.
- Dutta, S., & Kanungo, R. N. (1975). *Affect and memory: A reformulation*. Oxford: Pergamon.
- Duval, S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. New York: Academic Press.

- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Original work published 1885)
- Edel, L. (1978). A manifesto. *Biography, 1*, 1-4.
- Eich, E., & Metcalfe, J. (1989). Mood dependent memory for internal versus external events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 443-455.
- Eich, E., Rachman, S., & Lopatka, C. (1990). Affect, pain, and autobiographical memory. *Journal of Abnormal Psychology, 99*, 174-178.
- Eich, E., Reeves, J. L., Jaeger, B., & Graff-Radford, S. B. (1985). Memory for pain: Relation between past and present pain intensity. *Pain, 23*, 375-380.
- Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition, 8*, 157-173.
- Eich, J. E., Weingartner, H., Stillman, R. C., & Gillin, J. C. (1975). State-dependent accessibility of retrieval cues in the retention of a categorized list. *Journal of Verbal Learning and Verbal Behavior, 14*, 408-417.
- Ellmann, R. (1988). *Oscar Wilde*. New York: Vintage.
- Epstein, S. (1973). The self-concept revisited, or a theory of a theory. *American Psychologist, 28*, 404-416.
- Erdelyi, M. H. (1974). A new look at the new look: Perceptual defense and vigilance. *Psychological Review, 81*, 1-25.
- Erdelyi, M. H., Finks, J., & Feigen-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General, 118*, 245-254.
- Erdelyi, M. H., & Goldberg, B. (1979). Let's not sweep repression under the rug: Toward a cognitive psychology of repression. In J. F. Kihlstrom & F. J. Evans (Eds.), *Functional disorders of memory* (pp. 355-402). Hillsdale, NJ: Erlbaum.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge: MIT Press.
- Eriksen, K. (n.d). *Obituary for Big Daddy: A sociological parable*. Unpublished manuscript.
- Eysenck, M. (1982). *Attention and arousal*. Berlin: Springer-Verlag.
- Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522-527.
- Fienberg, S. E., Loftus, E. F., & Tanur, J. M. (1985). Recalling pain and other symptoms. *Milbank Memorial Fund Quarterly, 63*, 582-597.
- Filipp, S.-H. (Ed.). (1981). *Kritische Lebensereignisse* (pp. 3-52). Munich: Urban & Schwarzenberg.
- Fitzgerald, J. M. (1988). Vivid memories and the reminiscence phenomenon: The role of a self narrative. *Human Development, 31*, 261-273.
- Fleming, J. H., Darley, J. M., Hilton, J. L., & Kojetin, B. A. (1990). Multiple audience problem: A strategic communication perspective on social perception. *Journal of Personality and Social Psychology, 58*, 593-609.
- Fordyce, W. E. (1976). *Behavioral methods for chronic pain and illness*. St. Louis: Mosby.
- Fordyce, W. E., Lansky, D., Calsyn, D. A., Shelton, J. L., Stolov, W. C., & Rock, D. L. (1984). Pain measurement and pain behavior. *Pain, 18*, 53-69.

- Freud, S. (1960). *The psychopathology of everyday life*. New York: Dover. (Original work published 1901)
- Freud, S. (1965). *Interpretation of dreams*. New York: Avon. (Original work published 1900)
- Furlong, E. J. (1951). *A study in memory*. London: Thomas Nelson.
- Gallatin, J. (1980). Political thinking in adolescence. In J. Abelson (Ed.), *Handbook of adolescent psychology* (pp. 344–382). New York: Wiley.
- Garfinkel, I., Corbett, T., MacDonald, M., McLanahan, S., Robins, P. K., Schaeffer, N. C., & Seltzer, J. A. (1988). *Evaluation design for the Wisconsin Child Support Assurance Demonstration*. Report produced under contract with Institute for Research on Poverty, University of Wisconsin-Madison for the Wisconsin Department of Health and Social Services.
- Gauld, A., & Stephenson, G. M. (1967). Some experiments relating to Bartlett's theory of remembering. *British Journal of Psychology*, *58*, 39–49.
- Gelfand, H. (1983). *Analyses of independence of attribute recall on the Trace Attribute Inventory*. Unpublished manuscript, St. Bonaventure University, St. Bonaventure, NY.
- Gilbert, D. T., & Jones, E. E. (1986). Perceiver-induced constraint: Interpretations of self-generated reality. *Journal of Personality and Social Psychology*, *50*, 269–280.
- Gilligan, S. G., & Bower, G. H. (1984). Cognitive consequences of emotional arousal. In C. Izard, J. Kagan, & R. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 547–588). New York: Cambridge University Press.
- Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, *9*, 491–501.
- Graesser, A. C. (1981). *Prose comprehension beyond the word*. New York: Springer-Verlag.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*, 603–618.
- Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory & Cognition*, *17*, 35–47.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.
- Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 550–561.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.
- Halbwachs, M. (1980). *The collective memory* (F. J. Ditter, Jr. & V. Y. Ditter, Trans.). New York: Harper & Row. (Original work in French published 1950)
- Haley, N. J., Axelrad, C. M., & Tilton, K. A. (1983). Validation of self-reported smoking behavior: Biochemical analyses of cotinine and thiocyanate. *American Journal of Public Health*, *73*, 1204–1207.
- Hamilton, D. L. (1981). Cognitive representations of persons. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium, Vol. 1* (pp. 135–159). Hillsdale, NJ: Erlbaum.
- Hamilton, L. C. (1981). Self-reports of academic performance: Response errors are not well behaved. *Sociological Methods & Research*, *10*, 165–185.
- Hansen, R. D., & Donoghue, J. M. (1977). The power of consensus: Information derived from one's own and others' behavior. *Journal of Personality and Social Psychology*, *5*, 294–302.

- Hardin, C., & Banaji, M. R. (1990). *Affective intensity and valence in memory*. Paper presented at the annual meeting of the American Psychological Association, Boston.
- Harsch, N., & Neisser, U. (1989, November). *Substantial and irreversible errors in flashbulb memories of the Challenger explosion*. Poster presented at the meeting of the Psychonomic Society, Atlanta.
- Harvey, J. H. (1987). Attributions in close relationships: Research and theoretical developments. *Journal of Social and Clinical Psychology, 5*, 420–434.
- Harvey, J. H., Flanary, R., & Morgan, M. (1986). Vivid memories of vivid loves gone by. *Journal of Social and Personal Relationships, 3*, 359–373.
- Hasher, L., & Griffin, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 318–330.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372–1388.
- Hastie, R. (1987). Information processing theory for the survey researcher. In H.-J. Hippler, N. Schwarz, & S. Sudman. (1987). *Social information processing and survey methodology* (pp. 42–70). New York: Springer-Verlag.
- Hawthorne, N. (1962). *The scarlet letter*. New York: Penguin. (Original work published 1850)
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52*, 511–524.
- Hedlund, J., Meckel, D., & Remz, N. (1981). *Everyday and flashbulb memories: The effect of gender and attribute type on retrieval mode*. Trenton, NJ: Trenton State College.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Herrmann, D. (1985). Remembering past experiences: Theoretical perspectives past and present. In T. M. Shlecker & M. P. Toglia (Eds.), *New directions in cognitive science* (pp. 262–275). Norwood, NJ: Ablex.
- Higgins, E. T. (1981). The “communication game”: Implications for social cognition and persuasion. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium, Vol. 1* (pp. 343–392). Hillsdale, NJ: Erlbaum.
- Higgins, E. T. (1990). Personality, social psychology, and person–situation relations: Standards and knowledge activation as a common language. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 301–338). New York: Guilford Press.
- Hilgard, E. R. (1980). Consciousness in contemporary psychology. *Annual Review of Psychology, 31*, 1–26.
- Hill, P., Haley, N. J., & Wynder, E. L. (1983). Cigarette smoking: Carboxyhemoglobin, plasma nicotine, cotinine, and thiocyanate vs. self-reported smoking data and cardiovascular disease. *Journal of Chronic Diseases, 36*, 439–449.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.
- Hippler, H.-J., Schwarz, N., & Sudman, S. (Eds.). (1987). *Social information processing and survey methodology*. New York: Springer-Verlag.
- Hobbes, T. (1952). *Leviathan*. Great Books of the Western World, Vol. 23. Chicago: Encyclopaedia Britannica. (Original work published 1651)
- Holmes, D. (1972). Repression or interference: A further investigation. *Journal of Personality and Social Psychology, 13*, 163–170.

- Holmes, J. G., & Rempel, J. K. (1989). Trust in close relationships. In C. Hendrick (Ed.), *Close relationships* (pp. 187–220). Newbury Park, CA: Sage.
- Holtzworth-Munroe, A., & Jacobson, N. S. (1985). Causal attributions of married couples: When do they search for causes? What do they conclude when they do? *Journal of Personality and Social Psychology*, *49*, 1398–1412.
- Horowitz, M. J. (1976). *Stress response syndromes*. New York: Aronson.
- Horowitz, M. J. (1982). Stress response syndromes and their treatment. In L. Goldberger & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects* (pp. 711–732). New York: Free Press.
- Hosking, G. A. (1989). Memory in a totalitarian society: The case of the Soviet Union. In T. Butler (Ed.), *Memory: History, culture and the mind* (pp. 115–130). London: Basil Blackwell.
- Hunter, M., Philips, C., & Rachman, S. (1979). Memory for pain. *Pain*, *6*, 35–46.
- Huston, T. L., McHale, S., & Crouter, A. (1986). When the honeymoon's over: Changes in the marriage relationship over the first year. In R. Gilmour & S. Duck (Eds.), *The emerging field of personal relationships* (pp. 109–132). Hillsdale, NJ: Erlbaum.
- Huttenlocher, J., Hedges, L. V., & Bradburn, N. M. (1990). Reports of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 196–213.
- Huttenlocher, J., Hedges, L. V., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating dates of events. *Psychological Review*, *95*, 471–484.
- Hyman, H. H. (1954). *Interviewing in social research*. Chicago: University of Chicago Press.
- Ingram, R. E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin*, *107*, 156–176.
- Ingram, R. E., & Smith, T. W. (1984). Depression and internal versus external focus of attention. *Cognitive Therapy and Research*, *8*, 139–152.
- Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 179–236). Hillsdale, NJ: Erlbaum.
- Isen, A. M. (1987). Positive affect, cognitive processes, and social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 203–253). San Diego: Academic Press.
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.
- Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin*, *13*, 314–336.
- Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.
- James, W. (1950). *The principles of psychology* (Vol. 1). New York: Dover. (Original work published 1890)
- Jamison, R. N., Sbrocco, T., & Parris, W. C. V. (1989). The influence of physical pain and psychosocial factors on accuracy of memory for pain in chronic pain patients. *Pain*, *37*, 289–294.
- Jarvis, M. J., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C., & Saloojee, Y. (1987). Comparison of tests used to distinguish smokers from nonsmokers. *American Journal of Public Health*, *77*, 1435–1438.

- Jersild, A. (1931). Memory for the pleasant as compared with the unpleasant. *Journal of Experimental Psychology*, *14*, 284–288.
- Jobe, J. B., & Loftus, E. F. (1991). Cognition and survey measurement [Special issue]. *Applied Cognitive Psychology*, *5*(1).
- Jobe, J. B., & Mingay, D. (1991). Cognition and survey measurement: History and overview. *Applied Cognitive Psychology*, *5*, 175–192.
- Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General*, *117*, 390–394.
- Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*, 371–376.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.
- Jones, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, *34*, 107–117.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*, 1–24.
- Jonides, J., & Naveh-Benjamin, M. (1987). Estimating frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 230–240.
- Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge, England: Cambridge University Press.
- Karr, J. W. (1990). *Assessing the accuracy of conflicting memories for the same event*. Unpublished master's thesis, University of Waterloo, Ontario, Canada.
- Katon, W. (1984). Depression: Relationship to somatization and chronic mental illness. *Journal of Clinical Psychiatry*, *45*, 4–11.
- Keefe, F. J., & Block, A. R. (1982). Development of an observation method for assessing pain behavior in chronic low back pain patients. *Behavior Therapy and Research*, *13*, 363–375.
- Keefe, F. J., Brantley, A., Gerdenio, M., & Crisson, J. E. (1985). Behavioral assessment of head and neck cancer pain. *Pain*, *23*, 327–336.
- Keefe, F. J., & Gil, K. M. (1986). Behavioral concepts in the analysis of chronic pain syndromes. *Journal of Consulting and Clinical Psychology*, *54*, 776–783.
- Keefe, F. J., & Hill, R. W. (1985). An objective approach to quantifying pain behavior and gait patterns in low back pain patients. *Pain*, *21*, 153–161.
- Keefe, F. J., Wilkins, R. H., & Cook, W. A. (1984). Direct observation of pain behavior in low back pain patients during physical examination. *Pain*, *20*, 59–68.
- Keefe, F. J., Wilkins, R. H., Cook, W. A., Crisson, J. E., & Muhlbaier, L. H. (1986). Depression, pain and pain behavior. *Journal of Consulting and Clinical Psychology*, *54*, 665–669.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. L. Vine (Ed.), *Nebraska Symposium on Motivation 1967* (pp. 192–240). Lincoln: University of Nebraska Press.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*, 107–128.

- Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York: Wiley.
- Kennedy, D. M. (1970). *Birth control in America: The career of Margaret Sanger*. New Haven: Yale University Press.
- Kent, G. (1985). Memory for dental pain. *Pain, 21*, 187–194.
- Kent, G. (1989). Memory of dental experiences as related to naturally occurring changes in state anxiety. *Cognition and Emotion, 3*, 45–53.
- Kerns, R. D., Turk, D. C., & Rudy, T. E. (1985). The West Haven–Yale Multidimensional Pain Inventory (WHYMPI). *Pain, 23*, 345–356.
- Kitayama, S. (1990). Interaction between affect and cognition in word perception. *Journal of Personality and Social Psychology, 58*, 209–217.
- Klein, S. B., & Kihlstrom, J. F. (1986). Elaboration, organization and the self-reference effect in memory. *Journal of Experimental Psychology: General, 115*, 26–38.
- Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology, 65*, 190–193.
- Kleinsmith, L. J., & Kaplan, S. (1964). Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology, 67*, 124–126.
- Krauss, R. M. (1981). Impression formation, impression management, and nonverbal behaviours. In E. T. Higgins, C. P. Hermann, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium, Vol. 1* (pp. 323–341). Hillsdale, NJ: Erlbaum.
- Krosnick, J. A. (1992). The impact of cognitive sophistication and attitude importance on response-order and question-order effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 203–218). New York: Springer-Verlag.
- Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*, 201–219.
- Krosnick, J. A., & Berent, M. K. (1991, March). *Comparisons of party identification and policy preferences: The impact of survey question format*. Unpublished manuscript, Ohio State University.
- Kubie, L. S. (1952). Problems and techniques of psychoanalytic validation and progress. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science* (pp. 46–124). Stanford: Stanford University Press.
- Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self-other differences. *Journal of Personality and Social Psychology, 37*, 499–514.
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207*, 557–558.
- LaMotte, L., & Blair, E. (1988). On estimating distributional characteristics from categorized data. *Proceedings of the American Statistical Association Section on Survey Research Methods* (pp. 286–288). Washington, DC: American Statistical Association.
- Larsen, S. F. (1985). Specific background knowledge and knowledge updating. In J. Allwood & E. Hjelmquist (Eds.), *Foregrounding background* (pp. 25–36). Lund, Sweden: Doxa.
- Larsen, S. F. (1987). *Remembering autobiographical and non-autobiographical events*. Paper presented at the meeting of the Psychonomic Society, Seattle.
- Larsen, S. F. (1988). Remembering without experiencing: Memory for reported events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 326–355). Cambridge, England: Cambridge University Press.

- Larsen, S. F., & Plunkett, K. (1987). Remembering experienced and reported events. *Applied Cognitive Psychology, 1*, 15–26.
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer-Verlag.
- Leavitt, F. G., Garron, D. C., Whisler, W. W., & Sheinkop, M. B. (1978). Affective and sensory dimensions of back pain. *Pain, 4*, 273–281.
- Lewinsohn, P. M., & Rosenbaum, M. (1987). Recall of parental behavior by acute depressives, remitted depressives, and nondepressives. *Journal of Personality and Social Psychology, 52*, 611–619.
- Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 551–578.
- Lin, I-F. (1990). *Using respondents as proxies to estimate the characteristics of non-respondents*. Unpublished master's thesis, University of Wisconsin-Madison.
- Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376–404). San Francisco: Freeman.
- Linton, M. (1978). Real-world memory after six years: An *in vivo* study of very long term memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 69–76). London: Academic Press.
- Linton, M. (1979, July). I remember it well. *Psychology Today*, pp. 80–82, 85–86.
- Linton, M. (1982). Transformations of memory in everyday life. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 77–91). San Francisco: Freeman.
- Linton, M. (1986). Ways of searching and the contents of memory. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 50–67). Cambridge, England: Cambridge University Press.
- Linton, S. J., & Götestam, K. G. (1983). A clinical comparison of two pain scales: Correlation, remembering chronic pain, and a measure of compliance. *Pain, 17*, 57–65.
- Linton, S. J., & Melin, L. (1982). The accuracy of remembering chronic pain. *Pain, 13*, 281–285.
- Loftus, E. F., Fienberg, S. E., & Tanur, J. M. (1985). Cognitive psychology meets the national survey. *American Psychologist, 40*, 175–180.
- Loftus, E. F., Smith, K. D., Johnson, D. A., & Fiedler, J. (1988). Remembering “when”: Errors in the dating of autobiographical memories. In M. M. Gruneberg, P. E. Morris, & Sykes, R. N. (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 234–240). Chichester: Wiley.
- Luepker, R. V., Pallonen, U. E., Murray, D. M., & Pirie, P. L. (1989). Validity of telephone surveys in assessing cigarette smoking in young adults. *American Journal of Public Health, 79*, 202–204.
- Malcolm, J. (1990). *The journalist and the murderer*. New York: Knopf.
- Mandler, G. (1982). Stress and thought processes. In L. Goldberger & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects* (pp. 88–104). New York: Free Press.
- Mandler, G., Nakamura, Y., & Van Zandt, B. J. S. (1987). Nonspecific effects of exposure on stimuli that cannot be detected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 646–648.
- Mannheim, K. (1952). The problems of generations. In K. Mannheim, *Essays on the sociology of knowledge* (pp. 276–322). New York: Oxford University Press. (Original work published 1928)

- Marcus, S. (1984). *Freud and the culture of psychoanalysis*. London: Allen & Unwin.
- Markus, H. (1980). The self in thought and memory. In D. M. Wegner & R. R. Vallacher (Eds.), *The self in social psychology* (pp. 102–130). New York: Oxford University Press.
- Marquis, K. H., Marquis, M. S., & Polich, J. M. (1986). Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association*, *81*, 381–389.
- Martin, L. L. (1985). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology*, *51*, 493–504.
- Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology*, *59*, 27–37.
- Mathiowetz, N. A., & Duncan, G. J. (1988). Out of work, out of mind: Response errors in retrospective reports of unemployment. *Journal of Business and Economic Statistics*, *6*, 221–229.
- Mathiowetz, N. A., & Groves, R. M. (1985). The effects of respondent rules on health survey reports. *American Journal of Public Health*, *75*, 639–644.
- Matlin, M. W., & Stang, D. J. (1978). *The Pollyana Principle: Selectivity in language, memory, and thought*. Cambridge, MA: Schenkman.
- Mayer, J. D., & Salovey, P. (1988). Personality moderates the interaction of mood and cognition. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition and social behavior: New evidence and integrative attempts* (pp. 87–99). Toronto: Hogrefe.
- McCreary, C., Turner, J., & Dawson, E. (1982). Principal dimensions of the pain experience and psychological disturbance in chronic low back pain. *Pain*, *11*, 85–92.
- McDaniel, L. K., Anderson, K. O., Bradley, L. A., Young, T. D., Turner, R. A., Agudelo, C. A., & Keefe, F. J. (1986). Development of an observation method for assessing pain behavior in rheumatoid arthritis patients. *Pain*, *24*, 165–184.
- McDonald, J., & R. Moffit. (1980). The use of tobit analysis. *Review of Economics and Statistics*, *62*, 318–321.
- McFarland, C., & Ross, M. (1987). The relation between current impressions and memories of self and dating partners. *Personality and Social Psychology Bulletin*, *13*, 228–238.
- McFarland, C., Ross, M., & DeCourville, N. (1989). Women's theories of menstruation and biases in recall of menstrual symptoms. *Journal of Personality and Social Psychology*, *57*, 522–531.
- McKie, D. C., Prentice, B., & Reed, P. (1983). *Divorce: Law and the family in Canada*. Ottawa: Minister of Supply and Services Canada.
- Mead, G. H. (1964). *Selected writings* (A. J. Reck, Ed.). Indianapolis: Bobbs-Merrill.
- Means, B., & Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology*, *5*, 297–318.
- Means, B., Mingay, D. J., Nigam, A., & Zarrow, M. (1988). A cognitive approach to enhancing health survey reports of medical visits. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory for everyday life* (pp. 537–542). Chichester: Wiley.
- Means, B., Nigam, A., Zarrow, M., Loftus, E., & Donaldson, M. S. (1989). *Autobiographical memory for health-related events*. Vital and Health Statistics, Series 6, No. 2. Hyattsville, MD: National Center for Health Statistics.
- Means, B., Swan, G. E., Jobe, J. B., & Esposito, J. L. (1991). *Use of estimation strategies for reporting cigarette smoking*. Manuscript submitted for publication.
- Means, B., Swan, G. E., Jobe, J. B., Esposito, J. L., & Loftus, E. F. (1989). Recall strate-

- gies for estimation of smoking levels in health surveys. *Proceedings of the American Statistical Association Section on Survey Methods Research* (pp. 421–424). Washington, DC: American Statistical Association.
- Meltzer, H. (1930). The present status of experimental studies on the relationship of feeling to memory. *Psychological Review*, *37*, 124–139.
- Melzack, R. (1975). The McGill Pain Questionnaire: Major properties and scoring methods. *Pain*, *1*, 277–299.
- Melzack, R., & Torgerson, W. S. (1971). On the language of pain. *Anesthesiology*, *34*, 50–59.
- Menon, G. (1991). *Judgments of behavioral frequencies: An information processing perspective*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Menon, G., & Sudman, S. (1989, May). *Increasing the accuracy of respondents estimates of behavioral frequencies: A cognitive psychology perspective*. Paper presented at the meeting of the American Association for Public Opinion Research, St. Petersburg Beach, FL.
- Menon, G., Sudman, S., Bickart, B. A., Blair, J., & Schwarz, N. (1990, May). *The use of anchoring strategies by proxy respondents in answering attitude questions*. Paper presented at the meeting of the American Association for Public Opinion Research, Lancaster, PA.
- Meyer, C. B., & Taylor, S. E. (1986). Adjustment to rape. *Journal of Personality and Social Psychology*, *50*, 1226–1234.
- Miller, P. V., & Groves, R. M. (1985). Matching survey responses to official records: An exploration of validity in victimization reporting. *Public Opinion Quarterly*, *49*, 366–380.
- Molenaar, N. J. (1982). Response effects of formal characteristic of questions. In W. Dijkstra & J. van der Zouwen (Eds.), *Response behavior in the survey interview* (pp. 49–90). New York: Academic Press.
- Moore, C. M., Herrmann, D., & Vessot, P. (1989). *Kinds of retrieval processes as a function of the affect of an autobiographical memory*. Unpublished manuscript, National Institute of Mental Health.
- Moore, J. C. (1988). Self/proxy response status and survey response quality: A review of the literature. *Journal of Official Statistics*, *4*, 155–172.
- Moreland, R. L., & Zajonc, R. B. (1979). Exposure effects may not depend on stimulus recognition. *Journal of Personality and Social Psychology*, *37*, 1085–1089.
- Morton, J. (1991). The bankruptcy of everyday thinking. *American Psychologist*, *46*, 32–33.
- Mosely, R. R., & Wolinsky, F. D. (1986). The use of proxies in health surveys. *Medical Care*, *24*, 496–510.
- Moss, L., & Goldstein, H. (Eds.). (1979). *The recall method in social surveys*. London: University of London Institute of Education.
- Nahaylo, B., & Swoboda, V. (1989). *Soviet disunion: A history of the nationalities problem in the USSR*. New York: Free Press.
- Natsoulas, T. (1970). Concerning introspective “knowledge.” *Psychological Bulletin*, *73*, 89–111.
- Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 378–386.
- Neale, C., White, C., Herrmann, D. J., & Chaffin, R. (1982). *Attributes of autobiograph-*

- ical memory and childhood amnesia*. Paper presented at the meeting of the Eastern Psychological Association.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton Century Crofts.
- Neisser, U. (1982a). Memory: What are the important questions? In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 3–19). San Francisco: Freeman.
- Neisser, U. (1982b). Snapshots or benchmarks? In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 43–48). San Francisco: Freeman.
- Neisser, U. (1986). Remembering Pearl Harbor: Reply to Thompson and Cowan. *Cognition*, 23, 285–286.
- Neisser, U. (1988). Time present and time past. In M. M. Gruneberg, P. M. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 2. Clinical and educational implications* (pp. 545–560). New York: Wiley.
- Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about *Challenger*. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: The problem of "flashbulb" memory* (pp. 9–31). Cambridge, England: Cambridge University Press.
- Neisser, U., & Winograd, E. (Eds.). (1988). *Remembering reconsidered: Ecological and traditional approaches to the study of memory*. Cambridge, England: Cambridge University Press.
- Nelson K. (1983). The derivation of concepts and categories from event representations. In E. K. Scholnick (Ed.), *New trends in conceptual Representation: Challenges to Piaget's theory?* (pp. 129–149). Hillsdale, NJ: Erlbaum.
- Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 18–55.
- Neustadt, R. E., & May, E. R. (1986). *Thinking in time*. New York: Free Press.
- Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nolan-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, 101, 259–282.
- Norman, D. A., & Bobrow, D. G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, 11, 107–123.
- Norvell, K. T., Gaston-Johansson, F., & Fridh, G. (1987). Remembrance of labor pain: How valid are retrospective pain measurements? *Pain*, 31, 77–86.
- Olcott, M. B. (1990, Summer). The Lithuanian crisis. *Foreign Affairs*, pp. 30–46.
- Orleans, C. S., & Shipley, R. H. (1982). Assessment in smoking cessation research: Some practical guidelines. In F. J. Keefe & J. A. Blumenthal (Eds.), *Assessment strategies in behavioral medicine* (pp. 261–294). New York: Grune & Stratton.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Ostrom, T. M. (1987). Bipolar survey items: An information processing perspective. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 71–85). New York: Springer-Verlag.
- Ottati, V. C., Riggle, E. J., Wyer, R. S., Jr., Schwarz, N., & Kuklinski, J. (1989). Cogni-

- tive and affective bases of opinion survey responses. *Journal of Personality and Social Psychology*, 57, 404–415.
- Panzer, P. (1981). *The effect of repeated recall of autobiographical memories on the mode of retrieval strategies*. Unpublished manuscript, Hamilton College, Clinton, NY.
- Parsons, T., & Bales, R. F. (1955). *Family, socialization, and interaction process*. Glencoe, IL: Free Press.
- Pavio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76, 241–263.
- Pearce, S. A., Isherwood, S., Hrouda, D., Richardson, P. H., Erskine, A., & Skinner, J. (1990). Memory and pain: Tests of mood congruity and state dependent learning in experimentally induced and clinical pain. *Pain*, 43, 187–193.
- Pechacek, T. F., Fox, B. H., Murray, D. M., & Luepker, R. V. (1984). Review of techniques for measurement of smoking behavior. In J. D. Matarazzo, S. M. Weiss, J. A. Herd, & N. E. Miller (Eds.), *Behavioral health: A handbook of health enhancement and disease prevention* (pp. 729–754). New York: Wiley.
- Pennebaker, J. W. (1982). *The psychology of physical symptoms*. New York: Springer-Verlag.
- Pennebaker, J. W., & Lightner, J. M. (1980). Competition of internal and external information in an exercise setting. *Journal of Personality and Social Psychology*, 35, 167–174.
- Perlman, D., & Duck, S. W. (1987). *Intimate relationships: Development, dynamics, and deterioration*. Newbury Park, CA: Sage.
- Petrov, K. V. (1989). Memory and oral tradition. In T. Butler (Ed.), *Memory: History, culture and the mind* (pp. 77–96). London: Basil Blackwell.
- Phillips, D. P., & Feldman, K. A. (1973). A dip in deaths before ceremonial occasions: Some new relationships between social integration and mortality. *American Sociological Review*, 38, 678–696.
- Pierce, J. P., Dwyer, T., DiGiusto, E., Carpenter, T., Hannam, C., Amin, A., Yong, C., Sarfaty, G., Shaw, J., Burke, N., & Quit for Life Steering Committee. (1987). Cotinine validation of self-reported smoking in commercially run community surveys. *Journal of Chronic Diseases*, 40, 689–695.
- Propp, V. (1968). *Morphology of the folktale* (2nd ed.). Austin: University of Texas Press.
- Pyszczynski, T., & Greenberg, J. (1986). Evidence for a depressive self-focusing style. *Journal of Research in Personality*, 20, 95–106.
- Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of depression. *Psychological Bulletin*, 102, 122–138.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. *Psychology of Learning and Motivation*, 14, 207–262.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134.
- Rabinowitz, J. C., Mandler, G., & Patterson, K. E. (1977). Determinants of recognition and recall: Accessibility and generation. *Journal of Experimental Psychology: General*, 106, 302–329.
- Rachman, S., & Eyril, K. (1989). Predicting and remembering recurrent pain. *Behavioral Research and Therapy*, 27, 621–635.
- Raglan, Lord. (1936). *The hero: A study in tradition, myth, and drama*. London: Methuen.

- Rapaport, D. (1971). *Emotions and memory*. New York: International Universities Press. (Original work published 1942)
- Reder, L. M. (1987a). Beyond associations: Strategic components in memory retrieval. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 203–220). Hillsdale, NJ: Erlbaum.
- Reder, L. M. (1987b). Strategy selection in question answering. *Cognitive Psychology*, *19*, 90–138.
- Reed, J. (1978). *From private vice to public virtue: The birth control movement and American society since 1830*. New York: Basic Books.
- Reiser, B. J., Black, J. B., & Abelson, R. P. (1985). Knowledge structure in the organization and retrieval of autobiographical memories. *Cognitive Psychology*, *17*, 89–137.
- Reiser, B. J., Black, J. B., & Kalamarides, P. (1986). Strategic memory search processes. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 100–121). Cambridge, England: Cambridge University Press.
- Reuband, K. H. (1980). Life histories: Problems and prospects of longitudinal designs. In J. M. Clubb & E. K. Scheuch (Eds.), *Historical social research: The use of historical and process-produced data* (pp. 135–163). Stuttgart: Klett-Cotta.
- Reuband, K. H. (1988). Von äußerer Verhaltenkonformität zu selbständigem Handeln: Über die Bedeutung kultureller und struktureller Einflüsse für den Wandel in den Erziehungszielen und Sozialisationsinhalten. In H. Meulemann & O. Luthe (Eds.), *Wertewandel—Faktum oder Fiktion?* (pp. 73–97). Frankfurt: Campus.
- Reuband, K. H. (1992). Veränderungen in den familialen Lebensbedingungen Jugendlicher seit der Jahrhundertwende: Eine Analyse auf der Basis retrospektiver Daten. *Zeitschrift für Sozialisation und Erziehungssoziologie (ZSE)*, No. 2, 99–113.
- Revelle, W., & Loftus, D. A. (1990). Individual differences and arousal: Implications for the study of mood and memory. *Cognition and Emotion*, *4*, 209–237.
- Richards, J. S., Nepomuceno, C., Riles, M., & Suer, Z. (1982). Assessing pain behavior: The UAB pain behavior scale. *Pain*, *14*, 393–398.
- Richardson-Klaven, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, *39*, 475–543.
- Rintala, M. (1968). Political generations. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 6, pp. 92–96). New York: Macmillan and The Free Press.
- Rizek, R. L., & Pao, E. M. (1990). Dietary intake methodology I: USDA surveys and supporting research. *Journal of Nutrition*, *120*, 1525–1529.
- Robinson, J. (1976). Sampling autobiographical memory. *Cognitive Psychology*, *8*, 578–595.
- Robinson, J., & Young, J. (1980). Temporal patterns in smoking rate and mouth-level nicotine exposure. *Addictive Behaviors*, *5*, 91–95.
- Roche, P. A., & Gijsbers, K. (1986). A comparison of memory for induced ischemic pain and chronic rheumatoid pain. *Pain*, *25*, 337–343.
- Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043–1056.
- Roediger, H. L. (1991). They read an article? A commentary on the everyday memory controversy. *American Psychologist*, *46*, 37–40.
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677–688.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortion in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.

- Ross, L. (1984). Thoughts and research on estimates about past and future behavior. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 65–68). Washington, DC: National Academy Press.
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*, 341–357.
- Ross, M., & Fletcher, G. J. O. (1985). Attribution and social perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed.) (Vol. 2, pp. 73–122). New York: Random House.
- Ross, M., & Holmberg, D. (1990). Recounting the past: Gender differences in the recall of events in the history of a close relationship. In J. M. Olson & M. P. Zanna (Eds.), *Self-inference processes: The Ontario Symposium, Vol. 6* (pp. 135–152). Hillsdale, NJ: Erlbaum.
- Rubin, D. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, *19*, 736–755.
- Rubin, D. C. (Ed.). (1986). *Autobiographical memory*. Cambridge, England: Cambridge University Press.
- Rubin, D. C., & Baddeley, A. D. (1989). Telescoping is not time compression: A model of the dating of autobiographical events. *Memory and Cognition*, *17*, 653–661.
- Rubin, D. C., Wetzler, S. E., & Nebes, R. D. (1986). Autobiographical memory across the lifespan. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 202–221). Cambridge, England: Cambridge University Press.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Hillsdale, NJ: Erlbaum.
- Rumelhart, D. E., & Norman, D. A. (1983). Representation in memory. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Vol. 2. Learning and cognition* (2nd ed.) (pp. 511–587). New York: Wiley.
- Salovey, P. (1992). Mood-induced self-focused attention. *Journal of Personality and Social Psychology*, *62*, 699–707.
- Salovey, P., & Birnbaum, D. (1989). Influence of mood on health-relevant cognitions. *Journal of Personality and Social Psychology*, *57*, 539–551.
- Salovey, P., Jobe, J. B., Willis, G. B., Sieber, W. J., van der Sleen, S., Turk, D. C., & Smith, A. F. (1989). Response errors and bias in the recall of chronic pain. *Proceedings of the American Statistical Association Section on Survey Research Methods* (pp. 413–420). Washington, DC: American Statistical Association.
- Salovey, P., O'Leary, A., Stretton, M. S., Fishkin, S. A., & Drake, C. A. (1991). Influence of mood on judgments about health and illness. In J. P. Forgas (Ed.), *Emotion and social judgments* (pp. 241–262). Oxford: Pergamon Press.
- Salovey, P., & Rodin, J. (1985). Cognitions about the self: Connecting feeling states and social behavior. In P. Shaver (Ed.), *Self, situations, and social behavior* (pp. 143–166). Beverly Hills, CA: Sage.
- Salovey, P., Sieber, W. J., Smith, A. F., Turk, D. C., Jobe, J. B., & Willis, G. B. (1992). *Reporting chronic pain episodes on health surveys*. Vital and Health Statistics, Series 6, No. 6. Hyattsville, MD: National Center for Health Statistics.
- Sanger, M. (1938). *An autobiography*. New York: Norton.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501–518.

- Schaeffer, N. C., Seltzer, J. A., & Klawitter, M. (1991). Estimating nonresponse and response bias: Resident and nonresident parents' reports about child support. *Sociological Methods & Research*, 20, 30–59.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schare, M. L., Lisman, S. A., & Spear, N. E. (1984). The effects of mood variation of state-dependent retention. *Cognitive Therapy and Research*, 8, 387–408.
- Schooler, J. W., Clark, C. A., & Loftus, E. F. (1988). Knowing when memory is real. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 83–88). New York: Wiley.
- Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Personality and Social Psychology*, 12, 171–181.
- Schramm, W., & Danielson, W. (1958). Anticipated audiences as determinants of recall. *Journal of Abnormal and Social Psychology*, 19, 319–328.
- Schuman, H., & Presser, S. (1977). Question wording as an independent variable in survey analysis. *Sociological Methods & Research*, 6, 151–176.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Schuman, H., & Scott, J. (1987). Problems in the use of survey questions to measure public opinion. *Science*, 236, 957–959.
- Schuman, H., & Scott, J. (1989). Generations and collective memories. *American Sociological Review*, 54, 359–381.
- Schwartz, B. (1991). Social change and collective memory: The democratization of George Washington. *American Sociological Review*, 56, 221–236.
- Schwarz, N. (1987). *Simmung als Information*. Heidelberg: Springer-Verlag.
- Schwarz, N. (1990a). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 98–119). Beverly Hills, CA: Sage.
- Schwarz, N. (1990b). Feelings as information: Informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 527–561). New York: Guilford Press.
- Schwarz, N., & Bienias, J. (1990). What mediates the impact of response alternatives on frequency reports of mundane behaviors? *Applied Cognitive Psychology*, 4, 61–72.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513–523.
- Schwarz, N., & Clore, G. L. (1988). How do I feel about it? Informative function of affective states. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior: New evidence and integrative attempts* (pp. 44–62). Toronto: Hogrefe.
- Schwarz, N., & Hippler, H.J. (1987). What response scales may tell your respondents: Informative functions of response alternatives. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 163–178). New York: Springer-Verlag.
- Schwarz, N., & Hippler, H.-J. (1991). Response alternatives: The impact of their choice and ordering. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement error in surveys* (pp. 41–56). Chichester: Wiley.

- Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response categories: effects on behavioral reports and comparative judgments. *Public Opinion Quarterly*, 49, 388–395.
- Schwarz, N., Hippler, H.-J., & Noelle-Neumann, E. (1992). A cognitive model of response-order effects in survey measurement. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 187–201). New York: Springer-Verlag.
- Schwarz, N., & Scheuring, B. (1988). Judgments of relationship satisfaction: Inter- and intraindividual comparisons as a function of questionnaire structure. *European Journal of Social Psychology*, 18, 485–496.
- Schwarz, N., & Scheuring, B. (1992). Selbstberichtete Verhaltens- und Symptommhäufigkeiten: Was Befragte aus Antwortvorgaben des Fragebogens lernen [Frequency-reports of psychosomatic symptoms: What respondents learn from response alternatives]. *Zeitschrift für Klinische Psychologie*, 22, 197–208.
- Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 31–50). Chichester: Wiley.
- Schwarz, N., Strack, F., Hippler, H.-J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology*, 5, 193–212.
- Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, 55, 3–23.
- Schwarz, N., Strack, F., Müller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, 6, 107–117.
- Schwarz, N., & Sudman, S. (Eds.). (1992). *Context effects in social and psychological research*. New York: Springer-Verlag.
- Seltzer, J. A. (1990). Legal and physical custody arrangements in recent divorces. *Social Science Quarterly*, 71, 250–266.
- Senn, A. E. (1990). *Lithuania awakening*. Berkeley: University of California Press.
- Shannon, C., & Weaver, W. (1948). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Sharp, A., & Holmes, J. G. (1990, August). *Attributions in close relationships: A daily diary study*. Paper presented at the annual meeting of the American Psychological Association, Boston.
- Shils, E. (1981). *Tradition*. Chicago: University of Chicago Press.
- Sieber, W. J., Salovey, P., Jobe, J. B., & Willis, G. B. (1990, August). *Social cognition and reporting chronic pain episodes*. Paper presented at the annual meeting of the American Psychological Association, Boston.
- Sigel, R. S., & Hoskin, M. B. (1977). Perspectives on adult political socialization. In S. A. Renshon (Ed.), *Handbook of political socialization* (pp. 259–293). New York: Free Press.
- Sillett, R. W., Wilson, M. B., Malcolm, R. E., & Ball, K. P. (1978). Deception among smokers. *British Medical Journal*, 2, 1185–1186.
- Silver, R. L., Boon, C., & Stones, M. H. (1983). Searching for meaning in misfortune: Making sense of incest. *Journal of Social Issues*, 39, 81–102.
- Silver, R. L., & Wortman, C. B. (1980). Coping with undesirable life events. In J. Garber

- & M. E. P. Seligman (Eds.), *Human helplessness* (pp. 279–340). New York: Academic Press.
- Singer, J. A., & Salovey, P. (1988). Mood and memory: Evaluating the network theory of affect. *Clinical Psychology Review, 8*, 211–251.
- Skowronski, J. J., Betz, A. L., Thompson, C. P., & Shannon, L. (1991). Social memory in everyday life: The recall of self-events and other-events. *Journal of Personality and Social Psychology, 60*, 831–843.
- Skowronski, J. J., & Thompson, C. P. (1990). Reconstructing the dates of personal events: Gender differences in accuracy. *Applied Cognitive Psychology, 4*, 371–381.
- Smith, A. F. (1991). *Cognitive processes in long-term dietary recall*. Vital and Health Statistics, Series 6, No. 4. Hyattsville, MD: National Center for Health Statistics.
- Smith, A. F., Jobe, J. B., & Mingay, D. J. (1991). Retrieval from memory of dietary information. *Applied Cognitive Psychology, 5*, 269–296.
- Smith, A. F., Jobe, J. B., & Mingay, D. J. (1992). *Concerning the suitability of diary keeping to validate report of dietary information*. Manuscript submitted for publication.
- Smith, T. W., & Greenberg, J. (1981). Depression and self-focused attention. *Motivation and Emotion, 5*, 323–331.
- Smith, T. W., Ingram, R. E., & Roth, D. L. (1985). Self-focused attention and depression: Self-evaluation, affect, and life-stress. *Motivation and Emotion, 9*, 381–389.
- Smyth, M. M., Morris, P. E., Levy, P., & Ellis, A. W. (1987). *Cognition in action*. Hillsdale, NJ: Erlbaum.
- Snyder, C. R., & Higgins, R. L. (1988). Excuses: Their effective role in the negotiation of reality. *Psychological Bulletin, 104*, 23–35.
- Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36*, 941–950.
- Sorrentino, R. M., & Higgins, E. T. (1986). *Handbook of motivation and cognition: Foundations of social behavior*. New York: Guilford Press.
- Spence, D. P. (1982). *Narrative truth and historical truth: Meaning and interpretation in psychoanalysis*. New York: Norton.
- Spiro, R. J. (1980). Accommodative reconstruction in prose recall. *Journal of Verbal Learning and Verbal Behavior, 19*, 84–95.
- Stopes, M. (1923). *Early days of birth control* (4th ed.). London: Putnam.
- Strachey, L. (1918). *Eminent Victorians*. New York: Putnam.
- Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer-Verlag.
- Strack, F., & Schwarz, N. (1992). Communicative influences in standardized question situations: The case of implicit collaboration. In G. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 173–193). Newbury Park, CA: Sage.
- Strack, F., Schwarz, N., & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, affect, and mode of thinking. *Journal of Personality and Social Psychology, 49*, 1460–1469.
- Strack, F., Schwarz, N., & Nebel, A. (1987, March). *Thinking about your life: Affective and evaluative consequences*. Paper presented at the conference on “Ruminations, Self-relevant Cognitions and Stress,” Memphis, TN.
- Strauss, A. (1959). *Mirrors and masks: The search for identity*. Glencoe, IL: Free Press.

- Strube, G. (1987). Answering survey questions: The role of memory. In H.-J. Hippler, N. Schwarz, & S. Sudman, S. (Eds.), *Social information processing and survey methodology* (pp. 86–101). New York: Springer-Verlag.
- Sudman, S., & Bradburn, N. M. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, *63*, 805–815.
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.
- Sudman, S., & Ferber, R. (1974). A comparison of alternative procedures for collecting consumer expenditure data for frequently purchased products. *Journal of Marketing Research*, *11*, 128–135.
- Sudman, S., & Schwarz, N. (1989). Contributions of cognitive psychology to advertising research. *Journal of Advertising Research*, *29*, 43–53.
- Sudman, S., Wilson, W., & Ferber, R. (1976, June). *The cost-effectiveness of using the diary as an instrument for collecting health data in household surveys* (revised). Report to the National Center for Health Services Research. Urbana: University of Illinois, Survey Research Laboratory.
- Sweet, J., Bumpass, L., & Call, V. (1988). *The design and content of the National Survey of Families and Households*. NSFH Working Paper No. 1. University of Wisconsin-Madison, Center for Demography and Ecology.
- Tanur, J. (Ed.). (1992). *Questions about questions: Meaning, memory, expression, and social interactions in surveys*. New York: Russell Sage Foundation.
- Taylor, S. E. (1981). The interface of cognitive and social psychology. In J. H. Harvey (Ed.), *Cognition, social behavior and the environment* (pp. 189–211). Hillsdale, NJ: Erlbaum.
- Taylor, S. E. (1983). Adjustment to threatening events: A theory of cognitive adaptation. *American Psychologist*, *38*, 1161–1173.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210.
- Taylor, S. E., & Schneider, S. K. (1989). Coping and the simulation of events. *Social Cognition*, *7*, 174–194.
- Teachman, J. D., & Polonko, K. (1990). Negotiating divorce outcomes: Can we identify patterns in divorce settlements? *Journal of Marriage and the Family*, *52*, 129–139.
- Teske, K., Daut, R. L., & Cleeland, C. S. (1983). Relationship between nurses' observations and patients' self-report of pain. *Pain*, *16*, 289–296.
- Tessler, R., & Mechanic, D. (1978). Psychological distress and perceived health status. *Journal of Health and Social Behavior*, *19*, 254–262.
- Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, *10*, 324–332.
- Thompson, C. P. (1985a). Memory for unique personal events: Effects of pleasantness. *Motivation and Emotion*, *9*, 277–289.
- Thompson, C. P. (1985b). Memory for unique personal events: Some implications of the self-schema. *Human Learning*, *4*, 267–280.
- Thompson, C. P., & Cowan, T. (1986). Flashbulb memories: A nicer interpretation of a Neisser recollection. *Cognition*, *22*, 199–200.
- Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1988a). Reconstructing the date of a personal event. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 241–246). Chichester: Wiley.

- Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1988b). Telescoping in dating naturally occurring events. *Memory & Cognition*, *16*, 461–468.
- Thompson, D. J., & Tauber, J. (1957). Household survey, individual interview, and clinical examination to determine prevalence of heart disease. *American Journal of Public Health*, *47*, 1131–1140.
- Tobias, B. A., Kihlstrom, J. F., & Schacter, D. L. (1992). Emotion and implicit memory. In S.-Å. Christianson (Ed.), *Handbook of emotion and memory: Research and theory* (pp. 67–92). Hillsdale, NJ: Erlbaum.
- Tomkins, S. S. (1979). Script theory: Differential magnification of affects. In R. A. Dienstbier (Ed.), *Nebraska Symposium on Motivation 1978: Human emotion* (pp. 201–236). Lincoln: University of Nebraska Press.
- Tompkins, J. (Ed.). (1980). *Reader-response criticism*. Baltimore: Johns Hopkins University Press.
- Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 149–162). New York: Springer-Verlag.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*, 299–314.
- Trevor-Roper, H. (1978). *Hermit of Peking: The hidden life of Edmund Backhouse*. Harmondsworth: Penguin Books.
- Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, NJ: Prentice-Hall.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Tulving, E., & Thompson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, *87*, 116–124.
- Turk, D. C., & Flor, H. (1987). Pain > pain behaviors: The utility and limitations of the pain behavior construct. *Pain*, *31*, 277–295.
- Turk, D. C., Meichenbaum, D., & Genest, M. (1983). *Pain and behavioral medicine: A cognitive-behavioral perspective*. New York: Guilford Press.
- Turk, D. C., & Rudy, T. E. (1987). Towards a comprehensive assessment of chronic pain patients. *Behavior Research and Therapy*, *25*, 237–249.
- Turk, D. C., Rudy, T. E., & Salovey, P. (1985). The McGill Pain Questionnaire reconsidered: Confirming the factor structure and examining appropriate uses. *Pain*, *21*, 385–397.
- Turk, D. C., & Salovey, P. (1986). Clinical information processing: Bias inoculation. In R. E. Ingram (Ed.), *Information processing approaches to clinical psychology* (pp. 305–323). Orlando, FL: Academic Press.
- Turk, D. C., Wack, J. T., & Kerns, R. D. (1985). An empirical examination of the “pain behavior” construct. *Journal of Behavioral Medicine*, *9*, 119–130.
- Turner, B. (1991). Unpublished honors thesis, University of Waterloo.
- Turner, C. F., & Martin, E. (Eds.). (1984). *Surveying subjective phenomena*. New York: Russell Sage Foundation.
- Tversky, A., & Griffin, D. (1991). On the dynamics of hedonic experience: Endowment and contrast in judgments of well-being. In F. Strack, N. Schwarz, & M. Argyle (Eds.), *Subjective well-being* (pp. 101–118). London: Pergamon Press.

- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1982). Availability: A heuristic for judging frequency and probability. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 163–178). Cambridge, England: Cambridge University Press.
- Uhlenhuth, E., Haberman, S. J., Balter, M. D., & Lipman, R. S. (1977). Remembering life events. In J. S. Strauss, H. M. Babigan, & M. Roff (Eds.), *The origins and course of psychopathology* (pp. 117–134). New York: Plenum Press.
- Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, 76, 559–573.
- Underwood, B. J. (1983). *Attributes of memory*. Glenview, IL: Scott Foresman.
- Underwood, B. J., Zimmerman, J., & Freund, J. S. (1971). Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 87, 149–162.
- U. S. Bureau of the Census. (1988). *Who's helping out? Support networks among American families*. Current Population Reports, Series P-70, No. 13. Washington, DC: U.S. Government Printing Office.
- U. S. Bureau of the Census. (1990). *Child support and alimony: 1987*, by G. H. Lester. Current Population Reports, Special Studies, Series P-23, No. 167. Washington, DC: U.S. Government Printing Office.
- Verbrugge, L. M. (1980). Health diaries. *Medical Care*, 18, 73–95.
- Vesey, C. J., Saloojee, Y., Cole, P. V., & Russell, M. A. H. (1982). Blood carboxyhaemoglobin, plasma thiocyanate, and cigarette consumption: Implications for epidemiological studies in smokers. *British Medical Journal*, 284, 1516–1518.
- Wagenaar, W. A. (1969). Note on the construction of diagram-balanced Latin squares. *Psychological Bulletin*, 72, 384–386.
- Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, 18, 225–252.
- Walker, W. H., & Kintsch, W. (1985). Automatic and strategic aspects of knowledge retrieval. *Cognitive Science*, 9, 261–283.
- Warner, K. E. (1978). Possible increases in the underreporting of cigarette consumption. *Journal of the American Statistical Association*, 73, 314–318.
- Wasserstein, B. (1988). *The secret lives of Trebitsch-Lincoln*. New Haven: Yale University Press.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.
- Weber, A. L., Harvey, J. H., & Stanley, M. A. (1987). The nature and motivations of accounts for failed relationships. In R. Burnett, P. McGhee, & D. Clarke (Eds.), *Accounting for relationships: Explanations, representation and knowledge* (pp. 114–133). London: Methuen.
- Weiner, B. (1966). Effects of motivation on the availability and retrieval of memory traces. *Psychological Bulletin*, 65, 24–37.
- Wells, G. L., & Loftus, E. F. (1984). *Eyewitness testimony*. Cambridge, England: Cambridge University Press.
- White, R. T. (1982). Memory for personal events. *Human Learning*, 1, 171–183.
- White, S. H., & Pillimer, D. B. (1979). Childhood amnesia and the development of a

- socially accessible memory system. In J. F. Kihlstrom & F. J. Evans (Eds.), *Functional disorders of memory* (pp. 29–73). Hillsdale, NJ: Erlbaum.
- Whitten, W. B., II, & Leonard, J. M. (1981). Directed search through autobiographical memory. *Memory & Cognition*, 9, 566–579.
- Wicklund, R. A. (1982). Self-focused attention and the validity of self-reports. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior* (pp. 149–172). Hillsdale, NJ: Erlbaum.
- Willett, W. C. (1990). *Nutritional epidemiology*. New York: Wiley.
- Williams, M. D., & Santos-Williams, S. (1980). Method for exploring retrieval processes using verbal protocols. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 671–689). Hillsdale, NJ: Erlbaum.
- Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very long-term memory. *Cognitive Science*, 5, 87–119.
- Wood, J. V., Saltzberg, J. A., & Goldsamt, L. A. (1990). Does affect induce self-focus attention? *Journal of Personality and Social Psychology*, 58, 899–908.
- Wood, V., Wylie, M. L., & Sheafor, B. (1969). An analysis of a short self-report measure of life satisfaction: Correlation with rater judgments. *Journal of Gerontology*, 24, 465–469.
- Wyer, R. S., Jr., & T. K. Srull (1986). Human cognition in its social context. *Psychological Review*, 93, 322–359.
- Wyer, R. S., Jr., & T. K. Srull (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Erlbaum.
- Wyner, G. A. (1980). Response errors in self-reported number of arrests. *Sociological Methods & Research*, 9, 161–177.
- Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 2, 65–74.
- Young, C. M. (1981). Dietary methodology. In Committee on Food Consumption Patterns, National Research Council, *Assessing changing food consumption patterns* (pp. 89–118). Washington, DC: National Academy Press.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151–175.
- Zeller, A. (1950). An experimental analogue of repression: The effect of individual failure and success on memory measured by relearning. *Journal of Experimental Psychology*, 40, 411–422.
- Zinnecker, J. (1985). Kindheit, Erziehung, Families. In Jugendwerk der Deutschen Shell (Eds.), *Jugendliche und Erwachsene '85: Generationen im Vergleich* (Vol. 3, pp. 97–292). Opladen, Germany: Leske + Budrich.

Much of our knowledge about individuals' behavior, the state of society, or social change is based on direct verbal reports. From consumer behavior to health problems, and from the styles of parenting to the nation's unemployment rate or the prevalence of crime, social scientists and psychologists rely on respondents' autobiographical memory for testing theories of human behavior and offering advice on public policy. The recent revival of basic psychological research into autobiographical memory coincides with an increasing interest in cognitive aspects of the survey process in the survey research community. Presumably, the better we understand the cognitive processes used to answer retrospective questions, the better we shall be able to design questionnaires that facilitate respondents' performance. At the least, we should be able to get a better understanding of how well respondents are able to perform the tasks that are given them, and what tasks it is reasonable to expect a respondent to perform. These hopes have fueled much of the recent collaboration between cognitive psychologists and survey methodologists on this topic reported in *Autobiographical Memory and the Validity of Retrospective Reports*. This book presents a careful theoretical framework to explain autobiographical memory as well as the most recent empirical findings on retrospective reports of behavior, event dating and time estimation, and comparisons of self and proxy reports. While there is still much to learn, the book demonstrates that significant progress has been made in recent years in this important area.

Schwarz  
Sudman  
Editors

Autobiographical Memory and the  
Validity of Retrospective Reports

ISBN 0-387-94167-3  
ISBN 3-540-94167-3

