A New Agenda for Interdisciplinary Survey Research Methods

From the CENTERS FOR DISEASE CONTROL AND PREVENTION National Center for Health Statistics

CASM II

Proceedings of the CASM II Seminar

A New Agenda for Interdisciplinary Survey Research Methods:

Proceedings of the CASM II Seminar

CASM II

Edited by
Monroe Sirken, Thomas Jabine, Gordon Willis,
Elizabeth Martin, and Clyde Tucker

PREFACE

It was in 1993, the tenth anniversary of the CASM I Seminar, that I first began to think about the CASM II Seminar. The obvious success of the CASM I Seminar in promoting and advancing communication between survey researchers and cognitive psychologists and other scientists, and the benefits realized in developing cognitive methods for designing questionnaires were matters worthy of careful review and assessment. Also, I sensed a waning in the momentum generated by the CASM I Seminar; the moment appeared right to reinvigorate the CASM fostering effort, and compose a roadmap for the future of interdisciplinary survey methods research beyond the year 2000. With support of staff at the National Center for Health Statistics, especially Susan Schechter, and Tom Jabine, a consultant to the Center, a CASM II Seminar proposal was developed in mid-1995. The proposal profited from discussions that I had with Robert Groves and Graham Kalton.

Taking advantage of subsequent funding opportunities provided by the National Science Foundation (NSF) and the National Center for Health Statistics (NCHS), the CASM II Seminar became a viable project during the latter part of 1995. The seminar would not have been feasible without the support of Ed Sondik, Director of the NCHS, and Cheryl Eavey, Head of the Measurement and Statistics Program, NSF.

The seminar became a reality with the establishment of the CASM II Planning Committee comprising prominent interdisciplinary minded researchers. Norman Bradburn served as chair of the first meeting of the committee in early 1996. Other committee members were Murray Aborn, Robert Groves, Doug Herrmann, Tom Jabine, Betsy Martin, Susan Schechter, Norbert Schwarz, Judy Tanur, Roger Tourangeau, Clyde Tucker, Gordon Willis and myself. Guided largely by the experience of the CASM I Seminar, the Planning Committee developed the CASM II Seminar agenda. For example, plenary sessions presented commissioned papers assessing the past performance and guiding the future performance of CASM research. Also, working groups met daily to discuss and propose innovative research methods for improving survey capabilities responsive to selected high priority survey needs. Presentation of an oral history video tape of the original

contributors to the CASM movement was a unique feature of the CASM II Seminar.

Thanks to the efforts of several NCHS staff, especially Susan Schechter, Karen Whitaker and Barbara Butler, and the Native American Management Services, Inc., who made the logistical and travel arrangements, the CASM II Seminar convened at the Boar's Head Inn, Charlottesville Virginia, June 8-13, 1997, almost exactly 14 years after the CASM I Seminar met in St. Michaels, Maryland, June 15-21, 1983. Many thanks are due to each of the 50 or so persons who attended the CASM II Seminar. Every one of them contributed in many ways to the seminar's success by presenting papers, leading working groups, participating in and preparing working group reports, serving as rapporteurs for plenary sessions, etc.

It is impossible to name all the individuals who contributed to the CASM II Seminar, but I want to recognize and thank these individuals for the tasks noted: Judy Tanur, Norbert Schwartz, Roger Tourangeau, and Doug Herrmann for organizing the seminar's four plenary sessions; Betsy Martin and Clyde Tucker for organizing and overseeing the activities of the working groups; Norm Bradburn for agreeing at the last minute to chair a closing session where he summarized and conceptualized the thoughts that had emerged; Barbara Wilson and Paul Beatty for their work on the oral history project, and those who served as interviewers and interviewees; the Bureau of the Census for arranging the SIPP interviews, and Karen Bogen for summarizing SIPP interviews with seminar participants prior to the seminar; Susan Schechter, Gordon Willis and Tom Jabine for coordinating the work of publishing the seminar's results; Kenneth Prewitt and Judy Tanur for advice about possible funding sources for CASM research, and Ed Sondik, Cheryl Eavey, Kathy Wallman and Nancy Kirkendall, and members of the Federal Committee on Statistical Methodology, especially Bob Fay, Larry Cox, Alan Tupek, and David Williamson for their post seminar roles in successfully planning a future funding mechanism to support interdisciplinary survey methods research.

Shortly after the seminar, the CASM II Planning Committee members and NCHS staff presented preliminary summaries of the commissioned papers and the working group reports to the survey community - at

an invited session at the annual meeting of the American Statistical Association in Los Angeles, CA during August 1997, and at a half-day meeting in Washington D.C. during November, 1997 that was sponsored by the Washington Statistical Society and the Washington D.C. chapter of the American Association for Public Opinion Research.

The CASM Seminar gave rise to two companion volumes. In addition to these Proceedings, a monograph, *Cognition and Survey Research* will appear in the Wiley series Probability and Statistics: Survey Methodology Section. (Editors of the Wiley monograph generously assigned their royalties to a CASM research fund.) Readers are referred to the introductory chapter of these Proceedings for detailed information about the contents of both publications. Suffice to say here that the monograph contains 15 commissioned papers along with introductions to the four plenary sessions by their chair/organizers; these Proceedings contain the working group reports, abstracts of the commissioned papers, summaries of comments of the seminar participants, and selected other material.

In the Preface to the Proceedings of the CASM I Seminar (Jabine et al., 1984) Judy Tanur noted that the CASM I Seminar Proceedings might be considered an experiment in encouraging cross-disciplinary research. I feel very much the same way about the CASM II Seminar reports, and I feel that the experiment is paying off. Lessons learned from the CASM movement about effective ways of fostering interdisciplinary research combined with the knowledge accumulating from the experiences of other successful research fostering efforts will, I believe, contribute to the success of interdisciplinary survey methods research fostering efforts that are adopted in the future (the reader is referred to the Epilogue of these Proceedings for more on that topic). My thanks to Arnold Zellner, Robert Groves and Seymour Sudman for sharing their experiences about infrastructural arrangements for convening meetings to sustain research efforts.

Finally, the other Editors join me in thanking Pat Dean Brick who served as executive editor, and Karen Whitaker who served as editorial assistant, for improving the readability of these Proceedings.

Monroe G. Sirken
National Center for Health Statistics
April 7, 1999

Reference

Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (Eds.), (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines.* Washington, DC: National Academy Press.

TABLE OF CONTENTS

COMMISSIONED PAPER ABSTRACTS: SESSION 3

COMMISSIONED PAPER ABSTRACTS: SESSION 4

WORKING GROUP REPORTS

CONCLUDING SESSION

CURRENT ISSUES AND FUTURE DIRECTIONS

EPILOGUE

APPENDIX A

APPENDIX B

APPENDIX C

INTRODUCTION

Thomas B. Jabine

The CASM II Seminar

These Proceedings document the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology (CASM), held in Charlottesville, Virginia, June 8-13, 1997. The Proceedings include material that complements or further develops seminar presentations.

The CASM movement seeks to foster collaboration between survey practitioners and cognitive scientists in research activities that benefit both disciplines. Since its inception, the CASM movement has exerted considerable influence on the conduct and organization of survey research. By the mid-1990s, many of those active in this research felt the need to review accomplishments and consider directions for the future.

Organization of the Seminar

The format of CASM II was similar to that of CASM I, held in St. Michaels, Maryland in June 1983. In the CASM II Seminar, emphasis was given to commissioned papers, 15 of which were presented and discussed in four plenary sessions. The following topics were dealt with:

1. Mission and history of the CASM movement.

2. Effects of the CASM movement on cognitive theory and survey methods.

3. Potential contributions of other disciplines to the CASM movement.

4. Potential contributions of the CASM movement beyond questionnaire design.

Introductions to the plenary sessions, abstracts of the papers, and summaries of the discussions that followed the presentations are included in the Commissioned Paper Sections of these Proceedings. The complete papers will appear in a forthcoming monograph, *Cognition and Survey Research*, published by John Wiley & Sons.

Each seminar participant was also a member of one of eight Working Groups:

1. CASM in a Changing Survey Environment.

2. Exploring the Interview Process.

3. Different Disciplinary Perspectives on the Question and Answer Process.

4. Applying Cognitive Methods to New Areas of the Survey Process.

5. Income Measurement.

6. Integrating Cognitive Research into Household Survey Design.

7. Measurement of Disability.

8. Adapting Cognitive Techniques to Establishment Surveys.

These groups met periodically during the seminar, reported preliminary findings and recommendations to plenary sessions, and produced their final reports after the seminar. The Working Groups Reports Section contains these reports.

Bradburn (Concluding Session) summarizes the key issues identified during the course of the seminar. These points provided the foundation for an open discussion which is summarized in that section. Participants were encouraged to write down any post hoc thoughts and considerations, and those that were submitted are also included.

Other parts of these Proceedings were prepared several months after the seminar, at which time the working group reports had been completed and progress had been made in efforts to develop a mechanism for funding ongoing CASM research. The section on Current Issues and Future Directions by Willis, Sirken and Jabine attempts to enumerate and analyze the research topics and specific research proposals that emerged from the discussions at the seminar and from the working group reports. Finally, Sirken (Epilogue) proposes a road map for charting the future of

interdisciplinary survey research. It is our hope that the material presented in this volume will point the way to a productive agenda for the future of the CASM movement.

Appendices

Appendix A contains material from an oral history project which was undertaken to provide an archival record of interviews with those who played important roles in the early stages of the CASM movement. Interviews with 17 persons, some in groups and some individually, were videotaped in late 1996 and early 1997. A composite tape, about one hour long, with excerpts from the individual taped interviews, was created and shown on the first evening of the CASM II Seminar. A lightly edited transcript of that tape and information about the availability of all of the taped interviews appears in Appendix A.

Prior to the CASM I Seminar, arrangements had been made for most participants to be interviewed by regular Census Bureau interviewers in the National Center for Health Statistics' (NCHS) National Health Interview Survey. The goal of conducting these interviews was to provide a common starting point for subsequent discussions of the cognitive aspects of survey interviews. This strategy worked so well that it was repeated prior to CASM II, this time using the Census Bureau's Survey of Income and Program Participation (SIPP). The SIPP questionnaire had recently been converted to a computer-assisted personal interview (CAPI) mode, thus giving many of the participants their first experience with CAPI. In addition to providing a common experience for all seminar participants, the SIPP interview was of special interest to those who joined the Working Group on Income Measurement, one of the eight working groups organized for the seminar. All of those who were interviewed were invited to provide comments on their reactions to the interview experience. Appendix B contains the summaries of these comments.

Attendance at the CASM II Seminar was by invitation. Each participant had one or more designated roles, such as session chair, presenter of a paper, rapporteur, or cochair of a working group. There were 58 participants, more than twice the number at CASM I. There were many more invited papers and each of the eight Working Groups required two cochairs. Most participants attended all or nearly all of the seminar sessions with only a few attending one or two sessions. Short biographies of the seminar participants are presented in Appendix C.

Chronology of Important CASM Events and Activities

This section briefly summarizes the origin and accomplishments of the CASM movement and important events prior to the CASM II Seminar. For additional details, see Aborn (Session 1) and O'Muircheartaigh (Session 1) (also see their chapters and Tanur's introduction in *Cognition and Survey Research* (forthcoming), Tanur and Fienberg (1996), Jobe and Mingay (1991), the Preface to *Questions About Questions* (Tanur, ed., 1991), and *Thinking About Answers* (Sudman, Bradburn, and Schwarz, 1996). The report of the first CASM Seminar (Jabine, Straf, Tanur, and Tourangeau, 1984) describes relevant events prior to CASM I.

Exhibit 1 presents some milestones of the CASM movement, starting in 1978, emphasizing the movement's influence on U.S. federal statistical agencies.

Exhibit 1. Milestones in the CASM Movement

1978	Seminar on problems associated with the collection and interpretation of retrospective and recall data in Social Surveys, U.K.
1980	Workshop on Applying Cognitive Psychology to Recall Problems of the National Crime Survey, Washington, DC.
	Committee on National Statistics convenes Panel on Survey Measurement of Subjective Phenomena.
	Federal Committee on Statistical Methodology establishes Subcommittee on Questionnaire Design.
1981	Census Bureau establishes Center for Survey Methods Research.
1981-86	The Bureau of Social Science Research bases its experimental redesign of the National Crime Survey instrument on hypotheses and approaches derived from cognitive psychology.
1983	Advanced Research Seminar on Cognitive Aspects of Survey Methodology, (CASM I), St. Michaels, MD, June 15-21.
1984	Follow-up meeting to CASM I Seminar, Baltimore, MD, January 12-14.
	Zentrum für Umfragen, Methoden und Analysen (ZUMA) International Conference

on Social Information Processing and Survey Methodology, Germany.

NCHS experiment, supported by NSF, demonstrates utility of cognitive laboratory methods for designing and testing questionnaires of ongoing surveys.

Social Science Research Council establishes Committee on Cognition and Survey Research.

1985   NCHS establishes a Cognitive Research Staff and creates a permanent Questionnaire Design Research Laboratory (QDRL).

1986   NCHS, supported by NSF, establishes the National Laboratory for Collaborative Research in Cognition and Survey Measurement to advance interdisciplinary survey-oriented basic academic research.

1986-93   The Bureau of the Census and the Bureau of Labor Statistics conduct a collaborative program of cognitive research and testing in support of the redesign of the Current Population Survey, implemented in January 1994.

1987   ZUMA establishes formal program on "Cognition and Survey Methods."

1988   Bureau of Labor Statistics establishes Collection Procedures Research Laboratory.

NCHS establishes a new publication series "Cognition and Survey Measurement" for reports of experiments supported by the National Laboratory for Collaborative Research.

1993   NCHS establishes a Working Paper Series for summary reports of questionnaires tested in the Questionnaire Design Research Laboratory.

1994-7   The Bureau of Labor Statistics and the Bureau of the Census conduct cognitive and qualitative testing of alternative questions and answer formats for race and ethnicity which, together with field tests, are used by the Office of Management and Budget to support recommendations for revisions to Directive 15, the Government standard for reporting race and ethnicity.

1995   The Census Bureau adopts a pretesting policy that all new and revised questions fielded in its demographic surveys must be cognitively tested.

1997   Second Advanced Seminar on the Cognitive Aspects of Survey Methodology (CASM II), Charlottesville, VA, June 8-13.

The beginnings of both cognitive psychology and survey research preceded the CASM movement by many decades. Some of the pioneers in survey research were initially trained as psychologists and this orientation was reflected in their efforts to improve questionnaire design and reduce nonsampling errors. However, as Aborn (Session 1) points out, it was not until psychology and linguistics turned to theories of cognition that the idea of a mutually beneficial collaboration between survey researchers and cognitive psychologists began to take shape. Each of the first two conferences listed in Exhibit 1, one in the U.K. in 1978 and one in the U.S. in 1980, brought survey researchers, cognitive psychologists, and social scientists together to explore memory and recall as cognitive processes with major implications for the quality of survey data. The U.K. conference addressed the topic in a general framework covering surveys that collect retrospective data (Moss and Goldstein, 1979). The U.S. conference focused on the National Crime Survey, which asked respondents to report on past incidents in which they were victims of crime. The conference participants attempted to devise cognitively-informed methods for improving recall and reporting of victimizations (Moore, 1980). Following the conference, the program of research and testing to redesign the instrument for the National Crime Survey explicitly incorporated and evaluated experiments based on hypotheses derived from cognitive psychology (Loftus and Marburger, 1983; Biderman et al., 1986; Martin et al., 1986).

Also in 1980, the Committee on National Statistics convened a Panel on Survey Measurement of Subjective Phenomena, with participation by survey researchers, statisticians, cognitive psychologists, sociologists, and other social scientists. The goal of this panel was to review broad issues related to the collection of opinions and other subjective data in surveys (Turner and Martin, 1981, 1984, Vols. 1 and 2). Among the panel's recommendations were to "broaden the basic science component of survey design and practice," and to "organize ... an extensive interdisciplinary investigation of subjective aspects of survey questions." The panel noted that "only some of the disciplines with potential contributions to make to survey research are actually involved" and recommended a collaborative research program including cognitive science, linguistics, and anthropology, as well as other disciplines to improve the base of scientific knowledge about surveys (Turner and Martin, 1981, pp. 60-61).

Another significant event in 1980 was the establishment, by the Office of Management and Budget's Federal Committee on Statistical Methodology, of a Subcommittee on Questionnaire Design. The subcommittee's assignment was to review and evaluate methods currently use to develop survey questionnaires. The subcommittee's final report, *Approaches to Developing Questionnaires* (U.S. Office of Management and Budget, 1983), included chapters on unstructured individual interviewing, qualitative group interviews, and participant observation. At that time, such techniques were used in the early stages of questionnaire development by a few survey organizations; they later became the mainstay of the cognitive research laboratories established by the Bureau of Labor Statistics (BLS), the Census Bureau, and NCHS.

The CASM I Seminar in June 1983 was a landmark event in the sense that it gave a formal title and structure to the CASM movement. It initiated a deliberate effort to foster interdisciplinary research on the cognitive aspects of survey methods. The seminar and the follow-up meeting in January 1984, convened by the Committee on National Statistics with funding from the National Science Foundation, "... were the main elements of the CASM project, whose goal was to foster a dialogue between cognitive scientists and survey researchers and to develop ideas and plans for collaborative research" (Jabine et al., 1984, p.1). Planning for the CASM project and the CASM I Seminar was guided by the following four principles:

- The project should encourage the development of proposals for collaborative research by cognitive scientists and survey researchers.

- In addition to recall, which had been the focus of earlier meetings, the project should also consider other cognitive processes that take place in survey interviews, such as comprehension and judgment.

- A small group of experts from the relevant disciplines should meet for an extended period to explore the common ground between the cognitive sciences and survey research and to develop ideas to stimulate joint research.

- Participation in the project should prove beneficial to members of both disciplines.

Encouragement for the development of joint research proposals was provided by the announced interest of the National Science Foundation in funding the most promising proposals to emerge from the CASM I Seminar.

Twenty-two researchers participated in the week-long CASM I Seminar. Two general background papers (Tourangeau, 1984; Bradburn and Danis, 1984) were developed and sent to participants before the seminar. A third paper (Marquis, 1984), on the subject of record checks, was based on a presentation at the seminar. Many of the discussions at the seminar were focused on the National Health Interview Survey. Prior to the seminar, participants had been interviewed in that survey, and two interviews with volunteer respondents had been videotaped for viewing at the seminar. In addition to attending the plenary sessions, the participants met in small working groups to discuss three cognitive processes that take place in surveys—comprehension, retrieval, and judgment/response—and three of the principal topics included in the National Health Interview Survey questionnaire—utilization of health services, health conditions, and restricted activity. Each seminar participant was a member of two working groups, one from each category.

Following the CASM I Seminar, several of the participants began new research activities or developed proposals for new research along the lines discussed at the seminar. These activities, descriptions of which are included in the final report of CASM I, were the primary focus of a two-day follow-up meeting in Baltimore in January 1984. Parts of two videotaped interviews from the National Opinion Research Center's General Social Survey, which includes numerous questions about respondents' attitudes and perceptions, were shown and discussed. Subsequently, a critique of the assumptions of standardized interviewing, based on an analysis of these videotaped interviews, was published (Suchman and Jordan, 1990). The Baltimore meeting also included small-group sessions which discussed additional research proposals and means for sustaining the CASM movement following the conclusion of the CASM I project. The participants concluded that the project had more than met its initial goals: "...it had generated several promising interdisciplinary research activities and plans, and it had established an informal network of scientists who appreciated the benefits of collaboration between cognitive scientists and survey researchers" (Jabine, Straf, Tanur and Tourangeau, 1984, p.155).

During the 1980s, the effects of the CASM movement on the field of survey research were seen in the creation of new institutional mechanisms for basic and applied research, increased use of cognitive techniques for the development and testing of survey questionnaires, incorporation of cognitive theories and techniques in courses on survey research, and rapid growth of a formal CASM literature. Developments in the U.S. included the establishment of cognitive research units in three of the principal statistical agencies: the Census Bureau, the NCHS, and the BLS. Some work at the Census Bureau preceded the formal CASM movement, for example, Rothwell (1985) and others, as part of extensive efforts to improve the quality of self-administered questionnaires for the 1980 Census of Population, had used observers and other qualitative, small group methods with volunteer subjects who were

asked to complete test versions of the questionnaires. In 1981, research on census and survey questionnaires, which until then had been carried out by the Census Bureau's Statistical Research Division, was shifted to a separate organization unit, the Center for Survey Methods Research. Since then, the center has been active in applied cognitive research for major demographic surveys conducted by the Census Bureau. In 1995, the center was instrumental in establishing the Census Bureau's current policy that all new and revised questions fielded in demographic surveys must be pretested using the more comprehensive cognitively-based methods (Bureau of the Census, n.d.).

Immediately after the CASM I Seminar, the NCHS began to develop laboratory methods for routinely testing its data collection instruments and, with funding from the NSF, undertook an experiment to compare laboratory and field methods of pretesting survey questionnaires. The results suggested that the methods were complementary rather than competitive. Soon thereafter, NCHS's newly established cognitive research staff in the Office of Research and Methodology developed a combined laboratory and field procedure for designing, testing, and assessing survey questionnaires, and that procedure became standard for NCHS data systems and for many extramural studies conducted for other federal agencies. In 1985, the cognitive research staff was allotted dedicated space for establishing a Questionnaire Design Research Laboratory (QDRL). In 1986, a second NSF grant resulted in merging the QDRL into the newly established National Laboratory for Collaborative Research in Cognition and Survey Measurement. NSF supported the Collaborative Research Program that awarded research grants to university scientists and appointed visiting scientists to undertake research at NCHS (Sirken, 1991), and NCHS continued its support for the QDRL. In recent years, the NCHS has expanded the focus of the CASM movement beyond the data collection phase of surveys by sponsoring programs to investigate the cognitive aspects of statistical map reading (Pickle and Herrmann, 1995).

In 1988, the BLS established a Collection Procedures Research Laboratory, whose underlying goal is "... to improve through interdisciplinary research the quality of data collected and published by BLS." The scope of the laboratory's research is considered to cover "... all forms of oral and written communication in the collection and processing of survey data" (Dippo and Herrmann, 1991). Some of the laboratory's work is undertaken for other agencies and some is contracted to private survey research organizations. Since 1991, the laboratory has undertaken work to improve the design of data collection software. This work has included the creation and evaluation of expert systems to assist data collectors in the field as well as more traditional computerized data collection instruments. In 1996, the laboratory began investing in specialized equipment to conduct in-house software usability testing, including the usability of data collection software for the Internet.

Beginning in 1986, the BLS and the Census Bureau undertook a collaborative research and development effort to redesign the Current Population Survey. The program included laboratory and field studies to identify conceptual issues and comprehension problems. Cognitive methods were then used to evaluate the efficacy of revised instruments (see, e.g., Polivka and Rothgeb, 1993). Starting in 1994, the same two agencies, as part of a governmentwide program sponsored by the Office of Management and Budget (OMB), conducted cognitive and qualitative testing of alternative questions and answer formats for the collection of information about race and ethnicity. These studies and field tests were used by the OMB to support its recommendations for revisions to Directive 15, the government standard for reporting on race and ethnicity. The new standards, which the OMB issued in October 1997, will be used in the 2000 census and their use will be required by all federal agencies as of January 1, 2003 (Evinger, 1996, 1997).

Soon after the CASM I Seminar and follow-up meeting, the Social Science Research Council, with funding from NSF and the Russell Sage Foundation, formed a Committee on Cognition and Survey Research. The committee met regularly to encourage research and held eight workshops between 1986 and 1990. Its work was summarized in a volume entitled *Questions about Questions: Inquiries into the Cognitive Bases of Surveys* (Tanur, ed., 1992), which included sections on meaning, memory, attitude measurement in surveys, social interaction, and government applications.

A counterpart to the U.S. CASM movement developed in Germany at about the same time, beginning in 1984 with an International Conference on Social Information Processing and Survey Methodology organized by ZUMA (Zentrum für Umfragen, Methoden und Analysen), a government-funded social science methodology center (Hippler, Schwarz, and Sudman, 1987). That conference led to an active program of cognitive research in survey settings and a series of conferences on cognition and survey measurement. The establishment of cognitive laboratories in major U.S. statistical agencies has been duplicated in other countries, one example being Statistics Sweden's Measurement, Evaluation and Development Laboratory, established in 1989 (Bergman, 1995). In other countries, existing units in statistical agencies and organizations started to use cognitive testing methods for the development and revision of their census and survey questionnaires.

Thus, by the mid-1990s, the CASM movement could be viewed in many ways as a remarkable success. Cognitive testing methods had become commonplace, although by no means universal, in the development of questionnaires for population censuses and household surveys. A substantial amount of basic and applied

research had been conducted and findings could be found in published volumes and journals. The cognitive aspects of surveys had gained a prominent place in the curricula of university programs in survey methodology, with the Joint Program in Survey Methodology (a joint undertaking of the Universities of Maryland and Michigan, and Westat, Inc.) offering a course in "Social and Cognitive Foundations of Survey Measurement" and the University of Michigan's Summer Institute in Survey Research Techniques offering a course in "Cognition, Communication and Survey Measurement." Sessions on basic and applied cognitive research relating to surveys had become routine at annual meetings of the American Statistical Association. The anticipated benefits of the CASM movement to the cognitive sciences were less in evidence, but national household surveys were being used by researchers at the National Institute on Aging (C. Schooler, personal communication to T.B. Jabine, Jan. 15, 1998) and the University of Michigan's Survey Research Center (Herzog and Wallace, 1995) as a vehicle for study of the relationships between age and cognitive abilities.

Three considerations pointed towards a second CASM seminar: (1) a need for a formal review and evaluation of what had been accomplished to date; (2) a need to assess the cognitive implications of recent developments in survey technology and the environment for surveys; and (3) a desire to explore the potential benefits of extending the scope of the CASM movement in several directions. A primary objective of CASM II was to provide a forum where promising opportunities for new interdisciplinary research on all aspects of surveys could be uncovered and explored.

Under the heading of *review and evaluation*, answers were sought to the following kinds of questions:

- To what extent can it be demonstrated that the use of cognitive testing methods in developing survey questionnaires has actually improved the quality of the data collected using those questionnaires?

- Of the many cognitive tools that have been used in questionnaire development, which are the most effective, individually or in combination? How should they be used in combination with more traditional methods?

- What kinds of arrangements for incorporating basic and applied cognitive research activities within the structure of government and other data collection organizations have proved most effective?

One of the most significant changes in survey technology and environment since the CASM I Seminar has been the widespread adoption of CASIC (computer-assisted information collection) methods in major surveys. Use of these methods in face-to-face and telephone interviews and for self-administered questionnaires has lead to significant changes in the cognitive interactions between survey designers, interviewers and respondents. However the effects of these cognitive interactions on data quality are far from being fully understood. Other challenges for which the CASM movement may be able to contribute to solutions are the increasing volume, complexity, and sensitivity of the content of surveys, the growing distrust of government and consequent resistance to participation in surveys, and changes in the structure and living arrangements of U.S. families.

Past successes suggest that significant benefits might be derived by expanding the scope of the CASM movement in several directions:

- Seeking cooperation with other branches of the cognitive sciences and related disciplines, such as linguistics, expert systems and artificial intelligence.

- Extending applications of cognitive methods to surveys of businesses and organizations.

- Taking a cue from the NCHS Cognitive Aspects of Maps program, seeking applications not just in data collection, but in all phases of surveys, including initial conceptualization and design, data processing, and all modes of data dissemination.

References

Bergman, L. (1995). Pretesting procedures at Statistics Sweden's Measurement, Evaluation and Development Laboratory. *Journal of Official Statistics*, 11, 309-323.

Biderman, A., Cantor, D., Lynch J., and Martin, E. (1986). Final report of the National Crime Survey redesign program. Washington, DC: Bureau of Social Science Research.

Bradburn, N. and Danis, C. (1984). Potential contributions of cognitive research to survey questionnaire design. In T. Jabine, M. Straf, J. Tanur, and R. Tourangeau, (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 101-129). Washington, DC: National Academy Press.

Bureau of the Census (n.d.). Pretesting policy and options: Demographic surveys at the Census Bureau. Washington, DC: Bureau of the Census.

Dippo, C. and Herrmann, D. (1991). The Bureau of Labor Statistics' Collection Procedures Research Laboratory: Accomplishments and future directions. In *Seminar on quality of federal data*, Statistical Policy Working Paper 20. Washington, DC: Federal Committee on Statistical Methodology, U.S. Office of Management and Budget, 253-267.

Evinger, S. (1996). Results of CPS research on the measurement of race and ethnicity. *Chance, 9* , 53-56.

Evinger, S. (1997). Recognizing diversity: Recommendations to OMB on standards for data on race and ethnicity. *Chance, 10*, 26-31.

Herzog, A.R. and Wallace, R.B. (1995). Measures of cognitive functioning in the AHEAD study. *Journal of Gerontology*, Series B, Psychological and Social Sciences, 52B (Special issue), 37-48.

Hippler, H., Schwarz, N., and Sudman, S. (Eds.). (1987). *Social information processing and survey methodology.* New York: Springer-Verlag.

Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines.* Washington, DC: National Academy Press.

Jobe, J. and Mingay, D. (1991). Cognition and survey measurement: History and overview. *Applied Cognitive Psychology, 5*, 175-192.

Loftus, E. and Marburger, W. (1983). Since the eruption of Mount Saint Helens, did anyone beat you up? Improving the accuracy of retrospective reports with landmark events. *Memory and Cognition, 11*, 114-120.

Marquis, K. (1984). Record checks for sample surveys. In T. Jabine, M. Straf, J. Tanur, and R. Tourangeau. (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 130-147). Washington, DC: National Academy Press.

Martin, E., Groves, R., Matlin, J., and Miller, C. (1986). Report on the development of alternative screening procedures for the National Crime Survey. Washington, DC: Bureau of Social Science Research.

Moore, J. (1980). Report of the workshop on applying cognitive psychology to recall problems of the National Crime Survey. Washington, DC: Bureau of Social Science Research.

Moss, L. and Goldstein, H. (Eds.). (1979). *The recall method in social surveys*. Studies in Education (new series) 9. Windsor, Ontario: NFER Publishing Company.

Pickle, L. and Herrmann, D. (Eds.). (1995). *Cognitive aspects of statistical mapping*. National Center for Health Statistics: Cognitive Methods Staff, Working Paper No. 18.

Polivka, A. and Rothgeb, J. (1993). Redesigning the questionnaire for the Current Population Survey. Paper presented at the annual meetings of the American Economic Association, Anaheim, CA, January.

Rothwell, N. (1985). Laboratory and field research studies for the 1980 Census of Population of the United States. *Journal of Official Statistics*, *1*, 137-157.

Sirken, M. (1991). The role of a cognitive laboratory in a statistical agency. In *Seminar on quality of federal data,* Statistical Policy Working Paper 20. Washington, DC: Federal Committee on Statistical Methodology, U.S. Office of Management and Budget, 268-277.

Sirken, M., Herrmann, D., Schechter, S., Schwarz, N., Tanur, J., and Tourangeau, R. (Eds.). (1999). *Cognition and survey research*. New York: John Wiley & Sons.

Suchman, L. and Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85*, 232-241.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Tanur, J. (Ed.). (1992). *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.

Tanur, J. and Fienberg, S. (1997). Surveying an interdisciplinary CASM. In L. Salter and A. Hearn. (Eds.), *Outside the lines: Issues in interdisciplinary research* (pp. 102-109). Montreal: McGill-Queens University Press.

Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. Jabine, M. Straf, J. Tanur, and R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines.* (pp. 73-100). Washington, DC: National Academy Press.

Turner, C. and Martin, E. (1981, 1984). *Surveys of subjective phenomena* (Vols. 1-2). Cambridge MA: Harvard University Press.

U.S. Office of Management and Budget (1983). *Approaches to developing questionnaires*. Statistical Policy Working Paper 10. Washington, DC: Federal Committee on Statistical Methodology, U.S. Office of Management and Budget.

Looking Backwards and Forwards at the CASM Movement

Judith M. Tanur

Introduction

This is the first of the formal sessions of the CASM II Seminar, and the first of the introductions. The introduction is supposed to give a substantive overview of the presentations included in the session. Yet, the introductions, unlike discussions, perforce occur before the presentations, and thus the introducer must take care not to steal the thunder of the presenters. Nevertheless, I wish to explain what we hope this session will accomplish, say something about the presenters, and then let the presenters have their say and encourage you to respond.

As the Organizing Committee planned the seminar, and in particular this session, I found myself thinking of it as the "Janus Session," drawing on the popular image of Janus as looking back and forward at the same time. Indeed, this session will review the territory covered by the CASM movement and where we have not yet been and might usefully go. As I looked further into the content of the Janus myth, I found myself more convinced that Janus was the appropriate patron for this session, and perhaps for the entire seminar. He was the ancient god of beginnings and activities related to beginnings. His name was invoked when sowing grain. His blessing was asked at the beginning of the day, the month, the year—and perhaps the seminar. If we date the formal beginning of the CASM movement to the middle of the 1980s, we have had a decade and a half of work conducted under the rubric of CASM. In the survey research community, "cognitive" (as in cognitive testing, cognitive processing, cognitive interviewing) has become a household—or office—word. Indeed, at the recent American Association for Public Opinion Research (AAPOR) meeting the word appeared in the titles of two sessions, cognitive testing was described as a routine part of the pretesting of questionnaires used in reported empirical research, and several papers were devoted to the exploration of cognitive theories of responding to surveys. We have clearly made much progress. But we have also long worried about several issues. Is it a question of old wine in new bottles? Or, given that we are doing things differently, is there any hard evidence that we are doing them better? We use ideas mainly from cognitive psychology, ignoring the other cognitive sciences—can that situation be changed?

We are more likely to use tools borrowed from our sister discipline than to use cognitive theory to inform us of the tools or interpretation of results—would it prove profitable to take a more theoretical stance? The initial CASM Seminar visualized the partnership between survey research and the cognitive sciences as a two-way street, with researchers in the cognitive sciences using large-scale surveys as a means of testing the generalizability of their laboratory-based findings. Why do we see so little traffic in that direction, and what can we do about it? This session will expand on where we have come from, what we have missed along the way, and ultimately, where we might usefully go in the future.

Our first speaker is Murray Aborn who will talk about how far we have come. Although retired, Murray has not slowed down a bit. I have known Murray for many years and I served on his review panel at National Science Foundation (NSF). I saw first hand his care and carefulness in fostering good research in the social sciences. When I first joined the panel, the program was called Measurement Methods and Data Resources--and it was the funding source of not only cross-social science methodological research but also of such large and important data sets as the General Social Survey, the National Election Surveys, and the Panel Study of Income Dynamics. To ensure the safety of these important data resources, Murray arranged for their transfer to the disciplinary programs most interested in maintaining them. The acceptance of such a reduction of one's own power for the greater good of the social science research enterprise seemed then to me, and still seems, truly admirable. The program was then renamed Measurement Methods and Data Improvement. This change also allowed the program to focus increasingly on forging partnerships with researchers in the cognitive sciences and pursuing new ways of improving survey data. I was still on Murray's panel when the funding for the CASM I Seminar was being considered, as well as during the time that some of the follow-up proposals from the seminar were arriving. Of course, I had to excuse myself from the panel's deliberations over the CASM I proposal—not only was I then a member of the National Research Council's Committee on National Statistics, which was sponsoring the seminar, but I was slated to chair the seminar if funding was awarded. So I did not get to hear how vigorously Murray pushed the

panel and how vigorously they pushed him, but I have my suspicions. I did, however, get early and full reports from Murray and the other panel members about their decisions and deliberations and witnessed the wise guidance that they, and particularly Murray, exerted on the shaping of the proposed interdiscipline. This seems to constitute strong evidence for the claim I make that Murray Aborn is the Godfather of the CASM movement. Murray's talk is entitled "CASM Revisited."

Our second speaker, Colm O'Muircheartaigh, has had a distinguished research career in statistics, survey sampling, measurement error models, and cognitive aspects of surveys. I have not known Colm quite as long as I have known Murray—I believe I met him during the 1984-1985 academic year while he was an American Statistical Association (ASA)/NSF Research Fellow at the U.S. Bureau of the Census. Since then I have run into him often at meetings of ASA, ISI, AAPOR, and at activities sponsored by the SSRC Committee on Cognition and Survey Research. On those occasions I was consistently impressed by Colm's reports of his research on sampling and on cognitive aspects of surveys. I have been awed by his comments several times when I have heard him act as discussant for a paper or group of papers. A good discussant can point out what has not been done, what is missing. Hence Colm was a logical choice when looking for an author for a presentation tentatively entitled "Gaps in CASM Achievements." In some sense, Colm was being asked to be a discussant of the entire decade and a half of CASM research. His presentation is entitled "CASM: Successes, Failures, and Potential."

To return to the Janus image, Janus is also the god of entrances, of going in and coming out, which means he is the god of doorways, bridges, ferries, harbors, boundaries—and perhaps of bridges over interdisciplinary chasms. We are familiar with images of Janus showing two faces, one looking into the future, and one into the past. As we attempt here to broaden the scope of the CASM movement, perhaps we should also revive images of Janus from the reign of Hadrian which showed him with four faces—he could then look simultaneously at several disciplines united in the mutual attempt to improve survey theory and practice and to enrich the parent disciplines themselves.

Abstract

CASM Revisited

Murray Aborn

This paper presents a selected history of CASM from three different perspectives: (1) as the latest in a long train of efforts to exert control over sources of nonsampling error in surveys; (2) as an attempt to foster an interdisciplinary research enterprise involving collaboration between practitioners of cognitive science and survey statisticians; (3) and as a movement to alert the survey research and cognitive science communities of the gains in knowledge derived from an interactive relationship between the two. In each case, examples of forerunner research studies, projects, or programs are given, and the wisdom that emanates from these studies is discussed.

The seeds of CASM may have been planted early in the chronology of survey research, but CASM, per se emerged relatively recently, that is, 1983—as a result of intellectual and institutional forces at play in the previous decade. Activism on the part of the National Research Council's Committee on National Statistics and the National Science Foundation produced the first CASM Seminar (CASM I) which, after taking an inventory of the problems to be faced and the obstacles to be overcome, developed plans for building a two-way interdisciplinary bridge between survey research and cognitive science that would increase the store of information available to each. In creating those plans, CASM I is seen as having foreshadowed developments relevant to CASM's viability that are now coming to light. First, recognition of the importance to cognitive science of establishing collaboration between researchers in the "field" (so to speak) and those working in experimental research laboratories. Second, growing scientific and societal concern over the complex dilemmas created by advances in technology that selectively favor those possessing particular cognitive capabilities. Perhaps the most important factor in solving such dilemmas will be the ability to make populationwide assessments of cognitive skills. The current trend toward the support of research that responds to the needs of applied work is also seen as advantageous to CASM's future given its applied orientation and obvious relevance to current research needs.

CASM's effect on survey research has been little less than awesome considering the short time it has had to make its influence felt. Of particular importance has been the establishment of cognitive research laboratories at three foremost federal statistical agencies—laboratories which have steadily made significant contributions to the improvement of ongoing national surveys. While CASM's effect on cognitive science pales by comparison, survey researchers and cognitive psychologists working under the CASM banner have been remarkably successful in inducting CASM-engendered research studies into the cognitive literature. Notwithstanding all the positive indications for a bright future, the paper ends on a cautionary note by citing CASM's vulnerabilities.

Abstract

CASM: Successes, Failures, and Potential

Colm O'Muircheartaigh

As a starting point for evaluation of the achievements of the CASM initiative, this paper first reviews the history of survey research and its emergence from three different sectors—government, academia, and business—each with its own disciplinary bases. A framework for evaluation is presented, making use of models of the survey process and identifying three criteria—*representation*, *control*, and *realism*—that are essential to the successful outcome of any research undertaking.

The beginnings of the CASM initiative were noteworthy for efforts to overcome barriers to cross-disciplinary work and for successful attraction of funding for collaboration and research. Notable progress has been achieved in establishment of cognitive laboratories in survey organizations, and application of the methods of cognitive psychology to issues of question form and wording.

Cognitive psychology, with its focus on memory, language comprehension, inferential reasoning, and judgment, has provided a theoretical framework for applied research on questionnaire design, subsequently expanded to cover the social nature of the survey interview. This area of work has produced a large and distinguished literature and has transformed the nature of academic research on survey questioning. Examples include the inclusion/exclusion model of assimilation and contrast effects, Krosnick's model of the cognitive miser and satisficing, studies aimed at better understanding of nonresponse, and work of Dillman and colleagues on design of self-administered questionnaires.

The paper also reviews arguments against the use of standardized interview processes and survey questions as a method of generating information and notes the shift in emphasis from the interviewer as the sole focus of attention to a joint consideration of the interviewer and respondent as taking part in a communication process where each has both expectations and responsibilities. Structures provided by disciplines other than cognitive psychology, such as ethnography and the theory of social representations, provide a basis for better understanding of the interview process, but findings may be difficult to translate into instruments that will satisfy the needs of large—scale survey research.

Examination of the use of cognitive laboratories shows that they have become facilities for pretesting questionnaires for use in ongoing surveys, with success evaluated by the throughput of questionnaires in a documented and timely manner. There do not appear to have been systematic efforts to evaluate the methods used and the results of pretesting, and the laboratories have not made significant contributions to the development of general principles for question or questionnaire design. The cognitive laboratory is seen as an example of *Taylorism*, in which an innovation is accepted, but in a form that assigns to it a classic production line function and separates it from the rest of the organization, thus minimizing the two-directional flow of ideas.

Finally, the development of computer-assisted information collection (CASIC) procedures has transformed the possibilities of the survey interview, but these new possibilities are far from being fully exploited. Further advances in survey methodology will require three elements: recognition of the *diversity of the population* (across categories and over time); the need to ground interview questions in a *social* reality; and the *empowerment* of the interviewer.

Summary of Presentations and Discussion

Fran Featherston

CASM Revisted: Murry Aborn

The planners of the first CASM Seminar in 1983 aimed at a cross-fertilization between survey research and the cognitive sciences. They envisioned that survey researchers would use the principles and tools of the cognitive sciences to improve their surveys. In addition, they held out the hope that people in the cognitive sciences would begin to conduct research suggested by interesting problems encountered in survey research and use surveys to expand the generality of their laboratory findings. For example, the telescoping that occurs in responses to survey questions was a phenomenon that cognitive psychologists might want to study.

Aborn feels that the effect on cognitive psychology has been small. Although many articles on cognitive aspects of survey methodology have appeared in journals, he sees no firm evidence that more than a narrow segment of psychology is concerned with surveys. Furthermore, he sees little need for transfers in this direction. Instead, he feels that the CASM movement should concentrate on promoting the development of cognitive tools and perspectives in order to improve the conduct of survey research.

Participants from the earlier CASM Seminar challenged Aborn on this point. In fact, it was pointed out that he himself had been an advocate of the importance of the transfer from survey research back to cognitive psychology. The rationale for this stance was that a collaboration between the two disciplines would be long-lasting only if the exchange benefitted both disciplines. One discussant expressed the view that after so many years of collaboration, we should not stop pushing for the transfer from survey research to cognitive psychology, and believed that we may be at the beginning of an era where this can happen.

In assessing the current position of the CASM movement, one participant noted that there are practical obstacles to be overcome in moving forward. It is difficult to conduct the large-scale studies that are needed to generalize cognitive findings in survey research beyond the limits of small-scale experiments. Such large studies require extensive, long-term funding for periods of time that are incompatible with tenure and promotion decisions in university settings. Another obstacle for researchers is obtaining access to national data from surveys.

There was some discussion about how one would measure the transfer from survey research to cognitive psychology. One suggestion was that we should note the many articles on survey research that appear in psychology journals, as listed in Aborn's paper. Aborn and others dismissed this indicator on several counts. First, just the mere appearance of an article in a journal cannot attest to the article's acceptance by more than a small circle of researchers. Second, a clear effect on the discipline of cognitive psychology is difficult to assess, and such an assessment would require measuring how the concepts of the one discipline affected those of the other.

CASM: Successes, Failures and Potential: Colm O'Muircheartaigh

O'Muircheartaigh's proposes three evaluative criteria for survey research designs. The first criterion, *representation*, judges whether the population is properly covered. The second criterion, *control*, judges the extent to which randomization or other techniques control for confounding variables. Third, the criterion of *realism* judges the extent to which the study mirrors real life situations. Sampling statisticians, psychologists, and sociologists each place differing emphases on the purposes of collecting survey data. Even the relative importance of the purposes of collecting survey data differ among the three disciplines. In the discussion, one participant pointed out that the three criteria often conflict, so that we cannot optimize all three but must make trade-offs.

Another participant noted that in developing question wording, a dynamic emerges where conflicting values take precedence at different phases of the development process. The researcher, early in the process, tries to identify the key features of a phenomenon and then tries to ensure that question wording captures these features. As the deadline approaches, the operations staff force the researcher to settle for the final question wording (regardless of how well the final wording measures the phenomenon in terms of the researcher's notion).

In assessing where CASM stands today, O'Muircheartaigh feels progress has been made, but Taylorism—a kind of mindless empiricism—often dominates, specifically in the arena of the cognitive laboratory. To the extent that cognitive laboratories have become isolated within survey organizations, rather than being used as a tool to investigate the principles of survey research, he thinks that they are more often used to discover flaws in questionnaires, fix them, and then certify the questionnaires as cognitively tested. Audience members felt that this judgment was exaggerated and fails to recognize broader work done by the labs on the development of constructs. Even more broadly, the cognitive lab at the Bureau of the Census, for example, is used in conjunction with field testing and ethnography rather than as the sole arbiter of proper question wordings.

O'Muircheartaigh sees a threat that under a narrow definition of cognitive research, organizations will fail to see the need to continue having the cognitive labs. He feels that it is time to broaden the mandate of the labs (and perhaps to adjust their titles as well), to legitimize the broader mandate and to push beyond current survey practices, especially in the arena of acceptable interviewer behaviors. He also feels that disciplines other than cognitive psychology, especially anthropology and social representation, should be involved in that effort.

In general the audience agreed that interviewer behavior is an area that needs more attention. Interviewers can range widely in style. At one extreme is the robotic stance in which questions are read exactly as worded and little illumination is provided to respondents beyond the script of the interview. At the other extreme, an interviewer plays the role of the expert who elaborates in an individualized style for each respondent. In the middle is the more typical standardized interview in which interviewers and respondents subtly negotiate a meaning for an item, when needed, and then proceed to the next question.

Some negotiation is as simple as informing the respondent that it is acceptable to guess at the answer. Evaluation of the consequences of varying degrees of standardization should be included on the research agenda.

O'Muircheartaigh advocates more emphasis on the substantive content of surveys in our research, in the sense of developing ways of getting accurate answers from respondents who differ in myriad ways. One participant noted, however, that emphasis on how respondents answer simple questions aids us in understanding how people make judgments. Another felt that we often confuse the substance with the measurement. For example, when we construct categories of race and ethnicity, such categories are socially constructed and thus are arbitrary in their selection. In admitting that these race/ethnicity categories are constructed and arbitrary, however, we make a statement that is very different from saying that there is no concept of a racial identity.

Themes that were suggested for further research included:

1. What choices in our conduct of surveys will minimize the trade-offs among the evaluative criteria of representation, control, and realism?

2. What is the most appropriate level of standardization for the survey interview? How much flexibility can we give our interviewers in conducting survey interviews without sacrificing measurement?

3. What are the contributions that cognitive laboratories can make to survey research in: (1) the design of specific surveys; (2) improvements to the state of knowledge on questionnaire development; and (3) enhancement of the use of other methods of understanding survey processes, such as field tests, ethnography, and expert review?

Cognitive Research Into Survey Measurement: Its Influence on Survey Methodology and Cognitive Theory

Norbert Schwarz

Introduction

Welcome to the session entitled "Effects of the CASM Movement on Cognitive Theory and Survey Methods." The four papers in this session review conceptual issues, basic research, and methodological developments related to the cognitive and communicative processes underlying survey responding. As all researchers agree, answering a survey question entails several tasks: Understanding the question; recalling relevant information from memory; forming a judgment based on that information; formatting the judgment according to the response alternatives provided by the researcher; and reporting the answer (see Sudman, Bradburn, and Schwarz, 1996, ch. 3, for a review). While problems of question comprehension have received attention as a separate research issue, the other tasks have mostly been studied in the context of attitude measurement and autobiographical recall.

Overview

In the first paper, Michael Schober addresses how people make sense of questions. His paper reviews research on question comprehension, focusing on conversational aspects of the survey interview. In the second paper, Lance Rips and Michael Shum review basic research into autobiographical memory and highlight its implications for behavioral reports in surveys. Next, Roger Tourangeau addresses context effects in attitude measurement and identifies the conditions under which we can expect the emergence of assimilation (carryover) or contrast (backfire) effects. Finally, Gordon Willis, Theresa DeMaio, and Brian Harris-Kojetin provide an update on the methods used in cognitive laboratories.

Effect of CASM on Cognitive Theorizing and Survey Methodology

As is apparent from these reviews, the "bridge between disciplines" (Jabine et al., 1984) that CASM hoped to build has seen considerable traffic. Yet, most of this traffic has taken only one direction—from psychology to survey methods. Although psychologists picked up a number of interesting phenomena from the survey literature, this literature itself has relatively little to offer in terms of substantive theorizing about human behavior and cognition. This, of course, is not surprising: survey research offers a methodology, not a theory of human cognition and behavior. So far, however, cognitive psychologists have not made much use of the unique opportunities that representative sample surveys afford. Working within the paradigm of general information processing models, cognitive psychologists have focused on "inside-the-head" phenomena. They have trusted that one mind works pretty much like any other mind, making the use of representative samples an unnecessary luxury. The perceived luxury appears even more extravagant when costly procedures offer poor control when compared to the cognitive laboratory. But much as psychology's shift from behaviorism to information processing has rendered psychologists interesting partners for survey methodological work, psychologists' increasing interest in the "situated," "contextualized," and "culture-bound" nature of human cognition may eventually make sample surveys a more appealing tool.

Survey methodologists, on the other hand, have contributed their own share to the one-way traffic on CASM's bridge. From psychologists' perspective, much of the work conducted in the cognitive laboratories established at various survey centers is overly applied and lacks a coherent conceptual focus. As O'Muircheartaigh noted in his presentation at this conference, "The function of the cognitive laboratory is not really the development of principles of question or questionnaire design. It is to provide a facility for pretesting questionnaires for use in ongoing surveys. It is evaluated not by its contributions to general principles or knowledge, but by its throughput of questionnaires in a documented and timely manner."

This focus, in combination with time and funding constraints, results in a highly pragmatic research approach: A question drafted by the client is tested in cognitive interviews and "fixed." While the "fixing" may be informed by the theoretical principles discovered in basic research, the applied work rarely finds its way into the theoretical discussion. In many cases, the research design does not provide an opportunity to ask theoretically interesting questions: If we want to understand why the "fixed" question works, whereas the "nonfixed" question did not, we need to specify the characteristics of the cognitive task presented to respondents as well as the characteristics of the respective questions—and we need to test the effect of these characteristics across a range of similar questions and behaviors that meet the same criteria. Unfortunately, the daily reality of work in cognitive laboratories rarely affords researchers this luxury—with two unfortunate consequences.

On the one hand, the applied work is rarely cumulative in nature. Given that general principles are rarely identified, the "testing" and "fixing" has to start anew with each new question asked—hardly an economical use of resources. On the other hand, an enormous body of research, which addresses a rich set of issues across a wide range of substantive phenomena, remains without a noticeable effect on the theoretical development of the field. In fact, the archives of the various applied laboratories may host numerous anomalies that challenge theoretical models, yet we are unlikely to learn about them. Some (but not all) laboratories make their findings available in the form of technical reports that summarize their experiences with a given question, but given the frequent lack of appropriate control conditions and theoretical discussion it is often difficult to determine what has been learned. As Presser and Wellens (Working Group 3) recommend, a systematic, theory-driven analysis of these archival materials promises a rich harvest, with important theoretical as well as applied benefits.

In light of the extensive applied work done in cognitive laboratories, it is surprising that a systematic evaluation of the practical usefulness of cognitive laboratory procedures is still missing. Apparently, the face-validity of these procedures is sufficient to justify the resource allocations made in the absence of any hard evidence that the difference made by "cognitive pretesting" is a difference that matters in terms of actual survey data quality. Moreover, we know relatively little about the relative performance of different techniques, including verbal protocols, cognitive interviews based on extensive prompting, expert systems, or behavior coding (for reviews see Sudman et al., 1996, ch. 2; Schwarz and Sudman, 1996). In fact, many of the available techniques may not be routinely employed in many laboratories, reflecting that "over time, the principal activity of the cognitive laboratory has become the expanded interview," based on extensive prompting, as O'Muircheartaigh noted.

To evaluate the efficiency of laboratory techniques, we need comparative studies that address a set of key questions specified by Groves (1996, pp. 401-402) in a related discussion:

1. Is there evidence that the "problem" will exist for all members of the target population? Is evidence sought that different problems exist for groups for whom the question is more/less salient, more/less threatening, more/less burdensome?

2. Do multiple measures of the same component of the question-answer technique discover the same "problem" (convergent validity)?

3. When the "problem" is "fixed," does replication of the technique show that the problem has disappeared?

4. When the "problem" is "fixed," does application of other techniques discover any new "problems"?

5. Is there evidence that the "fix" produces a question with less measurement error than the original question?

We hope that CASM II will broaden the second lane on CASM's "bridge between disciplines" by fostering an increased theoretical orientation in the conduct of applied work and by highlighting the benefits that the sample survey has to offer to psychologists.

References

Groves, R. M. (1996). How do we know that what we think they think is really what they think? In N. Schwarz and S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 389-402). San Francisco: Jossey-Bass.

Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

Schwarz, N., and Sudman, S. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Abstract

Making Sense of Questions: An Interactional Approach

Michael F. Schober

To make sense of speakers' questions in ordinary extemporaneous conversation, addressees rely on at least two interactive resources beyond their individual cognitive comprehension processes. First, to arrive at an initial interpretation, addressees presume that questioners have followed a principle of *audience design*, basing the wording and framing of the questions on the knowledge, beliefs, and assumptions that both participants share. Second, addressees rely on *grounding* procedures to make sure they have understood the question. Because addressees' initial interpretations of questions are not guaranteed to match speakers' intentions, conversational participants can engage in additional conversational turns to reach agreement that a question has been understood as intended.

This contrasts with conversation in the typical standardized survey interview, where interaction is restricted in various ways (see, e.g., Suchman and Jordan, 1990). These restrictions deny respondents and interviewers the ordinary use of both audience design and grounding. Audience design in surveys more closely approximates the kinds of "generic" or community-based audience design used in literature, rather than the particular-audience design used in ordinary conversation. The grounding procedures available to survey interviewers and respondents are also severely restricted. In some respects, respondents are treated as conversational side participants or even overhearers; as psycholinguistic research has shown, overhearers and side participants comprehend language differently than addressees do.

These restrictions have several effects. As Schober and Conrad (1997a, in press) have described, actual survey interviews show evidence that both interviewers and respondents try to subvert the restrictions that have been placed on them: They deviate from the script in order to try to ground understanding and to achieve some measure of audience design. Second, even when they follow the script, respondents do not necessarily switch off their ordinary conversational reasoning. This can lead to some of the mysterious "response effects" in survey interviews.

Here I describe how an interactive approach can help explain three examples of response effects. (1) Response alternatives presented as part of a question can influence respondents' answers. This can be attributed in part to audience design: respondents assume that question wording is informative about what the survey research designers intended. (2) Respondents provide different answers depending on the order that questions appear in. Much as in unscripted conversation, respondents seem to interpret questions in light of their previous answers, as if the interviewer had understood those answers and framed the current question accordingly. (3) Different respondents can interpret exactly the same question in radically different ways. An interactive explanation is that in most standardized surveys, respondents cannot discover whether their initial interpretations of questions match the survey designers' intentions. As demonstrated in Schober and Conrad (1997b, in press), when surveys allow interviewers and respondents to ground their understanding of questions, response accuracy can increase substantially.

References

Schober, M. F., and Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 60,* 576-602.

Suchman, L., and Jordon, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85(409),* 232-241.

Abstract

The Respondent's Confession: Autobiographical Memory in the Context of Surveys

Michael S. Shum and Lance J. Rips

This paper examines current theories of autobiographical memory—memory for events that a person has experienced—and their implications for survey research. Memory for any event (personal or public) is memory for an event representation. What we retrieve about it depends on how we originally represented the event, on changes to the representation that occur over time, and on the way current cues match the representation. Theories of autobiographical memory differ in the way they organize event representations, but all theories imply some limits on survey questions' effectiveness in getting respondents to recall events. Surveys cannot always tailor questions to a respondent's encoding of an event. In some situations, however, surveys can take advantage of the overall effectiveness of certain prompts.

On logical grounds, the most effective prompt will depend on the goal of the retrieval task. If the aim is to retrieve properties of a particular (independently identified) event, then the most specific and faithful cues may work best. A cue is specific if it contains many details about the event; a cue is faithful if it correctly matches the event's representation in memory. For example, if respondents are asked to remember whether they received a vaccination during their last trip to their HMO, then the best cue will single out exactly that trip. Studies of autobiographical memory suggest that the most effective cues are often ones stating what went on (e.g., consulting an internist about an infection); less effective are cues specifying other people involved in the event (e.g., Dr. Wigton) or where the event took place (e.g., the HMO office on King Drive); least effective are cues mentioning the time of the event (June 25, 1994).

If the aim of memory search, however, is to find some event or other (or as many events as possible) within a larger class, then specificity must be sacrificed. When respondents are asked to retrieve all criminal victimizations experienced in the last two months, the question has to be general enough to cover the entire set. Nevertheless, it may be the case that people cannot recall particular incidents within the set unless the cue contains some threshold level of specificity. "Any criminal victimization" may not suffice for retrieval, and the respondents may have to elaborate the cue with further detail (robberies, burglaries, incidents when police were present) before it is useful. We present research suggesting that individuals' life roles, especially the calendars or schedules they live under, can aid them in retrieving personal memories under these circumstances.

Abstract

Context Effects on Answers to Attitude Questions

Roger Tourangeau

Survey researchers have known for some time that the order in which questions are asked can affect the answers obtained. Such context effects reflect the processes involved in answering survey attitude questions. Context effects arise because respondents are uncertain about how to interpret the questions, because they rely on whatever comes to mind quickly in formulating an answer, because their evaluations reflect salient norms or standards of comparison, and because they want to report information that is both new and consistent with other information they already have provided.

Context sometimes affects the overall direction of the answers and sometimes it affects the correlation between answers to different questions. Both the directional and correlational effects can involve either assimilation or contrast--shifts toward or away from the direction of the context items.

Two general models attempt to tie the diverse findings on the effects of context together. Schwarz and Bless's elegant inclusion/exclusion model attempts to specify the conditions under which assimilation or contrast effects will occur. That model focuses on three basic mechanisms underlying context effects--the incorporation of accessible information, the exclusion of information that is invalid or suspect in some way, and the use of anchors or standards in rendering a judgment. My belief-sampling model assumes similar mechanisms but emphasizes factors that lead to consistency in answers over time and across contexts. These include whether the answers reflect a homogeneous pool of underlying considerations, whether the same considerations are tapped each time a question is asked, and whether the considerations that are tapped are evaluated in the same way. Both models emphasize that context effects are the inevitable result of the nature of attitude questions and the process of answering them.

Abstract

Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques

Gordon Willis, Theresa DeMaio, and Brian Harris-Kojetin

The use of cognitive interviewing techniques to assess and to address sources of response error represents an important operationalization of the CASM movement. Typically, a specially trained interviewer conducts a one-on-one interview in which a draft survey questionnaire is administered to a recruited laboratory subject. The interviewer encourages the tested individual to "think-aloud" during the course of question answering, and also administers additional probe questions that purport to identify the key cognitive processes that influence responses to the survey questions. Based on observations made across a number of such interviews, cognitive interviewers or other survey researchers decide that particular survey questions, or questioning approaches, are problematic from a cognitive standpoint, and propose revisions which address these problems. Through an iterative process of cognitive testing and revision, a final questionnaire version is developed which is intended to produce less response error than did the original, when the survey is field pretested or conducted in a field environment.

Over the past ten years, cognitive interviewing techniques have become widespread, especially in U.S. federal agencies such as the Bureau of the Census, Bureau of Labor Statistics, and National Center for Health Statistics. Numerous publications have described these methods, and have presented case studies in which they have been used to develop particular survey instruments. Further, a few studies have endeavored to assess whether the results of cognitive testing are reliable, in terms of consistency with the results of other pretesting methods. In general, these reports have been positive. The authors have concluded that cognitive techniques are useful and should be applied widely to questionnaires in both population-based (in-person, telephone, and mail) and establishment-based surveys.

However, to date there has been limited evaluative research focusing on the demonstration of efficacy of cognitive interviewing techniques in terms of their application to the field survey environment. In other words, little attention has been paid to the key issue of validation. The current paper therefore addresses this gap by considering the "state of the art" of cognitive interviewing, with a focus on the demonstration of validity. We describe the basic methods that have come to be used, and develop a framework for evaluation of these methods. Second, we review the existing CASM literature in order to summarize the evaluation work that has been done. Finally, we propose studies that will further the process of assessing whether cognitive pretesting is truly a useful tool for improving the quality of a wide range of survey questionnaires.

Summary of Presentations and Discussion

Jaki Stanley

Audience Design and Grounding in the Interview: Michael F. Schober

Michael Schober compared the interactions that occur in ordinary conversation with those that occur in survey data collection. In conversation, remarks are designed for a specific audience. Grounding procedures, such as requesting clarification of a statement, help to ensure that all parties understand what is said.

Survey interviews involve two or more people—the researcher, the respondent and sometimes an interviewer—and the interactions between them are critical to the outcome. Each brings different knowledge, typical behaviors, and expectations to the interaction. However, questions tend to be designed for generic respondents and grounding techniques used in normal conversation are not always allowed. Words and phrases almost always have alternate interpretations, and without adequate grounding respondents may choose one that differs from that intended by the survey designer.

Much of the discussion of Schober's paper focused on the idea that a clear understanding of the researcher's intentions (versus the specific wording of the question), by both interviewers and respondents, is essential to successful data collection. Because question intentions are not always obvious, more emphasis is needed on clarifying concepts for interviewers. Some organizations have included concept testing for interviewers as part of their training, but this varies both within and across organizations. Some discussants felt that most organizations probably do not do an adequate job of explaining subject matter concepts to their interviewers. For example, for a survey question about the consumption of fats and oils, interviewers were directed to include peanut butter, but the rationale for this was never explained to them.

Another point discussed was that earlier models of survey research viewed questions as independent entities within the interview. It was agreed that interview participants view the interview as a cumulative process and that survey researchers and methodologists should be aware of this in developing questionnaires. It was observed that respondents exhibit more concern about repetitive questions when being interviewed than when they are completing self-administered questionnaires.

Several of the discussants said that we should be careful about the terminology we use for alternatives to traditional standardized interviewing. To call an interview practice "nonstandard" might imply that the interviewers are given total freedom to collect information in whatever way they choose. We might want to use the term "enhanced standardization" for a format in which interviewers are expected to follow standard protocols and procedures but are empowered to deviate from the written script when they consider it necessary for accurate data collection. The key will be to train interviewers when and how to deviate.

In many survey interviews there is another player—a computer. Do some respondent assumptions change when there is a computer involved in the interaction? Are respondents grounding with the interviewer or with the computer? It was suggested that ideally the grounding should be between the respondent and the researcher, with the interviewer (computer or human) acting only as the intermediary. However, depending on the mode of the interview, some of the assumptions do change. As noted above, respondents to self-administered forms do not appear to mind repetitive questions, but they do object to them when asked by an interviewer.

It was suggested that part of the role of the interviewer should be to encourage respondents to raise questions when they are unsure of what a question means or their ability to give an accurate answer. Interviewers now often use subtle cues to detect possible problems in question comprehension, such as not answering immediately or expressing uncertainty. But this is left entirely up to the skill and discretion of the individual interviewer.

There was some skepticism that such empowerment of interviewers was necessary or would produce data of better quality. It was noted that there are many situations where there is little or no interactional grounding involved, such as lectures, broadcasts, and newspapers. However, these kinds of interactions, while not including grounding in the interaction itself, are designed for specific audiences, with some expectations about their levels of knowledge and understanding. In addition, it is not always clear that audiences in these settings have consistent interpretations of the information they receive. In survey data collection, misinterpretations may have

serious effects on data quality. This is perhaps less likely to be the case in these other situations.

New views of the role of interviewers may necessitate new thinking not only about how they are trained, but also about the criteria used in hiring. But is "enhanced standardization" always desirable? Perhaps it is only justified for complicated situations with unusual data. Clearly, increased investment in training will be required for this approach. Which components of "enhanced standardization" are important, which are not? What is the next model for the survey interview? Clearly, much research will be needed to answer these questions.

Autobiographical Memory: Lance J. Rips

Psychologist Lance Rips described recent research on autobiographical memory and its possible implications for survey research. There are several different models of autobiographical memory, but they share the belief that its units are individual event descriptions and that retrieval of information about an event depends on how close the cues (or survey questions) are to the stored information. He showed results from four studies of college students which indicated that they recalled more events that occurred at boundary points between school terms and vacations than at other times of the year. Survey questions often ask for the number or frequency of events in a specified category, rather than details of an individual event. This presents a more difficult retrieval problem, and one to which memory researchers have give relatively little attention so far. In response to a question about how people do respond to event-class questions, Rips answered that it depends on the specific demands; sometimes it may be impossible. He thought that people with more life changes might recall events better, but said there was no empirical evidence for this supposition.

One participant described the use of a life history calendar in the National Survey of Family Growth in an attempt to obtain better recall and dating of events. Rips said that this has been done in several studies, sometimes with significant positive effects. When asked whether different techniques would be required for high versus low frequency events, he said that some studies of food consumption show that individual event recall does not work well for foods frequently consumed. In response to a question about evidence for hierarchical memory structures, it was his view that it would usually be more effective to begin retrieval efforts at a higher (more general) level of detail.

A participant asked whether the neuropsychological field may provide useful insights about how memory retrieval works. Rips answered that amnesia studies may provide some relevant information about where different kinds of information are stored in the brain.

Attitude Questions and Context Effects: Roger Tourangeau

It is commonly believed that responses to attitude questions are produced by a rough sort of sampling of the most accessible bits of relevant information that come to mind when a question is asked. There have been numerous studies showing that other information in the questionnaire or other questions can effect the way a single survey question is answered. To fairly evaluate responses to a survey question, it is probably not enough to see the particular question that was asked. One should also look at other questions in the survey instrument. For self-administered questionnaires, context effects can result from questions placed either before or after the target question. The stress on individual words in a question can also influence responses.

Context effects can be assimilation effects, in which respondents assume that the target question refers to the same topic as previous questions. They can also be contrast effects, in which respondents infer that the intent of the target question must differ from that of the preceding question.

A participant asked whether we should be concerned about the possibility of deliberate manipulation of context designed to bias data? Another participant felt that it would be much easier to manipulate answers with biased question wording than with context. However, effects due to context are less obvious. Perhaps both survey sponsors and respondents need to be made aware of these potential effects. Some types of instructions or transition statements might help respondents avoid these contextual biases.

Context effects may differ according to the mode of data collection. For example, paper self-administered questionnaires that allow respondents to look forward and backward at other questions may prompt context effects which differ from those associated with computer-controlled questionnaires that restrict information and flow. However, order effects have been observed in self-administered questionnaires. Context effects are sometimes thought of as synonymous with order effects; however, context is more than strict linear order.

A discussant asked whether Gricean principles of conversational interaction are likely to be relaxed in survey interviews. One such principle is the assumption that successive utterances are connected and contribute to an agreed-on purpose for the interchange. Tourangeau thought that this assumption on the part of respondents might be relaxed if the questions were "scattered," that is, did not follow a clear-cut topical outline.

Evaluation of Cognitive Interviewing Techniques: Gordon Willis

Cognitive laboratories and cognitive techniques such as think-aloud interviews, probing, vignettes, and retrospective verbal protocols have become part of the standard tools of questionnaire development. It is clear that problems in instrument design can be uncovered with these techniques. However, Willis stated that it is as yet unclear whether there are any benefits unique to these methods and whether they solve the problems they purportedly uncover. He recommended that we step back and ask what general principles about questionnaire design can be derived from our extensive experience in using these cognitive tools.

It was suggested that a three-pronged approach to questionnaire development might be more effective than any single approach. Theory, cognitive techniques (such as verbal protocol analysis), and behavioral measures (such as reaction time) could be used in combination to better evaluate survey instruments and data collection. Experimental studies have shown that reaction times increase for poorly constructed questions. However, increases in reaction time may also be due to increased thoughtfulness or other factors. Perhaps behavior coding could be used to supplement such studies, and not used strictly as a stand-alone technique.

In his paper, Willis made a distinction between the use of cognitive techniques for problem discovery and their use for problem verification. A discussant suggested that problems anticipated by a researcher are more likely to be verified when the researcher is also the cognitive interviewer. It is a common practice for the same person to assume both roles.

Part of the problem with evaluating how well cognitive interviewing works and generalizing from the results is the lack of documentation. The NCHS requires, following each major questionnaire testing activity, preparation of a working paper containing full documentation of what was done and the resulting actions. This process, although resource intensive, leads to increased comprehensive thinking about the process and its benefits and shortcomings.

Another participant observed that the theories and methods used in measurement of aptitudes might offer some ideas that would be useful in survey research. Aptitude tests rely on a series of items to measure the target quality; these items sample the domain of content. Reliance on a series of items might help to overcome bias caused by context effects in surveys. Willis agreed that it would be useful to explore these possibilities.

Casting a Wider Net: Contributions from New Disciplines

Roger Tourangeau

Introduction

After the original CASM conference, for which I had written a paper describing the relevance of the cognitive sciences to survey problems, Andrew Ortony remarked that neither my paper nor the first CASM conference in general had drawn on the full range of cognitive sciences, but rather had drawn almost exclusively from cognitive psychology. As much as I hated to admit it, Ortony's observation was basically correct. Futhermore, his statement could have been even broader. In some ways, CASM drew not from cognitive psychology in general, but rather from a couple of specific areas, from which it drew rather heavily—namely the psychology of memory and judgment.

Today's session is an effort to broaden the base of disciplines from which the CASM movement draws. It tries to represent five viewpoints that were either omitted entirely from the first conference or that played a secondary role there. They include linguistics, on which Charles Fillmore will give a talk; artificial intelligence, for which Art Graesser will be the spokesman (he will talk about his computational model QUEST); connectionism, about which Eliot Smith will give a talk; and ethnography, which will be represented by Eleanor Gerber. Last but not least, Robert Groves will consider the integration of insights from CASM-style research into statistical estimation procedures. A type of paradigm shift took place after the first CASM conference—a fundamental shift in the way we view survey errors. If a paradigm determines the questions people focus on, the methods they use in addressing those questions, and the concepts and theories they use in answering them, a shift has in fact taken place. We have a new vocabulary for describing survey errors, new tools for investigating them, new theoretical ideas for understanding the errors and reducing them.

Nevertheless, I think many of us feel that the acceptance of the new ideas has been superficial. As one friend put it, "cognitoriums" sprang up all over Washington after the first CASM conference, but it was not always clear what went on in the new cognitive laboratories. Researchers often claim to use "cognitive" methods in pretesting questionnaires, but in many cases they refer to methods like focus groups that have no particular cognitive basis. And then there is the great terminological shift which I (and others) have already alluded to. It is now virtually mandatory to sprinkle the word "cognitive" throughout methodological papers.

It seems that there are some serious obstacles to deeper acceptance of the new paradigm—and to more profound accomplishments by the movement. One obstacle has been the existence of a very successful paradigm preceding the cognitive one. Let me refer to the existing paradigm as the Hansen-Hurwitz-Bershad approach; it (and its descendants) are the reigning statistical approach to error. O'Muircheartaigh effectively critiqued this approach, noting that every time a new methodological problem is discovered, a new variance component is duly named and added to the model. Nonetheless, the old paradigm constituted a rich and interesting approach. It stands as one of the major milestones in the history of survey research—the development of a statistical notion of survey error and the use of this notion in the design of surveys and the estimation of survey results. Most importantly this earlier paradigm is embedded in survey practice; by contrast, the cognitive paradigm has only been superficially grafted on to survey practice.

In this context, it seems to me that it will be an extremely useful endeavor to grasp how this new notion of error relates to the existing notion and to reconcile those two viewpoints whenever possible. Grove's paper offers an important beginning in that direction.

A second barrier to the acceptance of the new paradigm and deeper practical accomplishments by its proponents is the lack of explicitness in many of the new ideas. It is not always clear what is being predicted or what is prescribed. Again, I believe that O'Muircheartaigh made the same point noting that people talk with great confidence about what findings mean, but then are absolutely unable to make predictions for the next survey. One way to address this lack of explicitness is to create models that yield definite predictions. That is an approach taken in artificial intelligence—where theories often take the form of a detailed and rigorous computational model that is, by its nature, forced to produce explicit predictions about what will happen. One such model relevant to the survey enterprise is the QUEST model developed by Art Graesser.

Aside from the existence of a prior paradigm which was in some ways superior to the new one and the lack of explicitness about implications of the new paradigm for survey practice, a third obstacle to the acceptance of the new ideas is the incompatibility of the cognitive approach with survey realities. One of these survey realities is cultural diversity. People are aware of the fact that survey questions often presuppose a conceptual framework that some respondents do not share. It is not clear that cognitive psychology provides useful tools for identifying or dealing with these kinds of conceptual incompatibilities. Instead, or in addition, we require a discipline that contributes more to our understanding of these conceptual incompatibilities (derived in part from cultural diversity). A better discipline to draw on for dealing with the issues created by cultural diversity might be cognitive anthropology or ethnography. Eleanor Gerber, a cognitive anthropologist, will talk about the role of ethnography in reducing survey error.

Another survey reality—closely related to cultural diversity—involves linguistic barriers. Much of the research inspired by CASM involves memory and judgment. But many of the problems in surveys, including Belson's disturbing results on the comprehension by survey respondents of everyday terms (an example from yesterday's discussion) involve comprehension of the questions. While there is a great deal of useful psycholinguistic research applied to surveys, there is also a basic discipline—linguistics—that deals with comprehension problems. Charles Fillmore presents a sample of the problems with survey questions identified from the point of view of linguistics.

Finally, even within psychology—the discipline which the CASM movement has drawn from most heavily—the movement has based much of its research on a relatively narrow set of ideas drawn from the study of memory and judgment, with some borrowing from psycholinguistics. In the meantime, since the CASM I conference, a whole new approach has developed within cognitive psychology that has been, at best, poorly represented within the CASM movement. That perspective is called connectionism. Connectionism starts from a model of how memory is structured and how recall is carried out. This model differs radically from the models that prevailed at the time of the CASM I conference. The connectionist models see memories as distributed across multiple units and represent learning as a change in the strength of the connections between these units. It is an interesting viewpoint, but has not influenced our movement much. Eliot Smith, who is a leading proponent and advocate of connectionism within social psychology, presents a introduction to the connectionist viewpoint.

Abstract

A Linguistic Look at Survey Research

Charles J. Fillmore

This paper urges the designers, evaluators, and interpreters of survey questions to pay close attention to the linguistic aspects of the context and the process of administering survey questionnaires and especially to the actual text of the questions.

Questions may contain words whose interpretations differ dialectally from region to region (the request to include, in a household description, any brother-in-law or sister-in-law that one shares a home with could yield different answers in different parts of the country. These terms may or may not include a spouse's sibling's spouse). The middle choice in a 5-point or 7-point scale might contain wording which entails a contradiction (the choice "I agree with both answers" cannot be seriously meant if the terminals of the scale are mutually contradictory); the fact that most interviewees do not notice the problem is itself a problem, since that tells us that we cannot be sure that there are not other intended parts of the survey that they do not notice. A set of response options might not include a choice appropriate for a given respondent (there may be no racial category which "best describes" a respondent's ancestry, and choices like "Other" or "Don't know" might not be appropriate). The order of response alternatives can impose an interpretation on the alternatives that might not be reflected in a public report of the questionnaire results that a group declare themselves "quite satisfied" in some respect is not informative unless one knows whether this was the highest, second highest, or third highest in the ordered list of responses). Words can have different meanings when paired with different antonyms (statements about preferring to pay in cash will have different values when the alternative is presented as paying by check or paying on credit as the latter allows payment by check). Questions can carry presuppositions which the respondent does not hold, without offering the respondent the opportunity to reject the presupposition. Passages in survey questions might be semantically confusing because of their grammars.

The author concludes that certain areas of expertise in the science of language could be of service to researchers and consultants in survey methodology.

Abstract

The Use of Computational Cognitive Models to Improve Questions on Surveys and Questionnaires

Arthur C. Graesser,Tina Kennedy, Peter Wiemer-Hastings, and Victor Ottati

Our goal is to defend the objective of building a computer aid to benefit designers of questionnaires. The computer aid would have particular modules that critique each question on various levels of language, discourse, and world knowledge. For example, the critique would identify cumbersome syntax, words that are unfamiliar to most respondents, ambiguous terms in the discourse context, questions that overload working memory, and questions that may appear to the respondents to be unrelated to the survey context. Some of these components are so complex, technical, or subtle that they are invisible to the unassisted human eye, including experts in questionnaire design and computational linguistics. The computer aid could be used as an automated tool if it succeeds in providing an accurate, reliable, and fully automated diagnosis of problematic questions. Alternatively, questionnaire designers could be trained to apply the analytical systems that underlie the tool.

Some empirical studies are reported that demonstrate the promise of a computer aid that adopted components of a cognitive computational model of question answering (called QUEST). For example, in one study we assessed how well adult novices were able to identify 12 problems with questions that frequently are asked in surveys. The forms included a U.S. census form, a dentist intake form, an application to graduate school, and a job application form. The 12 problems would be anticipated by the QUEST model. The adult novices could identify only 2 of the 12 problems with any reliability, so the QUEST model would be expected to have advantages over the pretesting of a questionnaire on a sample of novices. In another empirical study, we found that QUEST-based revisions of problematic questions ended up producing more reliable answers in a sample of college students. It is an open empirical question how well the computer aid would compare with analytical coding schemes for diagnosing problems with questions and with samples of expert questionnaire designers.

Our second goal is to describe some contemporary computer systems that have the potential to diagnose problems with bad questions on questionnaires. The computer models that we had in mind are grounded in cognitive science, a field that integrates contributions from artificial intelligence, computational linguistics, and psychology. Four computer models or technologies are discussed: computerized lexicons, syntactic parsers, simulators of working memory load, and latent semantic analysis. These components are already automated and could be tailored to a computer aid for questionnaire designers.

Abstract

The View from Anthropology: Ethnography and the Cognitive Interview

Eleanor R. Gerber

This paper examines the ways in which cultural anthropology has been applied in a survey context, and suggests areas in which anthropology might make further contributions.   Anthropology is often characterized by the ethnographic method. The use of ethnography to provide information about the beliefs and sociocultural practices of surveyed populations is reviewed. The kinds of populations that are the subjects of ethnographic studies are examined, comparing the use of ethnography in special populations and "mainstream" groups. (Survey methodology tends to use ethnography to learn about groups that are considered different or unreachable by survey authors.)

Qualitative research used in pretesting frequently becomes "back-up ethnography," in which culturally-based problems in question wording or concepts can be discovered. This occurs in part because survey authors may not realize the difficulties which technical terms and concepts pose for respondents. Cognitive interviewing is also used to assess variation in response between subgroups.   Research strategies applicable to ethnography and cognitive interviewing are compared, and the ways in which the two are incompatible are examined.   In particular, the necessarily narrow focus on the questionnaire context in pretesting makes it difficult to successfully combine effective anthropological investigation with cognitive interviewing.

Survey methodology and cognitive anthropology share a strong interest in the nature of categories and semantic analysis. A complete semantic analysis of the domain is usually unnecessary for pretesting questionnaires. However, it can be analytically useful to investigate some of the terms which do not appear in the questionnaire, but are in the domain naturally used by respondents.

Cognitive anthropology no longer focuses exclusively on semantic analysis. Concepts which stress the representation of complex cultural beliefs, such as "schemas" have become a theoretical focus of the field. This paper discusses the relevance of schemas to understanding the survey response process, using materials drawn largely from research on residence and race. The choice of a relevant schema for interpreting events, called "instantiation," may be relevant to respondent behavior. In particular, the variety of available interpretations and the consistency of the outcomes they produce may influence respondents' perception of "ambiguity" in survey questions. It is suggested that divergent interpretations cause respondents to look for guidance in the questionnaire, while a set of interpretations with consistent outcomes causes respondents to ignore definitions and information offered on the questionnaire.   Schemas consist of networks of association, and respondents may frame survey responses to elements of a schema which go far beyond the direct subject matter of the question.   For example, respondents treat the decennial census race question as more than a request for a relevant category. The question is interpreted as part of a complex American dialog on the subject of race, and this controls many response patterns.

Abstract

Survey Error Models and Cognitive Theories of Response Behavior

Robert M. Groves

Some results from experiments testing cognitive theories of survey response behavior identify consistent over- or under-estimation of objective phenomena. Survey statisticians would tend to label these evidence of "response bias." Other results show that response behavior is less stable, subject to more variability under certain circumstances. Survey statisticians would label these evidence of "response variance." A premise of this paper is that the quality of survey information is enhanced by measures of such response errors.

This review makes the following observations: (1) as with all of science, progress is denoted by greater understanding of constituent components of phenomena (e.g., we now have theories about the comprehension step of respondents related to measurement errors), (2) progress in inventing practical measures of nonsampling errors for routine survey use is impeded by clashes between viewpoints of experimental and observational studies (e.g., we lack a family of estimators incorporating various nonsampling errors), and (3) advances in understanding nonsampling errors will require use of model-assisted estimation tools now foreign to most survey analysts (e.g., most nonsampling error parameter estimates require some model of the behavioral phenomenon creating the error).

The paper is built around a set of results from CASM that appear to have relevance to different classes of measurement error models common to surveys. For each of these results a survey design is proposed that reflects the findings from the cognitive theories. Most often these involve some randomization or replication in the measurement. Estimators of survey statistics reflecting these designs are then proposed. These estimators either have different expected values, reflecting the insights from the cognitive theory, or have measurable variance properties that reflect uncertainties arising in the response formation or delivery steps.

Examples are offered to stimulate a bridging of the gap between cognitive theories of survey response behavior and the practical tasks of constructing survey estimators that reflect as much information about their error properties as possible.

Abstract

New Connectionist Models of Mental Representation: Implications for Survey Research

Eliot R. Smith

Conventional models of mental representations (memories for specific events, attitudes, judgments, etc.) have evolved over the years. New connectionist models question the traditional assumption that representations are discrete entities that can be stored in memory and retrieved independently.   In a connectionist model, meaningful representations are patterns of activation across a number of computational units, and multiple patterns are stored together (superposed) in a single set of connection weights. Connectionist models offer compelling accounts of how general knowledge is built up from specific experiences, and how memory retrieval processes can either access specific well-learned representations or make use of multiple representations simultaneously as constraints. These models also imply that people will have difficulty distinguishing facts that were learned from those that were inferred. However, connectionist memory systems have difficulty learning new memories while preserving old information.

To overcome this problem of interference, a connectionist model of dual memory systems that fits well with both psychological and neuropsychological data has been proposed by McClelland, McNaughton, and O'Reilly (1995). Separate fast-learning ("episodic") and slow-learning ("schematic") memory systems are postulated that interact in specific ways.   New information is learned in the fast-learning system, and gradually transferred in the process of "consolidation" to the slow-learning system, so that it can be nondestructively integrated with existing knowledge. This model implies that people can hold separate and even potentially contradictory knowledge in the two memory systems.   Information in the slow-learning system, including general impressions, associations, and overall judgments, is generally more easily and quickly accessed than information in the fast-learning system, which retains memories of specific incidents. Survey researchers must consider the type of representations (general judgments or impressions versus specific episodic memories) they want their questions to tap. They should also be aware that human memory is thoroughly reconstructive (rather than reproductive) in nature; any response that people give is likely to be influenced by many representations and by more than one memory system.

Reference

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

Summary of Presentations and Discussion

James Kennedy

Since the first CASM seminar, survey methodologists have increasingly come to rely on results from cognitive psychology to inform development and testing of survey questionnaires. This interdisciplinary change comprised a paradigm shift of sorts (some hesitate to call it a revolution). Yet, it has appeared to some observers that psychological theory may have been superficially grafted on to previously existing survey methods. In particular, it might seem that a very narrow range of phenomena from cognitive psychology is applied to survey methods, whereas the larger corpus of cognitive science is ignored. This third session of the CASM II Seminar sought to bring in researchers from other branches of cognitive science, including linguistics, artificial intelligence, and cognitive anthropology, and from statistics to share their insights with survey methodologists. This session also included a cognitive psychologist who described recently developed connectionist theories of memory.

A survey agency or researcher may propose a question about the prevalence or intensity of a phenomenon in a population, the answer to which is to be found by asking sampled members of that population. As respondents are usually sampled in such a way as to be representative of the general population, there is a good chance that they share little common experience with the bureaucrats, politicians, or academics who proposed the question in the first place. The result is frequently a dissociation between researchers' intentions and respondents' interpretations and a corresponding misinterpretation of survey responses by researchers. Thus, several papers touched on issues of surveying respondents who differ culturally from the survey designers, and from one another; similarly, survey designers from different professional subcultures may be at odds with one another.

Linguistics: Charles J. Fillmore

Linguist Charles Fillmore analyzed wording in items taken from various versions of the General Social Survey, as well as items from the National Health Interview Survey and the Current Population Survey. He pointed out many instances where the intent of an item's authors might be unclear to some respondents, violate relevant norms invoked in answering questions, or violate other conventions of language.

Fillmore noted that the surface interpretation of an utterance may differ from its apparent or actual intention. A respondent is likely to attempt to understand the reason a question is asked, and base his or her answer on that inferred motive, essentially ignoring the literal meaning of the item. Interviewers, as well, may be faced with the job of "construing the task," providing interpretations when respondents appear to be confused. Individuals might infer a question's intent from their knowledge of, or stereotypes about, the person asking the question—whether this is the interviewer or the organization conducting or sponsoring the survey. Cultural differences between survey designers and respondents may also lead to assumptions about point of view, author's presuppositions about the respondent and the knowledge he or she holds, and different definitions of terms between subcultures.

A participant noted that many of the problems that Fillmore identified by linguistic analysis were similar to those discovered in the statistical agencies' cognitive laboratories. In response to a question about "culture-neutral" language, Fillmore indicated that it might sometimes be impossible to compose wordings that contained no cultural biases. He further doubted whether computer programs could be written to identify linguistic difficulties.

Computational Cognition: Arthur C. Graesser

This last statement was contested by cognitive scientist Art Graesser, whose paper described methods he has used to conduct computer-assisted analyses of survey questions. Graesser's QUEST model of human question answering is designed to assist questionnaire designers in identifying 12 kinds of potential problems with questions. QUEST is based on a theoretical perspective which incorporates question interpretation, access to relevant information sources, pragmatics (common ground and goals of communicators), and convergence to relevant answers. In a laboratory experiment, individuals trained on the QUEST model identified a significantly

greater number of question problems than did untrained novices.

A computer aid for questionnaire designers can make use of several computational resources that are available free of charge through the Internet. These include lexicons containing large numbers of words, their meanings, and other relevant semantic features; syntactic parsers which translate strings of words into meaning representations; working memory management software which computes an item's load on working memory; and techniques for latent semantic analysis, which is a new method for computing world knowledge and deep comprehension of text. Graesser did not assert that computers can understand language, or that they can replace humans in the interpretation of semantic and syntactic problems with questionnaire items, but did express his belief that computational tools can assist survey questionnaire developers.

Some participants noted that the computational approach to question evaluation does not nullify earlier perspectives on question-wording issues, and perhaps does not even go beyond them. One asked what evidence there is that questions identified as "bad" lead to poorer quality of response, or the converse. Others voiced the opinion that question-wording problems are often of types that might escape analytic methods such as QUEST; for instance, simple combinations of common, well-defined words ("What have you done today?") are often the most difficult for respondents to understand and answer. Some discussants found the ideas promising and suggested ways to extend the developmental research. Such suggestions included comparing the 12 problem categories with classification schemes used by other survey researchers to evaluate survey questions and applying the QUEST model to questions that were used by Presser and Blair (1994) in their research on alternative methods of pretesting.

Cognitive Anthropology: Eleanor R. Gerber

Cognitive anthropologist Eleanor Gerber reviewed potential applications of anthropological theory and methods in survey research and compared them with the uses of cognitive interviews. Ethnography, defined as "the description of culture," can be used to elucidate language norms and schemas which respondents may bring to a survey interview. Schemas are knowledge structures that are learned, are shared, are relatively durable, and are thematic (can be applied in a number of contexts). Members of cultural subgroups are likely to share schemas which determine their interpretations of survey questions. In many instances respondents might have to choose among alternative schemas in answering a question. Both ethnographic research and cognitive interviews can be useful, but they are best kept separate.

In response to a question, Gerber recommended that survey agencies conduct ethnographic research early in the survey development process (although it does not always happen that way), to assess the potential effect of cultural differences on responses. Another participant thought it would be useful to do ethnographic studies for topics like income and employment, for which more or less standard content items have been developed and are used in numerous surveys. In the same vein, a discussant suggested that such studies would be useful for aspects of culture that are changing in ways that might require changes in the kind of questions asked about them. There was some discussion of whether and how the ethnographic approach of adapting questions to respondents' cultural perspectives might conflict with the goal of standardization of interviews and how such conflicts might be resolved. Finally, there was some discussion of how one might learn more about the schemas of survey designers or sponsors and how they may differ from those of respondents, with cognitive interviews of these "clients" suggested as a possible approach. It may also be useful to present findings from cognitive interviews with respondents to the survey sponsors.

Statistics: Robert M. Groves

Whereas previous speakers discussed implications of respondents' and interviewers' cultural backgrounds, Robert Groves examined the interests and goals of two different cultures that exist within the survey research community. One culture, consisting mainly of behavioral and social scientists, attempts to identify the behaviors that cause nonsampling error and thereby identify circumstances under which these errors can be reduced. However, such findings do not normally lead to measurable reductions of nonsampling errors. The other culture concerns itself with sampling error and, based largely on probability theory, develops designs that will achieve specified levels of error. Groves called for collaboration between cognitive scientists and statisticians in an effort to develop survey estimators that will take account of known sources of nonsampling error. Such estimators could be partially based on auxiliary data collected in the survey.

Discussants found Groves' recommendations intriguing, especially the idea of including auxiliary data items, of the type needed for the proposed estimators, in the main survey instruments. One participant mentioned that Robert Ferber had proposed using a secondary instrumentation many years ago, but his idea was not accepted at that time. Others observed that the estimators Groves described are model-based estimators. The data might not always provide a good fit for the models selected, and some data users might be less inclined to accept the results. However, another

discussant pointed out that some model-based estimates, such as the results of educational assessments, have been developed through collaboration between conceptual experts and statisticians, and that these results are fairly widely accepted. Some participants asked about extension of the proposed estimation procedures to analytical uses of the data, requiring estimation of statistics other than means and variances. Groves acknowledged that this is more difficult; at present he is considering only the latter.

Connectionism: Eliot R. Smith

In the final paper of the session, social psychologist Eliot Smith reported on a new paradigm in cognitive psychology called connectionism, which hypothesizes that memories or mental representations are stored as patterns across a large number of simple, interconnected processing units. These representations are reconstructed on the basis of all stored knowledge, rather than simply retrieved. A second feature of connectionist theory is the existence of two kinds of memory systems. The first is a fast-learning short-term memory, which encodes specific episodes for rapid recall, while the second is a slow-learning long-term storage system which integrates events and stores them as prototypes or "typical" event descriptions.

Connectionist theory challenges the traditional idea of memory as a kind of filing cabinet, from which records could be retrieved if one only knew where to look for them. Instead, Smith suggests that short-term episodic memories might be best retrieved through access of the appropriate memory system, for instance, by providing cues which suggest enumeration of specific incidents. This kind of effect is expected when questions are asked about specific, unusual events such as crime victimization. On the other hand, information stored in the long-term system is integrated, and details of individual incidents are probably not available.

Several questions were asked about how connectionist models (of which there are several, according to Smith) explain various aspects of memory and related cognitive processes. In response to a question about what determines whether a stimulus is directed to the slow- or fast-learning system, he responded that it would be based on a diagnosis of its novelty. To a question about how people might respond to inquiries about frequencies or rates, he indicated that they might use the long-term system to develop some kind of average. One participant asked about instances of long-term retrieval for which the details are very clear to the individuals ("flash bulb" memories). Smith said that such memories are not always as accurate as people believe, but that if identical events are presented frequently to the slow-learning system, their representations become more complete.

Potential Contributions of the CASM Movement Beyond Questionnaire Design: Cognitive Technology and Survey Methodology

Douglas J. Herrmann

Introduction

The previous sessions have focused on the history of the CASM movement and the contributions of cognitive psychology to the understanding and improvement of data collection in surveys. In this session we turn from our review of past accomplishments emanating from CASM I to consideration of how the CASM movement can extend its influence to other aspects of the survey process and how disciplines other than cognitive psychology might contribute to improving the quality and utility of survey research.

There have been many efforts in recent years to apply findings and research tools from the cognitive sciences to areas other than survey methodology, such as medicine, law, business, and governmental activities other than surveys. The planners of the conference recognized that the principles of cognitive science have relevance for all parts of the survey process, not just for data collection. For example, cognitive psychology could be adapted to aid sponsors of surveys in their formulation of a survey's mission. It could also be used to facilitate the formulation of particular concepts to be measured, data processing, data analysis, and the dissemination of data. The application of cognitive models to those aspects of surveys is only just beginning. The papers in this session represent the state of the art in moving into other aspects of the survey process.

Each paper serves as an example of what survey methodology might become. These papers will provide knowledge that some of you will want to incorporate into your own programs. For others, the papers will suggest ways in which research might be adapted to address aspects of the survey process that have yet to receive much investigative attention. Of course, this session will not address every possible way that cognitive psychology can be applied to survey research. However, it will surely elicit your interest in other parts of the survey process.

This session consists of four papers. The first presentation, by Mick Couper, applies to the data collection phase of surveys, and focuses on the computer-assisted methods of data collection that have largely replaced more traditional methods in the years following the CASM I Seminar. Couper draws on findings from the field of human-computer interaction and reports some results of recent cognitively-oriented investigations of computer-assisted methods for collecting data. Fred Conrad's paper suggests that models of cognitive processes can be usefully applied to all phases of survey development, processing, and data dissemination and provides several examples of how this is being done in the Bureau of Labor Statistics. Michael Friendly shows how sophisticated graphical presentation of categorical data may reveal patterns that are not apparent from traditional tabular presentations. Stephan Lewandowsky reviews findings from recent research on human cognitive processing of statistical maps and graphs and examines their implications for the communication of scientific data.

Cognitive research has revealed many of the bases of respondent error, but cognitive research into other phases of the survey process has just begun. Cognitive psychology and related disciplines hold the promise of solving a wide variety of survey problems. This session illustrates some of the paths to future development in the cognitive approach to survey methodology.

Abstract

The Application of Cognitive Science to Computer-Assisted Interviewing

Mick P. Couper

This paper reviews current developments in computer-assisted interviewing (CAI) and the contributions that cognitive science, through human-computer interaction (HCI) or usability research, can make to survey research. In the same way that cognitive psychology has had a profound influence on the field of questionnaire design, so too can human-computer interaction affect the development of computer-assisted survey instruments. This paper discusses some of the key theoretical and empirical findings in HCI and their application to computer-assisted interviewing. Methods of usability research and examples of the application of these procedures to the study of survey instruments are also presented. A particular focus is on end-user methods, including usability laboratories for the evaluation of survey instruments. A variety of theoretical developments and research findings in HCI with relevance to computer-assisted surveys are also discussed. The paper reviews a variety of studies in which HCI principles and procedures are being applied to CAI surveys. The paper ends with a look to the future role of usability research and testing in CAI and the challenges facing the field. The goal of this paper is to extend the contributions that cognitive science has made to survey research, and examine ways in which both theoretical insights and empirical research on cognition and communication can be used to improve the design of computer-assisted interviewing systems in survey research.

Abstract

Customizing Survey Procedures to Reduce Measurement Error

Frederick G. Conrad

Survey measurement error is typically thought of as the difference between the observed outcome of each participant's behavior and the "true value" of the information he or she is trying to produce. This paper presents an alternative in thinking about measurement error that focuses more on how participants perform tasks than on the outcome of those tasks, and evaluates performance by how much it differs from the way survey designers intend the task to be performed. By this view, measurement error occurs when participants' actions do not correspond to designers' plans.

A similar view of how people interact with computers has led to the development of "user models" that enable software designers to adapt their plans to particular users' knowledge and action. The purpose is to promote accurate and satisfying interaction with the software. Similarly, survey designers can reduce measurement error by adapting their plans to particular participants' knowledge and actions. Three examples are presented that involve software-supported survey tasks though the approach does not necessarily require computers.

The first example involves retrieving published survey results from a web site. Currently, statistical web sites present a single "view" of the site to all users. An alternative that might make it easier for different users to find the information they seek is to present different sets of initial pages, depending on what the users indicate their task is.

The next example involves the accuracy of standardized versus conversational interviewing. The goal of the former is to keep interactions consistent between interviewers by scripting as much as possible.

The goal of the latter is to promote the respondents' comprehension by allowing interviewers to clarify misconceptions and confusions, in effect, adapting the designers' plans to respondents' understanding. These techniques were compared in three versions of a computer-based questionnaire that varied in the availability of definitions for concepts in questions. Under certain circumstances, respondents were quite inaccurate when no definitions were available (analogous to standardized interviewing) but highly accurate when they could obtain definitions as needed (one version of conversational interviewing) and when the computer could offer definitions after prolonged inactivity by the respondent (another version of conversational interviewing).

The final example involves comparing products whose prices are figured into the Consumer Price Index (CPI) at the Bureau of Labor Statistics. The CPI reports the monthly price change for a fixed set of products. When a product becomes unavailable, a similar product is substituted. The comparability of substitute and original products is evaluated by commodity analysts (CAs) who are experts in different product areas. An experimental expert system (COMPASS) was developed to simulate the CAs' reasoning in order to detect CA oversights or "slips" and, in theory, improve the accuracy of CPI estimates. By allowing each CA to create the relevant knowledge base, COMPASS, in effect, individualized the plan for evaluating product comparability in each product area. In the experiment, there were numerous discrepancies between COMPASS and the CAs, largely due to slips by CAs.

Abstract

Visualizing Categorical Data

Michael Friendly

Graphical methods for quantitative data are well-developed and widely used in both data analysis (e.g., detecting outliers, verifying model assumptions) and data presentation. Graphical methods for categorical data, however, are only now being developed, and are not widely used.

This paper outlines a general framework for data visualization methods in terms of communication goal (analysis versus presentation), display goal, and the psychological and graphical design principles which graphical methods for different purposes should adhere to.

These ideas are illustrated with a variety of graphical methods for categorical data, some old and some relatively new, with particular emphasis on methods designed for large, multi-way contingency tables. Some methods (sieve diagrams, mosaic displays) are well-suited for detecting patterns of association in the process of model building; others are useful in model diagnosis, or as graphical summaries for presentation of results.

Abstract

Statistical Graphs and Maps: Higher-level Cognitive Processes

Stephan Lewandowsky

Statistical maps and graphs have become an indispensable tool for the analysis and communication of data. The re-expression of data in pictorial form capitalizes on one of the most sophisticated human cognitive processing capabilities: the ability to perceive, classify, and understand complex visual patterns that defy simple verbal description. The rising popularity of statistical displays has been accompanied by increasing research interest in the way people perceive and process graphical information. So far, research has mainly focused on low-level perceptual and psychophysical variables, for example, the degree of error associated with estimates of size, angular extent, or length in a statistical graph. This paper selectively reviews the literature on the perception of statistical maps and graphs to direct attention towards other, more complex cognitive judgments. Concerning statistical maps, the literature is reviewed with an emphasis on cluster detection and other complex tasks, for example examining a dynamic spatial process on the basis of partial and fuzzy information. Experiments are reviewed that aim to identify the most appropriate statistical map for use in various situations. Concerning statistical graphs, the literature review focuses on the perception of large patterns and trends, and in particular the comprehension of interactions involving two or more variables. It is known that people's understanding of experimental results differs dramatically with different representations chosen for line graphs; the implications for the communication of scientific data are reviewed and possible remedies suggested.

Summary of Presentations and Discussion

Paul Beatty

Most collaborative efforts between cognitive psychology and survey methodology have focused on questionnaire development and pretesting. Yet there is great potential for cognitive science and related disciplines to contribute to other aspects of survey methodology as well. The papers in this session discussed efforts to apply these disciplines to areas such as usability testing and data presentation, as well as to understand the roles of survey interviewers and respondents.

Usability Testing: Mick P. Couper

Usability testing has recently emerged from the field of human-computer interaction (HCI), a relatively new interdiscipline with a history similar to CASM's. The methodology is also a logical extension of cognitive interviewing. Whereas cognitive interviews focus on the effectiveness of the questionnaire, usability testing focuses on the effectiveness of the system that delivers it (for example, the CAPI version of the National Health Interview Survey). Thus, while cognitive interviewing has allowed researchers to pay more attention to respondent difficulties, usability testing pays attention to a much less studied area—difficulties that interviewers may have carrying out their tasks. Computer-assisted interviewing systems offer considerable advantages for data collection efforts, but may be designed with an "ideal" user in mind. It is important that such systems work well for *typical* users.

A video shown by Mick Couper provided examples of usability laboratory equipment and practice at the University of Michigan. Multiple cameras capture a wealth of information; the interviewer's hands on the keyboard, the computer screen, and a shot of the entire laboratory can be recorded simultaneously. At the same time, the computer keeps detailed records of interviewer activity, to the level of keystrokes. The amount of data collected from one interview can actually be overwhelming, which leads to one of the key challenges raised by usability testing: How does one effectively analyze all of these data?

Couper's presentation and the subsequent discussion touched on a number of other issues that need to be resolved. For example, should usability tests make use of novice or expert interviewers? Should the tests take place in a laboratory setting or a more realistic environment? Should the respondents follow scripts, which would facilitate comparisons across interviewers, or should the tests use actual respondents to achieve greater realism? How many interviews are needed to determine instrument usability? What standard of evidence should be required to indicate that there is a "usability problem"? Similar questions have bedeviled methodologists in cognitive laboratories for some time.

One issue raised was the difficulty of separating problems with the interviewing system from problems with the questionnaire. It was suggested that many problems do not strictly belong to one category or the other. Yet, usability testing (as described in this session) treats these as distinct problems and focuses only on the former. Couper agreed that problems may stem from both sources, but suggested that the sheer inflexibility of some authoring systems for CAPI and CATI questionnaires can be a serious problem, imposing almost arbitrary rigidity on the interviewing process. Furthermore, the systems themselves may directly contradict the intended goal of making questionnaires work better for respondents. He recommended that survey researchers attempt to influence the development of the authoring systems to increase flexibility. It was pointed out that the history of the development of software for variance estimation shows that if user-identified deficiencies are related back to the designers, they will correct them. However, this has not yet happened with CAPI/CATI authoring systems.

One participant suggested that it may be appropriate to re-evaluate the trade-offs involved with computer-assisted interviewing. Are the benefits worth the constraints, particularly if much interviewer attention is focused on providing an answer that satisfies the machine rather than a high-quality response? Other practical concerns were discussed, such as the ability to document questionnaire versions over time, especially with computer technology changing so rapidly. Comparing the usability of paper and computer-assisted instruments might be valuable.

Several participants were impressed by how clearly the video example illustrated usability problems. One participant was struck by the image of the interviewer, quite detached from the respondent while intently

engaging her computer screen. Others were surprised to learn that the interviewer, who appeared to be a novice, had actually used the same instrument for almost a year. The video examples heightened everyone's appreciation of usability testing, which may improve connections between computer and cognitive sciences while expanding the tools available in survey-oriented laboratories.

Understanding the Roles of Survey Participants: Frederick G. Conrad

Fred Conrad's presentation explored the roles of survey participants, defined broadly to include survey designers, data processors, and data users, as well as interviewers and respondents. He distinguished between "mental slips"—the usual domain of questionnaire designers in cognitive laboratories—and larger discrepancies between survey designers' plans and survey participants' strategies. His presentation called for developing models of how survey participants approach their tasks. This would resolve problems beyond the scope of individual questionnaire repairs.

Conrad argued that survey designers expect participants to use strategies that may work under generic but not specific circumstances (a point reminiscent of Couper's comment that interviewing systems may work for an ideal but not an actual user). This viewpoint implies that the burden falls on the survey designer to ensure that survey participants can actually perform their proposed tasks. Problems emerge when survey designers fail to provide a plan for what respondents and other survey participants should do; problems also emerge if the plan is defective, or if participants misinterpret the plan.

Conrad's proposals to model the actions of survey participants reminded one discussant of developments in the philosophy of engineering design: Rather than teach a person to adapt to a machine, adapt the machine to the person. Another discussant mentioned that we may learn that survey designers are asking participants to perform tasks that they probably cannot do well. Such findings could force us to reconsider the appropriateness of surveys to collect some types of data, rather than trying to force respondents to provide data that resemble what we hope to collect.

Conrad encouraged the viewpoint that measurement error could be considered as a difference between designer intention and participant action, rather than a deviation from some "true value." A discussant raised the point that designers rarely reveal their plan explicitly, usually offering information about sponsors or purposes of the data instead. Conrad's suggestions assume that the respondent has the ability to understand the designer's plan and adapt accordingly. One discussant

recommended that these assumptions be the focus of additional research.

Several discussants addressed the problem of investigators being too far removed from instrument authors. This can be due to the structure of a research organization, and can also result from researchers' desire to maintain control over their component of the survey process. Maximizing the connection between what the investigator wants and what the survey actually measures can be difficult for these reasons; nevertheless, it is an important challenge to pursue.

Data Presentation: Michael Friendly and Stephan Lewandowsky

Papers from Michael Friendly and Stephan Lewandowsky shifted the focus to the challenges of effective data presentation. As a survey must effectively communicate its purposes to respondents, so data presentation must communicate clearly to its audience (probably data analysts and data modelers for purposes of these papers). Researchers presenting categorical data have many choices to make, and Friendly mentioned sieve diagrams, mosaics, and four-fold displays as possible vehicles. The actual presentation choices should depend on the substantive point being made, as well as the characteristics of the audience. In any case, presentation techniques for categorical data are in their infancy compared to those for quantitative data.

Discussants observed examples of categorical data displays and asked if these had been designed with the help of usability testing. Although they had not, their designs were at least partially based on principles from cognitive psychology. Another discussant, assuming a devil's advocate position, suggested that simple tables with numerical values might be more informative than some of the graphical displays that had been presented. Friendly answered that traditional tabular presentation might in fact be more informative for individual data cells. However, graphical methods are much more effective for displaying overall patterns, relationships between variables, and differences across groups. He also pointed out that there are many other graphical methods that were not presented during his talk.

Lewandowsky's paper focused on the psychological implications of various presenting styles rather than on new methods for presentation. Specifically, his paper focused on the use of graphs and maps to display cognitively complex information, such as 2- or 3-way interactions (graphs) or the presence of clusters (for maps). Of course, the effectiveness of a presentation style depends largely on the data being conveyed. For example, a pie chart is no better than a table for comparing two quantities, but a pie chart is much more effective for displaying more sophisticated comparisons. For graphs, the challenge is perceiving information on a

third axis, but unfortunately, "3-D" graphs are not particularly effective. Other techniques have been proposed, but their effectiveness has not yet been established by experimentation. As for maps, they already use two dimensions, so color is a logical choice for expressing a third. Designers may choose between monochromatic scales, scales with two colors at opposite ends (green/red), or categorical use of colors (red, blue, green, etc.) The two-color scale is generally preferable for expressing quantitative differences.

Most of the discussion following this paper focused on issues related to choices of colors, for example, a map scale with red-yellow endpoints might be preferable to a scale with red-blue endpoints for various reasons. A general principle that emerged from this discussion was that deeper hues tend to universally suggest a greater presence of the characteristic of interest, regardless of the specific color used.

Finally, a discussant noted that cognitive psychology has contributed a number of valuable principles for creating maps. However, it is still important to encourage experimentation to determine the most effective presentation styles for new publications, taking into account the particular data content and the intended audiences.

WORKING GROUP REPORTS

Foreword to the Working Group Reports

Clyde Tucker and Elizabeth Martin

Throughout the conference, the working groups met to consider new areas of interdisciplinary work which may enrich scientific knowledge about surveys, and to recommend additional applications of cognitive methods to survey measurement problems. All conference participants were members of one of eight working groups. Each group was asked to develop two to four proposals for research related to its topic area. The working groups' preliminary reports and recommendations were presented and discussed at the conclusion of the conference. The final written versions were submitted for publication in these *Proceedings* approximately nine months after the conclusion of the seminar.

The research proposed by the working groups, described in the reports which follow, can be broadly grouped into three areas:

1. New research to apply current cognitive methods to address survey measurement problems (described in working group reports 5-8).

2. Research to address unresolved issues and basic methodological questions about the interview process itself (reports 2 and 3).

3. Research to adapt to and build on broader changes in the survey world—especially the tremendous changes in the organization and technology of survey data collection and distribution (reports 1 and 4).

In the first of these three areas, two groups offer research proposals to improve the measurement of specific concepts in household surveys which monitor socioeconomic and health trends in the population. Working Group 5 focuses on the measurement of income, and Working Group 7 addresses the particularly problematic measurement of disability. The working group on income measurement proposes studies ranging from the purely qualitative to sophisticated quantitative analyses. The group suggests ethnographic studies of the reports of income in various subpopulations, linguistic analyses of question wording, and experiments involving new questionnaire strategies. Like some of the other groups, the income measurement group suggests creating

better measures of data quality. The working group on disability confines its attention to measures of the performance of normal daily activities and the extent of limitations in on-the-job performance. The group considers sources of variability in the answers to questions about work and functional limitations, especially as a function of comprehension and recall difficulties.

The other two groups concerned with new applications of cognitive methods are quite different from the first two and from each other. Working Group 6 examines the limitations of the current cognitive methods used to improve surveys, and Working Group 8 suggests ways of extending the use of cognitive methods to the study of establishment surveys. After generating a number of performance measures, Working Group 6 describes a multiphase experiment to validate questionnaire evaluation methods and goes on to propose studies which incorporate findings from the cognitive sciences and other sources of information into models to more accurately estimate population characteristics. Working Group 8 proposes several studies to explore what the cognitive sciences and related disciplines offer for improving establishment surveys. These studies include examining techniques for gaining the cooperation of establishments and for identifying the proper respondents. The group suggests investigating the ways establishments access and use their records when completing a survey form. The role of expert interviewers also is discussed.

Two groups examined unresolved issues: one examined the interview process itself and the other looked at various disciplinary approaches to the question and answer interchange. Working Group 2 suggests a study to document the ways in which the current practice of standardization is implemented in the field and then discusses the measurement of the costs and benefits of allowing varying degrees of interviewer flexibility in different survey contexts. Working Group 3 explores the interview process from various perspectives: they suggest that (1) the results of studies applying cognitive science to surveys be collected in a systematic manner and archived in a database accessible on the World Wide Web, and (2) additional work be done in the area of survey anomalies. The group also proposes work to

develop an ethnography of the survey interview as seen both by respondents and interviewers.

One of the groups focusing on the future of survey research considers the effects of changing technology on survey questionnaire development and data collection (Working Group 1), and the other looks at how the cognitive sciences might influence the other parts of the survey process (Working Group 4). Participants of Working Group 1 believe that work is needed to understand respondent tasks in self-administered computer-assisted questionnaires as well as the changing role of interviewers in automated surveys. The new technology required for questionnaire design and testing also is discussed. In considering the other parts of the survey process, Working Group 4 concludes that additional cognitive research should be done to help subject matter experts refine concepts and measures. The application of the cognitive sciences to the development of data dissemination and visualization products is advocated by the group, and this would necessarily lead to studies designed to gain a better understanding of the needs and statistical literacy of data users.

Although the working groups cover a great diversity of topics, several common themes emerge from their discussions and recommendations. A dominant theme is the need for continued development and evaluation of data quality. This includes the development of improved measures of quality. A more thorough examination of the interview process, itself, is needed for both household and establishment surveys. Of particular interest is the contrast between standardized and flexible interviews. Unfortunately, the survey research community has not done a good job of documenting knowledge about survey methods and practice. The results of experiments are not reported in a systematic way, and there is no central repository for these results. Several of the reports point to this shortcoming. Lurking in the background is the growing realization that technology is changing the face of survey research. This is an area in which the cognitive and related sciences have a lot to offer with respect to evaluating the usability of new tools and their contributions to the survey world. Finally, as surveys and survey data become more commonplace, we need to know more about the public's understanding of the purposes and the results from surveys. We especially need to know their view of the survey task, and how they use the results.

WORKING GROUP 1[*]

CASM in a Changing Survey Environment

Kent Marquis and Dan Kasprzyk

Introduction

The future survey environment, full of new technologies and tools, will pose exciting challenges for both applied psychologists and survey research practitioners. A new CASM research agenda can help the survey profession meet the promises and challenges of the new technologies and can benefit the behavioral sciences by generating new, fundamental knowledge. In designing survey applications that exploit new technologies and tools, we can expect to experience (and learn from) many successes and failures. And we can also expect to acquire better quality data, covering a broader range of topics, more quickly, at less cost, and with much less cognitive burden on respondents.

The working group that produced this report was given a broad charge: to develop a CASM research agenda appropriate for the changing survey environment. The first and most central step in developing such an agenda is to identify the significant research problems in the changing survey environment. Three general classes of problems are discussed in the report: human-computer interaction in the self-administered questionnaire; the changing role of interviewers in computer-assisted data collection; and supporting the development of modern questionnaires with modern tools. The goal in each instance is to enumerate research proposals and to establish why they deserve consideration.

The New Survey Environment

The new survey environment is marked by:

- Increased collection and dissemination of data through a variety of computer applications:

  - Computer self-administered questionnaires (CSAQ), including Internet surveys;

  - Computer-assisted personal interviews (CAPI);

  - Computer-assisted self interviews (CASI) which occur as part of CAPI sessions; and

  - Audio-CASI, which includes an option to listen to a voice reading the questions.

- Increased use of color, icons, applets, multimedia, etc., in data collection applications.

- A changing role for interviewers in the CAPI setting, since now the third element in the three-way interaction is the computer, rather than the paper questionnaire.

- Different demands (cognitive and motor) on both interviewers and respondents in the CAPI setting and for respondents in the CSAQ setting. The demands will (a) increase the need to switch attention between multiple sources; (b) increase reliance on working memory, possibly leading to overload; (c) complicate the comprehension task by requiring users to infer the meaning of the interface and its design; (d) require participants to coordinate multiple activities, and (e) require participants to learn many new skills.

- Expectations that questionnaire designers will develop more questionnaires, but with less time for pretesting.

In the new and dynamic survey environment, three topic areas can be identified where a continuing research program is desirable. The issues generated by these topics evolve from the current status and use of technology in surveys. The most promising research topics are:

- A fuller understanding of respondent psychology in the survey setting, especially when self-administered questionnaires are used;

---

[*]Additional members of the Working Group are: Fred Conrad, Mick Couper, Art Graesser, and Jim Kennedy.

- A better grasp of the dynamics of interviewer-administered data collection that includes technological assistance for the interviewer; and

- The development and use of tools to create better data gathering environments, for both electronic questionnaires and paper questionnaires, using basic tools, such as checklists, and advanced computerized tools, such as context analyzers.

To be effective, almost any future survey application based on technology will require a combination of: appropriate computer/media technologies; good design based on improved models of human cognition and performance; empirical tests of the usability of new applications with interviewer and respondent users; and iterative revisions of new applications.

A research program that is successful in modeling human behavior might also influence the development of better, off-the-shelf, commercial technology that will further improve our surveys.

The Respondent, the Self-Administered Questionnaire, and the Interface with the Computer

We need to improve our understanding of respondent tasks in the self-administered questionnaire framework (whether CSAQ or CASI) and then develop designs to ameliorate and overcome the difficulties associated with the tasks. When the respondent is also the user, the host of problems specific to intermediary users (especially interviewers) are no longer at issue. Research should focus on the use of technology to help respondents accomplish their tasks in an optimal way. Research should identify ways to overcome the cognitive, motivational, ability and knowledge limitations respondents may have, and to facilitate retrieval and reporting of information. Measurement of the effects of the different strategies is desirable. Ultimately, we seek to perfect today's psychological models that simulate human perception, thought, and performance on survey tasks.

The Questionnaire Interface

"Usability design and evaluation," which is an interdisciplinary field combining cognitive psychology, human factors engineering and (in some instances) computer science, promises to spur important advances in the design of questionnaires by focusing on the "human-computer interface." Our understanding of interface issues is already being informed by excellent models such as LUCID (Logical User-Centered Interaction Design) and GOMS (Goals, Operators,

Methods, and Selection rules). To what degree must we specialize these models to account for the way interviewers and respondents use an electronic questionnaire? How do interviewers and respondents differ in the way they interact with their instruments?

Common research issues for which we seek general solutions include:

- Designing forms for both novice and expert users.

- Bringing novices to expert status quickly.

    These issues are important because some respondents continually use a particular instrument and become expert users. They remain experts for much of the time they use a particular instrument. Therefore, it makes sense to design such questionnaires for expert users, but only if doing so does not increase the burden on novice users.

- Managing task complexity.

- Measuring and coping with the cognitive and motor demands of questionnaire interfaces (these demands include attention, learning, working memory, problem solving, comprehension, decision-making, and motor control).

    These issues will require that we sharpen our ability to analyze survey tasks, identifying the likely bottlenecks, and finding ways to avoid them through better design.

- Giving the user a cognitive map to navigate through a complex questionnaire.

- Identifying meaningful navigational and performance feedback and supplying it in a rational way.

    Navigational feedback is particularly important in survey questionnaires – relative to, say, spreadsheet or word processor applications. Instrument design imposes or assumes a particular sequence of user actions (say entering responses). If the user is uncertain about his or her current position in this sequence because of lack of feedback, numerous usability problems may result. Other types of performance feedback could also be incorporated. For example, the computer could report that the respondent is taking a particularly long time and offer to help. Such functions would force designers to confront issues of standardization: What kind of information can be provided to users without biasing response?

- Integrating additional sensory inputs.

- Guidelines for using color.

- Portraying complex meaning with visual icons.

These issues are currently unresolved in the world of human-computer interaction. Conrad (1997) provides a discussion of modeling users, such as respondents to CSAQs, to reduce measurement errors.

For CSAQs, we need to learn interface design principles that predict and control respondent attention, learning, motivation, understanding, and ability to formulate appropriate answers. For example, presenting questions via screen and headphones together may enhance respondents' concentration and perception of privacy, resulting in better data, especially about sensitive topics. Are there principles for combining multiple sensory inputs to achieve highest quality responses? What are the important qualities of voice audio? How can we use them to enhance performance? Do some voice features cause social desirability response biases and, if so, can they be filtered out or their effects neutralized?

At this early stage in research (Shneiderman, 1996), it seems that learning and performance are enhanced when everything needed for a task is on the screen simultaneously (perhaps in summarized form). How do we get around the intimidating nature of such screens for the novice and how do we help the experienced user manage the complexity? Early results also suggest that respondents will require control over navigational features to sustain attention and motivation. If we do not force respondents to follow a standard question-by-question path, we risk creating nonstandard conditions and partially completed questionnaires. What degree and kind of control must we relinquish to the respondent, why, and how can interface designs minimize the expected negative consequences for standardization and task completeness?

Replacing the Interviewer

New technology will inspire practitioners to implement electronic, self-administered questionnaires as a principal data collection mode. For those converting from interviewer-administered surveys, what is the effect of removing the interviewer? What effects did the interviewer have, which ones should we recreate by technology, and which should we avoid? Should electronic questionnaires have a "personality"? Should designers develop electronic questionnaires that pass the Turing test, that is, should such questionnaires be sufficiently human-like that respondents cannot tell whether they are being interviewed by a machine or a person? Respondents are more likely to report sensitive behaviors to a computer than a human interviewer, especially when this increases the privacy of their response (Turner et al., 1992). This is an instance in which human-like qualities in a questionnaire may not be desirable. However, a questionnaire that can detect respondent uncertainty and provide clarification might improve data quality. What are the effects of voice and other cues and how extensive are those cues in an interviewer-administered interview? What about interviewers' pauses, probes, feedback, and pace? Are the effects of these cues positive and could the positive ones be recreated in the electronic, self-administered form? Can audio files replace the human voice and can a video camera, or an animated character such as an avatar generate interviewer effects? (See Reeves and Nass, 1996, for some controversial claims on these issues.)

It is recognized that interviewers contribute to obtaining high cooperation rates. By contrast, self-administered questionnaires sent electronically might appear as cyber junk mail (spam) to potential respondents and result in low response rates. How will technology assume the persuasive functions now performed by the interviewer? It is said that, with some experience, human interviewers can make any paper questionnaire "work," despite severe problems in wording, navigation instructions, or task specification. We are discovering that users of electronic forms break off or stop participating when even slight problems arise or the task becomes uninteresting. Good interviewers redirect the respondents' wandering attention and motivate them to continue with tedious, difficult response tasks. Other interviewers appear to negotiate the level of effort necessary to "get through" the questions. Technology's potential to replace the attention-motivation-directing functions of the interviewer needs to be assessed.

The Interviewer-Assisted Survey

Although the number of self-administered questionnaires may increase in the future, there will almost certainly be demand for interviewer-assisted surveys. Two broad research areas for such surveys are: (1) developing improved procedures, with implementation of appropriate technical assistance, and (2) improving the overall technology of computer-assisted personal interviewing.

Technology for Better Interviewing Practice

Survey practitioners are currently debating the shortcomings of "standardized" interview practice. Alternative approaches, which can differ along several dimensions, may achieve better results. The

standardized interviewing paradigm presents all respondents with exactly the same stimuli (question meaning, wording, and sequence). Hallmarks of standardized interview operations are (1) exposing each respondent to exactly the same questions and definitions, allowing little room to "negotiate" the meaning of key concepts and terms, and (2) a navigational plan to direct respondents through identical sequences of questions to accommodate order requirements and context effects, and to "branch" respondents around questions that do not apply to them. Designers of interviewer-administered questionnaires are still trying to implement the standardized paradigm on laptop computers for face-to-face and telephone modes, largely within a DOS-based environment. Some of these implementations have proved expensive and some have taken longer than expected. Furthermore, by today's standards, the character-based interfaces and restricted navigation possibilities are sometimes disappointing.

New technology, such as graphical user interface software, should help. But the research agenda is broad and will call into question either the basic paradigm or the ways in which it is now being operationalized. Using new technology, is it possible to redesign the interview to achieve the goals of standardization and simultaneously develop improved operational methods? For example, can interviewing software be developed to clarify the meaning only when needed as opposed to having the question designer build complex definitions into the basic questions? Further, could the software relieve the interviewers of the need to learn and recite esoteric definitions at exactly the right moments? Is it possible to relax the requirement that each interview adhere to a precise navigational plan throughout the form? This becomes increasingly feasible with electronic as opposed to paper forms. Software can be designed to keep track of the questions that have been asked, regardless of order, in a way that human interviewers may not be able to do. Can we develop models of how interviewers and respondents perform questionnaire tasks? And if so, could this information be used to tailor our designs to what users know and want to do?

Upgrading CAPI Technology

Technology promises to make the household survey interviewer's job easier and more professional. The household survey interviewer could concentrate more intently on tasks that humans do best, and allow the computer to handle the details of navigation, recording, searching for definitions, etc.

A perplexing problem in the federal statistical system, however, is that the full array of technological aids has not been incorporated into existing CAPI systems. Existing technology is DOS-based, without a graphical user interface, and may be at odds with what users expect. At worst, the advanced technologies may

only have placed additional demands on interviewers. As a result, interviewers may pay less attention to respondents, serving more as intermediaries who read questions and record information electronically. There is also a potential for less social interaction between respondents and interviewers, due to interviewers focusing too much on the computer.

Interviewers must often balance several complex tasks simultaneously. We need alternative interfaces to improve interviewer effectiveness in complex task situations (reduce costs, improve data quality, and timeliness). As we approach research on interviewer-assisted interviews, we are aware that most of the human-computer interface literature deals with individuals working alone or in long-term work groups. Little has been published about how to design for the human intermediary, in particular, the one who performs a helper-facilitator role (or a similar role such as the help line attendant in the software firm or the directory assistance telephone operator).

Other potential research topics within the interviewer-assisted collection task include the following broad areas:

- Because interviewers coordinate multiple tasks and technologies, designs that optimize the performance of simultaneous tasks need to be developed. Development studies should focus on how interviewers handle multitasks (attending to needs of respondents, communicating with respondents, carrying out the correct survey tasks, making decisions about what actions to take, and physically interacting with the computer).

- Are there alternative input technologies that may be superior to the conventional keyboard for the multitasking interviewer? Candidates include pen computers, palm sized keyboards, and voice recognition.

- System errors, such as, an application "freezing," a computer "crashing," or unexpected occurrences on the screen can have devastating effects on an entire interview. Researchers need to develop principles of error recovery based on laboratory experiences, and from these individual experiences work toward general principles for designing error-recovery behavior in a computer-assisted interview.

- Quantify the benefits and drawbacks of alternative designs for the particular computer-interviewer interfaces (how they look, how they are organized, the ease with which one moves around the instrument). Develop and test interface guidelines and style sheets. Proselytize for empirical, quantitative usability testing with a set of showcase studies.

- Develop and evaluate training to ensure effective computer-interviewer-respondent interaction and performance. How much training can be eliminated by designing intuitive interfaces? How much training can be handled by simple, on-line assistance features? What kinds of training should be reserved for manuals, classroom instruction, on-line tutorials, and automated "wizards"?

- Develop a better understanding of the role and effects of interviewers in CASI settings (when respondents are self-administering questions of a sensitive nature).

For a fuller discussion of issues in computer-assisted personal interviewing, see Couper (1997).

Designing Tools and Managing the Survey Research Environment

Current questionnaire design processes can be improved considerably. In adopting computer-assisted collection technologies, we have increased instrument complexity and development costs for uncertain gains in reducing skip errors, making better use of previous answers, carrying out real-time edits, and performing calculations. Research is needed to go beyond these improvements and to develop better ways of designing questionnaires, to automate the review of complex instruments, and to develop additional tools that aid the performance of interviewer and respondent tasks. Research can also provide insights into how to organize and manage the survey organization to take maximum advantage of new technical knowledge.

One especially promising research area is use of technology to improve the design of questionnaires, electronic or otherwise. One goal would be to develop a computer model to evaluate questions at the levels of language, discourse, and world knowledge. The model would identify cumbersome syntax, rare words, ambiguous terms, misleading presuppositions, memory overload conditions, and out-of-context items. Most of these components are subtle, so they can sometimes elude questionnaire designers. Even an imperfect computer capability could probably surpass our current ability to detect some kinds of problems. We need to continue to improve language analyzers, such as QUEST, to be used as an aid in identifying problem questions. (For more discussion, see Graesser et al., 1997).

Other research in this area might include:

- Development of general software tools to identify improper and unrealistic branching (skip pattern analyzer).

- Development of graphical interfaces, a common language, and other cooperation tools to facilitate communication among participants in the questionnaire design process.

- Investigation of electronic tools, such as intelligent agents, to assist respondents in retrieving and recognizing their electronic information (possibly held by others, e.g., bank records) needed to answer survey questions.

- CAPI interviewing applications, at least in the federal government, must catch up with current computer technology, such as using graphical user interface tools (i.e., a Windows rather than DOS environment).

- Studies of alternative organizational structures for survey R&D, design, and implementation of electronic data collections. How should an agency organize to promote usability and to avoid being overwhelmed by "information technology" considerations? How can cooperation be improved between programmers, questionnaire developers, and sponsors in the development of specifications?

References

Conrad, F.G. (1997). Modeling survey participants to reduce measurement error. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Couper, M.P. (1997). The application of cognitive science to computer-assisted interviewing. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Graesser, A.C., Kennedy, T., Wiemer-Hastings P., and Ottati, V. (1997). The use of computational cognitive models to improve questions on surveys and questionnaires. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Reeves, B. and Nass, C. (1996). *The media equation: How people treat computers, television and the new media like real people and places.* New York: CSLI and Cambridge University Press.

Shneiderman, B. (1996). *Designing information-abundant websites.* Technical report, CAR-TR-824, CS-TR-3634, ISR-TR-96-40, Human-Computer Interaction Laboratory, University of Maryland. Also to appear in *International Journal of Human-Computer Studies.*

Turner, C.F., Lessler, J.T., and Gfroerer, J.C. (Eds.). (1992). *Survey measurement of drug use: Methodological studies* (DHHS publication No. (ADM) 92-1929). Washington, DC: Government Printing Office.

WORKING GROUP 2[*]

Exploring the Interview Process

Nora Cate Schaeffer and Patricia Royston

Introduction

Despite the wide range of topics initially discussed, most of our proposal topics fall under two headings: (1) increasing the interviewer's flexibility in assisting the respondent with comprehension and reporting, and (2) increasing the interviewer's and the respondent's motivation to obtain high quality data. These issues were motivated by the common observation that practices of standardization may interfere with respondents' ability to understand survey questions and provide accurate answers (see, e.g., Mishler 1986; Suchman and Jordan, 1990; Schaeffer 1991). Such speculations have received additional support in the recent work of Schober and Conrad (1997a; 1997b). Furthermore, the level and style of standardization may also affect respondent's motivation to provide accurate information (Dijkstra, 1987; van der Zouwen et al., 1991).

Our topics have implications for both current interviewing modalities (such as computer-assisted or paper face-to-face or telephone interviews) and technologies that are not yet widely used (such as computer-assisted self-administered interviews and touchtone data entry). Thus, the issues addressed by these topics will remain relevant for the foreseeable future. We did not attempt to develop an integrated research agenda, but instead assembled a set of topics or issues and then selected two topics involving the behavior of the interviewer or style of standardization for more detailed development.

Standardization and Interviewer Flexibility

Suchman and Jordan's criticisms of standardized interviewing have received considerable attention among survey researchers, and their suggestion that researchers attempt a "more collaborative" interview in which the "questionnaire ...is, at least, visually available to both parties" (1990, p. 240) has been widely discussed. Others (e.g., Mishler 1986) have suggested that standardized interviews be replaced by the collection of narratives. No

published research, however, has yet compared the results obtained with procedures like those sketched by Suchman and Jordan to those provided by the current practice of standardized interviewing. The set of practices we currently refer to as standardized interviewing (which actually vary across interviewing organizations as well as across interviewers within an organization), grew up at least partially in response to the findings of Hyman et al. (1954) that when the behavior of interviewers was not controlled, there was a substantial component of variation in answers that was due to the behavior of the interviewer. Thus, ideally, any research on modifying standardization would attempt to determine an "optimum" degree of interviewer flexibility by considering the trade-offs between costs (interviews that take more time, the resulting increase in break-offs, fewer questions, increased interviewer training, the need for more highly skilled interviewers) and benefits (improved validity and reliability of data, less frustration and burden for respondents).

One starting point is to describe the current practice of standardization. Fowler and Mangione (1990) have provided a concise statement of one practical method of standardization, and it is widely invoked as a standard. However, social research organizations vary widely in their actual implementation of standardization (for example, practices of "verification" or "probing" differ across interviewing organizations). Even for organizations subscribing to an identical canon, actual practice can nonetheless vary widely.

Research on this topic also needs to consider methods of modifying standardization separately for questions about events and behaviors, on the one hand, and questions asking for judgments and evaluations on the other. Much of the work reported by Hyman et al. (1954) concerned opinions and other subjective questions, for which both the object evaluated and the evaluative dimension may be vague or ambiguous, and thus relatively susceptible to influence by the behavior of the interviewer. For questions about events and behaviors, communicating the survey concept and stimulating the respondent's memory may be goals that

---

benefit from "tailoring" by the interviewer. Thus, this discussion and the resulting proposals were largely limited to issues that concern reporting of events and behaviors.

Studies of the "interactional substrate" of the interview (e.g., Schaeffer and Maynard, 1996; Clark and Schober, 1992) provide a useful framework when developing studies of the effects of increasing interviewer flexibility. Two relevant portions of the "substrate" include the skills that respondent and interviewer deploy, as ordinary interactants, to communicate or diagnose comprehension problems or to propose an answer and communicate its adequacy or the source of a potential problem with an answer. So, for example, respondents may indicate problems with questions by pausing, by providing "reports" instead of formatted answers, and so forth (Schaeffer et al., 1993). Such studies build on an awareness that much of what we call "cognition" is fundamentally interactive.

If such techniques are to be explored, additional questions need to be answered, including the suitability of these techniques for a diverse range of respondents and interviewers.

Motivating Interviewers and Respondents

The factors or influences that motivate respondent participation and willingness to provide accurate information in surveys still remain something of a mystery, and an area that might profit by collaboration between survey researchers and psychologists (e.g., Groves, Cialdini, and Couper, 1992). Work by Cannell and his associates (see review in Cannell et al., 1981) suggested that having respondents sign commitment statements, explaining to them the importance of accuracy, and providing systematic feedback could significantly increase, and presumably, improve, reporting of health care episodes. However, the effectiveness of some of these techniques was not replicated in telephone interviews (Miller and Cannell, 1982). Furthermore, the ways in which interviewers are evaluated or compensated may send ambiguous messages about the importance of data quality; these structures sometimes encourage speed over data quality, for example.

Research Proposal 1: Describe the Range of the Current Practice of Standardization and Interviewer Flexibility in Household Surveys Conducted by Major Survey Organizations

Goals

1. Describe how interviewers are currently trained and the rules of standardization they are supposed to follow.

2. Describe actual interviewing practices and how they accord with principles of standardization.

3. (Opportunistically) Describe the relationship between the practice of standardization/flexibility and data quality.

Rationale

A realistic benchmark is necessary to test the effects of varying the degrees of flexibility introduced into standardized interviewing. If current interview practice already incorporates different kinds of conversational flexibility, we might be able to use existing data to see effects of different kinds of flexibility on data quality.

Methodology

1. Collect from three government, three private, and three academic organizations the official policies regarding standardization. This would be done by:

   a. Collecting written policies and training materials,

   b. Observing interviewer training, and

   c. Documenting procedures for interviewer monitoring and quality control.

2. Conduct focus groups with interviewers from the nine organizations. Agenda will include interviewers' commentary on their organization's rules—what does and does not work—and interviewers' recommendations.

3. Collect data systematically on actual practices. Possible data sources include:

   a. Observe interviews,

b. Transcribe recordings of interviews (e.g., randomly sample, transcribe, and behavior code interviews), and

c. Conduct surveys of interviewers about their practices and beliefs about standardization, and their preferences for interviewing style.

4. Classify survey policies and actual practices (e.g., standardized, scripted definitions on demand, improvised definitions on demand, scripted definitions on demand and unsolicited, etc.).

5. Use the transcripts to conduct further studies of the interactional substrate of the standardized interview to identify interviewer and respondent behaviors that are likely to be associated with data quality.

6. In carrying out step 1 above, investigate availability of indicators of data quality and, if possible, analyze these data by level of standardization. However, this issue may be difficult to address productively in a nonexperimental study.

Possible indicators of data quality could include:

- Item nonresponse (don't know/refusals)

- Break-offs (partial interviews)

- Measures of interviewer variability

- Debriefing/reinterview data

- Other measures of data quality (both bias and variability)

These projects should consider that actual practice and adherence to rules of standardization depend on the characteristics of the particular survey, its sample, and the specific survey questions. This study should be descriptive and is intended to provide background for other investigations. The study could also be expanded to include descriptions of the norms and practices of interviewing across different types of interviews, such as ethnographic interviews, clinical interviews, therapeutic interviews, police interviews, etc. As Betsy Martin (personal communication) suggests: "Documenting the norms of different types of interviews might yield useful insights into how the interviewer's purpose is construed, and how aspects of interviewing are controlled in different settings."

Data collection organizations to survey would include: Census, BLS, NCHS, RTI, NORC, Westat, ISR, University of Massachusetts, and the University of Maryland.

Research Proposal 2: Measure the Effects on Data Quality of Varying Degrees of Interviewer Flexibility Within the Standardized Format.

Objective

This study examines effects on data quality, completion rates, and costs of varying degrees of interviewer flexibility within a standardized interview using a standardized instrument. Measures of cost include interview time, training time and training costs. Measures of benefits include increased respondent satisfaction, possible reduction in bias, and possible reduction in interviewer variance.

Rationale

Although standardization decreases interviewer variability, it has long been suspected of decreasing the validity of at least some responses. This could arise because interviewers are not allowed to deviate from the rules of standardization except under highly restricted circumstances and in highly restricted ways, even when they suspect that the respondent may not understand a question. Furthermore, in many (or most) standardized interviews the interviewer may not be adequately prepared to explain survey concepts and provide respondents with definitions and explanations; standardization provides some protection from interviewer variation that results from inadequately trained interviewers.

Methodology

The issues described above could be investigated by varying the following features: the instrument, the respondents' situation, the interview setting, the interviewing technique, the instructions given to respondents, and the training of interviewers. Within each of these features there are several design options:

Instrument: Use an established instrument from a major survey (e.g., CPS) or a set of questions selected to raise specific problems in interviewing, either from one instrument or from several instruments.

Respondents' situations: Use hypothetical situations (vignettes), or personal experience.

Settings: Laboratory versus field study.

Techniques used by interviewers: Each interviewer uses only one interviewing technique versus each interviewer uses more than one (within-interviewer measurement).

Instructions to respondents: Respondents could receive explicit training on how to be a flexible respondent. Or respondents could receive no explicit training.

Interviewer training: How much? Role-playing?

There are benefits and drawbacks to each option. For example, vignettes allow the researcher to construct situations in which effects can be detected. They also allow the fit between questions and respondents' circumstances to be controlled, but they lack ecological validity. Without ecological validity, the distribution of possibilities in the population may not be represented. Laboratory settings increase control but decrease ecological validity. One useful approach would be to begin with small-scale laboratory studies and then move to a larger field study.

Experimental Treatments

The suggestions provided below for experimental treatments could be augmented by discoveries made in the course of Study I described above. Interviewers would be trained to carry out interviews involving different kinds of flexibility. First, all interviewers would ask the questions exactly as scripted. Then, interviewers would do one of four things in response to requests for clarification:

1. Use only "neutral" probes (e.g., Fowler and Mangione's version of standardization).

2. Read official definitions only when respondents explicitly request help.

3. Improvise explanations based on official definitions only when respondents explicitly request help.

4. Read and/or improvise official definitions not only when respondents explicitly request help but also whenever interviewers think respondents need help.

Measures of Data Quality

1. Response accuracy: In a laboratory study one can measure response accuracy directly (the extent to which answers match what official definitions require). To evaluate response accuracy in a field study, one needs either record checks or proxy measures, such as response change across interviews or subsequent ratings (by questioning respondent) of fit between answers and definitions.

2. Interviewer variability (measured by assigning replicates of the sample to interviewers).

3. Respondent behavior (e.g., how many questions they ask, how quickly they answer) and ratings of satisfaction with or frustration with the interview.

Other Issues to Examine

Can all interviewers be trained to do flexible interviewing well? This issue is particularly important for surveys that require a large national field force. Interviewing is relatively low-paid work, and the pool from which interviewers are recruited may not have the skills required by more flexible interviewing techniques.

How do respondents know when they need help, and when are they willing to ask? What are the different ways in which respondents ask for help?

Can reporting be improved by allowing respondents to control the order in which they answer questions (e.g., in a sequence of questions about an event)?

References

Cannell, C. F., Miller, P. V., and Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389-437). San Francisco: Jossey-Bass.

Clark, H.H., and Schober, M.F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 15-48). New York: Russell Sage Foundation.

Dijkstra, W. (1987). Interviewing style and respondent behavior: An experimental study of the survey interview. *Sociological Methods and Research, 16,* 309-334.

Fowler, F.J., Jr. and Mangione, T.W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error.* Newbury Park: Sage.

Groves, R.M., Cialdini, R.B., and Couper, M.P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56,* 475-495.

Hyman, H.H. and others. (1954). *Interviewing in social research.* Chicago: The University of Chicago.

Miller, P.V. and Cannell, C.F. (1982). A study of experimental techniques for telephone interviewing. *Public Opinion Quarterly, 46*, 250-269.

Mishler, E.G. (1986). *Research interviewing: Context and narrative.* Cambridge, MA: Harvard University Press.

Schaeffer, N. C. (1991). Conversation with a purpose—or conversation? In P. P. Biemer, R. M.

Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman (Eds.), *Measurement errors in surveys* (pp. 367-391). New York: John Wiley & Sons.

Schaeffer, N.C. and Maynard, D.W. (1996). From paradigm to prototype and back again: Interactive aspects of "cognitive processing" in survey interviews. In N. Schwarz and S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass. (Revised version of CDE Working Paper 94-17.)

Schaeffer, N.C., Maynard, D.W., and Cradock, R. (1993). *Negotiating certainty: Uncertainty proposals and their disposal in standardized interviewing*. (CDE Working Paper 93-25.)

Schober, M.F. and Conrad, F.G. (1997a). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 60*, 576-602.

Schober, M.F. and Conrad, F.G. (1997b). Does conversational interviewing improve survey data quality beyond the laboratory? *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 910-915.

Suchman, L. and Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85(409)*, 232-41.

Van der Zouwen, J, Dijkstra, W., and Smit, J.H. (1991). Studying respondent-interviewer interaction: The relationship between interviewing style, interviewer behavior, and response behavior. In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman, (Eds.), *Measurement errors in surveys* (pp. 419-438). New York: John Wiley & Sons.

WORKING GROUP 3[*]

Different Disciplinary Perspectives on Cognition in the Question and Answer Process

Stanley Presser and Tracy Wellens

Introduction

The Working Group's recommendations cover three areas: an archive of cognitive testing findings; promoting cognitive science attention to survey measurement; and conceptions of the survey interview.

An Archive of Cognitive Testing Findings

One of the major outcomes of CASM I was the testing of questionnaires in the laboratory. Cognitive testing is now a routine phase of survey development in some organizations, and hundreds of questionnaires have been tested in this way. But though such testing has undoubtedly improved many individual survey questionnaires, it has so far contributed little to a general understanding of the question and answer process.

Testing is usually done to improve a particular questionnaire, with little concern for amassing experience and developing general principles. Moreover, the results of cognitive tests are typically squirreled away in unpublished technical reports that are rarely accessible to others.

Our first proposal is to collect the reports of these tests, code them, and archive the resulting database. The archive would facilitate systematic research into issues such as: What characteristics of questions are identified by cognitive interviewing as engendering particular problems? What testing features are associated with discovering different problem types? What sorts of solutions are adopted in response to various classes of problems?

These analytical goals would determine what information should be collected from the cognitive labs in compiling the database, how the information should be coded, and how the database ought to be organized. Analyses of the database would then provide the groundwork for the development of better theories of the question and answer process.

The database could also inform researchers whether items similar to those they intend to use in a new survey had been tested in the past, and, if so, how the items fared. This would not only reduce the likelihood of conducting cognitive interviews on items already diagnosed as showing problems, it would also allow some projects to start with the improved items that had been the end product of an earlier survey's testing. This in turn would allow research on how revisions recommended by an earlier project's testing fared when the "improved" items were themselves tested in the lab by a later project. Finally, the archive would promote learning across labs and potentially lead to improvements in practice.

For all these purposes, it would be desirable to make the archive available on a World Wide Web site.

Promoting Cognitive Science Attention to Survey Measurement

Another important outcome of CASM I was the influence of survey findings on research in the cognitive sciences. Thus, work on event dating was influenced by the finding of telescoping in panel surveys, and the study of judgmental processes was shaped by several context effects from attitude surveys. The research on basic cognitive processes stimulated by these survey findings in turn led to advancing the field of survey methods.

Additional work of this sort might be stimulated by making survey findings more widely known among cognitive scientists. A description of survey puzzles and anomalies could be written up for researchers in relevant fields. These descriptions might be publicized through papers given at meetings of cognitive scientists, articles in newsletters of cognitive associations, and small-scale conferences like those sponsored by the Social Science Research Council panel after the first CASM meeting.

Ideally, the choice of findings to be publicized (or to be the focus of a conference) would be informed by consideration of features of survey problems that might make them attractive to cognitive scientists. Our group discussed whether there were aspects of telescoping and context effects that distinguished them from the many

---

other response effects that have not been taken up by cognitive scientists. Although we were not able to identify such aspects and tentatively concluded that it was not clear before the fact that these particular findings would stimulate cognitive research, further work on this issue would be valuable.

An example of the kind of finding we have in mind is the result from the 26-city National Crime Survey that showed victimization reports for a 12-month recall period were significantly increased when the questions were preceded by a series of attitudinal items about crime. Although reports were not validated, the results suggest that giving respondents an opportunity to think generally about crime improves recall of specific experiences with crime. Is there a basic psychological principle at work here, that is, does providing opportunities to process material in a domain (e.g., by making attitudinal judgments) before having to engage in a recall task produce greater opportunity for triggering links to relevant memories? Questions like this might serve as a fruitful line of inquiry for collaborative effort between survey researchers and cognitive psychologists.

Conceptions of the Survey Interview

Our third proposal calls for a program of research into social representations of the survey question and answer process. We know a fair amount about how responses to a question are influenced by respondents' understanding of preceding items. But we know little about how responses are affected by respondents' understanding of the larger survey context. These include issues such as: Why do people think they are being interviewed? What do they think will be done with their answers? Do people have scripts or schemas about the survey interview? What assumptions do they make? And do they assume a survey interview is like a normal conversation, or that it is more like a test?

Similarly we know only a limited amount about these questions from the interviewer's perspective. How do interviewers think about the purposes and aims of survey interviews? What assumptions do they make about the rules they are supposed to follow? How do they understand the meaning of the confidentiality assurances they are instructed to give respondents? And so on.

Examination of these issues would benefit from the perspective of qualitative and interpretive cognitive disciplines like anthropology. In a sense we need an ethnography of the survey interview from both the respondent's and the interviewer's points of view.

Such ethnographies could have important implications for many aspects of data quality. For example, respondents frequently appear unwilling to take at face value the reference periods included in behavioral survey items. Surveys ask "Did you go to church in the last seven days?" "Did you see a physician in the last six months?" "Did you eat any fruit yesterday?" We suspect that respondents report events from outside the reference period (as well as failing to report some events that did occur during the period) because they assume the survey intent is to discover what kinds of people they are, not what they did during some arbitrary period of time—a goal that may seem pointless to many respondents. Thus respondents who hate fruit and almost never eat any may assume that they shouldn't report the melon served yesterday at a friend's home and eaten only to be polite. Probing into respondents' understandings of the purposes of survey interviews may help us better understand this issue, as well as many other aspects of the question and answer process.

WORKING GROUP 4[*]

Applying Cognitive Methods to New Areas of the Survey Process

Cathryn Dippo and Cynthia Clark

Introduction

Four areas of the survey process where the application of cognitive science could be expanded were singled out for discussion: concept development, classification and coding structures, statistical information dissemination, and statistical literacy.

Concept Development

Not all concepts are initially defined by sponsors to a level sufficient for precise measurement using a sample survey. In terms of the model presented by Groves (1989, ch. 1) the *underlying characteristic* may be vague or ill defined. Or, it may be that there is no useful measurement history or mapping of the characteristic into a measurement tool. Consider, for example, the concept of a *job opening*. A position for which a vacancy announcement has been printed is clearly a job opening. But, what happens if the employer finds among the applicants two extremely well-qualified candidates and decides that with their available budget they will hire both? Or, what if none of the candidates completely meets the job requirements, but two of the candidates have complimentary qualifications and each is looking for part-time work? This example serves to illustrate difficulty in measuring the concept.

All too often sponsors approach survey organizations without having considered the complexities of the construct of interest, given the diversity of the target population (for example, the meaning of *hours worked* to someone who is self-employed). Sometimes the sponsor comes with a set of questions that they wish to have administered, but without a statement of the measurement goals and some idea of the types of statistical analyses to be performed on the resulting data. In each of these cases, the survey organization has to expend considerable resources if the sponsors' ultimate measurement goals are to be attained. The process by which the survey methodologist aids the sponsor in honing the meaning of the construct to be measured is

not well-defined. A model which could be used by potential sponsors in preparing themselves for discussions with survey developers and by survey developers in working with sponsors could be a very valuable tool in reducing the time needed to move from construct to measurement; improving the communication between sponsors and developers; and ensuring that measurement results meet sponsors' needs. Such a model should reflect the role of research for the purpose of adding knowledge about survey measurement processes and for providing feedback to sponsors on the quality of the data products. This information should aid in the refinement of the constructs themselves in light of findings and observed measurement problems. We propose research directed toward developing a general model that could be applied to devise and refine survey concepts to ensure measurability.

Classification and Coding Structures

One particular form of concept development that is especially difficult relates to classification systems. In almost all federal establishment surveys, industry classification plays an important role. Significant resources have gone into developing a new industry classification system over the last few years, and there has been an attempt to overhaul the occupational classification system used in many household and establishment surveys. Price index programs like the Consumer Price Index and many health surveys also depend upon classification systems of products or health conditions like disabilities.

In addition, many questions are designed to include answer categories. In some instances respondents are asked to select an answer category from a list of possible categories; in other instances interviewers are expected to select a category based upon information provided by a respondent who has no idea what the answer categories are. In other instances the number or complexity of the categories is such that the interviewer is asked only to collect the open-ended text from the respondent, and the

---

classification of responses into categories is done by a centralized coding staff. This is almost always the case when respondents' open-ended answers need to be classified using a standardized coding scheme such as the Standard Occupational Classification (SOC) system or the North American Industrial Classification System (NAICS). In this situation, the separation of the classification task from the data collection task often results in incorrect or inadequate information being recorded, forcing the development of extensive coding rules to cover innumerable possible (but not necessarily very probable) situations.

The way standardized classification systems (or "coding frames") are currently developed is often intuitive and subjective. The categories are usually derived from a construct in a particular discipline—the NAICS is based on an economic theory of production—without the benefit of empirical research into the way coders interpret the categories or, for that matter, the way end users of the coded data interpret the categories. Typically, a small group working within a particular framework creates systems of mutually exclusive categories which supposedly are homogeneous within and heterogeneous between, based upon some set of assumptions about the concept to be classified. A different working group would not necessarily end up with the same categories, even if it started with the same set of assumptions. (Traditional practice also defines other methods, for example, developing a classification scheme based on a sample of responses.)

Cognitive psychology has used a set of multivariate techniques for converting people's implicit and internal conceptual structures into explicit and external conceptual structures. One such method is **M**ulti**D**imensional **S**caling (MDS) that visually arrays in n-dimensional space a set of concepts according to their similarity (e.g., Rips, Shober, and Smith, 1973, used MDS to depict the relation among people's concepts of mammals). This method was also used to explore people's notions of the job categories in the Standard Occupational Classification (SOC) system (Conrad and Tonn, 1993). SOC users were asked to judge the similarity of a set of occupations. When their ratings were submitted to MDS, the occupations were plotted according to the amount of education they required (nurse and teacher at one end of the scale, cashier and stock clerk at the other) and also according to the abstractness of the concepts involved in the work (underwriter and librarian at one end, carpenter and repairer at the other).

The proposed research is intended to evaluate methodologies that may be used to enumerate a set of categories and the criteria to be used in classification. We propose using several methods (including techniques designed to uncover mental taxonomies) such as multidimensional scaling to develop a classification system for a particular domain and then to evaluate the resulting schemes using measures such as intercoder reliability, accuracy with respect to expert judgments, and subjective ratings of user satisfaction. (For an example of work addressing mental taxonomies, see Rips and Conrad, 1989.)

Statistical Information Dissemination and Visualization

The potential role of cognitive sciences in the area of analysis and dissemination of sample survey data is extremely broad and diverse. To focus the discussion, the working group concentrated on research associated with dissemination and visualization via an Internet web site (http://www.fedstats.gov), where the primary audience is the general public, rather than academic researchers, public policy analysts, or other regular users of federal statistics.

Research on this topic refers to two sets of questions related to making the results of surveys and other large-scale data collection programs available and accessible to information seekers (broadly defined).

- What sorts of data displays (tables, charts, and graphs) make it easier for information seekers to carry out tasks (lookup, making comparisons, finding patterns and relationships) associated with public databases? For concreteness, we consider databases where the primary structures are: time series (e.g., Consumer Price Index), geographic (unemployment rates), or multiway frequency tables (e.g., General Social Survey). Various improvements to the traditional ways in which such data have been displayed have recently been proposed, for example, micromaps, time series with aspect ratios chosen to enhance perception, mosaic displays and fourfold displays, and revisions to tabular displays. It is proposed to adapt and extend the cognitive methodologies used successfully in the study of simple graphical displays and of statistical maps to the study of these newer forms of data display. Some subsidiary or derivative questions are: How to design data displays to show the uncertainty, quality, or metadata associated with some quantitative summary? How to design displays suitable for audiences with different levels of statistical sophistication? What possibilities are provided by dynamic or interactive displays? What are the design properties of a good user interface for visual display?

- Survey data are distributed in many different forms, but the possibilities and challenges presented by the Internet create some of the most interesting opportunities for CASM research. It is appropriate and potentially useful to focus here on the newly

initiated Fedstats site. We advocate the development of projects to answer some of the following questions: What sorts of navigational and search facilities make it easy and efficient for users to find the information/data they are seeking? How can we provide data retrieval and data display methods on a web site that can adapt to the specific questions or level of expertise of the information seeker? How should metadata be linked to data in order to promote their use? How can data from multiple surveys with relevance to the same topic be put in context? What types of on-line analyses beyond precomputed tables would naive users of statistics find helpful?

Potential research methods include laboratory studies, Web surveys, and analysis of user logs. Experiments could be designed to evaluate how well alternative visual displays or user interfaces aid users in predefined tasks of information seeking and comprehension.

Statistical Literacy

Improving statistical literacy is a necessary concern of those who produce statistics for use by the general public. With the increased availability of statistics via the Internet and the current administration's goal of making computers and the Internet accessible to every student, we now have an opportunity to lay the foundations for ensuring that future generations are increasingly statistically literate.

This research project is related to the topic of dissemination of information from government and academic surveys. Specifically, it focuses on an important precondition for effective dissemination: do members of the public understand the different uses they might make of statistical information in their own lives, and can they be educated in the more effective use and interpretation of statistics?

The first set of research questions concerns people's current ideas about statistics. What do people currently feel they need or want to do with government statistics (such as unemployment, CPI, crime, or health data) or attitude survey findings? As one example, when people contemplate a move to a new community do they consider looking up information about that location to make a more informed decision? Do elementary, secondary or college students think of using government statistics or attitude data when writing reports or preparing papers for classes?

There are several subsidiary questions about the way in which people consider using statistics, even when they consider them relevant. Does the general public understand the difference between descriptive statistics (appropriate, for example, for comparing the crime rate

in different communities) and statistics that involve control for potentially confounding variables (appropriate for making inferences about causes of crime)? Do people distinguish between carefully collected and highly aggregated statistical data and vivid anecdotes or individual examples? There is relevant psychological research which illustrates that when people reason about quantities and probabilities they do not always take available statistical data (such as base rates) into account, or apply relevant statistical rules (for example, conjunctions of independent facts can be no more likely than either fact by itself). (See Kahneman and Tversky, 1982, for an overview.)

Research methods that might be used to answer these questions include survey questions (e.g., asking people whether they have ever accessed or used government statistics and for what purpose) and focus groups.

Finally, a naturally related set of research issues involves ways of educating the public to improve understanding and use of statistics. Can people effectively be taught the difference between descriptive and partial numerical summaries, or the role of statistics in causal inference, for example? Can they be led to realize that statistical summaries should carry more weight in their decisions than one or two vivid anecdotes?

Methods to approach these questions would involve working with educators and related professionals (librarians and curriculum developers) to see what could be done in the elementary and secondary school context. Further, research could be devoted to developing brief on-line tutorials on relevant issues that could be accessed over the Internet. Perhaps government statistical web sites could include such tutorials to guarantee ready accessibility. Principles from cognitive science, particularly educational psychology, can be used to help develop educational methods. Such principles include presenting statistical content in the context of substantive material (such as geography or history) rather than in abstract, content-free form and tailoring presentation methods to the audience in terms of their prior knowledge and cognitive style (e.g., visual versus verbal learning style).

References

Conrad, F. and Tonn, B.(1993) Intuitive notions of occupation. In *Proceedings of International Conference on Occupational Classification* (pp. 169-178). Washington, DC: Bureau of Labor Statistics.

Groves, R.M. (1989). *Survey errors and survey costs.* New York: John Wiley & Sons.

Kahneman, D. and Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, and A.

Tversky, (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 493-508). Cambridge: Cambridge University Press.

Rips, L.J. and Conrad, F.G.(1989). Folk psychology of mental activities. *Psychological Review, 96*, 187-207.

Rips, L. J., Shoben, E. J., and Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*, 1-20.

WORKING GROUP 5[*]

Income Measurement

Martin David and Jeffrey Moore

Introduction

Data on income and assets are collected in greater or lesser detail in many major surveys such as the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the Survey of Consumer Finances, and the new Survey of Program Dynamics designed to measure the effects of welfare reform. These data are critical to the formulation and evaluation of economic policies for taxation, income transfer programs, health care and education.

Issues and Problems

Collection of complete, accurate data on income and assets encounters many obstacles that lead to error. Unit and item nonresponse are high. Levels of nonresponse are higher at both extremes of the income distribution. Sensitivity to divulging financial information is thought to contribute to nonresponse. Response error is large. Errors in reporting receipt and amount of income are substantial. Comparisons to benchmarks estimated from administrative data suggest that income from all sources combined is understated by about 10 percent. Errors are especially high for certain types of income – means-tested transfer income, self-employment income, dividends, interest, and interfamily transfers. Comparison of particular sources of income from the 1990 SIPP to benchmarks reveals understatement of dividend income by more than half and overstatement of pensions from state and local governments by more than one-sixth (Moore, Stinson, and Welniak, 1997). These comparisons are problematic because of the need to adjust for differences in coverage and concept, but they indicate an underlying problem in reporting income that cannot be ignored.

Not much is known about *causes* of response error. Respondent burden appears to be large, especially for longitudinal surveys. Several kinds of burden are present, especially when interviews are long. The reference period or timing of an interview may make it difficult to retrieve information. Unfamiliar terms are used. Respondents are asked for information that is not available to them (e.g., payments made by others for health care or insurance).

Behavior of interviewers and respondents contributes to errors. When assistance is requested, interviewers may not be able to provide helpful clarification of questions to respondents with complex financial affairs. Respondents may smooth reports within the reference period. This behavior is probably present in most surveys, but is most readily measured in longitudinal surveys. It results in more changes of status between interview reference periods than within these periods (the seam effect).

An additional problem is that measures of income may be distorted by the failure to distinguish the income received by individuals as trustees for the benefit of others, and income accruing to the individual's own benefit.

Despite all of the foregoing, many people give high quality data on particular items, particularly wages and salaries and old-age benefits from Social Security.

Benefits

If income is poorly reported, estimates of total income and the distribution of income are faulty. Improved measures of income can lead to better estimates of income that is not captured in administrative or economic account data. Better design of tax and transfer legislation would follow from an understanding of the real economic position of families, particularly those at lower income levels.

The credibility of statistics on poverty could be improved if underreporting were adequately measured. Estimates of bias and mean squared error could be used to produce a clearer understanding of the gap between income and poverty thresholds. Improved measures would also improve understanding of the economic situation of subpopulations. More accurate aggregates for sources of household income would be obtained, since

---

[*]Additional members of the Working Group are: Norman Bradburn, Brian Harris-Kojetin, Thomas Jabine, Jim Press, and Ed Welniak.

available administrative data do not necessarily relate to the household population. Perhaps the most important benefit would be that better measurement at the individual and household level would clarify associations between income, behavior that procures income, and consequences of income for saving and spending.

Research Program

Cognitive research on eliciting financial information promises substantial insight into existing measurement errors and offers the hope of reducing those errors. First, the precise meanings of words used in eliciting information may be unclear to the respondent. For example, many persons reported take-home pay instead of gross earnings in the Survey of Income and Program Participation prior to 1996. Second, questions may call for information that is cognitively complex, as when persons receiving means-tested welfare are asked to classify the income according to some legal description. Cognitive research could ascertain when the information is unavailable to the respondent. Third, cognitive research may uncover motives for distorting answers, while providing frameworks that give positive incentives for full and truthful reporting.

Qualitative Research Using Ethnographic Methods

Objective

Qualitative study of respondents and their cultural context using ethnographic methods can improve collection of income and asset data. Improvements will be sought not just for individual survey questions, but for all features of information collection designs, including:

- Reference periods and timing of data collection in cross-sectional and longitudinal surveys,

- Choice of respondents,

- Choice of mode(s) of data collection,

- Amount of detail,

- Interviewer training,

- Uses of respondents' own records, and

- Use of records held by administrative agencies or financial agents.

Types of Information Sought

Intensive interviewing can produce data on schemas and cognitive frameworks used by various types of households and individuals to manage money. Answers to important questions can be obtained:

- How do persons deal with receipts and expenditures within the household unit?

- How do persons track income and expenditures for budgeting and tax purposes?

- What terminologies do persons use to label different types and sources of income and different kinds of assets?

- Ancillary information is also important. Do respondents have prior experience with surveys? Are they motivated to respond or to withhold information? What are respondents' views of the utility of various kinds of surveys? What income types and amounts do respondents fail to report?

Proposed Research Design

We recommend two types of ethnographic interviews:

Conventional approach, general population. These would be modeled on the ethnographic studies related to race/ethnicity and household rostering that were conducted in preparation for the 2000 census. Subjects would be selected purposively. Subjects should be drawn from several race/ethnic groups, households with high and low income levels, and the self-employed. Special groups, such as recently widowed persons might also be included.

Ethnographic interviews of respondents in the CPS Income Supplement or SIPP. For SIPP respondents, interviews might be conducted after the initial wave, the wave with the first tax round-up supplement, and the final wave. A cash or other incentive for participating in these ethnographic interviews is recommended. The conditioning effects of multiple interviews would have to be addressed. These ethnographic interviews would focus more directly on the interview experience, especially the parts that deal with income.

Expected Benefits

We expect that the proposed ethnographic studies would provide valuable background information on which to base more focused experiments designed to improve information collection systems for income and assets. Improvements may be realized in the form of reduced mean squared errors, or lower costs. If

procedures that cost more per household produce better data, trade-offs between sampling and nonsampling errors can be evaluated. If properly timed, the ethnographic studies could be very helpful in efforts to maintain the quality of data on transfer income during the transition period for welfare reform.

Linguistic Analysis of Questionnaires

Objective

Apply the lexicon and heuristics offered by Graesser (1997) to identify linguistic "problems" with income and asset measurement in major surveys used by the federal government (Consumer Expenditure Survey, Survey of Consumer Finances, Current Population Survey, and the Survey of Income and Program Participation).

Graesser indicates that ambiguities in meanings, complex referents, and other linguistic problems arise in many federal surveys and administrative data collections. He has already tested federal income tax questions and a number of survey instruments for these problems. A computerized lexicon is already available. Computer algorithms to analyze linguistic structure have already been tested in other contexts.

Linguistic analysis can identify questions that tax the cognitive abilities of respondents, questions that entail a memory overload, and questions where the answers cannot be retrieved because of semantic or syntactical ambiguity. Awkward constructions and words with meanings that are ambiguous or infrequently encountered in common use can also be identified.

The proposed analysis, which requires no "field work," will aid survey methodologists in identifying when substitutes for existing questions should be considered. The grounding of later questions in meanings established from earlier dialogue can also be determined. The March income supplement to the Current Population Survey would appear to be an ideal candidate for linguistic analysis, as "redesign" is planned.

Alternatives to "Exact" Dollar Amount Reporting

In the absence of records, some respondents offer no answer; others estimate a value. If the questionnaire protocol embeds a strategy for estimating the value needed, it may elicit a range of values from persons who would otherwise not respond.

The primary objective of this research is to identify question sequences that reduce nonresponse to survey income items. Reducing respondents' cognitive burden and reducing the sensitivity of the information provided may be important factors in the success of such sequences. The hypothesis is that both don't know and refusal nonresponse may arise from an insistence on the reporting of precise dollar amounts. Precise amounts are not available to many respondents. "Don't know" becomes the only truth they can report. Estimation of amounts may not be permitted by the questionnaire structure, and the work required to produce a precise amount may be difficult and time-consuming. In addition, insistence on precision may exacerbate sensitivity, especially in the absence of perceived justification for that precision (either with regard to the overall survey task, or for perceived "trivial" income sources).

Questionnaires that permit estimation may encourage persons who give exact answers to offer less precise information (Kennickell, 1996). Research will need to balance reduction in error among nonrespondents against possible increases in error among respondents to current questionnaires. The effect of different strategies on respondents' perceptions of burden and sensitivity, as well as their perceptions of the importance of expending effort to produce high quality responses also needs to be measured. (See Sirken, Willis, and Nathan, 1991 for an example.)

Field experiments are required to measure the actual effect of the alternative strategies on nonresponse (unit and item), frequency of record use, field costs, and, with appropriate validating information, measurement errors. Research should also investigate the additional "analytical burden" which would result from the imprecision of income range reports, although this work, too, would be guided by a comparison of survey reports and validating data.

Experts believe that responses are not missing at random. If a substantial part of nonresponse to exact amount questions can be converted to coarse classes describing the amount, bias may be substantially reduced.

Adapting Protocols to Populations That Use Different Cognitive Frameworks

Objectives

1. To discover subpopulations in which mean squared error (MSE) of income reports can be significantly reduced.

2. To create data collection protocols that reduce mean squared error, or achieve existing mean squared error at lower cost, by tailoring protocols to the identified populations.

3. To validate the conceptual framework used to locate the populations by cognitive interviewing.

We assume that income is measured with a panel design. Screening information is acquired initially and subsequently updated. Screening allocates the sample to different interviewing protocols. Questions,

explanations, and reference period may vary over protocols. The design elicits measures at different points of time. Because the characteristics of respondents change over time, protocols assigned to them may also change over time.

The intent is to discover protocols that reduce respondents' exposure to questions that fail to take account of the individual's circumstances and to adapt question sequences to the cognitive frameworks used by respondents in thinking about money and finances. Both kinds of changes can reduce burden on respondents. Adapting question sequences to the respondent's circumstances has the potential to increase comprehension, motivation, and accessibility of answers. Reducing burden may avert failure to answer questions (eliminating a source of nonignorable nonresponse). Questions "tailored" to the respondent may increase the likelihood that the respondent supplies a "true" answer. "True" answers need to be established by the comparison of survey reports with external records.

The experimental design required is summarized by the diagram below. Two phases in the analysis need to be explained. Phase I describes the underlying experimental design. Phase II recovers the best estimators from the observed data.

Phase I. The control is the existing design, *U*, a universal protocol. Screening questions are used to partition the experimental group into *A* and *B* treatment groups (possibly more). Different protocols are applied to the *A* and *B* treatment groups. Each treatment produces a measured mean squared error, $MSE_{T|G}$, for the income variables of interest; *T* indexes the treatment; *G* indexes the group to which treatment is applied. The mean squared error for each group is established by a design that includes validation of reports by record checks.

The outcomes of the experiment include *desired* effects, $MSE_{A|A}$ and $MSE_{B|B}$, where gains may be obtained by tailoring questions to subpopulations. Because the screening mechanism is fallible, the *A* treatment is sometimes applied to members of the *B* group, and conversely. For that reason, outcomes also include $MSE_{A|B}$ and $MSE_{B|A}$, where the wrong protocol is applied to a respondent. If tailoring the survey design to particular populations is to be successful, the desired effects must reduce the sum of $MSE_{A|A}$ and $MSE_{B|B}$ below the sum of $MSE_{U|A}$ and $MSE_{U|B}$. It is also necessary that this reduction in error is not offset by $MSE_{A|B}$ and $MSE_{B|A}$ in the inappropriately treated population. These two considerations imply that mean squared error for the whole population under the multiple protocol design must fall below the mean squared error for the whole population under a uniform protocol.

Phase I: Screening

True Target Population		
Treatment applied	*A*	*B*
Q_A experiment	$MSE_{A|A}$	$MSE_{A|B}$
Q_U control	$MSE_{U|A}$	$MSE_{U|B}$
Q_B experiment	$MSE_{B|A}$	$MSE_{B|B}$

Phase II: Postinterview stratification

Correctly treated?　　$p_A < 1$　　　$p_B < 1$

Phase II. The second phase of the design identifies which respondents and households are misclassified by screening questions. This analysis estimates the probabilities of appropriate treatment. Improvements in the survey design can therefore be made in two ways: Increasing the number of treatments or increasing the accuracy of the assignment of the population to alternative treatments.

Screening questions should identify the respondent's (or household's) principal income sources and the absence of significant assets. Questions in the screening phase that identify record-keeping practices, attitudes toward government, and understanding of economic concepts (cash flow as opposed to gross income) would also be desirable. Effort should be given to identifying items that correlate to "threat" posed by, or "social acceptability" of, reporting some kinds of income, for example, ticket scalping.

Cognitive interviewing techniques can be used: to get respondents' evaluations of their confidence in the correctness of their answers to questions; to obtain postinterview evaluations of the difficulty of the task; and to verify hypotheses that are generated by ethnographic and cognitive analyses of the questionnaire protocols (i.e., findings from the first two parts of the research program).

Modeling the Perceived Quality of Answers and Willful Response Error

The objective of this research is to develop methods of scoring interviews for data quality in ways that will be useful for modeling in the presence of refusals to supply information or falsification of responses.

Income questions induce high rates of nonresponse. The quality of answers provided by some respondents is also in doubt because of demonstrated omissions and false positive responses. Reluctance to provide income information leads respondents to refuse, claim they do

not know, provide partial information, or intentionally distort their responses to varying degrees.

Data quality ranges from distortion/inaccuracy, to nonresponse, to accurate reporting. Although data quality is typically derived from external records, it may be possible to construct a measure using information from the survey itself supplemented with information from the interview. Tucker (1992) used information about the reports in the *Consumer Expenditure Diary Survey* in a latent class analysis to construct an ordered set of data quality categories for those data. Further research should be conducted along these lines to evaluate the feasibility of constructing an ordered set of data quality categories or a continuum for income data and its usefulness.

Locating a respondent on a continuum of data quality will require investigation into indicators of data quality and respondent behavior and motivations regarding the reporting of income. Ethnographic research would seem particularly useful to gain a better understanding of the concerns respondents have about providing income information, as well as what would help alleviate those concerns. Further exploration is needed into how respondents provide income information and how they feel about giving it, and the strategies they employ to avoid providing income data (e.g., refusing, fabrication, rounding, omitting some sources).

Interviewer ratings and respondents' assessments of the quality of their answers can be used as additional measures of data quality. Measures of quality can be used to create a model of the probability of error (Sirken, Willis, Nathan, 1991). Those probabilities can be incorporated into estimation as proposed by Groves (1997) and Bollinger and David (1997).

Priorities

Ethnographic research and linguistic analysis of existing questionnaires should have the highest priority. Few designers and users of income questions are adequately informed as to the cognitive problems that face typical respondents to income and asset questions, and many questions clearly pose unnecessary memory loads or other cognitive problems.

A continuing program of validation studies is also essential. Estimates of mean squared errors cannot be calculated without such studies. Determining variation of mean squared error in the population is a necessary part of any improvement program.

Questions that might be used to sort the population into groups with different types of problems in responding to the questionnaires need to be evaluated against a criterion that is more substantive than current practice, which can be caricatured as "more income reported is more accurate reporting."

References

Bollinger, C.R. and David, M.H. (1997). Modeling discrete choice with response error: Food Stamp participation. *Journal of the American Statistical Association, 92*, 827-835.

Graesser, A.C. (1997). The use of computational cognitive models to improve questions on surveys and questionnaires. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Groves, R.M. (1997). Survey error models and cognitive theories of response behavior. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Kennickell, A. (1996). Using range techniques with CAPI in the 1995 Survey of Consumer Finances. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 440-445.

Moore, J., Stinson, L., and Welniak, E. (1997). Income measurement error in surveys: A review. Background paper prepared for the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Sirken, M.G., Willis, G.B., and Nathan, G. (1991). Cognitive aspects of answering sensitive survey questions. *Proceedings of the International Association of Survey Statisticians, International Statistical Institute,* 130-131.

Tucker, C. (1992). The estimation of instrument effects on data quality in the Consumer Expenditure Diary Survey, *Journal of Official Statistics, 8*, 41-61.

WORKING GROUP 6[*]

Integrating Cognitive Research into Household Survey Design

Judith Lessler and Jennifer Rothgeb

Introduction

Our discussions and subsequent suggestions for research projects were based on the assumption that the goal of using cognitive research in the design of household questionnaires is to improve the overall survey measurement process, that is, to produce survey estimates which have improved reliability and validity. We recognized that cognitive research could be used to address many of the components of the survey measurement process, such as decisions to respond to the survey, interviewer training, and so on. However, in order to focus our discussion, we narrowed our consideration to the cognitive processes of respondents as they complete the question-answering task and the concomitant effect on the quality of the measurements. In addition, we further narrowed our focus to demographic household surveys.

We also recognized that while cognitive laboratory methods have potential to identify highly accurate questioning strategies for improving recall, understanding, and other cognitive processes of respondents, these methods often take time and may be costly to administer. In addition, we noted that despite our best attempts to design questions that respondents can answer accurately, it is not possible to remove all error from survey responses. Responses will always be subject to errors. However, we further recognized that the degree to which respondent reports are subject to error will depend on the nature of the question, the respondent's particular situation, and the respondent's cognitive abilities and efforts. For example, questions on illness are harder to answer, and, thus subject to greater error, if the recall period is longer, if the respondent has experienced many similar episodes of illness during the reference period, or if the respondent finds recall difficult. Thus, all of our survey measurements have a more or less probabilistic nature. This raises the possibility that cognitive laboratory research focused on a specific response task as well as general measures of cognitive abilities can be used to develop alternative measurement strategies that can be embedded in surveys to permit adjustment for errors associated with responses to questions.

Within this context, we felt that one of the most important next steps in the CASM movement is to address two basic research questions, namely:

1. Does cognitive laboratory research improve the questionnaires used in household surveys?

2. Can findings from cognitive science be used to construct measurement strategies that make use of multiple sources of information and models to more accurately estimate characteristics of the household population?

Assessing the Quality of Survey Responses

We recognized that in order to assess the quality of the survey responses we would need some outcome measures that were independent of those gathered during cognitive assessments. Such independent measures can include:

- Missing data/item nonresponse,

- Inconsistent data,

- Nonsubstantive responses (DK, refusal, uncodable),

- Break-off rates,

- Response variance/reinterview studies,

- Respondent debriefings questionnaires,

- Level of reporting,

- Values from external sources, and

- Distribution of responses.

---

[*]Additional members of the Working Group are: Paul Beatty, Frances Chevarley, Theresa DeMaio, Steven Lewandowsky, and Gordon Willis.

Proposed Research Projects 1: Effectiveness of Cognitive Laboratory Pretesting Methods

Survey researchers use a variety of cognitive laboratory methods including:

- Intensive cognitive interviews making use of think-alouds, probing, and paraphrasing

- Focus groups

- Interaction coding

- Expert appraisals

- Vignette classification

- Sorting and rating tasks

- Small-scale laboratory experiments

Little systematic research has been conducted on the relative effectiveness of these methods either in terms of their cost and the degree to which they allow a researcher to identify poor questions and make improvements (for a review, see Willis, DeMaio, and Harris-Kojetin, 1997). The goals of this research are to determine which methods are the best predictors of errors and identify their relative costs. We propose to evaluate questions using the following overall research design:

- Questions that were not developed using cognitive laboratory methods would be evaluated using the above methods.

- Based on the results of the laboratory research, predictions would be made as to the quality of the resulting measurements.

- Questions would not be revised.

- A field test would be conducted.

- Using one or more of the independent measures of question quality described above, the data would be analyzed to determine whether the predicted results were achieved.

Because of the need to ensure that the various cognitive laboratory evaluations are independent from each other, separate teams of researchers would conduct separate evaluations. The following table illustrates a possible design. In this design, the entire set of questions has been divided into three parts: *A*, *B*, and *C*. This would allow for the use of three teams[*] and testing of three different cognitive laboratory methods.

Cognitive Method	Research Team 1	Research Team 2	Research Team 3
Intensive cognitive interviews	A	B	C
Expert appraisal	B	C	A
Focus groups	C	A	B

Under this design, each research team uses each method and every part of the questionnaire is evaluated by three different methods.[**] An additional component of the evaluation would include examining time and cost data related to the various laboratory methods in an effort to identify the most cost-effective methods.

Proposed Research Project 2: Using Theories and Findings of Cognitive Science and Models in Estimation

The objective of this research is to examine the potential to produce model-based estimates that adjust for inaccuracies. To accomplish this, we would identify problematic question-respondent interactions using the theories and findings of cognitive science. Additionally, we would need to build models that relate the quality of responses to the characteristics of the cognitive task. Information gathered in prior laboratory studies, gathered in the survey, or gathered as a follow-up would then be used to produce the model-based estimates that adjust for inaccuracies in aggregate statistics or adjust individual reports for inaccuracy, as proposed by Groves (1997). An example of how this research could be carried out is provided below for a question on hours worked.

Sample question: How many hours do you work in an average week?

---

[*]This design would also be appropriate for comparing different cognitive research laboratories.

[**]Possible variants to this research could include using previously evaluated questions in the initial cognitive laboratory assessments or using a few questions in the field test that are constructed to be poor based on the theory that underlies the cognitive assessments.

This question would be evaluated in the laboratory research to identify sources of error. For example, the research might uncover the following:

- Use of an inaccurate heuristic which results in overestimation or underestimation.

- Failure to exclude long lunches, late arrivals, early departures, etc.

- Other exceptions, such as regularly taking Friday afternoons off.

Once laboratory research is conducted, then characteristics that might be associated with each error (e.g., self-employment, age, cognitive functioning, etc.) could be explored. Then a model would be built for a study (external data record check, diary, intensive interviewing) in which the "accurate values" of an item of interest can be determined. These data would then be used to predict the error in an estimate, and the predicted error could be used at the micro-level to adjust the estimate.

Another interesting variation of the research would be to include in the instrument general measures of cognitive functioning, such as simple digit-span tests, interference tests, recognition of famous names and faces and examine how these relate to quality of response. The types of cognitive functioning tests that might be appropriate could be constructed based on an analysis of the response task. For example, different measures of cognitive functioning might be more or less appropriate depending on whether the chief difficulties were associated with comprehension, recall, estimation, or making fine distinctions as to timing of events.

We encourage persons interested in these proposals to work collaboratively with others in government, private research organizations, and academic environments to conduct such research. Sharing ideas, resources, and talent can only enhance the quality of the final research products.

References

Groves, R.M. (1997). Survey error models and cognitive theories of response behavior. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

Willis, G., DeMaio, T., and Harris-Kojetin, B. (1997). Is the bandwagon headed for the methodological promised land? Evaluating the validity of cognitive interviewing techniques. Paper presented at the Second Advanced Seminar on the Cognitive Aspects of Survey Methodology, Charlottesville, VA.

WORKING GROUP 7[*]

The Measurement of Disability

Nancy Mathiowetz and Ron Wilson

Introduction

The measurement of disability is a particularly complex estimation issue, due to the variety of definitions of disability that exist and are used in different surveys and to the complexity of the conceptualization of disability. Estimates of the number of persons with disabilities in the United States vary significantly, depending upon the definition and the question items used to enumerate disabilities. Furthermore, terms such as impairment, disability, functional limitation, and handicap are often inconsistently used, resulting in different and conflicting estimates of prevalence. Attempts to measure not only the prevalence but also the severity of the disability further complicate the measurement process.

The complexity of the conceptualization of disability can be illustrated by an example. Disability is often defined in terms of environmental accommodation of an impairment; hence, two individuals with the same impairment may not share the same perception of the disability. For example, an elderly individual who has mobility limitations but lives in an assisted-living environment which accommodates those limitations may not see him or herself as disabled. The same individual living in a second story apartment building with no elevator may have a very different perception of the disability.

Disability data are used both to categorize individuals for the purpose of determining benefits and to provide population estimates that are used by policy analysts and decision makers. Although reliability at the individual level is desirable for both population estimates and benefits determination, it is critical when used to classify individuals as eligible for social programs. However, the assessment of the reliability of disability measures is often difficult because the effects of impairments on functioning can change over time as health status and social contexts change.

From a cognitive perspective, the measurement of disabilities offers particular challenges as well as opportunities for exploring the limits of working memory. The challenges include understanding how individuals encode information about disabilities and how the encoding of the information varies according to whether the individual is describing self-limitations or the limitations of another person, for whom the respondent may or may not be a care giver. Literature cited below suggests that the current battery of questions used to measure disability are plagued by comprehension problems. The semantic and lexical complexity of some of the measures offer opportunities for cognitive scientists to explore the limits of working memory. For instance, at what point does the question provide more information than the respondent can reasonably process?

Measures of Interest

The working group began by reviewing the concepts of "impairment," "disability," and "handicap." (We recognize that recent literature has moved away from the latter two terms to the use of such terms as "performance" and "activities" in the discussion of "disability" and terms referring to the interaction of persons with the environment instead of the term "handicap." Nevertheless, for ease of communication and because of our lack of familiarity with the new language, we continued to use the old terminology in our deliberations.) We did, however, use the World Health Organization's definitions. "Impairment" is the loss or abnormality of a physiological function or a body part or structure. "Disability" is a loss or abnormality in performance of an activity by a person (a new draft of the WHO definitions is recommending the term "activity"). "Participation" (handicap) is the nature and extent of a person's involvement in life situations resulting from the interaction of impairment or disability with environmental factors. The working group chose to focus on the concept of "disability" and examined a wide variety of measures that have been used in the decennial census, health and employment surveys, and as criteria for qualifying for social programs. These measures vary, depending on the focus of the measurement, the age

---

[*]Additional members of the Working Group are: Monroe Sirken, Roger Tourangeau, Clyde Tucker, and Barbara Wilson.

group of interest, the specificity of the disability, and the severity of the disability. For example, the specific questions associated with the measurement of activities of daily living (ADLs), a widely used set of questions among surveys of the elderly, vary from a single question covering whether anyone in the family has "difficulty" with various activities such as walking, dressing, getting out of a chair or bed, etc., to a long series of questions which focuses on a single individual, with separate questions for each type of activity and separate follow-up questions to determine the degree of assistance necessary to perform the function.

The group decided to concentrate on two measures:

- Activities of Daily Living (ADLs) and
  Instrumental Activities of Daily Living (IADLs)

- Limitations in the kind or amount of work.

These measures were chosen based on the frequency with which they are used in both health and other surveys, the diversity of forms in which they are administered, and the coverage of both the elderly, nonworking as well as employment-age population. Each of these measures has been administered as a single "screener" question as well as in a series of detailed questions. We realized that in focusing our attention on these two measures, we were, for the most part, ignoring disability measures targeted at children.

Measurement Error: Literature

Estimates of employment-related limitations and functional limitations appear to be subject to a large degree of variability, both across different survey instruments and within individuals across time. In many cases, comparisons across surveys or across time for individuals are confounded by differences in question wording, context, respondent changes, and changes in the mode of administration. For example, the 1990 Decennial Census (for the most part a self-administered instrument) included the following item:

> Does this person have a physical, mental, or other health condition that has lasted for 6 or more months and which-
>
> a. Limits the kind or amount of work this person can do at a job?
>
>   o Yes o No

The Content Reinterview Survey (CRS) (interviewer administered) included the same question. The census question was preceded by questions concerning the person's ancestry, native language, and military service;

the CRS question was preceded by several questions focusing on limitations in activities. Inconsistent responses were evident in both directions (yes to no and no to yes). Approximately one-third of the respondents who answered "yes" to the decennial census question responded "no" to the CRS question (U.S. Bureau of the Census, 1993). Mathiowetz and Lair (1994) found that significant proportions of elderly respondents reported "improvement" over time with respect to the number and severity of functional limitations as measured by activities of daily living; in multivariate models examining the probability of improvement, these improvements tended to be a function of methodological factors as opposed to substantive factors. Ofstedal et al. (1995) found similar discordance in measures of functional limitations among respondents in the Longitudinal Survey of Aging. Although functional status is a dynamic state and therefore change over time is not necessarily indicative of a "bad question," the literature cited above does suggest that the amount of change evident in these data is theoretically unexpected.

Potential Areas of Research

The research cited above and other research studies in this area do not clearly identify the source(s) of variability in individual or population estimates of disability. Among the possible sources for the differences cited above are:

- Question wording, including differences in the number of questions used to determine the existence of limitations,

- Context effects,

- Shifts in respondent (e.g., from self- to proxy or from one proxy to another proxy), and

- Mode of data collection.

In addition, examination of the wording of the individual items led the working group to raise questions as to problems of comprehension due to the use of inherently vague terms (e.g., "difficulty" bathing), the perceived intent of the question (does bathing mean getting in and out of the bathtub or the ability to keep oneself clean), as well as the potential limits of working memory when respondents are asked to consider whether anyone in the family is limited in one or more of several functional areas.

Three general areas of future research were proposed by the group. These areas included:

- Determining the source of variability in responses to both the work limitation and functional limitation questions;

- Examining comprehension difficulties concerning the measurement of employment related and functional limitations; and

- Exploring limits of working memory using these questions.

Understanding Sources of Variability

The working group proposed three avenues of research to identify the source or sources of variability in responses over time to these disability measures. These included:

1. Meta-analytic review of the literature and further analysis of current data. Several federally funded longitudinal studies include one or both of the measures of interest, including the National Health Interview Survey, the Longitudinal Survey of Aging, the Survey of Income and Program Participation, the Medical Expenditure Panel Survey, and the Health and Retirement Survey. To fully understand the factors that affect the measurement of disability, we need to document when the variability in estimates became evident.

2. Take advantage of current programs (e.g., reinterview programs) to test under what conditions variation in responses arises and to determine whether variation exists when mode, respondent, and context are controlled.

3. New experimentation. The comparison of the census and content reinterview survey is confounded by changes in mode of data collection, context, and possibly, changes in respondent. A simple split-ballot experiment in which only context is changed could address the context effect question.

Comprehension

The working group proposed the use of various cognitive methods to examine comprehension difficulties with disability measures, including the use of paraphrasing, vignettes, and think-aloud interviews.

Working memory limitations

Measures of disability, especially single screener-type measures, appear to press the limits of working memory. These items could be used as test items to determine the capacity for processing at multiple levels (e.g., across members of the family and across different functional areas).

The group recognized the importance of conducting basic research on the cognitive aspects of measuring disability, and therefore, advocated undertaking research projects that are conducted as interdisciplinary research with collaboration between survey methodologists and cognitive scientists.

Group Discussion

Following presentation of the working group proposal at the end of the CASM II Seminar, the following items were discussed:

- *Working Memory.* Almost all disability related questions appear to have started as a series of questions in which the focus for each question was a single area of limitation concerning a particular household member. Over time, these items have often been collapsed for one of two reasons:

  1. Parsimony to save time and interviewing costs. In general population surveys, the frequency of "yes" responses for any one of these items tends to be less than 5 percent; hence as a time-saving measure, instrument designers have compressed the items both vertically and horizontally, focusing on family-level questions which encompass multiple limitations.

  2. Fear over conditioning effects in surveys in which data are collected for multiple members of the family. Here the concern is that once the respondent has responded to a set of items for one person in the household he or she learns that a "yes" response will result in additional questions.

- *Defining the construct of interest.* Several participants noted the importance of defining the construct of interest, especially for these measures where there are many different approaches to the measurement process leading to estimates of very different constructs. The specific comments included:

  1. The selection of the appropriate construct(s) for a specific survey or study will depend on the purposes for which the results are to be used.

The constructs should be made explicit at the beginning.

2. If the research is relevant to disability benefits or the rights of disabled persons, representatives of the eligible population should be consulted about selection of the construct(s) to be used.

3. A clear statement of the constructs and measurement instruments used in the research should be given to data users.

References

Mathiowetz, N. and Lair, T. (1994). Getting better? Change or error in the measurement of functional limitations. *Journal of Economic and Social Measurement, 20*, 237-262.

Ofstedal, M.B., Lentzner, H., and Weeks, J.D. (1996). Evaluating alternative ways to measure change in functioning: Subjective versus objective indicators in the longitudinal study of aging. In R. Warnecke (Ed.) *Health survey research methods: Conference proceedings* (pp. 9-14), Hyattsville, MD: Public Health Service.

U.S. Bureau of the Census (1993). *Content reinterview survey: Accuracy of data for selected population and housing characteristics as measured by reinterview (Evaluation and Research Reports, CPH-E-1).* Washington, DC: Department of Commerce.

WORKING GROUP 8[*]

Adapting Cognitive Techniques to Establishment Surveys

David Cantor and Polly Phipps

Introduction

Data collection in establishment surveys, conducted predominantly by federal statistical agencies, has two separate historical streams. The first is the use of paper forms for regulatory and statistical purposes, recording the inputs, outputs, and prices of units in many sectors of the economy. This is a tradition that largely does not use verbal questions to elicit data, but instead uses data labels (e.g., "total employees," "total retail sales in March," "square footage of building"). The second is the use of subject matter specialists to obtain data by interviewing business and farm operators. Often these substantive experts record responses on a form that contains no questions. The expert is free to formulate questions to elicit the data in ways that fit the expertise and production processes of the sample firm. For both of these reasons, the measurement process in establishment surveys is often quite different from that in household surveys.

There are characteristics of establishment surveys that are important to consider when developing a research agenda. One revolves around the issue of *who is authorized to approve participation in the study and who can best provide the information*. The person who maintains and uses the record system is typically the best person to provide the data. However, the record keeper may not be the person who can authorize release of information about the company and may have to ask someone at a higher level of authority. The *use of records* to respond in an establishment survey affects the response process. Rather than relying on purely cognitive processes to provide an answer, as is the case in most household surveys, an establishment survey respondent has to develop the response within the context of his/her record system (e.g., see Edwards and Cantor, 1991). The format and level of sophistication of the record system is likely to vary greatly by such factors as the size and industry classification of the organization. Consequently, there may be a greater need for an establishment survey to tailor the response task to the size and complexity of the business.

Related to issues of the record keeping system, establishment surveys *involve many technical terms* that may not be appropriate on a household survey. The establishment survey respondent is typically a record keeper who understands much of the jargon that might be used by the designers of the survey. However, technical terms may mean quite different things across organizations. Consequently it is important for the survey designer to keep the definitions of these terms as clear as possible. For surveys that use sample selection with probability proportionate to the number of employees, there will be large establishments that are *certainty selections*. For continuing government surveys (e.g., Current Employment Statistics program; Occupational Safety and Health Survey), this means that some organizations are permanently part of the survey and are continually asked to respond to other surveys as well. The approach to obtaining responses from these firms may be quite different from that for smaller organizations that may only be surveyed occasionally.

Respondents to many establishment surveys *may be users of the information* produced by the data collection. For example, businesses may be particularly interested in surveys that provide information on monthly economic conditions, especially in a local area, to assist in business planning. This interest may affect both their willingness to respond and the type of data that are chosen when responding. A final factor that can affect the response process is *whether the question is factual or attitudinal* (e.g., based on record information vs. an opinion on the effects of a particular law on administrative burden). Both the selection of an appropriate respondent and the processes used in the response to a survey request may be quite different, depending on the type of question.

The development of a research agenda for establishment surveys requires the application of several theoretical streams. This is where the use of the principles developed by CASM I is especially relevant. There is clearly a need to use not only theories of cognitive psychology, but also principles that explain how organizations, such as business establishments, function. Cognitive psychology is relevant to how the

---

individual respondent may process the information request and ultimately provide the information. However, organizational theories are also important in understanding the response process. These theories would provide information on how the level of complexity of the organization (vertical and horizontal) might affect the way in which a survey request is handled. Relevant factors might include, among other things, how information is communicated within the organization and the relationship of the organization to external bodies (e.g., the government).

The following four proposals suggest a broad range of research topics to be explored for establishment surveys. Research on these topics will potentially enrich the body of knowledge related to the survey process for establishment surveys. They deal with interactions of individuals when they are agents of an organization or establishment rather than representatives of their households, families, or themselves. To limit research on cognitive processes only to reports on information about individuals and individual behavior would restrict research to a very narrow set of information. Many areas of concern touched on at CASM II, such as conversational norms and the use of ethnography, are also fruitful areas of research within the establishment survey context.

An important concept to keep in the forefront is that establishment surveys, which form the basis of many of the key indicators of our economy (such as employment, inflation, natural resource and agricultural crop availability), involve a complicated stream of communication. Despite this, however, the design of establishment surveys has received relatively little attention from survey researchers and may have the greatest potential for improvement through reduction of nonresponse and measurement errors.

Research Program Idea 1: Authority Versus Capability

The first proposed area of study focuses on survey participation and response quality when one actor grants access to participate and another provides survey data.

Motivation

As organizations grow, individual roles become more specialized. It is common in establishment surveys that the knowledge or records containing data relevant to survey items are located at a different level of the establishment than the level with authority to grant access to establishment information. Those with authority to grant access may be unaware of whether the information sought exists in the establishment.

Hence, it is not clear whether entry into the sample establishment should be at the level where the information is held or where those with the power to act on the request are located. Complicating this further is the fact that some organizations divide record systems into specialty units (e.g., personnel, payroll, shipping, sales, accounting), so relevant survey data may lie in multiple systems.

Since total survey quality is a function both of nonresponse and measurement error, establishment survey designers have little guidance in measuring the error trade-offs of alternative strategies.

Research Questions

- What organization structures and cultures lead to greater empowerment of actors at the level where the data are held?

- What features of the survey request are most salient to top managers (chief counsels, heads of public relations), who have the authority to grant survey requests?

- Are there any characteristics of the establishment that signal its propensity to participate in the survey? If so, could protocols be developed that cater to the individual participation propensities and thus maximize the chance of participation?

- Once permission has been obtained, what processes are most effective in gaining commitment of the designated reporter in the establishment?

- Since establishment record systems are observable, what survey procedures produce high measurement quality and high respondent commitment, by way of the survey takers' examination of the record system?

- In longitudinal surveys, what feedback to the original authority who granted access is desirable as a means of maintaining high response rates and measurement quality?

- In longitudinal surveys, what feedback to the designated reporter is desirable for response rate and measurement quality?

Research Methods

The key challenge is the identification of design principles that apply to large classes of organizations and survey topics. Therefore, some importation or invention of theories is necessary:

1. *Collaboration Between Organization Theorists and Survey Methodologists*

Some initial work would focus on identifying principles or key organization attributes from current organizational theories that predict the location of databases in the establishment and the distribution of authority to provide data to an external requestor. We would expect that establishment size would be a key attribute, as well as the nature of links between the target establishments and others (through loose alliances, contractual relationships for data storage, corporations, etc.). This research might succeed in identifying typologies based on size and structure that could be used to mount survey designs that would be more effective at gaining participation or providing accurate data.

2. *Experiments with Alternative Approaches to Sample Establishments*

The dependent variable would be participation rates to surveys on different topics. Experimental treatments might involve varying the person or unit in the organization who serves as the initial contact (e.g., establishment manager, person designated for external relations, legal department, data manager, or database user unit). The hypotheses to be tested would concern the relationship between the role of the respondent within the organization and the decision to participate in the survey.

Research Program Idea 2: Implications of Using Expert Interviewers

The second proposed area of study is conversational pragmatics between expert interviewers and expert respondents on technical survey data.

Motivation

Establishment survey interviewing may employ "expert interviewers," trained to assist respondents in providing data as desired by the survey. Interviewers often have subject matter training, are versed in alternative record systems used to store relevant data, and are empowered to obtain the survey data in ways they judge best for each sample unit.

Most establishment surveys obtain quantitative data on technical matters involved in running an economic unit (i.e., a business, a corporation). The data sought are often key to operations of the unit. However, they may be classified in a manner not identical with survey needs, may not be updated on a schedule needed by the survey, or may be aggregated at a level unsuited to the survey.

Some survey items request expert judgements from the establishment staff, judgements that may need definitional refinements from the interviewer.

For all of these reasons, the interactional style of establishment interviews appears to be quite distinctive from that of most household interviews.

Research Questions

- What is the process of establishing common ground on technical issues in such interviews?

- What conversational markers are used by interviewers to intervene with additional technical definitions, to seek further information from the respondent regarding informational resources he/she has?

- What questions are posed by expert respondents to reveal mismatches between records available and survey needs?

- What are the consequences of failure of expert interviewers to resolve such mismatches? How do respondents cope with inadequate information?

- What short-hand communicative styles exist on technical matters once the common ground is established?

- Is the process of verification of receipt of information in expert exchanges similar to those in other exchanges?

- Does a more collaborative style of data acquisition exist for such exchanges, and if so, does it increase data quality?

- What level of technical knowledge is necessary for interviewers?

- In what survey and interaction types are expert interviewers most necessary?

- What are the cost implications of using expert interviewers?

Research Methods

The research is probably best mounted in at least three phases.

1. Qualitative Preliminary Work

The initial research would be a set of qualitative investigations aimed at mapping the conversational patterns of such expert interchanges. The focus of these investigations would be developing hypotheses about how common ground is established and what conditions might cause conversational breakdowns. The ultimate dependent variable in this research is the quality of the responses provided by the respondent. It would be assessed through intensive reinterviews with respondents, with the assistance of their records.

2. Experiments on Small Samples in the Field

We believe an experimental phase is necessary to identify the components of expert interaction that influence accurate reporting by the respondent. These experiments would examine a component of the conversational interaction from the first step investigations and selectively omit it from or alter its form in the conversation. We expect that this research will differ from much prior CASM work because it should sometimes use the expert interviewers as subjects (e.g., to examine what utterances by respondents communicate the need for use of interviewer technical knowledge for clarification of the respondent task), and sometimes use the expert respondents as subjects (e.g., to examine what prompts them to reveal technical mismatches between their records and their requested data).

3. Implementation and Evaluation of Interviewing Protocols

One current weakness in expert interviewing as presently conducted in establishment surveys is that it has not been codified; indeed, to some it is not even definable at this point. The final step in the research program would be the identification of: (a) technical information useful under different circumstances in an interview; (b) cues provided by respondents or by the nature of their record systems that indicate the need for the application of that knowledge; and (c) the most effective methods for expert interviewers to apply that technical knowledge. The findings would be used to develop an interviewing protocol and eventually training materials. This last step would include field experiments in survey settings to assess alternative features of the interviewing protocol.

Research Program Idea 3: Record-Assisted Retrieval

Motivation

The purpose of this research is to investigate two components of the "cognitive" model for establishment surveys: records formation/encoding and information retrieval processes. Little is known about records content and how respondents do or do not use records in answering establishment survey questions. Earlier work using response-analysis surveys focused on identifying errors in survey answers, with lesser emphasis on the source of the errors, for example, whether records data were available or not, mismatches between survey definitions and records content, respondent comprehension errors, use of records whether they do or do not fit survey definitions, or respondent estimation in place of or in coordination with records use.

Research Questions

Basic questions in the encoding/records formation processes include: What is in the records and how does the information fit survey definitions? Basic questions in information retrieval include whether and how records are used in the survey answering process, the respondent's relation to the records system, and circumstances under which records, estimation, or a combination may be preferred.

Research Methods

We envision carrying out site visits with establishment respondents, and at a later point, possibly, experiments where the match between records and survey definition is explored. When the two do not mesh, we are interested in how respondents formulate answers, for example, does a respondent simply report the records information regardless of definitions, leave the item blank, or does he or she use some type of estimation to account for the mismatch. If respondents use an arithmetic formula to adjust for differences, perhaps that algorithm should be incorporated within a survey estimation system.

Research Proposal Idea 4: Willingness of an Establishment to Participate and Implications for Nonresponse and Data Quality

Motivation

There is very little information about operational design features that may affect an organization's willingness to participate in a survey. The purpose of this research would be to investigate the extent to which response rates and data quality may be affected by a variety of methods that could be used to increase survey participation. These methods would be based on conceptual models of the response process within an organizational survey. The research would draw from both organizational theory to provide information on how survey requests are handled, and social-psychological theories to explain why an individual may or may not respond within an organizational context.

Research Questions

The specific issues to be examined would include: (1) the importance of emphasizing the utility of the survey; (2) the design of the data collection form; (3) the use of incentives; and (4) the involvement of the interviewer in the data collection (e.g., mail vs. telephone).

Research Method

The research would proceed in several stages. The first stage would be a series of interviews with previous survey respondents and nonrespondents. These interviews would explore issues such as how the organization typically handles survey requests and factors that surround decisions to participate (both organizationally, as well as individually). The second stage of the research would be a series of experiments that varied design features found to be important in the first stage of the research. These features might include: (1) use of prenotification; (2) emphasis on the utility of the survey; (3) use of incentives for the respondent (i.e., the person who eventually provides the data for the survey); and (4) the mode of interview. The primary outcome variables would be the response rates across treatments and the accuracy of the information provided. The accuracy of the responses would be measured in several ways. One would be to visit a subset of the organizations and conduct an in-person interview that reconstructed the retrieval process used to arrive at the survey response (e.g., was it taken directly from records, was it adapted, was it estimated without records). The second measure of accuracy could be taken from a response-analysis survey that asked respondents about the processes they used to provide the information (e.g., how were records used; did they read the instructions; etc.).

References

Edwards, W.S., and Cantor, D. (1991). Toward a response model in establishment surveys. In P. Biemer, R. Groves, G. Forsman, L. Lyberg, N. Mathiowetz, and S. Sudman, (Eds.), *Measurement error in surveys* (pp. 211-236). New York: John Wiley & Sons.

CONCLUDING SESSION

Models of the Survey Process

Norman Bradburn

I am now doing two things that I promised myself I would never do. One is to be the last person on a multiday program, and the other is to try to summarize what has gone on in a very complex set of presentations and discussions. It is an impossible task, but I will do what I can.

I am going to start by elaborating on a model that is slightly different from the one Colm attributed to me in his presentation on the first day (O'Muircheartaigh, Paper Session 1), a model that was developed prior to the first CASM Seminar (see Exhibit 1). The one I show here is not the traditional version. It is a social model rather than a cognitive model. Our idea at the time was that there was a micro-social system consisting of the interviewer and the respondent. One thing Colm did not mention in his model, but an important thing, I think, was that the task was specified by the person called the investigator. That is an unusual situation in the sense that a social interaction is conducted in a context determined by a third party who is not present. Translation obviously is going to be a problem. Also included in my model are environment and culture, because we did have some awareness of their possible effects, although we did not talk about them. The notion is that not only are things going on in this little micro-environment, but the norms that govern this interaction are to some extent drawn from the larger culture.

What CASM I did, I think, was to elaborate the notion of the cognitive nature of the task. What we did in the book that Sudman and I wrote on response errors (Sudman and Bradburn, 1974) was to think about the interview more as a social situation, without considering the cognitive demands of the task, although we did talk a little about memory problems. The CASM I Seminar elucidated, from the point of view of communication and internal cognitive processes, what the respondent is asked to go through to answer the interviewer's questions. That was the major contribution of the first CASM Seminar.

Now, the question is what have we added here at CASM II and where are we going (see Exhibit 2)? One thing, it seems to me, is that very early on we elaborated on the difficulty of communicating meaning, both the meaning intended by the investigator and what the task is all about. Comprehension of the question is one of the main problems. The original model treated the

investigator as if that were the person who designed the questionnaire. Now we talk about the questionnaire designer as independent of the investigator, which is consistent with what Colm said about Taylorization, that is, there is a greater division of labor among the participants in the process, and communication among them becomes increasingly diffused. There is often a large gap, at least in government surveys and many other places, between the investigator and the questionnaire designer. Constructs may not be very well specified initially or they may simply grow vague as they undergo the process of being translated into measures. There is more role specification and division of labor, especially as computers play a greater role—I'll come back to that shortly. A pervasive, underlying assumption is that there is some minimum sharing of meaning across respondents and investigators. The underlying assumption of survey measurement is that there is sufficient common understanding of the constructs being investigated that you can obtain measures relevant to these constructs. This is one step that it seems to me we have added.

Another thing that has emerged repeatedly throughout the conference is a greater emphasis on the interviewer's role in the data collection process. Until recently what I call the "robotic theory" of the interviewer has prevailed. That is, the interviewer's role has been narrowly constrained in order to reduce variability in interviewer performance. The new approach views the interviewing task as a conversation and then asks what are the characteristics of a conversation? One of them is that it tends to be a dialogue. The logic of the conversational nature of the interview leads to various proposals to empower respondents to ask questions and interviewers to answer them in some sensible way. Of course interviewers need to be trained (at some additional cost) to respond in ways that are helpful for improving the quality of data rather than hurting it. The point is to make the interviewer a real participant in the conversation rather than someone who simply relays verbatim what the investigator wants.

A third point (see Exhibit 3) introduces a new actor into the model, the computer. The computerization of much of the process has added substantial complexity and a whole new set of elements into the system, with mixed results. We have talked about the many positive things that the computer can do, but there has also been

discussion about the computer controlling the situation in ways that we do not really like, perhaps in fact making the interviewer's job harder in some respects. Computer-assisted interviewing has tempted investigators (and they yield to that temptation very easily) to enlarge and complicate the questionnaire and to demand responses that may well go beyond respondents' cognitive capacities. The computer enables checking and routing to contingent questions in a way that interviewers have a great deal of trouble doing on their own. One of the questionnaires that NORC developed for computer-assisted interviewing could not be administered with a paper-and-pencil version—it has become so complicated you could not possibly do it without a computer. We expressed some concerns about the computer taking control of the task. Have we been lured into situations where we are now so dependent on computers and computer systems that as investigators we are in danger of losing control?

There were mixed messages about the value of computers, ranging from the highly skeptical to much more positive views about how they might be used. Obviously we need to give much more attention to the human-computer interaction. Some of the problems we are facing now are passing problems and can be solved. We are only now beginning to understand that we are doing something radically different by using computers—it is not just that they are helping us do something better, quicker, or more reliably than before. Actually whether it is quicker or not depends on what part of the process you think about. Those familiar with computer applications in interviews know that the length of time it takes to get a new questionnaire up on the computer is much greater than that needed to produce a paper-and-pencil version. That is a very big addition to the model, and it has its pluses and minuses. The question of whether computer-assisted interviewing aids or subverts the movement toward more flexible interviewers is unresolved at the moment. I think that in the short run it may be going in the opposite direction, but as people have mentioned, there are a lot of things one can do to change this. And yet I feel that some survey researchers envision the use of the computer as a means of eventually eliminating the interviewer. There seems to be an odd subtext in all of these technological cycles either to eliminate or to downplay the role of the interviewer. We are in the brief period when we are trying to upgrade the role of the interviewer. Let's grab onto that. Interviewers play a crucial role in the survey process and should not be ignored.

The fourth item, one that doesn't change the model so much, is the realization or articulation of the fact that questionnaires are usually designed for generic respondents. Interviews, on the other hand, are for specific respondents, which raises the possibility or desirability of adapting questionnaires for individuals or groups of individuals. Several of the research projects suggested by the working groups involve the possibility

of not having a single standardized questionnaire for everybody. Questionnaires or questionnaire modules can be tailored to different groups. We have talked for a long time about adapting the wording of questions to particular individuals; now we are thinking about changing not just the wording of some questions but maybe whole blocks of questions or whole questionnaires. The difference here, of course, is that the computer would enable having alternative versions of the questionnaire that could be called up as needed. Some obvious problems that we recognize are the screening and routing problems, that is, how do you design the up-front stuff to get the right questions to the right individuals? These are practical problems we can work on.

Another theme that affects questionnaires, although this is not something new, is the notion that mental representation of events is a constructed process. We have heard about new theories of memory—how memories are stored, how they are retrieved, and so forth. Whether these connectionist theories will really hold up and what their implications are for the way we design questionnaires is not clear yet, but I think it is something we will want to pursue.

I want to take up one theme which was not elaborated on very much, although we touched on it while discussing the report of the Working Group on the Measurement of Disability. It is the assumption that there are constructs whose meanings are sufficiently commonly understood that we can actually measure them. It seems to me that there are some constructs whose meanings are not widely shared, at least not shared enough to measure them. What we are doing with such constructs is different from what we ordinarily think of in terms of other kinds of measures. Along the lines of what Betsy Martin said, there may be indicators like disability that have political meaning or other kinds of meaning that should be thought about in different ways from the way we think about measuring income. Eleanor Gerber mentioned ethnicity. Ethnicity may not be a concept or it may be a bundle of concepts that are so intertwined and so different in the way different people perceive them that it really is not possible to measure them in any consistent sense—you have to do something else. Basically, you have to fall back on some kind of arbitrary definition, negotiated with those who will be using or affected by the data. One thing we didn't talk about, that might be worth thinking about, is distinguishing between how well something is measured in terms of how the data are obtained and the form in which the data are reported. A major issue in the revision of the OMB directive on reporting race and ethnicity was: What is the minimum number of categories one has to have to report race/ethnicity? That is a different question from the way we ask about race and ethnicity or the way we get the data. This is an important distinction for constructs that are particularly hard to measure (see Exhibit 4).

The final thread to add is the potential effect on data analysis that arises from modeling measurement errors on the basis of social science theory, using measures of the quality of reporting that go beyond dichotomous measures of response and nonresponse. Estimators could be based on the recognition that not all reports are equally good and adjustments could be made either for particular kinds of variables or even for particular kinds of respondents. Such adjustments are analogous to current practice in adjusting for unit and item nonresponse, based on some social theory about how the report was generated or the characteristics of the reporter. For other parts of the survey process, analysis and reporting, we discussed how the choice of graphical and other methods of presentation are affected by the intended audience and goals. Much of what was said addressed the importance of graphical methods for analysts, giving them a greater ability to detect complex patterns.

I should note that I have now separated analysis from data collection in the model. We have talked about the application of cognitive methods in all phases of data collection, including not only the interview, but subsequent processing—editing, coding, and the other steps prior to reporting and analysis. Thus, we have moved away from a fairly simplistic social model that included only an interviewer and respondent to a much more complicated model in which we introduced the computer as a new and almost independent actor. We mentioned the problem of ensuring that the person(s) who are programming the computer understand the constructs and goals that are communicated to them so that they work in harmony with the ultimate goals of the survey. We now have a framework that illustrates the need for collaboration among people with a wide variety of interests, knowledge, and skills.

Reference

Sudman, S. and Bradburn, N. (1974). *Response effects in surveys: A review and synthesis.* Chicago: Aldine.

EXHIBIT 1. PRE-CASM

EXHIBIT 2. NEW ELEMENTS


```
┌─────────────────────────────┐
│        INVESTIGATOR         │
├─────────────────────────────┤        ┌─────────────┐
│   QUESTIONNAIRE DESIGNER    │───────▶│  CONSTRUCTS │
└─────────────────────────────┘        └─────────────┘
                                              │
                                              ▼
        ┌──────────────┐              ┌─────────────┐
        │     DATA     │◀─────────────│  MEASURES   │
        │  COLLECTION  │              └─────────────┘
        └──────────────┘
              │
        ┌──────────┐
        │   TASK   │
        └──────────┘
         CONVERSATION

┌──────────────────┐                   ┌──────────────────┐
│   INTERVIEWER    │                   │   RESPONDENT     │
└──────────────────┘                   └──────────────────┘
```

ENVIRONMENT/CULTURE
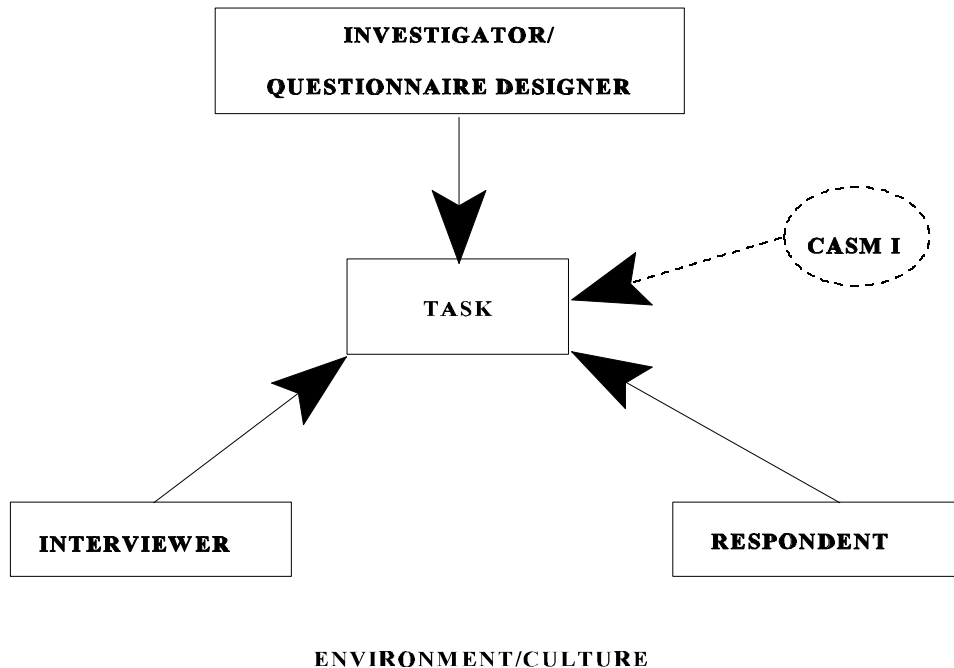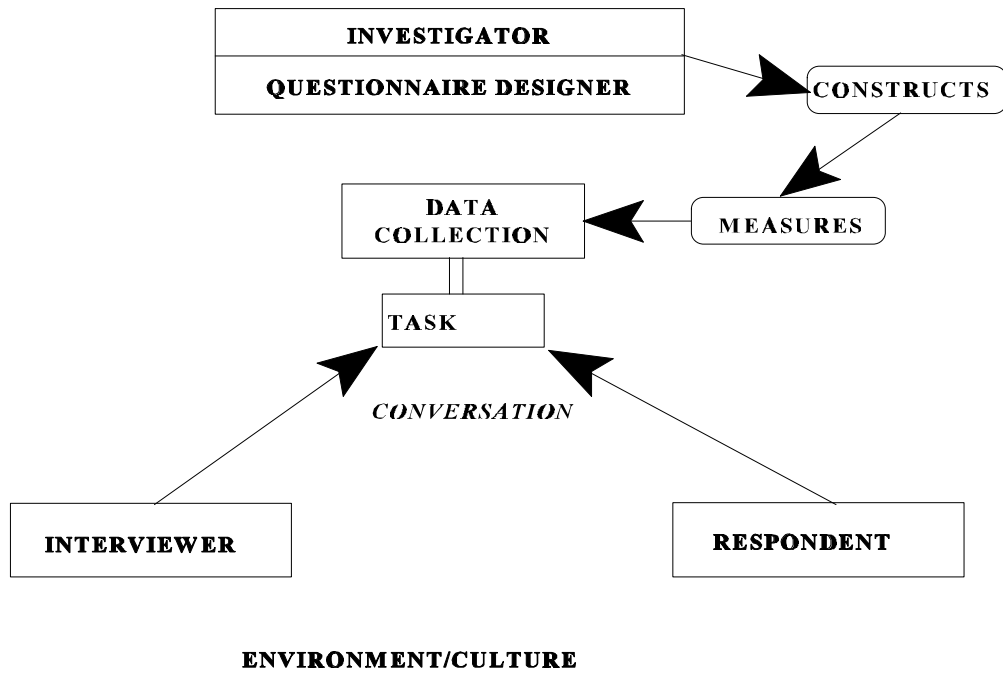
EXHIBIT 3. ENTER THE COMPUTER


```
┌─────────────────────────────┐
│        INVESTIGATOR         │
├─────────────────────────────┤        ┌─────────────┐
│   QUESTIONNAIRE DESIGNER    │───────▶│  CONSTRUCTS │
└─────────────────────────────┘        └─────────────┘
                                              │
                                              ▼
        ┌──────────────┐              ┌─────────────┐
        │     DATA     │◀─────────────│  MEASURES   │
        │  COLLECTION  │              └─────────────┘
        └──────────────┘
   ┌──────────────┐   ┌──────────┐
   │   COMPUTER   │──▶│   TASK   │
   └──────────────┘   └──────────┘
   ┌──────────────┐              ┌──────────────┐
   │ INTERVIEWER  │◀────────────▶│  RESPONDENT  │
   └──────────────┘              └──────────────┘
```

ENVIRONMENT/CULTURE
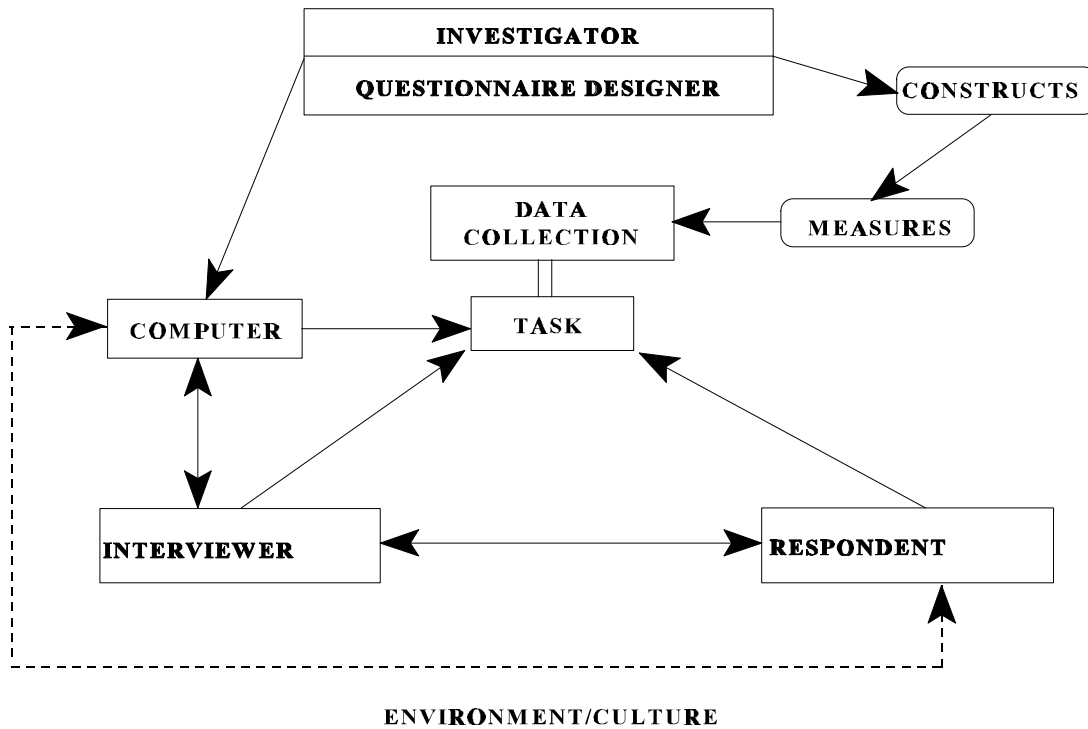
EXHIBIT 4. INSTITUTIONAL CULTURE

```
DISSEMINATION            ┌─────────────────────────────┐
                         │        INVESTIGATOR         │──────►  CONSTRUCTS
    ▲                    │                             │
    │                    │   QUESTIONNAIRE DESIGNER    │
  ANALYSIS ─────────────►└─────────────────────────────┘              │
    ▲                                                                  ▼
    │         ┌──────────────┐                              ┌────────────────┐
    └─────────│     DATA     │◄─────────────────────────────│    MEASURES    │
              │  COLLECTION  │
 Multi-media  └──────────────┘
 (Aesthetics)       │
  ┌───────────┐  ┌───────┐
  │ COMPUTER  │─►│ TASK  │
  └───────────┘  └───────┘
    ▲                ▲        ▲
    │                │        │
    ▼                │        │
┌──────────────┐  ┌──────────────┐
│ INTERVIEWER  │◄►│  RESPONDENT  │
└──────────────┘  └──────────────┘
                         ▲
         (dashed line)   │

           ENVIRONMENT/CULTURE
```

ENVIRONMENT/CULTURE
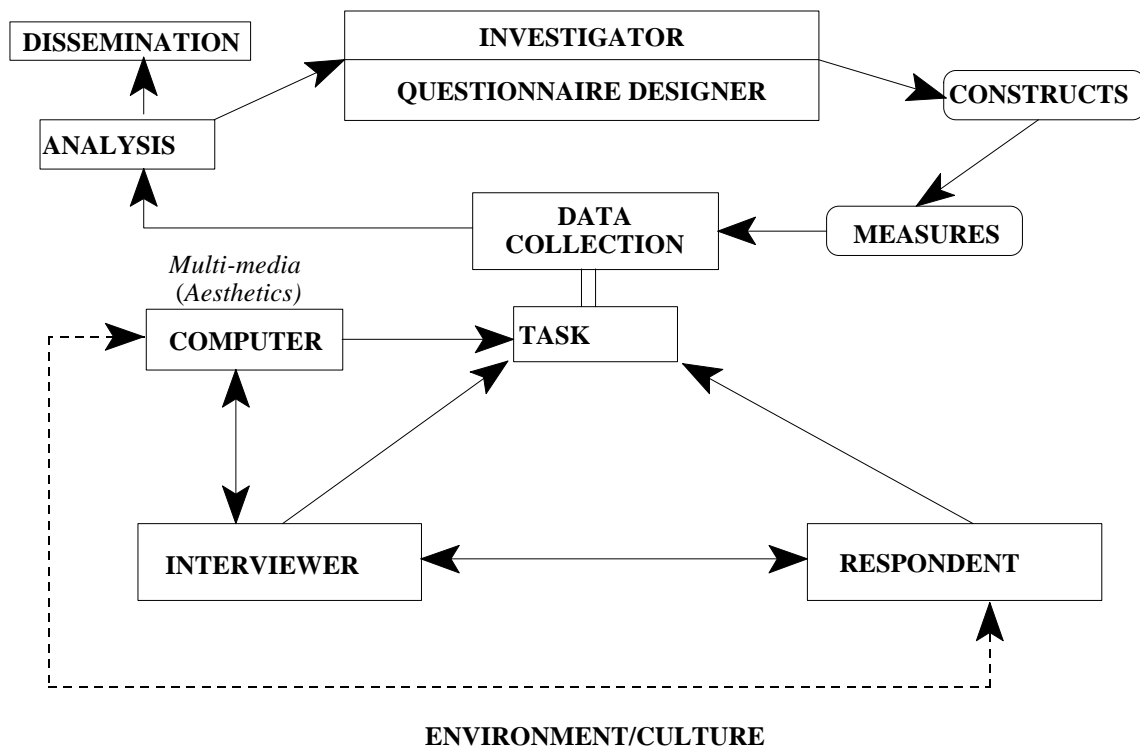
SUMMARY DISCUSSION

Thomas B. Jabine

This section summarizes the discussion generated by Norman Bradburn's overview of the themes that emerged at the seminar. Participants who had postseminar "afterthoughts" were encouraged to record and submit them to the seminar organizers. This summary is based on comments made at the session or sent in subsequently. It focuses on three areas: evaluation of what the cognitive laboratories at BLS, Census, and NCHS have accomplished to date; recent and potential future changes in the environment in which surveys are conducted; and ideas for future research.

The Cognitive Laboratories

A major outcome of the CASM I Seminar was the establishment of cognitive laboratories at BLS, NCHS, and the Census Bureau. Several participants commented on the missions, organizational arrangements, and accomplishments of the labs. So far, a major part of their work has been to apply cognitive tools in the development or redesign of questionnaires for major surveys, especially household and other demographic surveys. As one participant put it, an important result of CASM has been to bring questionnaire designers closer to the level of respondents—we now pay attention to how respondents think about the topics covered by our surveys. Economic surveys have received less attention, as have other aspects of surveys, such as interviewer training. For some surveys (and the 1990 census), the labs have conducted ethnographic and cognitive research that has influenced the formulation of basic constructs like residence, labor force status, race, and ethnicity. The labs have also undertaken or sponsored basic research on cognitive processes in surveys, but at present have limited resources for this purpose. One participant urged that the labs expand their missions to cover experimentation, validation studies, and analysis. Other comments focused on the relationship between cognitive labs and the survey organizations they serve. Not surprisingly, recommendations for changes in constructs and questionnaire items are not always accepted. Working relationships with subject matter specialists in sponsoring agencies can be particularly problematic.

Changes in the Survey Environment

As described by Bradburn in his summary review, new technologies are changing the nature of the cognitive processes that occur in surveys. Computer-assisted interviews can be conducted in several different modes: in-person with the interviewer using a laptop, by telephone, or in a self-administered format. These modes have many widely advertised advantages over paper-and-pencil interviews. One cognitive scientist suggested that a laptop computer would be an excellent vehicle for recently developed "priming" methods which offer the possibility of more reliable indicators of prejudices and stereotypes by measuring the relative rapidity of response between "prime" and "target" words. But there are also some potential problems with the new technologies. Telephone interviews cannot use visual aids, unless they are distributed in advance. New telephone features like caller ID may have negative consequences for response to telephone surveys. The introduction of computers in survey interviews has changed the two-way interaction between interviewers and respondents into a more complex three-way interaction that now includes the computer. Questionnaire developers sometimes find that the limitations of the authoring systems used for CATI and CAPI questionnaires place constraints on their ability to implement changes needed to resolve problems identified in cognitive testing.

Although definitive measures are lacking, there are some indications of a decline in the willingness of people to participate in surveys and in their confidence in the validity and utility of survey results. Several factors may be implicated in such trends: general distrust of government, politicians, and "big business"; growing concerns about individual privacy; and the increasing complexity and burden imposed by survey interviews, especially in longitudinal surveys. Increasing difficulties in conducting censuses and surveys have led to proposals for greater use of administrative records to provide information needed for the formulation and evaluation of social and economic policies.

Proposed Topics for Research

More complete coverage of the research proposals developed at the seminar is provided by the working group reports that preceded this summary. The topics presented here were mentioned by participants in the final session of the CASM II Seminar or subsequently submitted as afterthoughts.

Standardized Versus Conversational Interviewing

Several participants favored further experimentation with greater use of conversational styles in survey interviews covering behavioral and other factual topics (see Conrad's paper for one example). Interviewers would receive training on the survey concepts and definitions so that they would be able to provide clarification when requested by respondents. The survey interview would be treated as an interactive cognitive processing task which attempts to achieve maximum understanding between interviewers, respondents and investigators. Experimentation would help define and describe the differences between standardized and conversational interview formats. One participant mentioned that there could be feedback to survey designers in surveys conducted via the Internet. Another suggested that the computer should be regarded as a partner in the survey process and that surveys could be made more enjoyable by the use of "aesthetically seductive" multimedia features.

Broad Social Aspects of Survey Research

One participant urged a broad examination of the social framework in which surveys take place and the views of the public about surveys, as a basis for developing efforts to promote greater participation in surveys. Another proposed a program of periodic ethnographic studies and focus groups for similar purposes (see also the reports of Working Groups 3 and 5).

A Computer Aid for Questionnaire Designers

Graesser expressed his intention to continue work on development of a computer aid for use in evaluating survey questions, along the lines described in his paper (Paper Session 3). Although some expressed skepticism about how well such a system would be able to detect problems, another participant believed that such an aid might be of some value to the many organizations that undertake surveys without any access to sophisticated questionnaire development staff and facilities.

Model-Based Analysis of Surveys

A participant supported research on the incorporation, in survey estimates, of information about known or measurable response errors, as proposed by Groves (Paper Session 3). The same participant called for more research on the causes of incorrect reporting, including poor cognitive facilities, poor record keeping, and deliberate misreporting.

Brain Scanning

A participant wondered whether, in the future, brain scanning techniques might help to understand the cognitive processes of survey interviewers and respondents. This was considered unlikely by another participant, who said that brain-scanning techniques were not well developed and would not be able to track the complex cognitive processes that occur in surveys.

Data Preparation and Processing

Although the presentations at CASM II went well beyond the realm of cognitive psychology and the data collection phase of the survey, one participant felt that more attention should be given to the steps that occur between data collection and data dissemination. Procedures for manual and computer-assisted editing of individual records and prepublication review of aggregate data could benefit from collaborative research with many of the disciplines represented at the seminar. These processes can have a significant effect on the quality of survey results and, as noted by one person, in many surveys they account for a substantial fraction of total costs.

CURRENT ISSUES AND FUTURE DIRECTIONS

A Summary of Research Proposed at CASM II

Gordon Willis, Monroe G. Sirken, and Thomas B. Jabine

Introduction

A major goal of the CASM II Seminar was to identify promising research avenues for the future. Participants articulated a number of ideas, largely structured according to the themes of the four commissioned paper sessions, and of the Working Groups (see Tucker and Martin's Working Group Introduction). However, the total set of research proposals developed, either formally or informally, span a wider spectrum than the scope implied by the conference's formal organization. Therefore, this section lists the set of proposals for potential research directions that emerged through a more thorough review of the Proceedings materials. We also discuss several summary themes that underlie, or are implied by, the various proposals. The goal of this section is to identify the broad issues that will likely impact CASM research initiated beyond the seminar.

From our review of the commissioned papers, Working Group reports, and discussions contained in this Proceedings, we identified 24 proposed topics for CASM research:

1. Evaluation of alternative methods for identifying problematic survey questions and for making improvements to those questions. In particular, several participants recommended the comparison of cognitive interviewing techniques with other methods such as expert review and behavior coding (Lessler and Rothgeb, Working Group 6; Willis et al., Session 2).

2. Development of an archive of cognitive testing results that can be used as a body of findings to aid in the design of questionnaires, and to summarize the general types of findings that have been obtained (Presser and Wellens, Working Group 3).

3. Translation of the results of cognitive testing into general principles of questionnaire design (Willis et al., Session 2).

4. Development of computational models for evaluation of survey questions, using algorithms that are based on an understanding of the nature of comprehension problems in questions (Graesser et al., Session 3).

5. Adaptation of models, theories, and findings from the field of linguistics in evaluating survey questions (Fillmore, Session 3).

6. The application of ethnographic models of respondent behavior to the development of survey questionnaires (Gerber, Session 3).

7. Focusing cognitive techniques on the development of concepts and classification systems that are used as a basis for the development of survey questions (Clark and Dippo, Working Group 4).

8. Cognitively-oriented study of the organizational aspects of survey design and implementation, focusing on communication among the various key players: survey sponsors, subject matter specialists, questionnaire designers, and cognitive laboratory staff (Bradburn, Seminar Synopsis; Jabine, Discussion Summary; Kasprzyk and Marquis, Working Group 1).

9. Social representation of the survey interview process; study of how respondents, interviewers, and members of the general public view the survey processes and results (Jabine, Discussion Summary; Presser and Wellens, Working Group 3).

10. Cognitive aspects of survey participation, and in particular, incentives to participate in surveys, in order to reduce survey nonresponse (O'Muircheartaigh, Session 1).

11. The "usability," or human factors aspects, of Computer-Assisted Personal Interview (CAPI) and Computer-Assisted Telephone Interview (CATI) instruments (Couper, Session 4; Kasprzyk and Marquis, Working Group 1).

12. Cognitive study related to the appropriate degree of standardization of interviewer behaviors, to provide optimal levels of interviewing flexibility (Royston and Schaeffer, Working Group 2).

13. The cognitive aspects of data preparation and processing, to avoid errors at postcollection stages (Jabine, Discussion Summary).

14. The cognitive aspects of data presentation through tables, text, maps, and graphs to enhance reader comprehension (Friendly, Session 4; Lewandowsky, Session 4).

15. Evaluation of the statistical literacy of target audiences, to aid in the communication of the messages intended in statistical publications and documents (Clark and Dippo, Working Group 4).

16. The development and application of cognitive techniques that are specially targeted to establishment surveys (Cantor and Phipps, Working Group 8).

17. Cognitive study of particular topic areas that present serious measurement challenges, such as income and disability (David and Moore, Working Group 5; Mathiowetz and Wilson, Working Group 7).

18. Development of new theoretical models of human memory and knowledge representation (Shum and Rips, Session 2; Smith, Session 3).

19. Study of question context effects that influence survey responding (Tourangeau, Session 3).

20. Development of statistical error models of the cognitive processes involved in survey responding and use of these models to develop improved survey estimators (David and Moore, Working Group 5; Groves, Session 3; Lessler and Rothgeb, Working Group 6).

21. Development of survey questions that are "tailored" to different respondent groups (a point made generally in several discussions).

22. The use of survey designs to conduct population-wide assessment of cognitive abilities (Jabine, Discussion Summary).

23. Modeling of the behavior of respondents as they are administered survey questions (Conrad, Session 4; Schober, Session 2).

24. Brain imaging technology as a potential tool for studying and validating responses to survey questions.

The following discussion attempts to move beyond the simple listing presented above, and considers several issues inherent to this set of proposals. We believe that attention to these underlying issues will be critical to the future of CASM research.

The Need to Focus on Stages of the Survey Process Other Than Data Collection

As pointed out by Tucker and Martin with respect to the Working Group Reports (Foreward to the Working Group Reports), the research proposals fall naturally into two groups. Several (1-6 above) are *intensive*, and involve a specific stage of the survey development and administration process—questionnaire design—that has been studied extensively since the time of CASM I, and that has largely defined the CASM movement. However, a second subset of proposals (e.g., 7-11; 13-15), are *expansive*, and attempt to shift our view outward toward stages or areas of the total survey process other than question design (e.g., prequestionnaire concept development, cognitive aspects of respondent participation, and data preparation, processing, and presentation).

We do not regard the future of CASM with respect to these possible directions to be an "either-or" proposition. Clearly, there is no profit in abandoning the current focus on the cognitive aspects of questionnaire-based data collection, because the existence of response error associated with the design of survey questions remains a vexing problem. On the other hand, there are additional facets of the survey process that have gained increased importance over time, and that are therefore ripe for attention by CASM researchers. Already, for example, work on CAGM (the Cognitive Aspects of Graphs and Maps) has spawned several publications that assess the cognitive aspects of graphical data presentation (Herrmann and Pickle, 1994; Pickle and Herrmann, 1995).

Note further that several proposed research areas are somewhat novel not because they focus on "new" stages of the survey process previously unemphasized by CASM researchers, but rather because they represent aspects of questionnaire design and data collection beyond those emphasized at the time of CASM I. For example, the development of CAPI/CATI instrument usability testing, and even dedicated usability laboratories (somewhat similar to existing cognitive laboratories), has been stimulated by the increasing importance of computer-based questionnaire administration. Overall, then, we feel that future CASM-based efforts should target stages of the survey process not typically studied by cognitive researchers, and also retain a strong focus on questionnaire design and data collection, but in a manner that recognizes recent changes in the ways in which questionnaire data are collected.

Introduction of New Methods for Questionnaire Testing

Because of its generally recognized potential for improving data quality, CASM researchers will likely continue to utilize the cognitive laboratory to test and develop survey questionnaires. Several proposals (4-6) advocate the adoption of novel methods for such purposes (e.g. use of computational, linguistic, and ethnographic models of the respondent's question-answering process). Within the areas of CASM that emphasize question testing, there is need for close collaboration between those whose focus is the evaluation and refinement of existing methods, and those who propose to improve questionnaires through the introduction of new techniques. It is likely that new methods that are applied to questionnaire design and evaluation could be incorporated into the existing infrastructure for questionnaire testing. From an evaluation methodology perspective, it is meaningful to determine how these new methods can be successfully integrated with existing techniques, rather than attempting to assess the "inferiority" or "superiority" of each method.

The Need for Methods Development Applicable to Novel Areas of the Survey Process

As mentioned above, apart from the development of new methods within the area of questionnaire design, seminar participants focused on methods development appropriate to additional survey stages, or areas. Research applications of CASM to these novel areas differ in the degree to which they may require the development of methods that also are novel. Several research areas may require protocol development, equipment, training, and staffing that differ substantially from the current "CASM toolbox." For example, usability testing of CAPI questionnaires has been characterized by the adaptation of existing cognitive interviewing methods, but with a much greater emphasis on recording and processing equipment (computers, multiple cameras, devices for integrating digitally-based monitor images with standard video and audio signals). Usability study also requires significant data reduction, for purposes of concise reporting, that extends beyond the requirements typically experienced in "standard" cognitive interview studies. Presumably, it will take time to develop the appropriate methods, and to determine which of these can be adapted from the current cognitive laboratory paradigm.

The cognitive study of maps and graphs (CAGM), of statistical literacy, of respondent reactions to the survey situation itself, and of a number of other potential areas also might rely on current cognitive interviewing-based techniques as a point of departure. Yet, each may also entail a specialized set of techniques, which may require development time and effort, especially to the extent that some of these (such as CAGM) depend on cognitive theories that focus on processes other than the answering of survey questions. Therefore, even if work in these areas is begun in earnest, it may be some time before measurable positive results will be observed. Researchers who delve into these areas must be careful not to underestimate the resources that will be necessary to develop or adapt the key methods that support these new areas of research.

Inclusion of Disciplines Other Than Cognitive Psychology in CASM Research

As Tourangeau (Session 3, Introduction) states, the CASM I Seminar focused almost exclusively on potential applications to surveys of cognitive *psychology*, and did not extend to other cognitive *sciences*, let alone to disciplines outside of cognitive science. The proposals from CASM II address this point by taking into account a much wider range of disciplines, including statistics, linguistics, and computer science. CASM is clearly becoming a much more interdisciplinary endeavor (Sirken and Schechter, 1999), and this extension presents considerable challenges to researchers who must strive to maintain a common focus in terms of language, procedures, and the integration of results from divergent fields. This increased diversity simultaneously presents promise, in the form of additional viewpoints, but also potential problems, if these views are expressed in language that results in researchers "talking past" one another.

Communication Between Basic and Applied CASM Researchers

CASM itself has been defined as an applied research area, in that its focus is application of a general discipline (cognitive psychology) specifically to the applied domain of respondent surveys. Nevertheless, within this orientation, there is a wide range of focus between basic theory, on the one hand, and strict application, on the other. Several of the proposals listed above (e.g., 18-20) emphasize basic research, in that they advocate the academic study of theory-oriented aspects of CASM. Other proposals, and especially those that describe potential for development of specific cognitive methods for questionnaire evaluation, tend to be less focused on theory, and emphasize more the practical aspects of questionnaire pretesting.

A recurrent theme through the CASM II Seminar was the need for improved communication between those who conduct basic cognitive research in order to develop cognitive principles related to the survey response

process, and those who engage in cognitive interviewing studies in order to evaluate and to improve features of specific data collection instruments. On the one hand, there has to date been limited flow of ideas from basic researchers to practitioners (or "pretesters"). On the other hand, pretesters are often intensely focused on the empirical testing of individual questions or questionnaires, rather than making use of general cognitive "laws" that might direct questionnaire design.

Furthermore, cognitive laboratory practitioners have typically reported their results on a questionnaire-by-questionnaire basis using a variety of formats and publication outlets. However useful these reports are for the intended purposes, they have not provided a useful knowledge base of results that may have more general applicability. Hence the proposal was made that cognitive lab results be placed in archives that are generally accessible to researchers (proposal 2 above). A complementary proposal would be that general principles of question design that are derived through basic cognitive research could be tested more frequently within the context of questionnaire development and evaluation, in order to determine the range of applicability of those principles.

Summary

It is possible that further development of either new CASM methods, or extension of CASM research into new areas may result in greater variation in the focus, viewpoints, and methods used by researchers. Researchers must recognize that the notion of interdiscipline no longer pertains strictly to the reconciliation of two camps consisting of cognitive sychologists and survey methodologists, but that CASM itself is beginning to spawn its own subareas that require active movement towards interdisciplinary research. To this end, active attempts (such as the CASM II Seminar) to bring together the emerging branches of CASM research may promote development of the field in such a way that it is able to continue to represent a core approach having a commonly accepted language, set of methods, and identifying characteristics. With the successful coordination of a variety of CASM research projects of the type summarized above, we believe that the future will be characterized by a scientific evolution that continues to be successful. The Epilogue of these Proceedings examines the future of the CASM movement in more depth, and proposes some mechanisms for facilitating the continuation of collaborative CASM research.

References

Herrmann, D, and Pickle, L. W. (1994). *A cognitive subtask model of statistical map reading*. National Center for Health Statistics: Cognitive Methods Staff, Working Paper No. 11.

Pickle, L. W., and Herrmann, D. (1995). *Cognitive aspects of statistical mapping*. National Center for Health Statistics: Cognitive Methods Staff, Working Paper No. 18.

Sirken, G., and Schechter, S. (forthcoming). Interdisciplinary survey methods research. In *Cognition and survey research*. New York: John Wiley & Sons.

EPILOGUE

A Roadmap for Survey Research Methods Beyond the Year 2000

Monroe G. Sirken

A roadmap is an extended look at the future of a chosen field of inquiry composed from the collective knowledge and imagination of the brightest drivers of change in that field.
- Editorial "Science Roadmaps" in Science (Galvin, 1998)

Introduction

Driven by the need to solve practical survey problems, applications of the theories and methods of cognitive psychology have significantly altered survey methods during the past 15 years. Furthermore, they have the potential for having even greater benefits for survey methods during the years ahead. These conclusions were reached at the CASM II Seminar that met in June 1997. It should come as no surprise that there was much enthusiasm and overwhelming agreement at the CASM II Seminar to continue and expand the scope of interdisciplinary research on the cognitive aspects of survey methods (CASM). Hence, it seems appropriate in this Epilogue of the CASM II Seminar Proceedings, to turn attention to the future of CASM research, and briefly consider the roadmap for interdisciplinary survey methods research beyond the year 2000.

Compared to other survey research interdisciplinary endeavors, CASM has been singular in viewing its efforts as an experiment in fostering interdisciplinary research (Tanur, 1984). The experience of CASM and other interdisciplinary research fostering efforts (Olkin and Sacks, 1988) suggest that preparing a road map for interdisciplinary survey methods research beyond the year 2000 poses the following questions:

- What kinds of innovative research are needed to address priority survey needs?

- What funding opportunities are available to support innovative survey research?

- What infrastructural arrangements are there to sustain innovative survey research efforts?

In seeking answers, it is relevant to review how these questions were addressed at the CASM I and CASM II Seminars, and to update developments during the 15 months that have elapsed at this writing since the CASM II Seminar.

Responding to Needs For Interdisciplinary Survey Methods Research

The CASM movement did not occur spontaneously. It was and remains a *deliberate* effort to foster interdisciplinary research on the cognitive aspects of survey methodology. The movement seeks to improve the methods of the survey measurement process while contributing to the knowledge base of the cognitive sciences.

It was probably prudent in 1983 at the CASM I Seminar to narrowly focus CASM research on the cognitive aspects of questionnaire design (Jabine et al., 1984). Otherwise, it probably would have taken much longer to successfully demonstrate the utility of CASM research to improve survey methods. It was similarly prudent in 1997 at the CASM II Seminar to broaden the scope of CASM research to all stages of the survey measurement process, and to extend CASM research collaborations to multiple scientific disciplines.

Opening the door for other disciplines to serve as co-participants in the CASM movement was a significant step in the right direction but insufficient to address many survey needs that require the tools and expertise of the behavioral, computer, statistical, and other sciences even more than those of the cognition sciences. (Sirken and Schechter, 1999). Hence, I propose *expanding CASM into a movement that fosters research on the interdisciplinary aspects of survey methods (IASM), and seeks solutions to survey problems at intersections of multiple disciplines, irrespective of the disciplines involved.*

I foresee the IASM effort as maintaining the integrity of CASM as a distinct subfield of survey research, while simultaneously fostering research on the behavioral, computer, and statistical aspects of survey methods, and with a sharp focus on research at intersections of multiple disciplines. Also, I foresee enhanced funding opportunities for interdisciplinary survey methods research, including CASM research, with the expansion of CASM into an IASM movement.

Since its inception, CASM has been essentially a domestically fostered effort. In retrospect, it is surprising that the notion of fostering an international CASM

movement was not discussed at the CASM II Seminar. Some of the earliest efforts to build bridges between cognitive psychologists and survey researchers and foster interdisciplinary survey methods research occurred in England (Moss and Goldstein, 1979; Sirken, 1986), and in Germany (Hippler, Schwarz, and Sudman, 1987). More recently, cognitive research laboratories have been established in official statistics agencies of several European countries. *It is especially timely and important now*, a*s CASM evolves into an IASM movement, to foster an international program in interdisciplinary survey methods research.*

Funding Opportunities In Interdisciplinary Survey Methods Research

Cultivating funding opportunities is often the most challenging task in fostering interdisciplinary survey methods research. Fortunately this was not CASM's initial experience. A year or so before the CASM I Seminar convened in 1984, Murray Aborn, then Head of Measurement Methods and Data Improvement Program at the National Science Foundation (NSF), established the first funding opportunity in cognition and survey methods research. Aborn's program funded the CASM I Seminar and awarded grants to several research proposals that evolved from the CASM I Seminar. Unfortunately, Aborn's program expired about 1990, and funding opportunities dedicated to CASM research did not exist when the CASM II Seminar convened in June 1997.

During the CASM II Seminar, Cheryl Eavey, Head of NSF's program in Measurement and Statistics (MMS), encouraged participants to send CASM research proposals to MMS, and at the conclusion of the seminar, she offered to establish a new MMS funding opportunity in survey methods research contingent on a consortium of federal agencies matching MMS' funding. Several months of negotiations brought forth late in 1998, an entirely new kind of arrangement for funding survey methods research. This funding opportunity is being sponsored jointly by the NSF and the Interagency Council on Statistical Policy (ICSP), a committee of federal statistical agency heads chaired by the Chief Statistician of the Office of Management and Budget (OMB).

At this writing, plans to administer this funding opportunity are as follows. Overall management of the program will be NSF's responsibility. After research proposals are reviewed by the MMS review panel, meritorious proposals will be reviewed for potential value to federal statistics by the Federal Committee On Statistical Methodology, a permanent OMB committee comprising researchers from federal agencies.

The NSF/ICSP funding opportunity seeks to expand and accelerate the development of new and innovative approaches in survey methods research. Priority will be given to research proposals (1) having broad implications for the survey field (2) having potential for creating fundamental knowledge of value to federal surveys, and (3) involving multiple discipline collaborations among the social, behavioral, cognitive, computer, statistical and other sciences. FY 1999 is the demonstration year for the NSF/ ICSP funding opportunity, and it is anticipated that funding will be renewed for two additional years if quality research proposals are forthcoming.

Though the emergence of this innovative funding opportunity is most satisfying and most welcome, this is hardly the time to indulge in self-congratulations. Reliance solely on NSF/ICSP funding would create a false sense of security, and could postpone efforts to foster longer term IASM research funding. *It is most important to foster efforts that stimulate researchers to respond creatively to the NSF/ICSP funding opportunity and submit quality research proposals responsive to federal statistics needs. Professional societies have important roles in stimulating and supporting these efforts.* Submission of quality research proposals could virtually assure renewal of the funding opportunity during FYs 1999 and 2000, and might possibly lead to its extension after FY 2001 when the existing NSF/ICSP arrangement is programmed to expire.

Sustaining Interdisciplinary Survey Methods Research

Responding to needs for innovative survey research and cultivating funding opportunities are requisites to fostering a interdisciplinary survey methods research program. But sustaining a vibrant interdisciplinary research program ultimately depends on obtaining the active participation of researchers in the various scientific disciplines. Hence, it is essential to foster activities that attract the attention of these researchers to the opportunities and stimulate their active participation in the interdisciplinary effort. This objective can be met by fostering a series of interdisciplinary meetings that compose and update roadmaps, and present and discuss research findings (Galvin, 1998).

Though not the first interdisciplinary survey methods research fostering effort in this country (American Statistical Association, 1978), CASM has fostered, far more interdisciplinary research meetings than any other effort which helps to explain why it has survived the longest. During the 14-year period between the CASM I and CASM II Seminars, for example, papers on a variety of CASM related topics were presented at four major CASM conferences (Schwarz and Sudman, 1992; 1994; 1996; Schwarz, Park, Knauper, and Sudman, 1999) that were supported mainly by the Universities of Illinois and Michigan. And ways of continuing and enhancing the CASM movement were discussed and studied in a series of eight CASM

workshops that were sponsored by the Social Science Research Council and funded by NSF (Tanur, 1992). This listing is illustrative only and not intended to be exhaustive. It excludes many presentations of CASM papers at numerous conferences and meetings.

In the past, CASM meetings were organized essentially on ad hoc bases by different individuals and groups who took advantage of emerging opportunities. However, ad hoc planning is essentially no planning at all. There has not been and there is not now a permanent infrastructural arrangement to sustain CASM meetings. With the scope of CASM rapidly expanding, it makes much less sense now that it did at the onset of the CASM movement at the CASM I Seminar to be content with ad hoc approaches in planning and arranging future seminars, conferences and workshops. Though CASM II Seminar participants strongly favored convening a CASM III Seminar sometime in the future, unfortunately, they did not discuss the infrastructural arrangements.

Certainly, ad hoc approaches to arranging future meetings should be encouraged. For example, at the 1999 Joint Statistical Meetings ASA's Survey Methods Section is sponsoring an invited session "The Changing Face of CASM Research in the Coming Decade" and a special contributed session "Survey Research at the Intersection of Statistics and Cognition." Ad hoc meeting are welcomed adjuncts to regularly scheduled meetings, but not as substitutes for them.

Planning and organizing future IASM meetings should be undertaken now, while the impetus of CASM II Seminar is still fresh, and the NSF/ICSP funding opportunity is about to get underway. The IASM meetings would stimulate research proposals for the NSF/ICSP funding opportunity, and presentations and discussions of NSF/ICSP funded research projects would add substance to the IASM meetings.

Much can be learned about meeting arrangements from experiences of other survey research related fostering efforts such as the Seminar in Bayesian Inference in Econometrics and Statistics (SBIES) (Berry, Chaloner, and Gerweke, 1996), and the International Workshops on Household Survey Nonresponse (Lyberg, 1996). Possible scenarios for IASM meetings are (1) alternating the host establishment between universities and official government agencies to encourage participation by researchers in both sectors, (2) shifting meeting sites among different countries to ensure international participation, (3) publishing peer reviewed monographs and proceedings to draw the best and brightest to participate, and (4) co-sponsoring meetings with SBIES and other research groups to stimulate interdisciplinary participation.

To consider these and other scenarios for periodic IASM meetings, I am proposing *to convene a small working group of interdisciplinary minded individuals to explore the needs for periodic IASM meetings, to prepare a proposal to solicit funding for IASM*

meetings, and possibly to serve as the steering committee for the first IASM meeting.

Summary Remarks

This Epilogue raises and considers questions that need to be answered in composing the roadmap for interdisciplinary survey methods research beyond the year 2000. In addressing these questions, several extensions are proposed to the CASM roadmap that was drawn at the CASM II Seminar:

- Proposal 1. Expand CASM into an international effort that fosters research on the interdisciplinary aspects of survey methods (IASM) and seeks solutions to survey problems that are at intersections of multiple disciplines irrespective of the disciplines involved.

- Proposal 2. Stimulate researchers and professional societies to participate in the research opportunities offered by the NSF/ICSP funding opportunity.

- Proposal 3. Establish mechanisms for convening periodic IASM meetings, and for publishing peer reviewed proceedings and monographs.

- Proposal 4. Convene a working group to examine alternative IASM meeting scenarios, prepare a funding proposal for IASM meetings, and possibly serve as the planning committee for the first IASM meeting.

Personally, I foresee many challenges ahead in fostering interdisciplinary survey methods research beyond the year 2000. However, the experience and success of the CASM experience persuade me that the vast untapped potential benefits remaining more than justify the proposed IASM fostering efforts.

References

American Statistical Association (1978). *Conference on the transfer of methodology between academic and government statisticians.* Washington, DC: American Statistical Association.

Berry, D., Chaloner, K. And Geweke, J. (Eds) (1996). *Bayesian analysis in statistics and econometrics.* New York: John Wiley & Sons.

Galvin, R. (1998). Science roadmaps. *Science, 280,* 803.

Hippler, H. J., Schwarz, N., and Sudman, S. (Eds). (1987). *Social information processing and survey methodology*. New York: Springer-Verlag.

Jabine, T.B., Straf, M.L., Tanur, J. M., and Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

Lyberg, L. (1996). *The nonresponse workshops - what have we accomplished so far?* Unpublished paper presented at the International Workshop On Household Nonresponse, Rome.

Moss, L. and Goldstein, H. (Eds.). (1985). *The recall method in social surveys.* Portsmouth, NH: Heinemann, 1985 (originally published by the University of London Institute of Education in 1979).

Olkin, I. and Sacks, J. (1988). *Cross disciplinary research in the statistical sciences*. Report of a Panel of the Institute of Mathematical Statistics. Washington DC: National Science Foundation.

Schwarz, N. and Sudman, S. (Eds). (1992). *Context effects in social and psychological research*. New York: Springer-Verlag.

Schwarz, N. and Sudman, S. (Eds). (1994). *Autobiographical memory and the validity of retrospective reports.* New York: Springer-Verlag.

Schwarz, N. and Sudman, S. (Eds). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

Schwarz, N., Park, D. Knauper, B and Sudman, S. (Eds.). (1999). *Cognition, aging, and self-reports.* Philadelphia, PA: Psychology Press.

Sirken, M.G. (1986). The recall method in social surveys. In Book Reviews of the *Journal of the American Statistical Association, 81*, 871.

Sirken, M. G. and Schechter, S. (1999). Interdisciplinary survey methods research. In M.G. Sirken, D. Herrmann, S. Schechter, N. Schwartz, J. M.Tanur, and R. Tourangeau (Eds.). *Cognition and survey research*. New York: John Wiley & Sons.

Tanur, J. M. (1984). Preface. In T. Jabine, M. Straf, J. Tanur and R. Tourangeau (Eds). *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. ix-xi). Washington, DC: National Academy Press.

Tanur, J. M. (Ed.). (1992). *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.

APPENDIX A

Cognitive Aspects of Survey Methodology: An Oral History

Barbara Wilson and Paul Beatty

Note: The following is an edited transcription of a videotape that was itself abstracted from a number of individually recorded interviews made prior to the CASM II Seminar. The videotape was shown to participants the first night of the seminar. The editors have made minor modifications to the transcript, in order to facilitate the translation of spoken speech into prose, and in particular, to restore and clarify speakers' intended meanings that may have been well-expressed in the spoken interview, but were lost in the literal transcription. Every effort has been made to convey the original meaning, as interpreted by the editors. Those interested in obtaining an original copy of the CASM II videotape, or the individual video-recordings on which this was based, should write or call NCHS:

ATTN: CASM II Seminar Videotape Archive
Office of Research and Methodology
Room 915, NCHS
6525 Belcrest Road
Hyattsville, MD   20782
(301) 436-7111

Introduction

During the last 15 years, Federal statistical agencies have been involved in an ambitious and unusual interdisciplinary effort commonly known as "CASM," - Cognitive Aspects of Survey Methodology. Survey methodologists and cognitive psychologists are now engaged in numerous collaborations, whereas only a short time ago, there was very little communication across these disciplines.

The CASM movement did not just happen—there was a deliberate attempt to foster it. This video tells part of the story of how this happened: the events that led to the first interdisciplinary "CASM" seminar in 1983, the goals and interests of some of the participants, and the agenda they set for the future. We explore why these collaborations emerged when they did. We also look at what CASM has accomplished, and what work still needs to be done. In short, we look at where CASM has been, where it is, and where it needs to go.

The seminar known as "CASM I" was arguably a landmark event in terms of breaking down communication barriers across disciplines. But it took a while for this idea to catch on.

Leading up to CASM I

Murray Aborn, the National Science Foundation, saw important changes in several disciplines during the early 1970s. *Cognitive psychology was beginning to emerge. Linguistics was coming into its own, moving from the humanities to the sciences. Somewhat to his surprise, survey practice seemed untouched by these developments.*

Murray Aborn: It struck me that what I had known of surveys from my days back in the fifties, at the National Institutes of Health, had not changed. State-of-the-art survey research had not been affected by the things that were happening in the most pivotal disciplines that form the underpinnings of survey methodology. So, in the early seventies, I attempted to bring together, in CASM I fashion, survey researchers with cognitive scientists and psychologists, linguists, and other cognitive scientists, in the hope that some of this would rub off. That is, I hoped that they would be interested in utilizing the things that these new disciplines had to offer.

Unfortunately this meeting didn't happen in the 1970s mainly out of widespread skepticism that anyone would be interested in crossing outside of disciplinary boundaries. Yet the idea didn't die. A few years later, talk of these collaborations began again, this time to address specific concerns about the quality of data from the National Crime Survey.

Aborn: There was considerable skepticism about data that was dependent totally on self-report and recall, and sometimes on recall from considerable periods back. The

Bureau of Social Science Research arranged a workshop, to which I was invited, to try to bring together cognitive psychologists to see what could be done to repair the damage, for the lack of a better word, done to the victimization surveys, and to redesign them.

Stephen Fienberg: Al Biderman, who led the redesign group, organized a workshop here in Washington in the fall of 1980, which brought together a group of cognitive psychologists, including Endel Tulving, Beth Loftus, and a few others—mainly those involved in the redesign project—to see what cognitive psychology might have to offer in thinking about the issues that we were concerned about in the National Crime Survey.

Although this workshop preceded the CASM I seminar by several years, some consider it to be the genesis of CASM.

Judy Tanur: Al, I think of you as truly the founding father of the CASM movement.

Albert Biderman: How about grandpop?

Tanur: Okay, whichever. But you were there. So I want you to talk about the seminar that you organized in 1980.

This seminar offered new hope for developing improved questioning strategies, to maximize the quality of recalled information. To develop truly novel approaches, Biderman helped to assemble an interdisciplinary consortium that mixed academic psychologists with federal statisticians.

Biderman: We had survey researchers, including a lot of people from the Census, and others from some agencies who were doing large scale surveys. Monroe Sirken was there.

Monroe Sirken: For me, it was an eye-opener. I knew nothing about the cognitive revolution in psychology. We had reached the point where we were showing sampling errors to several decimal places and couldn't say anything about response errors.

    I felt, intuitively, that we working on all these problems of memory, and that there was a whole science devoted to understanding it. Why weren't we getting together and trying to learn from each other?

Stephen Fienberg: What was very clear to me was that everybody who had studied victimization had related ideas. What the cognitive psychologists were doing was telling us about a systematic way to think about some of those ideas.

    But, to make the fit with the survey environment was far more complicated than you could hope to do, sitting at a table, in two days. And, since everybody had their own research agenda, the question was, what could you

do to influence the research agendas, to get people to take each other seriously, over a longer period of time?

Murray Aborn's program at the National Science Foundation encouraged proposals to expand this research agenda. Miron Straf, on the Committee of National Statistics, was involved in early proposals. However, there was still resistance.

Miron Straf: We submitted a proposal. It was resoundingly rejected at first, because we had planned for full-fledged study by a panel of experts. NSF reviewers thought that was a bit premature. But, they encouraged us to revise our plan and have a workshop or conference-like setting, which we did. And, they provided the funds for us to carry it out.

Stephen Fienberg: By then I was chair of the Committee on National Statistics, at the National Academy of Sciences, and I suggested that we do a follow-up activity that was broader than the one that was tried in 1980.

Thomas Jabine, then with the Committee on National Statistics, was also involved in the conference plans.

Tom Jabine: I was asked to help organize it, and I worked with Miron Straf (who was at that time the assistant to Ed Goldfield, the Director of the Committee), Roger Tourangeau, and Judy Tanur, who was the member of the Committee on National Statistics in charge of the seminar.

Judith Tanur was one of the committee members who named the seminar—and subsequent research activity—as "CASM."

Judy Tanur: When the Committee of National Statistics first started thinking about this project, the working title was Cognitive Aspects of Survey Questionnaire Design. As we thought about it more, the notion of making it broader became more and more attractive. It particularly became attractive when we realized that if we called it "Cognitive Aspects of Survey Methodology," we had a wonderful acronym that would stress not only what we meant, but also the enormous chasm between the disciplines that we were trying to bring together.

The CASM Seminar

In 1983, the CASM I seminar convened in St. Michaels, Maryland.

Tom Jabine: The whole idea was to have the participants, who were relatively small in number (about 25), meet for a fairly extended period, and to not only have formal

sessions, but to have opportunities for various kinds of recreation and informal gatherings.

Miron Straf: Our first trepidation was that we were bringing together two very different cultures. We had lined up people very carefully, who were committing an entire week of their lives. I think we began on a Wednesday evening and ended on a Tuesday morning, so they literally were going to be away from their homes and offices and families for an entire week. They all made that commitment. We squirreled them away in a retreat location, so they couldn't be distracted. They were forced to be together, and we thought it would take about a day to break the ice between the group of cognitive scientists and the group of survey researchers, who really hadn't communicated with one another. Actually, it only took a few hours in the first evening to do that. By then, people were really talking with one another.

Stephen Fienberg: People told me that the CASM seminar wouldn't work—it would be a failure—nobody would pay attention. There would be no follow-up.

We enlisted some really good people to come and participate in that workshop. And it was one of the most exciting weeks of my research career.

University Research Following CASM I

CASM I opened up the possibility of new collaborations, but that was only the first step. These new research ideas needed support, which the National Science Foundation helped to provide.

Norman Bradburn: Murray Aborn had played a critical role in developing both NSF support for research, and also in getting various people across a wide range of social sciences interested. So, when I came back after the St. Michael's conference, I got busy and wrote proposals. We then had about five or six years of good funding to do a lot of follow-up studies.

Many participants pursued new research based on ideas that crystallized at CASM I.

Elizabeth Loftus: It definitely has affected interaction among and between people who would not have collaborated otherwise. It also provided ideas and opportunities for people to get together and do things that they wouldn't have been able to do by themselves. I collaborated for a number of years with Judy Tanur and Steve Fienberg. We wrote a number of articles together. There was no way that without the CASM introduction, I would have had the opportunity to work with such smart people in fields other than psychology.

I got some funding to study people's recollections of their hospitalizations and their medical procedures. We tried to improve people's memories for their medical procedures. We compared their memories to what we found in the medical records. It was during that project that I found out how awful it can be when you have to rely on medical records and doctor's handwriting to try to make sense of anything.

It was through these efforts and experiences that I learned about forward telescoping. That wasn't in my vocabulary before the early 1980s, and I got the idea that landmark events could help stop forward telescoping. We had an incredible event happen in the state of Washington—the eruption of Mount St. Helens— and we decided to use this as a landmark to see whether we could stop the forward telescoping that would occur when people were trying to recall their crime victimization.

NCHS began to support a great deal of research in this area, through a program initiated and directed by Monroe Sirken. Judith Lessler, who managed some of the early NCHS contracts, had thought that laboratory work could contribute to survey research—but this idea hadn't been well received at first. This program allowed her to pursue some of those ideas:

Judy Lessler: I was at the American Statistical Association and I was chatting with Monroe Sirken, and we had lunch together. I thought that we really ought to focus more on errors by doing laboratory type studies. He started telling me about having gone to the CASM conference, and how important that was, and that cognitive psychology then would form a way to do some of those studies, because you would be looking at what's going on inside the head. That's where a lot of the errors are occurring.

He told me that he was thinking of applying for the grant from Murray Aborn at NSF. He asked whether, if he got it, I would be interested in coming to help him run it I thought, sure, I'd love to do that.

The program was organized to focus on the cognitive problems plaguing particular surveys...

Monroe Sirken: ... particularly those surveys that had very important policy implications. So we proceeded to develop a series of contract topics for which we requested proposals, in which we identified certain problems that had nonsampling error consequences that were serious.

The Development of Cognitive Laboratories

The NCHS program contained another innovative element: the establishment of the first cognitive laboratory in a federal setting.

Judy Tanur: Although some of us were skeptical about whether NSF would be willing to fund such a thing at a federal agency, Monroe did an enormously persistent job

in getting such a laboratory funded, and I think his persistence and vision was a major step, and a major influence in having the CASM movement be as successful as it was. That laboratory became the model for similar laboratories and to other federal agencies.

Roger Tourangeau: It seems to me that the most dramatic impact, and practical achievement of the CASM movement, has been the radical change in how survey questions are pretested. I think it appears revolutionary in some ways because the vocabulary has all changed. But it in fact was evolutionary. It used to be that in private organizations like NORC there was some small-scale intensive interactive reviewing of survey questions with respondents. But for the most part, that didn't happen in the large government statistical agencies.

It seems that in the space of a year or two all the government statistical agencies set up cognitive laboratories—sometimes not just using the standard set of survey tools. They adopted these methods of testing.

It's sort of the "atom bomb" of the CASM movement—the one very dramatic methodological change that I think is directly traceable.

The Spread to Other Agencies

Research programs emerged in other statistical agencies as well. Elizabeth Martin leads an interdisciplinary staff at the Census Bureau that does a great deal of CASM-related work.

Betsy Martin: We have people who come with backgrounds in cognitive psychology and social psychology, and in anthropology, including cognitive anthropology. In the past, we've had people from political science, sociology, and survey methodology. Occasionally we have mathematical statisticians on our staff. So it's really all the social sciences, as well as psychology—it's very interdisciplinary. We have a mix of people who are both quantitatively and qualitatively oriented. I personally find it very interesting and gratifying to be able to interact with people from different disciplinary perspectives.

A lot of the work that we've done has been on questionnaire testing in the decennial area, to some degree in the economic area, and also on the demographic survey side on questionnaire design and testing. We've also done a lot of other work that really isn't in questionnaire testing. For example, in the decennial census, we looked at procedures for enumerating homeless people, and then did field work.

Another group formed at the Bureau of Labor Statistics. Janet Norwood was commissioner of the Bureau of Labor Statistics at that time.

Janet Norwood: I decided that it was really necessary to formalize this kind of work, and I knew that some work was going on at Census, and had been for some time. I also knew that Monroe Sirken had done wonderful things at NCHS.

I had a little difficulty when I went to the Congressional Subcommittee on Appropriations, because, of course, they look at what you're doing, and they wanted to know why the Bureau of Labor Statistics had hired psychologists. I thought about that for a moment and said "Well, Mr. Chairman, I want to assure you that it is not because we are engaging in therapy for our staff; it is because we believe that this is one discipline which has an important effect upon survey design. And there are other disciplines which we will include as well."

The way I tried to explain it was that, if you didn't do this kind of research, you'd have to go out and do a great deal of experimentation in the field. Field work is tremendously expensive. It's much cheaper to figure out, in advance, what may or may not work, and then go out and test it. That way, you're using the money much more effectively.

Clyde Tucker, as director of this program at BLS, sought to make this sort of developmental work increasingly routine.

Tucker: We were able to institutionalize laboratory testing in our survey design process. It's an up-front part that most survey managers are now seeing as necessary and useful to them, prior to actually finalizing a survey instrument. I think they've come to count on it, particularly with new surveys.

Questionnaires are now often integrated into computer-assisted data collection. Labs have adapted to these changes with the emergence of usability testing.

Betsy Martin: We're going to be doing usability testing of some of the Internet surveys in the economic area. We're testing surveys that have been sent out through the Internet, and having respondents answer them on their computers. There is a lot of work with the technology, in terms in how (the questionnaire) gets there, downloading it, and whether people can find their way through an instrument that's on the computer.

On a tour of the interviewing facilities in the BLS cognitive laboratory, Clyde Tucker described how BLS is testing usability of computer-assisted surveys.

Tucker: Our newly created usability lab is for testing software and computer-assisted survey information collection instruments. We can have our user sit at the console and monitor and use the software while we're videotaping it. From one camera, we videotape their

face. We have what we call a scanning converter which allows us to capture the screen images that they're working with at the same time. Eventually we'll have another camera that will capture their use of the keyboard. So, we'll have all three images simultaneously that we can use to evaluate the usability of our software for data collection.

Altogether, the CASM I seminar seems to have had a substantial impact.

Tom Jabine: Of all the different projects I worked on, for about a fifteen-year period now, the CASM I conference stands out more than any other in terms of the result, in terms of the effects that it's had. Of course, as Monroe and Murray are pointing out, the outstanding effect is the creation of the cognitive laboratories.

Roger Tourangeau: I often have felt that there is a lot of excitement at conferences like these, but there is rarely much in the way of follow-up, or any concrete achievements. Often, years later you can recognize that something important has come out of it, but usually it is subtle or indirect.

This was a real exception. Both at NCHS and at NSF, there was a real effort to create, to consolidate what had happened, and to institutionalize it and to create projects that would keep the movement going. I'm not sure whether I expected that. But, in fact, the consequences of the first seminar were quite important.

Merging with Previous Work in Psychology of Survey Response

Yet, these developments did not occur in a vacuum. As influential as CASM I was in federal agencies, academic survey researchers had also been thinking about psychological aspects of survey response. University researchers had laid a groundwork that allowed CASM to prosper... and in turn, CASM provided the university researchers with a new audience for their work.

Seymour Sudman: Through the seventies, Bradburn and I were interested in issues of response effects, issues of how sensitive questions impacted on respondents, and how people would react to those. We did a lot of work in the area of threatening questions.

Certainly, we saw some of the things that we did as having psychological implications. We were obviously aware that issues of memory, issues of self-presentation, and issues of understanding the question, related to answering questions. Although we didn't call it cognitive psychology, we certainly were aware that psychology played an important role in understanding what was going on.

Norm Bradburn: I had been working in areas of survey methodology before this. On the basis of the kind of review that we had done in the early seventies, I had come to the conclusion that the tasks involved in interviews, particularly the questionnaires and the way questions were framed, were most important in determining what effects there may be on the answer. There had been a lot of demonstration of effects—you could show effects, but there was no theory that was particularly useful. So, I found it extremely helpful, bringing cognitive theory into an area that was basically devoid of theorizing.

Prior to CASM, psychologist Norbert Schwarz had noticed that survey researchers sometimes failed to focus on relevant psychological variables when looking at potential influences on responses.

Norbert Schwarz: The influence of question wording like "forbid" and "allow" would be analyzed as a function of education, or social status, or things like that, which seemed completely wrong, given what we had learned about how people think. I thought that one really has to analyze it at the level of language and cognition and communication.

For example, some of his research focuses on how respondents infer meaning from the content and context of survey questions. Even in a survey context, people follow certain conversational norms.

Schwarz: In particular, when an utterance is ambiguous, the listener is supposed, and expected, to draw on the context. That's what we do all the time. It's only when we write survey questions that we think it's somehow strange, an odd effect, and a problem, if people draw on the context. This is not the problem at all. If there is any problem, or anything that's surprising, it is when something is context independent. Because that's how we make sense of language—the same utterance means different things in different contexts.

Charles Cannell, one of the founders of the Institute for Social Research at the University of Michigan, had been involved in research on the psychology of survey response for decades.

Charlie Cannell: We did quite a bit of this in earlier times. We didn't dress it up in cognitive terms, but that's what it was. For example, there are four studies that we did. One looked at whether, if you ask a question in somewhat longer version, where you explain what the question is, and why you're asking it, and so forth, you get better responses. We found that kind of a question is more powerful than the short question. Why? Well, because it makes the respondent think. It gives him a clear notion of the context in which we're coming.

As early as the 1950s, Cannell was working on problems such as underreporting of hospital visits—these explorations necessarily drew heavily from psychological ideas.

Cannell: Overall, there was approximately a 15 percent failure to report a hospitalization, for which we had a record. Then we began to think: Why is it that these are not being reported? Again, to shortcut an awful lot of data, there were three correlates of underreporting that we located.

One was, "How long before the interview was the hospitalization?" The curve of reporting dropped off as you went from one month to two months, and down to eleven and twelve, the reporting was really bad.

The second was that the reporting varied according to how serious the hospitalization was. If it was for a major surgery or something, the reporting was a lot better. If it was something minor, it wasn't reported well.

And the third variable was self-image, or social desirability threat. You could classify conditions by how embarrassing they would be for a person to report it. And that was a beautiful predictor of whether a person reported it.

What we did, at that point, was to decide that the respondent really had very little incentive to report, or to spend a lot of time devoted to reporting—that this was a motivational issue. We then asked: How do you get the respondent to report accurately? And how do you get him to accept his role? And how do you do this, that, and the other thing, to stimulate his activities?

So how does CASM fit in to this—is it a breakthrough? Or, old ideas for a new audience? Or a logical step forward?

Roger Tourangeau: I think it was an incremental advance but at the same time it was a breakthrough. I think the work of Cannell and Fowler and Oksenberg and Kent Marquis is an important precedent. The work that Norman Bradburn and Seymour Sudman had been doing together also was important. Norman's training was as a psychologist—he had been using social psychological concepts. So I think it was an incremental extension of that prior work.

The reason why I think it was a breakthrough, though, is that I don't think either Norman or Charlie really had much background in the latest cognitive models. So, when I came along, with people like Beth Loftus and others who had also been involved earlier, we were really able to apply the latest cognitive models to these problems.

Betsy Martin: I think that the CASM movement has done a very good job of promotion and making coherent a set of ideas, some of which originated with the first CASM conference. But there was work going on independently in these different areas—before the CASM

conference, actually. So as a coalescence of a set of intellectual ideas, I think it is pretty influential.

Why CASM Took Root When it Did

Perhaps the recent work on the cognitive aspects of surveys can be seen as a merging of the previous groundwork with the new conceptual framework, ideas, and opportunities that followed CASM I. That still leaves the question—why did this happen when it did?

One factor may have been the fact that cognitive psychology itself was newly emerging, which created potential that had not existed before.

Norm Bradburn: One of the things I had noted was that there had been a great advancement since I was in graduate school. So, I learned a whole lot about cognitive psychology. And I thought that it might have implications for what we do in surveys, because I had been doing stuff on survey methodology in any case. There were also other things going on—work at the National Research Council on measuring subjective phenomenon, that had a cognitive flavor.

Since psychologists and statisticians have very different goals, and there are difficulties communicating across disciplines, it took an explicit push to overcome these barriers.

Stephen Fienberg: The cognitive psychologists had lots of wonderful things to say, and they were usually generalizing from experiments done with ten people in a room, usually undergraduate psych majors. They would make generalizations of a sort that, as a statistician, I would be quite loathe to even think about. The question is, what was the applicability of those generalizations to the kind of problems we were studying?

Clyde Tucker: I think there's still some tension between those who favor doing large-scale testing with representative samples, versus relying on information from small-scale, either field or laboratory experiments, where the number of cases is quite small. There are types of research that have long been accepted in psychology, but have not been accepted in survey research, particularly among statisticians.

Fienberg: Things don't happen simply by having a great idea—you have to promote it a little. And I think CASM was a great idea, but if it hadn't been promoted, it would have withered and the impact would not have been there.

Monroe Sirken: I also think there was a readiness for it. It was recognized that there was a need to do something, but there was a frustration in not really knowing what to do. This seemed to be something that could be done.

Nobody knew whether it would work out. But I always felt that it would, and as a matter of fact it was quite successful.

Accomplishments of the CASM Movement

Before turning to goals for CASM II, and thinking about how to support research in this area, we should ask: What is the current state of research in the cognitive aspects of survey methodology? And what has changed in the last 14 years? Certainly, there has been growth in both practical applications and theoretical knowledge.

Norbert Schwarz: When you look at the applied side, it's very clear that how pretests are being done has changed. How people develop questionnaires has changed. The level of attention to what could be going on in survey interviews has changed. And most big survey organizations have some kind of cognitive laboratory now. At the theoretical level, we have clearly made progress. We have a better understanding of how people answer questions, of how the survey interview is influenced by conversational processes, context effects, and attitude judgments. We have a better idea of autobiographical memory and it's implications on what we can and cannot report.

Some argue that the CASM movement has helped provide a more realistic perspective of the problems involved in interviewing humans; we now have a new pragmatism in data collection.

Norm Bradburn: It's become part of the common sense of the survey world. It's just been incorporated into everything that people do.

Monroe Sirken: I think that the CASM movement "unstiffened" government approaches. Before the whole approach, questionnaire design and interviewing were structured as highly as possible. I think that posture went over into the way the researchers worked—they were pretty stiff themselves. It could very well be that when we decided that we really had to find out what was going on in a respondent's mind—that this had an unstiffening effect.

Janet Norwood: I think probably what's most important is the recognition that if you are collecting data, you have to pay some attention to the respondent.

There were a lot of places where, it seemed to me, we could do better, because my belief is that a lot of the research that has been done on data collection is done in order to reduce interviewer variance. That's fine—I'm all for it. But if you reduce interviewer variance by creating cognitive problems with the survey instrument, I'm not at all sure that you're doing the right thing.

Charlie Cannell notes that there has been a major rethinking of the interviewer's contribution to survey error.

Cannell: I think that the traditional idea of what's wrong with questions was that the interviewers were doing a bad job: "It's the interviewers fault. If you only trained those interviewers right, you'd be all right." Well, that, of course, is not the case.

Herb Hyman's made a famous comment that, "Our problems with our questions are not with the interviewers, but with ourselves." So the idea that there needed to be more care and thought in the development of questions, and of thoroughly testing the questions, was a very significant part of the operation that had been sadly neglected.

Perhaps another benefit is that researchers have started to consolidate what they learned about the response process, and move research forward systematically.

Roger Tourangeau: I think the kinds of models that have come out of the CASM movement really allow for there to be accumulation of research results, and to show the relationship between a finding on one question and a finding on another. Look at the context effects area—prior to the eighties, there was a list of 30 or 40 examples, but it was hard to know what to make of them, or what general lessons were there.

Seymour Sudman: We understand better why people misremember dates, and how they misremember dates. We understand much better than we did—this is one of Norbert's major achievements, how context affects work, and how they work in different sorts of ways, so that sometimes you see results and sometimes you don't, depending on whether the effects balance themselves out or not.

Tourangeau: I think we're almost to the point now, where you really could say it would be wiser to encourage people to estimate in a certain situation, and this would lead to more accurate answers. Or, it might be wiser to try and get them to recall every specific incident and sum that up. I think we're almost to the point, in certain areas, where we really have good theories that lead to very concrete recommendations about how to construct a set of questions. But that's the exception rather than the rule.

Sudman, Bradburn, Schwarz, and colleagues have gone further, by sponsoring a number of interdisciplinary conferences in recent years, to further enhance understanding of the response task. These conferences have not only added a great deal to available literature, but also fostered even more collaborations.

Seymour Sudman: In the initial one we were interested in bringing survey researchers and cognitive psychologists together. That has been our aim ever since, although the topics have changed. Sometimes we bring in people from linguistics. We've also looked for other areas, because, of course, surveys involve language. The intent has been to introduce people to each other who might not otherwise know each other—to let them hear each other in a fairly intimate setting, and let them talk over drinks and meals, and get to know each other's work. Perhaps they'll learn, as I have, from interactions with people in the other disciplines.

Shortcomings of the CASM movement

Many researchers hoped that cognitive psychology would receive as much from survey research as survey research received from cognitive psychology. Some have expressed disappointment that this may not have happened.

Judy Tanur: In CASM I we talked about a two-way street, where the insights, methods, theories of the cognitive sciences would be useful in improving survey research practice. We also hoped that surveys would come to be seen by people in the cognitive sciences as laboratories to test their theories. I have not seen very much of that. I would hope that we could think of ways of helping to implement that two-way street, following CASM II.

Yet, there has been some transfer back to cognitive psychology.

Norbert Schwarz: The survey interview has confronted cognitive researchers with a much richer social reality and social situation, in which people do their thinking and communicating, judging, and so on. That has broadened the agenda for basic research. But, what have we learned? We've learned quite a bit about basic research.

Norm Bradburn: I've been surprised. I didn't think (CASM) would have too much of an effect on psychology, but there are certain areas where at least some people have picked it up. The area of telescoping is one, because it's a phenomenon that is very puzzling to psychologists, and so it's quite interesting. It fits in with some other things that people obviously had been worrying about. My evidence for this is that, when I was working with Steve Shevell—who is purely a psychologist, with no particular interest in surveys—we published a paper on autobiographical memory in Science. It turns out that an enormous variety of people read Science. What surprised me was how quickly articles began to appear in psychological journals about telescoping. Because this happened so fast, this suggested to me that people had a lot of data sitting around and they didn't know what question the data addressed. All of a sudden they said "aha, the phenomenon is called telescoping."

There are other concerns about what has not been accomplished. For example, although cognitive laboratories offer new possibilities, they may be misused as shortcuts. These new tools were never intended to take the place of systematic, follow-up research.

Judy Lessler: I've heard people say, "We know this question is right because it was validated in the cognitive lab." This typically means that they did four or five think-aloud interviews, found some problems with the question, and changed it. I think that you've got to go farther.

Stephen Fienberg: The fact that something worked in an experiment that cognitive scientists did on undergraduate psych majors doesn't make the idea fully portable into the National Health Interview Survey, or the National Crime Survey, or in changing how we measure unemployment. For that, you have to do systematic work. You have to ask much more carefully how you make the progression from insights that you get from cognitive interviews, and very exploratory ideas in the labs, to working your way, scientifically, through building up systematic theory at the interface.

I don't see enough of that second part: the careful testing and real experimentation, with randomization, and full structures, bringing to bear what we, as statisticians, have been pushing in other ways, and indeed in the design of the surveys themselves. The intermediate step is what is being lost.

Another problem is a lack of perspective—which errors are most important to address?

Roger Tourangeau: I don't think people have a good sense of the relative magnitudes of the different sources of errors in a survey. People spend a lot of money on increasing sample sizes to reduce sampling error. It is just not clear whether that is a good investment relative to, say, doubling the money spent on pretesting a questionnaire. Likewise, within nonsampling errors, it's not clear whether you should spend more money getting a higher response rate or more money on, perhaps, new data collection methods, such as using diaries to reduce recall burden. Again, the relative magnitudes of the different types of errors are very poorly understood. It seems to me the a great leap forward might come if we had a better sense of what we ought to be worried about.

CASM may also focus too strongly on applications such as pretesting techniques, and not enough on an

intellectual agenda. Norbert Schwarz feels that CASM's accomplishments— and viability for the future— should be measured by how much our understanding of global issues has increased.

Schwarz: I don't think the right level of thinking is the level of *technique*. I think the right question for CASM research into the cognitive and communicative aspects of survey responding is: Has that research done anything good? I think the right level to look at is to ask: How do we think about survey responding now, and how did we thought about it ten years ago? If a technical by-product of this is a better pretest, then that's great—hopefully that would be the case. But if all we get is a more sophisticated research methodology, I think we have missed the boat.

If we address the bigger issues, and develop models for how people think about the world, how they evaluate political issues, how they answer questions, how they make sense of language, and all the big questions, then presumably we'll know how to write a good question, too.

Even then, if that's all it does, I think the CASM movement will soon be dead, as an intellectual endeavor. If it is going to survive as an intellectual endeavor, I think it has to view the bigger picture. (CASM) has to raise the broad questions, which are about human reasoning, human communication, and so on, and the survey interview provides a wonderful setting to study those.

New Challenges for the CASM Movement

In terms of goal setting for the future—where should our energies be directed? There is no shortage of ideas. Many discuss the possibility of expanding into other disciplines such as linguistics and anthropology. Another idea is expanding the application of cognitive science into areas other than questionnaire design.

Clyde Tucker: This would include conceptual development, prior to developing the actual questions, attention to analytical aspects of the data, once (the survey) is completed, and also looking at presentation—that is, how we present data. This is certainly an area in which cognitive psychology has been important—looking at the best way to present data to people, and how best they can use it and understand it.

Miron Straf: I've seen CASM being used in other ways than just on questionnaires. It has been used by NCHS, for example, on an atlas of mortality, which has to convey a lot of information through graphical displays and presentations.

Other issues include understanding the limits of survey response.

Norm Bradburn: I think we need to understand better what people can actually be expected to do with any degree of accuracy. Just because we can figure out a way to deliver a set of complex questions to them, this doesn't mean they can answer them.

Judy Lessler: I think that we have shown that there is a limit to what you can learn by asking questions. People always knew that, but it has become more salient. Also, we realized that there are some things that just aren't in memory. Either they were never there, or they've been lost over time. What is really needed is to understand the characteristics of questions, and then to relate that to the estimation at the end—in particular, to the modeling at the end. We need to then go back and see if we can make adjustments, and build mathematical models to deal with this.

And, to continue to develop theory about the response process.

Judy Tanur: I think that what was missing beforehand, and I think is still missing, is a real theoretical basis for the survey process. I think we have bits and pieces of it. The CASM movement has helped to offer those bits and pieces. I think we have a little better understanding. I think we still don't have a good theory of response effects or context effects. As I say, I think we're moving toward it. I would like to see considerably more theoretically-based research going on.

And also building bridges to new disciplines.

Norm Bradburn: I want to push into some other fields where I think exciting things have been going on. The one which I think is most relevant is linguistics—in particular, socio-linguistics, and the analysis of language. We need to bring in the tools of conversational analysis, socio-linguistics, and so forth. That's an area where I think some interesting things have been going on in the last ten years or so.

CASM may also open up the possibility of re-defining the role of the interviewer, to make survey interviewing a true collaboration with the respondent.

Seymour Sudman: My thoughts on that are a little bit radical. The old notion that the interviewer must be limited to asking the question exactly as written, and no way else, and to repeating it, and then simply letting the respondent interpret as they want— I think is a mistake.

I think that, for behavioral questions, it makes sense to try to be as explicit as possible. If people ask you what the question means, tell them what the question means, as the researcher wants it to mean. Then, you get the answer that you want, rather than assuming you have it.

Norm Bradburn: We still are strongly influenced by the earlier view that you want to keep the interviewer as inactive as possible—what I call the "robotic theory of interviewing," and that the perfect interviewer is a robot. But, if you can, you should get the interviewer to actually provide information that would help the respondent, rather than withholding this information.

The Role of CASM II in Meeting These Challenges

Will CASM II help to meet these challenges, and what are the objectives of CASM II?

Tom Jabine: One objective is to step back and see where we are. I would call this evaluation. Are the changes that have occurred in the process of questionnaire development really paying off? But the more important aspect, I think, is expansion of the CASM movement. In CASM I, the "second party," if you will, was mainly the cognitive psychologists. But there are many other related disciplines that have made great progress in recent years, and that should be brought into this effort—for example, expert systems, artificial intelligence, and a number of others.

Most important is the possibility of extending the application of the cognitive sciences and related disciplines to other aspects of surveys, beyond data collection. This starts with the planning of the survey, and then data collection, and it goes on to the review and processing of the data, and to the presentation of the results to users.

Monroe Sirken: I have felt all along, once I saw the light, so to speak, that the cognitive sciences were the sciences that we needed in order to break through on nonsampling errors, just as the mathematical and statistical sciences were the sciences that we depended upon to develop our sampling theory. From that viewpoint, I felt that it wasn't enough to let this go on by itself, and that it really was important for us to come together again and re-dedicate this as a fostered movement. I'm hoping that CASM II will serve that function.

APPENDIX B

SIPP Interview Summary

Karen Bogen

In preparation for the CASM II Seminar, the Census Bureau offered seminar participants a chance to be respondents to a survey interview. The interviews, conducted by regular Census interviewers (called field representatives, or FRs), involved the administration of the questionnaire for the Survey of Income and Program Participation (SIPP), and were conducted using Computer-Assisted Personal Interviewing (CAPI). These interviews had two main purposes: (1) to give participants exposure to the cognitive aspects of a survey interview, from the point of view of the respondent, in a way that this would be fresh in their minds at the time of the Seminar, and (2) to provide background material for the Working Groups on Income Measurement and Exploring the Interview Process.

The CASM II participants' experience was somewhat different from that of most regular SIPP respondents. Most participant interviews were conducted at the respondent's place of business, during the working day, because Census FRs conduct most of their "real" interviews in the evening, and could not devote that time for this extra assignment. Under these circumstances, many of the CASM II respondents did not have at their disposal the records that they might ordinarily refer to in order to respond to questions about income, assets, and other topics. A second difference is that, for most regular SIPP respondents, the initial interview is merely the first of a series of interviews administered at four-month intervals over a minimum of three years.

In spite of these important deviations from the regular SIPP interviews, we concluded that the CASM II interviews served their intended purposes very well. All of the SIPP respondents were sent a self-administered debriefing questionnaire, and invited to submit comments on their experiences as survey respondents. The 29 sets of comments that we received were abstracted in order to produce a report on which the one contained here is based, containing a summary for each debriefing question. This report was well-received at the Census Bureau, and was of particular interest to Field Division staff in charge of the administration of SIPP interviews.

The Interview Process

Could this interview be conducted in a less structured manner? Please explain.

Seminar respondents had a mixed reaction to the idea of a less structured SIPP interview—some thought it was feasible, but others did not. Those who thought it was feasible felt that a dialog clarifying the meaning and purpose of particular questions, and an initial open conversation about the topics of interest might provide a useful context for subsequently answering specific questions. Others were skeptical, in part because of the nature of the content (the collection of detailed facts about income), and in part because the SIPP would be difficult to analyze if the data were collected in a less structured manner. One respondent felt that structured lists of types of income were important for soliciting this information; lacking these, it is easy for respondents to fail to report particular sources of income.

One respondent related her agency's consideration of a plan to train interviewers to extract information from respondents' records, in lieu of conducting respondent interviews, and suggested this approach for the SIPP. She felt that her SIPP interview could have been done almost entirely from her records, in her absence. However, she acknowledged that this arrangement lacks practicality, given that respondents want to be in control of sensitive information such as income, and are unlikely to turn over their records for the interviewer to complete the survey.

The use of records was also the theme of another participant's opinion. He felt that, if accuracy of amounts reported is of concern, it is important to encourage the use of records. However, record use may necessitate a greater degree of flexibility in the way that information is recorded, and therefore imposes a less-structured approach.

Were you unsure how to answer some items? Was the interviewer able to help?

Almost all participants acknowledged some lack of confidence for at least a few items. One participant reported feeling unclear for about 5-10 percent of the questions, and although the interviewer offered some help, this was not sufficient, because of the requirement that she stick to the interview script.

On the other hand, respondents generally felt the interviewer was able to help by clarifying the purpose of the question or by guiding the respondent on what to include or exclude.

Was there any interviewer reaction to your inability to provide some information? Did he/she encourage/ discourage your taking the time to get the information immediately or taking the time to try to recall the information?

There appeared to be somewhat mixed responses about the interviewers' reactions, or perhaps in some cases, an underlying tone to the interviewers' messages. Although many participants reported that the interviewer was patient and did not rush them (e.g., waited for them to get records), others seemed to sense some feeling of being rushed, and that they were taking an inordinately long time to look up each answer. For example, one respondent said it was clear that the interviewer expected him to retrieve and use records, because she read such a request, and remained silent until he agreed to use them. Later, however, he had the feeling she was quite happy to have him produce quick estimates, from memory. Another respondent reported what might be impatience on the part of the interviewer, who defaulted to a "what if I were to call back later" strategy. However, this might be seen as an acceptable practice, given that most of what the respondent could not answer depended on records which were only available at home.

To what degree did the interview have a "conversational" tone? What effect did this have on the interview? Did the tone of the interview influence your responses?

Respondents were generally complimentary in their comments about the conversational tone of the interview. One respondent felt that this tone made it easier to ask questions and to take the time to search records, and left the impression that the questions were reasonable and appropriate. Another respondent felt that the conversational tone encouraged her to try harder. A third reported that the FR started out clearly reading a script, which made the respondent uncomfortable, and so the FR adopted a more conversational manner as the interview proceeded.

On the other hand, a few respondents did not feel that the interview tone was conversational. One said that the interview was dominated by the computer, which at times seemed to ask questions that surprised the interviewer. Another respondent noted that the flow of the interview was frequently interrupted by the interviewer's focus on operating the computer. A third respondent felt that the awkward wording of some questions resulted in a nonconversational tone, and a fourth felt that the interview tone was not a particularly important factor, compared to issues such as whether some of the questions made sense.

Did the interviewer give you the impression that questions (for clarification, for example) were welcome, or that they were an intrusion? How did the interviewer answer the questions? What was most helpful? What "message" did you get from the content and other aspects of the interviewer's handling of your questions?

Most respondents said that clarification questions were welcome. A few respondents commented on the interviewers' neutrality, remarking that they did not seem personally interested in the respondents' answers but did seem to want to get the best, most accurate answer.

Did the interviewer in any way encourage or discourage your use of records? How?

For most participants, who were interviewed at their offices, this was irrelevant. The interviewer likely did not expect that respondents interviewed at work would have records. One respondent who was interviewed at her office offered to get her checkbook to look something up (the only records she had at the office), but she was not encouraged to do so; the interviewer told her that record use was more common at the second interview.

For those few respondents who were interviewed at home, there was quite a lot of cooperation from interviewers, with respect to the use of records. One respondent reported that the interviewer told her the types of records she would likely need when she called to set up the appointment, and encouraged record use at the start of the interview. Another respondent commented that her interviewer read a statement asking her to pull together records and waited patiently while she did so. However, the respondent felt that the statement did not give much information on the types of records she would need—the interviewer was able to provide more guidance.

General comments about the interview process

One respondent commented on the importance of the interviewer, who added legitimacy to the project and also helped turn an otherwise mundane activity into something more important and worthwhile. The interviewer functioned as a collaborator, working with the respondent to answer the questions. This respondent noted that his interviewer used an apologetic tone when asking about food stamps recipiency, for example, and in fact admitted that she would not ask those questions in affluent homes, for fear of annoying the respondent.

Another respondent felt that many of the problems encountered in the interview could have been alleviated if the interviewer knew the "big picture" before heading down a particular path, and if the respondent similarly had a "big picture view" before answering some of the questions (this is akin to the idea of starting the interview with an unstructured discussion). The respondent gave as one example the fact that she found herself going down the self-employed path of the labor force section (she answered "Yes" because she fills out self-employment tax forms), even though the a later "moonlighting" question would have been more appropriate for her sporadic consulting income.

Another interesting comment concerned interviewer assumptions; the respondent clarified that, on the whole, she did not think that these assumptions were bad (for example, because the respondent gave an address that didn't have an apartment number, the FR asked, "And that's a private house?"). Another respondent noted a remarkable degree of deviation from strict standardization, particularly in the area of question reading, and in providing substantive help on requests for clarification. The interviewer even occasionally suggested answers (for example, when the respondent had trouble deciding whether his university was a for-profit or nonprofit entity). The respondent said he felt most comfortable when the interviewer was flexible, and most put-upon when she couldn't answer his request for clarification and left the interpretation up to him.

Finally, one respondent noted that, despite knowing something about survey research and about SIPP, he found it uncomfortable to be a survey respondent. He absolved the interviewer of blame, but reported that he was not at ease during the interview and resented being asked to furnish information which, under ordinary circumstances, he regards as private.

Recall and Knowledge

What strategies did you use to recall the information that was asked?

A common response to this question was "guessing." However, in most cases, respondents offered as informed guesses as were possible. For example, one participant reported that he estimated an amount per month, and multiplied by 12, to produce an annual amount.

Several respondents also noted that "recall" was so immediate and automatic for some items that they were unaware of any strategy at all. Some respondents used specific strategies for particular questions, noting that strategies were item-specific. For example, one mentioned knowing ancestry for three of four grandparents, but guessing between several choices for the fourth one. Also, for a question concerning weeks away from work in the past four months, one respondent said she first thought about holidays (e.g., Christmas) and then about trips she had taken. Another respondent said that her strategy was to visualize forms; likewise, another said she tried to retrieve from memory her most recent bank statement to report her interest earnings. One respondent noted that he looked at calendars, while others mentioned reliance on records or having recently completed tax forms.

Were there any instances where "recall" was irrelevant because the information simply was not represented in memory?

In response to this question, many respondents raised the issue of reporting proxy data, especially, but not limited to, adult children living at home—with respect to earnings data, details about others' jobs, and asset income. One respondent also mentioned being asked in great detail about a nonrelative currently living in his household. Many other respondents mentioned their lack of knowledge of income information in the period requested, and suggested that annual information was in memory, but other time increments were not. Another respondent mentioned that some income was reported only at the end of the year and that reporting of this information required access to records, rather than to memory. Similarly, a respondent suggested that dividend information simply could not be answered without detailed information from several bank statements and brokers.

Several respondents noted trouble with the questions on employer and establishment size (number of persons employed by the employer at all locations, and the number at the location where one works). Two respondents felt that the ranges used in the response categories helped a great deal, and one thought they should be included as a part of the original question, rather than simply offered when one starts struggling to provide an answer. Finally, another respondent mentioned that items asking for long-ago dates of events that have no bearing on current life (such as when one first worked six straight months at a job or business) may not be well-represented in memory.

Which items or types of information posed particular recall problems? How did the sequence of questions make it easier or harder for you to remember or report your answer? What could be done to make it easier for you to recall?

To improve recall, several respondents suggested that it would have helped to have been told ahead of time what would be asked (retirement plans, etc.), as this would have allowed them to refer to records and be better prepared to provide answers. Note that in production interviewing, FRs are fearful of this approach, because it may lead the respondent to refuse the interview. Another respondent felt that the greatest burden was remembering several months back, particularly for information from multiple record sources. He suggested that it would be useful to either limit the survey to questions that can be answered from a single source of information or type of record, such as tax returns for the relevant year, or to ask about only the most recent month so that details are fresh. One respondent suggested that, for an ethnic origin question, it might be easier to answer if the northern European groups were more clustered and not listed singly.

Several items were noted as particularly difficult (or impossible) to recall: date last paid; amounts received from various assets (some of which could not be reported even with records, such as separate amounts for dividends from stock, mutual funds, reinvested dividends); the difference between a money market account and a savings account; some historical questions (such as, "In how many of these 33 years did you NOT work 6 straight months?" and "During the time since 1963 that you have worked, have you generally worked 35 hours or more per week?"). One respondent felt that recall would be easier and data quality improved by asking for less detail, and not asking respondents to report the result of complex mental calculations. Finally, two respondents noted particular difficulty reporting gross pay, indicating that they could have instead reported net pay.

Several participants felt that the most serious failure of the interview does not involve recall difficulty, but rather, is asking for information that was too trivial to be of interest (e.g., interest on checking accounts), and trying to induce the respondent to report proxy information.

Data Quality

How would you rate the quality of the data you provided for yourself? What about for others in your household?

Not surprisingly, most respondents felt more sure about the quality of their own data than about the quality of the data they provided for others in the household, which one respondent reported for her husband as "wild guesses." Several respondents used general verbal descriptions for their data, calling it "good" quality or "pretty or fairly" good, or "pretty sure" or "pretty close"; two even said "high quality." One respondent felt that the variation in quality depended on the closeness of the detail asked for in the question to the detail represented in her records; if the detail asked for in the interview matched that in the record, data quality was good. If there wasn't such a match, quality was poor.

A few respondents felt that the areas with least accuracy were the ones that would least impact any analysis, and cited, as an example, interest on a checking account. One of those respondents said that he would have liked to have had a way of conveying his lack of confidence in particular pieces of information he reported.

One respondent also mentioned that she did not report "under the table" income, and said she would not disclose this in a real interview either. She acknowledged that such an omission would lower the quality of the data she provided. One respondent admitted that data quality suffered because the simplification of her reports, particularly about a complicated banking situation, simply made the interview easier for her, as a respondent. In lieu of explaining all the complexities, she lumped a number of sources together, knowing she was underestimating the actual value.

Did you ever consult pay stubs, bank statements or other records to answer the questions?

Respondents who were interviewed in their homes reported using their records at a generally higher rate than the SIPP is accustomed to. However, as mentioned earlier, most respondents were at work when the interview was administered, and therefore did not use records.

Were there any answers that you or you and the interviewer together had to compute "off instrument"? Did any problems arise because of this?

Several respondents cited "off-instrument" calculations, including the use of a calculator to add up monthly amounts that were asked for the four months together. No one noted any problems arising as a result of these computations.

As a rule, do you think the survey instrument is aiming for EXACT answers to the questions that ask for amounts, or do you think the aim is for more ballpark amounts?

Of the respondents who answered this question directly, five said "ballpark," two said "exact," and two others felt that the FR aimed for exact amounts, but would accept ballpark. A few other respondents said that the instrument must be looking for estimates. One felt the "estimate" message came from the interviewer's lack of encouragement of record use, and others suggested that this message was conveyed by the instrument itself, given that some of the questions are impossible to answer exactly, even with records.

General comments about knowledge, recall and quality

One respondent felt that despite having fairly full knowledge of his work history, income, and assets, he could only offer approximations, because the exact amounts simply were not in memory. With respect to proxy information, this respondent learned later that he was off by 50 percent or more on some income amounts, had misclassified some assets entirely, and made errors in the work history series.

Another respondent felt that some items, such as wages for herself and her husband, were simple to recall. But she also felt that their case was unusually simple (salaried employees with once-a-year changes in wages). Likewise, another respondent who felt she could answer most questions easily also admitted to having a relatively simply financial situation. One respondent commented that she was "dismayed" at the amount of estimation that goes into the responses, and the consequent lack of validity. Another noted the artificiality of four-month reference periods, and the fact that this period is not especially meaningful to respondents. He felt that this tends to produce error because the respondent does not take this seriously, and even if trying to, he or she may have difficulty recalling information without producing telescoping errors.

Question Wording

Did you need to ask for clarification of any questions? If yes, which questions?

Many respondents knew they had asked for clarification on at least some questions, but could not recall which questions those were. One respondent recalled wanting to report "Less than" some amount, not knowing the exact amount but knowing it was no higher than some amount. Another respondent needed clarification on a question about "other sources of income" because she was thinking about only wages when she heard "income." One respondent recalled asking for clarification about how to report capital gains distributions from mutual funds. Another asked for

clarification of "line of work" when asked how long he was in this occupation or line of work.

General comments about question wording

One respondent pointed out that there are "mismatches" between the task required of the respondent and the question wording. For example, for a question on assets, she was given a showcard with a list of assets and asked to indicate which assets she "owned." However, the FR then went on to ask a yes/no question concerning each of the listed assets. Therefore, it was not clear whether this was intended to be a question asking "which ones," or a series of "yes/no" items.

Another respondent had several general comments about question wording. He felt that some of the questions failed to communicate the degree of precision desired in the response. He suggested that in real-world conversation, the respondent decides how to answer a question depending on the context of the conversation and the perceived intention of the question asker. For example, if asked "Where do you live?" one might answer "The U.S." if at an international meeting, but would give our exact address if asked by a postmaster. He felt that the survey questions sometimes violated such norms or failed to provide sufficient cues for the respondent.

This participant also noted that interviewers often spontaneously reword questions to improve communication, without even realizing they are doing this. So, when asked how well the questions work, the interviewers may report that they seem fine - simply because they take for granted that they "fix them up a little" when they ask them. He suggested that SIPP survey designers consider using behavior coding procedures to track rewordings that do not change question meaning. He also pointed out that interruptions by the respondent are frequent in government surveys, and that these may not simply be a function of respondent behavior, but also of the interviewer. Interviewers unwittingly "signal" that an interruption is acceptable by "trailing off" in their reading of long questions (decreasing both their reading volume and enunciation).

An additional respondent felt that, overall, the language level seemed too complex. Phrases such as, "assets that provide income" and "gross amounts," and words like "royalties," do not seem like they would be readily recognized by the general population.

Income Measurement and Other Content Issues

Were there subject areas that you deemed "off-limits" for the interview?

Most respondents said not, but two said they refused to give their Social Security Numbers. Another said that she did not report her "unofficial income" (referring to money not reported to the IRS). Another said that income and expenditures are private, but she did not say she refused to answer. Finally, a respondent said that the subject of previous marriages was inappropriate.

Were there parts of the interview that were irrelevant to you? Did that affect other parts of the interview in any way?

A few respondents thought that questions on Welfare and program participation were irrelevant. One respondent found it peculiar to be asked about Medicaid and energy assistance immediately after answering items that would seem to verify that she was not eligible for those programs. Despite this, no participant reported that the irrelevant questions affected the rest of the interview. In fact, several respondents said that it was easy to answer "No" to those questions, and to then move on.

Were there questions for which it was unclear what types of income were supposed to be reported or for what time period?

One person mentioned a specific item here—her husband's income includes a military housing allowance which he reported as part of his salary. Another respondent felt that a question on monthly income question was unclear. By example, he said, "A person on a salary, or hourly with same number of hours each week, 'earns' the same amount each month, give or take a day or two. However, if the question is interpreted as paychecks received and they are paid every two weeks, it could vary significantly, since some months have two paydays and some three."

General comments about income measurement and other content issues

Another respondent pointed out her "foolish mistake" of stating that her 15- year-old son had income, which consisted of an occasional $20 for babysitting and yard work. This led to numerous questions intended for self-employed individuals, and a set of questions which were awkward and inapplicable and which the interviewer finally decided to answer himself (he presented them to the respondent in the form of statements). Two other respondents had almost the same experience. One had a teenage son who works between two and four hours per week and gets paid a token amount (less than minimum wage) for the work. The respondent had to answer many detailed questions about this work, and wondered whether there was a way to exclude this income. The other respondent also had a teenage son for whom he attempted to report babysitting work and income totaling at most $100, over a four-month period. The respondent suggested that it is impossible to report when exactly the money was earned, or the exact total. These respondents tended to regret having reported such small sources of income.

Several respondents felt that the interview sought to categorize sources of assets (and even some other sources of income) too specifically—where 3 or 4 categories might have been adequate, 10 or 12 were used. One respondent suggested that more thought needs to be given to the level of detail actually needed for analysis.

General Comments about the Interview

What are your immediate reactions to the interview? Were there positive features about the interview or the interview experience? What could be done better? What could be done to improve the interview?

General comments touched on the artificial nature of the SIPP interview, and in particular, the ways in which it was different from a true survey respondent's experience (including being interviewed at work, without access to records, by interviewers who knew the interview wasn't "real," and with the possibility that someone who they knew might overhear their responses).

Several respondents had strong negative reactions to the SIPP interview and made comments such as:

"...one of the most difficult, burdensome respondent experiences I have ever had."

"For this particular study, at least as implemented, the machine (CAPI) seems to have increased rather than reduced both the interviewer and respondent burden."

"It's hard to realistically imagine respondents' willingness to participate or provide high quality information."

"This interview schedule is very dry for respondents. It contains little, if any, context in order to provide motivation for the person who has to supply the data. This lack of motivation could lower response rate, but it could also decrease the effort that the

respondent exerts in order to provide accurate answers to the questions. As a respondent, I felt little loyalty to the project. I did not bother to look up records, nor did I have much concern that answers I gave for my husband's income had little basis other than guesswork. ... Few explanations are offered as to how personal data will be handled..."

Another respondent (actually, the wife of a respondent but who completed her own part of the interview) said that the interviewer's explanations were inadequate—for example, she did not understand why she had to be present during her husband's interview, nor did she think the interviewer's opening explanation of data uses was adequate.

Regarding positive features of the interview, several respondents mentioned the interviewer ("nice," "experienced and professional," and "I would hire her"), and several did mention the instrument ("relatively easy to answer and didn't take very long"; "it went quite smoothly"; "good use of showcards"). Overall, respondents reacted favorably to the use of show cards.

Regarding possible improvements, one respondent suggested that there be an "informative dialogue that clarifies technical constructs and terms." This seems related to another respondent's suggestion that we work "to improve ambiguous questions." This respondent felt that interviewers should have more basic information available to help answer respondents' questions, and perhaps on-line help. Other suggestions for improvement stressed the need for more emphasis on accurate and complete reporting using records, the need to address the issue of deliberate lying, and the need to cut out questions that are not essential to the goals of the survey.

Another type of suggestion was related to technical problems with the CAPI instrument, and in particular, interviewers' mentioning that there were inappropriate screens coming up. The resultant interruption disrupted the flow of the interview, and it was suggested that this might cause breakoffs in the interview.

Was length of the interview an issue for you? In what way?

There was mixed reaction to the issue of interview length, though there were fewer complaints than one might expect. Respondents who lived in households with more than two adults found the length to be significantly more onerous than persons with just one or two adults in the household. Of the respondents who answered this question, five said that it wasn't too long, and only one said it was very tedious. Note, however, that respondents who felt that the interview was unacceptably long may have been those who did not have time to complete the interview debriefing questionnaire. One respondent, who didn't think the overall length was excessive, had some

concerns about how long it takes to reconstruct income from records. Another felt that if someone found the subject matter to be acceptable, they probably would find the length to be acceptable, as well.

There were two complaints that it took a very long time to get to relevant questions (one respondent estimated 15 minutes spent on name, address, family, etc. before getting to the income and asset questions). Another noted that long surveys produce what he calls the "Just say no" phenomenon: Respondents learn that responding "Yes" to a screener question triggers additional questions, so they learn to avoid these. He admitted to doing this in his SIPP interview. Worse, the interviewer caught him (she stopped and asked "Is it really NO, or are you saying NO to avoid questions, like my real respondents do?" - he reported that he was).

One suggestion for decreasing length was to focus many of the questions on the household, rather than on the individual. The respondent suggested that this option should be available within the instrument and, presented to the primary respondent as a choice—especially with respect to items on food stamps, ADC, etc.

Comments about CAPI

One respondent felt that at times the machine was "randomly generating" questions. In contrast, one respondent felt as if his interviewer acted as though the questions were randomly generated, as evidenced by the confusion and difficulty she had with some of the questions. Another respondent noted that her interviewer was hopelessly stuck several times, making statements such as "Error? I guess something's wrong." Finally, another respondent noted "technical difficulties" that slowed things down, which she found "mildly aggravating."

One interviewer and respondent had a discussion about pre-CAPI versus CAPI interviewing. The interviewer told the respondent that in the pre-CAPI era, she been in the habit of making a list of the types of assets reported in the initial interview, and gave this to the respondent to facilitate preparation for subsequent interviews. However, under CAPI, the interviewer is unable to do this; once the interview is completed, she cannot re-access the questions. The respondent wondered whether it is possible to build in a feature that would make it possible to produce such a list.

Assorted comments about future interviews, respondent burden, level of detail, and suggestions for improvement:

One respondent wondered why others agree to be interviewed multiple times. She asked:

"While I know dependent interviewing tends to result in an underreporting of change, what's the goal of this instrument? If the interviewer came back to me in three or four months with a sheet summarizing my responses and asked me about any changes of over 10 percent, I would be far more inclined to participate than to sit through the same set of questions (and more questions with the supplements). Is it necessary to chase down every penny of change over time? How are analysts using these data? I think these questions need to be re-addressed in light of the burden of this instrument."

Another respondent felt that substantial improvements could be made by (1) reducing the level of detail for some questions, such as assets, (2) eliminating inapplicable questions, and (3) improving wording and flow for some questions.

One respondent, who had the impression that the survey instrument was not "friendly" to the interviewers, suggested that the Census Bureau conduct modified cognitive laboratory interviews in which SIPP interviewers are respondents.

The inaccuracy of her own reports made one respondent wonder what she would have done at subsequent interviews to be consistent in her inaccuracy. She had lumped together several bank accounts to make her reporting easier, and wondered if she'd remember to do so the next time.

The demographics section seemed "ponderous" to one respondent. In particular, she wondered why her address was verified twice (one was a home address, the other mailing address, which for most people, is the same), why "living quarters" weren't defined, and why her maiden name was of interest. She (as well as two other respondents) also thought that the Social Security Number was asked too early in the interview, and that the SIPP might obtain better response if it were asked later.

One wording change suggestion in the labor force section was to ask about "your main employer" first, instead of asking about "one of your employers." Another suggestion was to leave more to the interviewer's discretion, in order to eliminate sequences that ask kind of work ("oral surgeon"), followed by what the individual does on his job ("oral surgery").

One FR reported to a respondent that she encounters resistance when she asks about asset holdings in later interviews. People remind her that they've already told her, that nothing has changed, and they often refuse to repeat the information. Given the amount of estimating that goes on, asking for new asset estimates in each wave may create artifactual change.

One respondent acknowledged that the SIPP interview is structured in a way that must represent a compromise among the concerns of data users, data collectors, and respondents. Yet he wondered whether the interests of respondents had been given sufficient weight. For them, minimizing burden—defined not only in terms of duration but also in the cognitive difficulty of the content— would be of great interest. He also felt that, for some respondents, the level of burden would also be a function of the impressions they receive from interviewers about the importance of providing accurate information. If an interviewer does not stress accuracy and seems willing to accept rough estimates, some respondents might feel they are wasting their time. Specifically, he noted items related to income from interest, dividends, and other income-producing assets (as well as self-employed business expenses). As did several other respondents, he wondered whether it is really necessary, for analytical purposes, to know the amount of such income for each week or month.

One suggestion for maintaining the appropriate balance among the needs of researchers, data collectors (program managers and high-level officials of the data collection agency), and respondents was to convene a SIPP panel with representatives from all three groups who would be interviewed and asked to comment on their experiences, with special attention to issues of burden and data quality. For longitudinal or repeated cross-sectional surveys, panel members might submit to at least a full year of interviews.

Another respondent suggested adding some examples of how data might be used in news articles or other publicly available forums that the respondent might recognize. In particular, it would be useful to explain why the respondent's help will improve the world. In general, if one is asking the respondent for assistance, one should also provide more explanation of why we need the assistance. Note that this type of information is typically given in the "doorstep phase" of a SIPP interview, but was not important for the CASM II participant interviews, because interviewers did not need to "sell" the survey to CASM II respondents.

One respondent suggested that there be opportunities built into the interview to let the respondent obtain information, rather than only providing it. One could ask, for example, "Would you like to know more about the Census Bureau?"; or, "Would you like to know more about this project?" Although the interviewer is prepared to answer questions from recalcitrant respondents, the cooperative ones should also be given a chance to learn more about the organization and the project itself. Simple information might help the respondent "warm up" to the tasks requested. For example: "The Census Bureau is located in Suitland, Maryland, not far from Washington, DC"; "There are X interviewers like myself that ask these same questions for X projects going on this year"; "The Census Bureau conducts a survey of every citizen of the United States every ten years." Perhaps some of this material could be used to break up the interview, and to provide some temporary relief from respondent burden (e.g. "We're about halfway through the interview now. Let me ask you a quick trivia question before we continue. How often do you think the

Census Bureau interviews the entire U.S. population-- every year, every five years, every ten years, or never?").

Several respondents urged that the SIPP designers provide more rationale for why questions are asked. For example: "In this series of questions we ask about the work that people do for money. These questions will help us understand the types of work that Americans do and how much they work to earn a living." Such introductions should help the respondent feel more like part of a conversation, and less like a machine.

One respondent was very concerned about the level of language used, and suggested analysis of the questions for readability, and checking of words with respect to degree of common usage. If words are not commonly used, one could explain them to the respondent, or set a context that gives the respondent an idea of what the term means. She felt that it seems presumptive to read or show an item and to assume that if a respondent doesn't recognize it, that it must not be applicable.

Finally, one respondent suggested that researchers spend some time in the field observing Census Bureau interviewers. In particular, the suggestion was that Census staff go "to the heartland" to observe the reactions of typical respondents and to ask them to suggest improvements to the survey instrument. Likewise, she suggested that SIPP questionnaire designers take advantage of the interviewers' knowledge and experience, and ask them for suggestions to improve the interview.

APPENDIX C

BIOGRAPHICAL SKETCHES OF CASM II PARTICIPANTS

Note: Biographical sketches list participant's professional affiliations at the time of the CASM II Seminar.

Murray Aborn

Murray Aborn retired from federal service in 1990, after a one-year sabbatical spent as a consultant to the National Center for Education Statistics and as Visiting Scholar at American University, Washington, D.C. At that time he was both Senior Scientist and Program Director for Measurement Methods and Data Improvement, in the Division of Social and Economic Sciences, National Science Foundation. Prior to joining NSF, Dr. Aborn was Executive Secretary of the Behavioral Sciences, Pharmacology and Toxicology, and Public Health Training Committees of the National Institute of General Medical Sciences, and before that Executive Secretary of the Behavioral Sciences Study Section, Division of Research Grants, National Institutes of Health. His career included employment in the Department of Defense as a Research Psychologist in the Air Research and Development Command, as a Research Scientist in the Special Operations Research Office, and as Education Specialist in Economic Mobilization in the Industrial College of the Armed Forces. His two-and-a-half wholly academic years were spent on the faculty of the Psychology Department of Michigan State University. His publications include an edited volume on scientific communication in the information age, research articles in psychological and linguistic journals, and contributed papers in statistical society proceedings. Dr. Aborn holds the Ph.D. from Columbia University, is an elected member of Sigma Xi, and is a Fellow of three professional associations and two scientific societies. Although now retired, he remains active as research consultant, reviewer, and writer.

Paul Beatty

Paul Beatty is a survey methodologist in the Office of Research Methodology at the National Center for Health Statistics. He received a B.A. in English and Statistics, and an MA. in Applied Social Research, both from the University of Michigan, Ann Arbor. At NCHS, he works with researchers to develop and test survey instruments on phenomena such as self-assessed quality of life, toxic occupational exposures, quality of clinical breast examinations, and various health behaviors and conditions. In addition to questionnaire design and pretesting, his research interests focus on the psychological basis of survey response, and developing methods that balance the communication and measurement requirements of survey data collection. He has also worked at the Census Bureau, the Institute for Social Research at the University of Michigan, and Amnigon Enterprises, a marketing research firm near Detroit.

Bennett Bertenthal

Since January 1997, Dr. Bertenthal has been Assistant Director of the Social, Behavioral, and Economic Sciences Directorate of the National Science Foundation. Prior to his appointment at NSF, Dr. Bertenthal was a Professor of Psychology at the University of Virginia and Director of the Developmental Psychology Training Program. He began at the University of Virginia in 1979 and became a full professor in 1990. From 1988 to 1990, he was an associate editor of the journal *Developmental Psychology*. He was a member of the Human Development and Aging Review Panel at NIH from 1991 to 1996, and served as chair from 1994 to 1996. He was also Chair of the Program Committee for the 1997 Meeting of the Society for Research in Child Development, and is currently a Member-at-Large for Division 7 of the American Psychological Association.

Dr. Bertenthal received a B.A. in psychology from Brandeis University in 1971, an M.A. in developmental psychology from the University of Denver in 1976, and a Ph.D. in developmental psychology from the University of Denver in 1978. He was a postdoctoral fellow at the Brain Research Institute of UCLA Medical School from 1978 to 1979. Dr. Bertenthal is the author of over 70 publications on perceptual and cognitive development, developmental methodology, visual processing of motion information, and nonlinear modeling of posture and gait.

Karen Bogen

Karen Bogen has an M.A. in Applied Social Research from the University of Michigan and has 13 years of survey experience in a variety of settings. Since joining the Census Bureau's Center for Survey Methods Research seven years ago, she has worked on numerous questionnaire design research projects for the Survey of Income and Program Participation (SIPP) and is currently working on the redesign of the March Income Supplement to the Current Population Survey (CPS). She is also working on revisions to the March CPS and SIPP, and development of a new SIPP supplement, to address changing data requirements in light of recent welfare reform.

Norm Bradburn

Norman M. Bradburn, Senior Vice President for Research and former NORC Director, has served as Provost of the University of Chicago and is the Tiffany and Margaret Blake Distinguished Service Professor in the Department of Psychology and also the Harris Graduate School of Public Policy Studies, University of Chicago. A survey methodologist, Dr. Bradburn chairs the Committee on National Statistics of the National Academy of Sciences, National Research Council. From 1992-1994 he was chair of the Panel to Evaluate Alternative Census Methods for Census 2000 and Beyond, another committee of the National Academy. During 1988 to 1992, he chaired the National Academy's Board on International Comparative Studies in Education.

Dr. Bradburn is past president of the American Association of Public Opinion Research. He is a member of the American Academy of Arts and Sciences and is a fellow of the American Statistical Association. He has written extensively on cognitive aspects of survey response, asking sensitive questions, and recall errors in surveys. He is the author, with Seymour Sudman, of four books on survey methodology. In 1995 he, with co-authors Seymour Sudman and Norbert Schwarz, published a sixth book, *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*.

In 1985 Dr. Bradburn received the Helen Dinerman Award given by the World Association of Public Opinion Research for distinguished contributions to survey methodology, and in 1987 Bradburn and Sudman shared the American Association for Public Opinion Research Lifetime Achievement Award.

David Cantor

David Cantor is a Senior Research Associate at Westat, Inc. He has a masters degree in statistics and Ph.D. in Sociology from the University of Illinois, Urbana-Champaign. His methodological interests include the measurement and reduction of error in surveys and administrative record systems. Over the past 15 years, he has assisted in the design and evaluation of a number of federal surveys, including the National Crime Victimization Survey, the Current Population Survey, the Survey of Income and Program Participation, the Consumer Expenditure Survey, the Current Employment Statistics program and the National Medical Expenditure Survey.

Lynda Carlson

Lynda T. Carlson is Director of the Statistics and Methods Group, Energy Information Administration, U.S. Department of Energy. She received her Ph.D. in Political Science at the University of Illinois. Prior to becoming Director of the Statistics and Methods group for the agency, she was responsible for all the end use energy consumption surveys at EIA. In this capacity she was responsible for the use of exploratory techniques in the development of the Commercial Buildings Energy Consumption Survey, which involved collecting data from building owners and managers, a rarely surveyed group. As head of the Statistics and Methods group she is responsible for the ongoing development and implementation of high quality statistical and analytical techniques in all EIA surveys. EIA surveys are primarily establishment based and the use of cognitive techniques for this population will be an new endeavor for EIA.

Fran Chevarley

Frances M. Chevarley is a statistician in the Division of Health Interview Statistics (DHIS). She has a masters degree and Ph.D. in Mathematics from Duke University; she also did a post-doc in Biostatistics at the University of Washington, Seattle. At NCHS she worked first in the Division of Vital Statistics, where she analyzed mortality data. Currently she works in the Survey Planning and Development Branch of DHIS where she has been working with researchers on the evaluation of the CAPI National Health Interview Survey. In addition to her evaluation interests, her analytical interests include cancer epidemiology, health services research, health issues for women with disabilities, and women's reproductive health.

Cynthia Clark

Dr. Cynthia Z. F. Clark is Associate Director for Methodology and Standards at the Census Bureau with responsibility for statistical, methodological, technological, and computing research; for the application of survey and statistical methodology to the Bureau's economic, demographic, and decennial statistics programs; and for maintaining quality standards. Prior to coming to the Census Bureau in 1996, she spent six years at the National Agricultural Statistics Service. Previously she was Assistant Chief, Agriculture Division, responsible for statistical research and methodology for the census of agriculture program, a statistical policy analyst at the Office of Information and Regulatory Affairs in OMB, an economic statistician at the Office of Federal Statistical Policy and Standards in the Department of Commerce, and a mathematical statistician in the Statistical Research Division of the Census Bureau.

Dr. Clark has been an adjunct professor at American University, and an instructor at Drake University and the University of Denver where she has taught in the Mathematics and Statistics Departments. She received a Ph.D. in statistics from Iowa State University in 1977, an M.S. in statistics from Iowa State University in 1973, a M.A. in mathematics from the University of Denver in 1964, and a B.A. in mathematics from Mills College in 1963.

Fred Conrad

Frederick Conrad is a research psychologist in the Office of Survey Methods Research at the Bureau of Labor Statistics. In general, his work concerns improving the quality of survey data by applying ideas and techniques from cognitive psychology, cognitive science and human-computer interaction. Recently he has explored strategies to estimate the frequency of events; the use of expert systems to reduce survey classification errors; the costs and benefits of standardized versus conversational interviewing; procedures for coding verbal protocols in survey pretests; and usability of computer-based data collection and data dissemination tools.

He has a Ph.D. from the University of Chicago and was a postdoctoral research associate at Carnegie-Mellon University. Before joining the staff at BLS, he worked in the Artificial Intelligence Research Group at Digital Equipment Corporation.

Mick Couper

Mick Couper is an assistant research scientist in the Survey Research Center at the University of Michigan and a research assistant professor in the Joint Program in Survey Methodology. He holds a Ph.D. in Sociology from Rhodes University (South Africa), an M.A. in Applied Social Research from the University of Michigan and an M.Soc.Sc. from the University of Cape Town (South Africa). He has over 15 years' of experience in the design, implementation and analysis of survey research on a variety of topics and using different methods. He has taught several different graduate level courses in survey methodology. His current research focuses on various issues in survey data collection, including nonresponse, the role of the interviewer, and computer assisted interviewing. His research in the latter area has focused on usability and interface issues in the design and implementation of computer-assisted surveys.

Martin David

Martin Heidenhain David is professor of economics at the University of Wisconsin-Madison. His research has explored a variety of problems relating to taxation and transfer programs, involving extensive work in collecting, analyzing, and managing complex data and the design of information systems to support such data. His books include two that report on major data collection efforts, Income and Welfare in the United States (McGraw-Hill, 1962), and Linkage and Retrieval of Micro-economic Data (Lexington Books, 1974). He has served on the National Research Council (NRC) Panel on Statistics for Family Assistance and on the Panel on Research Strategies in the Behavioral and Social Sciences on Environmental Problems and Policies. He is a fellow of the American Statistical Association. He serves as adviser to the Statistics of Income Division of the Internal Revenue Services; was a member of the Committee on National Statistics of the NRC where he contributed to Panel reports on the Survey of Income and Program Participation and Privacy and Confidentiality; and a member of the Advisory Board of the German Socio-economic Panel Study, which is conducted by the Deutsches Institut für Wirtschafts-forschung. He received an A.B. degree from Swarthmore College and M.A. and Ph.D. degrees from the University of Michigan.

Terry DeMaio

Theresa J. DeMaio is a social science statistician at the Center for Survey Methods Research, U.S. Bureau of the Census. During her 21 years at the Census Bureau, she has conducted research in two areas—survey nonresponse and questionnaire design. Her current interests are questionnaire design for the decennial census and the development and evaluation of questionnaire pretesting methods. She has also been involved in using cognitive research methods to test and revise demographic and economic questionnaires.

Cathy Dippo

Dr. Cathryn Dippo is currently the Assistant Commissioner for Survey Methods Research at the Bureau of Labor Statistics. Her staff of mathematical statisticians and behavioral scientists conduct theoretical and applied research projects aimed at improving methods used in the Bureau's household and establishment surveys. She was responsible for the establishment of the Behavioral Science Research Laboratory at BLS and led the joint effort of BLS and Census to develop and implement a new computer-assisted data collection instrument for the Current Population Survey. Her current activities are aimed at improving the availability of and access to survey methodological metadata, especially via Internet, and the development of Internet-accessible statistical information-seeking analytical tools for use by the general public. She is a Fellow of the American Statistical Association (ASA), former officer of the Survey Research Methods Section of ASA, and former President of the Washington Statistical Society. She is a co-editor of the recently published book on *Survey Measurement and Process Quality.*

Cheryl Eavey

Cheryl Eavey is Program Director of the Methodology, Measurement, and Statistics Program at the National Science Foundation. She also is the Coordinator of the NSF Human Dimensions of Global Change initiative. Dr. Eavey earned a B.S. in mathematics and political science from Valparaiso University and an M.A. and Ph.D. in political science from Michigan State University. She has been at the Foundation since 1993. Dr. Eavey's research interests include the experimental testing of game-theoretic solution concepts and the role of non-self-interested motivations and institutional constraints on the outcomes of group decision making processes.

Fran Featherston

Fran Featherston is a Senior Social Science Analyst at the U.S. General Accounting Office (GAO) in Washington, D.C., where she has worked since 1985. Her job responsibilities include designing and implementing surveys for reports issued to the U.S. Congress. Her survey work emphasizes mail surveys, including establishment surveys as well as surveys of individuals. In addition, she conducts data analyses of the survey results, conducts focus groups, and teaches research courses for GAO's Training Institute. Fran's research interests include respondent burden and mail survey administration. Her current work looks at the effects of cognitive pretesting on mail survey design.

Before coming to GAO in 1985, Fran worked for the State of Washington for three years as a Research Specialist with the Office of the Administrator for the Courts. In 1982, Fran received her Ph.D. from the University of Michigan's Political Science Department, using data from the Detroit Area Study for her dissertation.

Charles Fillmore

Charles J. Fillmore is Professor of Linguistics (Emeritus) at the University of California. Throughout his career (10 years at the Ohio State University and 25 years at the University of California) his research and teaching interests have focused on syntax and semantics, both general and descriptive. His semantic interests have centered on questions of text understanding and lexical analysis; he is currently director of a three-year NSF-sponsored research project in computational lexicography, designing a lexical resource that makes explicit the semantic and syntactic combinatorial properties of several thousand English words.

Eleanor Gerber

Eleanor Gerber received her Ph. D. in anthropology from the University of California, San Diego, in 1975. Her theoretical area of specialization is cognitive anthropology. She taught anthropology and interdisciplinary social science to undergraduates until joining the Census Bureau in 1992. Since then, she has specialized in questionnaire design and development. Research in which she has participated for the Census Bureau has included studies of the way respondents understand residence, roster strategies for self-administered questionnaires, the use of vignettes in cognitive interviewing, coverage research, and enumerating clients at shelters and soup kitchens.

Art Graesser

Dr. Arthur Graesser is presently a full professor in the Department of Psychology and an adjunct professor in Mathematical Sciences at The University of Memphis. He is currently a co-director of the Institute for Intelligent Systems. In 1977 Dr. Graesser received his Ph.D. in psychology from the University of California at San Diego. He was a visiting researcher at Yale University in 1983, Stanford University in 1984, and Carnegie Mellon University in 1991. Dr. Graesser's primary research interests are in cognitive science and discourse processing. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, memory, expert systems, and human-computer interaction. In addition to publishing approximately 160 articles in journals and books, he has written two books and has edited four books.

Bob Groves

Robert M. Groves is Director of the Joint Program in Survey Methodology, based at the University of Maryland, an NSF-sponsored consortium of the University of Maryland, University of Michigan, and Westat, Inc.. He is a Professor of Sociology at the University of Michigan and Research Scientist at its Institute for Social Research. At the Michigan Survey Research Center he is a member of the Survey Methodology Research Program. From 1990-92 he was an Associate Director of the U.S. Census Bureau, on loan from Michigan. He is the author of *Survey Errors and Survey Costs* (Wiley, 1989) and (with R. Kahn) of *Surveys By Telephone* (Academic Press, 1979); chief editor of *Telephone Survey Methodology* (Wiley, 1988); one of the co-editors of *Measurement Errors in Surveys* (Wiley, 1991); and author of many articles in survey and statistical methodology.

Dr. Groves has over 20 years of experience with large- scale surveys. He has investigated the impact of alternative telephone sample designs on precision, the effect of data collection mode on the quality of survey reports, causes and remedies for nonresponse errors in surveys, estimation and explanation of interviewer variance in survey responses, and other topics in survey methods. His current research interests focus on theory-building in survey participation and models of nonresponse reduction and adjustment.

Brian Harris-Kojetin

Brian A. Harris-Kojetin received his Ph.D. in social psychology from the University of Minnesota. He is a Research Psychologist in the Office of Survey Methods Research at the Bureau of Labor Statistics. His research focuses on issues of survey data quality with particular emphasis on survey nonresponse and proxy reporting. He has further research interests in interviewer performance, the evaluation of interviewer training procedures, and interviewer-respondent interactions, as well as motivational and organizational issues surrounding technological changes, such as the transition to computer-assisted survey information collection.

Doug Herrmann

Dr. Herrmann was trained as an engineer at the U.S. Naval Academy (B.S., 1964) and as a psychologist at the University of Delaware where he obtained an M.S. (1970) and a Ph.D. (1972) in Experimental Psychology and at Stanford University where he engaged in postdoctoral study on the mathematical modeling of thought processes. Subsequently, he taught at Hamilton College (Clinton, New York) until 1989, where he was a full professor and Chairperson of the Psychology Department. He also was a research fellow at: England's Applied Cognitive Psychology Unit in Cambridge, the University of Manchester, and at the National Institute of Mental Health. In 1990 he became the founding Director of the Collection Procedures Research Laboratory at the Bureau of Labor Statistics, after which he served as the Special Assistant on Cognitive Psychology to the Associated Director for Research and Methodology at the National Center for Health Statistics. Since 1995 he has been Chairperson of the Psychology Department at Indiana State University. Dr. Herrmann has written or edited several textbooks concerned with the psychology of

memory. He was co-editor of the *Journal of Applied Cognitive Psychology*, 1987-1991, and currently is the Editor of the new journal *Cognitive Technology*.

Tom Jabine

Thomas B. Jabine is a statistical consultant who specializes in survey research methods, response errors, sampling, and statistical policy. He was formerly Statistical Policy Expert for the Energy Information Administration, Chief Mathematical Statistician for the Social Security Administration, and Chief of the Statistical Research Division of the Bureau of the Census. He received a B.S. in mathematics and an M.S. in economics and science from the Massachusetts Institute of Technology. He has taught several courses in questionnaire development and has numerous publications in the fields of survey methodology and sampling. He is a fellow of the American Statistical Association, an elected member of the International Statistical Institute, and served as President of the Washington Statistical Society in 1979-80. As a consultant to the Committee on National Statistics, he attended the first CASM conference in 1983. He participated in the planning and follow-up activities for the conference and was a co-editor of its final report, *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines.*

Dan Kasprzyk

Daniel Kasprzyk is the program director for the Education Surveys Program in the National Center for Education Statistics at the U.S. Department of Education. He is involved in the design, implementation, and analysis of a number of cross-sectional survey programs funded by NCES, including: The Schools and Staffing Survey, a periodic system of surveys of school districts, principals, schools, and teachers, the National Study of Postsecondary Faculty, the Private School Survey, and the National Household Education Survey, a random digit dial survey of the U.S. population. He is also the Department of Education's liaison to the Committee on National Statistics Panel on Estimates of Poverty for Small Geographic Areas. Prior to taking a position at NCES, he spent over ten years in various capacities working on the Survey of Income and Program Participation and its predecessor, the Income Survey Development Program. He has a Ph.D. in Mathematical Statistics from The George Washington University.

Jim Kennedy

James Kennedy received his Ph.D. in Social Psychology from the University of North Carolina, Chapel Hill, in 1992. He works in the Statistical Methods Group (SMG) of the Office of Compensation and Working Conditions (OCWC). Most of his work at BLS centers around development of a new integrated wage survey for OCWC. For instance, he is conducting research into replacing subjective methods with empirical methods for reviewing data, and with constructing valid definitions of occupational levels. He has developed methods for modeling complex survey data sets with neural networks and cluster analysis, in order to build editing and review utilities. He has developed a social simulation algorithm for nonlinear optimization, the particle swarm algorithm, and is interested in other examples of adaptation in natural and artificial systems. His background is in experimental laboratory research, measurement and formation of attitudes, and interpersonal social influence.

Judy Lessler

Judith T. Lessler is the Director of Research Triangle Institute's Statistical Research Division.   Dr. Lessler is a leader in the development of cost-effective survey designs. Her experience spans theoretical investigations on frame construction and multiple frame designs, sample design. and nonsampling errors. She has published widely on the statistical theory of surveys. Dr. Lessler has devoted much of her career to developing innovative solutions for survey design problems and has also been a leader in the initiative to use cognitive laboratory methods for studying measurement error and for redesigning questionnaires. She collaborated with Monroe Sirken and others on the initial study exploring the use of cognitive laboratory methods for designing survey questionnaires, and this project led to widespread use of cognitive science in survey research. More recently, Dr. Lessler has participated in the development and testing of audio computer-assisted self interviewing (Audio-CASI) for gathering information on sensitive questions.

Dr. Lessler develops the design, sample sizes, and sample allocations based on precision constraints for a variety of research studies. She also designs procedures for imputation of missing data, and examines measurement error by assessing inconsistencies in data. In addition to her expertise in mathematical statistics, Dr. Lessler has also been actively involved in questionnaire development, cognitive testing, and the design and analysis of field experiments that compare alternative survey methods.

Stephan Lewandowsky

Stephan Lewandowsky obtained his B.A. from Washington College, Chestertown, MD in 1980. He obtained his Ph.D. from the University of Toronto in 1985. From 1990-95 he was Assistant and then Associate Professor of Psychology at the University of Oklahoma, before moving to the University of Western Australia in Perth, where he now holds the position of Associate Professor. Stephan has been conducting research on several basic issues in human cognition and memory, in particular computer modeling of serial and associative recall and categorization. His applied research interests include the perceptual and cognitive processing of statistical graphs and maps.

Kent Marquis

Kent H. Marquis received his B.A. in Psychology from Yale University, and his Ph.D. from the University of Michigan in Social Psychology. Since 1985, he has worked as a Research Psychologist at the U.S. Bureau of the Census. Between 1981-1985, he was the Chief of the Center for Survey Methods Research at the Census Bureau. He has also worked at the Michigan Survey Research Center as a Study Director, as a Senior Social Scientist at The Rand Corporation, and as Associate Director of the Statistics Research Division at the Research Triangle Institute. His professional interests include new data collection methods, total survey error, human factors and usability evaluation, sensitive topics, measurement, response error, nonresponse, interviewer-respondent interaction, proxy response, recall, income reporting, incentives to participate, heuristic response strategies, record checks, behavior coding, reinterviews, panel measurements, interrogation, suggestibility, question length, and eyewitness testimony.

Nancy Mathiowetz

Nancy A. Mathiowetz is an Assistant Professor in the Joint Program in Survey Methodology at the University of Maryland. Prior to joining JPSM she was the deputy director, Division of Statistics and Research Methodology, at the Agency for Health Care Policy and Research. A graduate of the University of Wisconsin (B. S. in Sociology) and the University of Michigan (M.S. in Biostatistics; Ph.D. in Sociology), her primary research interests have been in the measurement and reduction of nonsampling error. She is co-editor of *Measurement Errors in Surveys* (John Wiley & Sons, 1991) and associate editor of the *Journal of Official Statistics.* She has published articles on various topics related to assessing the quality of survey data, including the effects of length of recall and the impact of mode and method of data collection on survey estimates. This work has appeared in journals such as *Public Opinion Quarterly*, the *Journal of Business and Economic Statistics*, and the *Journal of Official Statistics.* Her current research efforts are directed at understanding the impact of measurement error on substantive economic models and the development of a model to assess the cost-error tradeoffs related to pursuing difficult-to-interview individuals.

Jeff Moore

Jeff Moore received a B.A. from the University of Minnesota, and his M.A. and Ph.D. in Psychology from the University of Michigan. He is currently a Research Psychologist at the Center for Survey Methods Research, in the Statistical Research Division of the U.S. Bureau of the Census. He interests include: survey response error, self/proxy response effects, and questionnaire design and evaluation.

Colm O'Muircheartaigh

Since 1991, Colm O'Muircheartaigh has been the Director of the Methodology Institute at the London School of Economics and Political Science, UK. He has previously directed the Joint Centre for Survey Methods at Social and Community Planning Research, in London, and has held positions as a Research Fellow at the U.S. Bureau of the Census, at the Institute of Public Administration, Dublin, Ireland, and at University College Cork.

His major professional interests and activities include survey sampling, measurement error models, and cognitive aspects of survey methods. He has been a survey practitioner since 1971 and a university teacher since 1968. With respect to the former, has served as a consultant to a wide variety of public and commercial organizations, including the International Education Association, OECD, British Household Panel Survey, UK Serious Fraud Office, US Bureau of Labor Statistics, Italian Statistical Institute, Chinese State Statistical Bureau, United Nations (FAO, UNESCO), the Commission of the European Communities, UK Monopolies and Mergers Commission, BBC World Service, AGB, UK Law Society, and Westat. With respect to the latter, he has held short-term appointments include Visiting Professor at the Universities of Padua, Perugia, Florence, Bologna, and Visiting Research Scientist and Visiting Associate Professor at the University of Michigan; short courses have been offered at a variety of institutions, including the Institute of Social Studies in The Hague, Georgetown University's London Programme, the UK Department of

Education and Science, the UK Market Research Society, and the University of Michigan/University of Maryland/Westat Joint Program in Survey Methodology.

Polly Phipps

Polly Phipps is a Sociologist in the Office of Survey Methods Research at the Bureau of Labor Statistics. Her interest in survey research dates back to graduate school at the University of Michigan, where she spent several years working as the Associate Director of the Detroit Area Survey. While at the Bureau of Labor Statistics, her research has focused primarily on establishment surveys. She has investigated measurement error across numerous establishment surveys and is currently conducting research on how establishment characteristics and alternative modes of data collection affect data accuracy. She is also studying the effect of organizational and individual characteristics on establishment survey nonresponse. Her other research includes a project investigating cognitive processes in reporting time use and a field experiment testing alternative 24-hour time use diaries.

Jim Press

Dr. S. James Press received his Ph.D. in statistics in 1964 from Stanford University. He has been in the Department of Statistics at the University of California, Riverside for the last twenty years, and served as its Chairman for seven of those years (1977-1984). His current title is Distinguished Professor. Prior to 1977 he was affiliated with the Universities of British Columbia (1974-1977), the University of Chicago (1966-1974), and the University of California, Los Angeles (1964-1966). He spent one-year leaves at the London School of Economics and Political Science, and University College London jointly (1970-1971), Yale University (1972-1973), and Stanford University (1984-1985). He was employed full-time at The Rand Corporation (1964-1966), and has consulted there part-time for about twenty-five years. Prior to 1964 he spent ten years in the aerospace industry as an engineer/mathematician/ physicist. His research over the years has been motivated largely by problems in the social sciences, including questionnaire design and the use of statistical methods in: focus groups, community attitude surveys, jury research, and problems in civil litigation; and questionnaire design and analysis applied to problems in a military context. His current research interests are focused on Bayesian methods and multivariate analysis applied to sample surveys, meta-analysis, and factor analysis. He is the author, coauthor, or coeditor of 14 books, research monographs, and manuals, and about 150 papers in refereed journals. He was a member of the CASM I conference.

Stan Presser

Stanley Presser is a professor in the Joint Program in Survey Methodology and in the Sociology Department at the University of Maryland, where he directs the Maryland Survey Research Center. His research interests center on questionnaire design and testing, the accuracy of respondent reports, and ethical issues, such as confidentiality and informed consent.

Lance Rips

Lance Rips is a cognitive psychologist with research interests in the areas of human inference, concepts, and autobiographical memory. He received his Ph.D. from Stanford University and has taught at the University of Chicago and Northwestern University, where he is currently professor in the Psychology Department and in the School of Education and Social Policy. His research has focused on the cognitive processes people use to determine when events occurred and on the effects mental calendars and schedules have in prompting recall of personal events. Other research interests have to do with people's understanding of two person arguments, their ability to interpret evidence, and their manner of assigning burden of proof. Rips is the author of *The Psychology of Proof*, a study of human deductive reasoning, and he is currently completing a book with Roger Tourangeau and Kenneth Rasinski on cognitive processes in survey responding.

Trish Royston

Patricia Royston is the Deputy Director of the Division of Information and Analysis, Office of Planning, Evaluation, and Legislation, Health Resources and Services Administration. She serves as the Agency reviewer for all primary data collection activities sponsored by HRSA, and frequently consults with program staff on the design of surveys. Prior to 1990, she worked in the Office of Research and Methodology, National Center for Health Statistics, where she had primary responsibility for establishing and managing the Questionnaire Design Research Laboratory,

which was the first permanent laboratory in a federal statistical agency to apply cognitive research methods to questionnaire design.

Nora Cate Schaeffer

Nora Cate Schaeffer is Professor of Sociology at the University of Wisconsin, Madison,   where she teaches courses in survey research methods and conducts research on issues in survey design and questionnaire development. She has taught questionnaire design at the Summer Institute of the Survey Research Center at the University of Michigan and through the University of Michigan-University of Maryland-Westat Joint Program in Survey Methodology. She has served on the Panel to Evaluate Alternative Census Methods for the National Research Council of the National Academy of Science, on the American Statistical Association Technical Advisory committee on the Survey of Income and Program Participation, on the National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences, and on the governing Council of the American Association for Public Opinion Research. She has been a member of the editorial boards for *Public Opinion Quarterly*, *Sociological Methods and Research*, and *Sociological Methodology*.

Her articles have appeared in *Journal of the American Statistical Association*, *Journal of Marriage and the Family*, *Public Opinion Quarterly*, *Sociological Methods and Research*, *Sociological Methodology*, and edited volumes.

Susan Schechter

Susan Schechter received her BA at the University of Maryland and her MA in Human Development Research at Antioch University. She began her Federal career at the Census Bureau, and then worked for the Department of Defense at the Army Research Laboratory. In 1992, Ms. Schechter joined the National Center for Health Statistics as Survey Statistician on the Cognitive Methods Staff. At NCHS, she directs questionnaire development and testing for a variety of surveys, and has conducted many cognitive interviews and focus groups with subjects recruited in the Questionnaire Design Research Laboratory. She has worked on the National Immunization Program, the National Survey of Family Growth, the National Health Interview Survey, and the Behavioral Risk Factors Surveillance System Survey. Ms. Schechter also currently directs several extramural research efforts with university scientists and researchers in private industry. Of note are two projects, one that is investigating racial and ethnic self-identification of mixed-race women and a second that is investigating the types of cognitive functioning questions that can be asked of elderly survey respondents on the telephone. She has authored publications and spoken at national conferences and government seminars about ways to reduce survey response error through improvements in questionnaire design.

Michael Schober

Michael Schober is an assistant professor of psychology at the Graduate Faculty of the New School for Social Research. He received his Ph.D. from Stanford University and taught at Occidental College for two years as a Pew Foundation postdoctoral fellow. His research interests include the mental processes involved in conversational coordination, perspective taking in conversation, how people comprehend disfluent speech and how people interact with computers. In collaboration with Fred Conrad at the Bureau of Labor Statistics, where he was recently a Senior Research Fellow, he has been studying how interviewers' deviation from standardized scripts affects survey response accuracy.

Norbert Schwarz

Norbert Schwarz is Professor of Psychology at the University of Michigan and Research Scientist in the Survey Research Center and Research Center for Group Dynamics of Michigan's Institute for Social Research. Prior to joining Michigan in 1993 he was on the faculty of the University of Heidelberg, Germany, and directed the Cognition and Survey Methodology Program at ZUMA, Mannheim, Germany. His research interests focus on human judgmental processes, including their implications for survey research. His recent publications in this area include *Cognition and Communication* (Erlbaum, 1996), *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology* (with S. Sudman and M. Bradburn; Jossey-Bass, 1996) and *Answering Questions* (edited with S. Sudman; Jossey-Bass, 1996).

Monroe Sirken

Monroe Sirken is a senior research scientist at the National Center for Health Statistics (NCHS). With a background in mathematical statistics and the social sciences, he conducts interdisciplinary survey research at the intersection of those disciplines. His current interests include network sampling, integrated sample survey design, cognition and survey measurement research, and modes of fostering interdisciplinary research. He is a recent recipient of the Roger Herriot Award for innovations in federal statistics. He organized and directed the CASM II Seminar. He has a B.A. and M.A. from UCLA, and a Ph.D. from the University of Washington. He was a Social Science Research Council Post Doctoral Fellow at the Statistics Laboratory, University of California, Berkeley. He taught at the University of Washington Medical School, and in biostatistics departments at Schools of Public Health at the University of California, Berkeley, and at the University of North Carolina, Chapel Hill. His government service began at the Bureau of the Census, but NCHS has been the venue for most of his career.

Eliot Smith

Eliot R. Smith is Professor of Psychological Sciences at Purdue University. He received his Ph.D. from Harvard University, and joined the Purdue faculty in 1982. The author of over 60 articles and book chapters, Smith's current research interests center on social cognition and new connectionist models of mental representation and process. He has served as Associate Editor of the Attitudes and Social Cognition section of *Journal of Personality* and *Social Psychology*, and has chaired the Social, Personality, and Group Processes grant review committee of the National Institutes of Mental Health. He is a Fellow of both the American Psychological Association and the American Psychological Society, and serves on the editorial board of several major journals including *Psychological Science*. Smith is coauthor (with James Kluegel) of the monograph *Beliefs About Inequality*, and also of two textbooks: *Research Methods in Social Relations* (with Charles Judd and Louise Kidder) and *Social Psychology* (with Diane Mackie).

Ed Sondik

Edward J. Sondik was appointed as Director for the National Center for Health Statistics (NCHS) in April 1996. NCHS, part of the Centers for Disease Control and Prevention, is the principal health statistics agency of the United States with responsibility to monitor the nation's health. At NCHS, Dr. Sondik directs a wide-ranging program of research and analysis in health statistics and epidemiology.

Dr. Sondik also serves as Senior Advisor to the Secretary of Health and Human Services on Health Statistics, providing technical and policy advice on statistical and health information issues. Before Dr. Sondik's appointment as NCHS Director, he served in the National Cancer Institute in a number of positions including Acting Director and Acting Deputy Director, Deputy Director of the Division of Cancer Prevention and Control, and Associate Director of the Surveillance Program. He received B.S. and M.S. degrees in electrical engineering from the University of Connecticut and a Ph.D. in electrical engineering from Stanford University. Prior to his federal government career, Dr. Sondik served on the faculty of the Department of Engineering-Economic Systems at Stanford University.

Jaki Stanley

Jaki S. Stanley is a psychologist currently working for the National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture. She received an ScB in Cognitive Science from Brown University. She also received an MA and Ph.D. in Experimental Psychology from the Catholic University of America. During her career at NASS she has worked in questionnaire design and development, interviewer training, pretesting and cognitive interviewing. She has recently become head of the Survey Quality Research Section in NASS' Research Division where her work will be focused on reducing respondent burden and increasing survey data quality. Although she lives and works in Virginia, her work with agricultural survey respondents (farmers, ranchers and agribusinesses) has taken her throughout the country.

Miron Straf

Miron L. Straf is the Director of the Committee on National Statistics at the National Academy of Sciences, National Research Council. Previously, he was on the faculties in the Departments of Statistics at the University of California, Berkeley, and the London School of Economics and Political Science. He holds a Ph.D. in Statistics from the University of Chicago and Bachelor's and Master's degrees in Mathematics from Carnegie Mellon University. Dr. Straf is a fellow of the American Statistical Association and the Royal Statistical Society and an elected member of the International Statistical Institute.

The Committee on National Statistics contributes to a better understanding of important national issues by working to improve the statistical methods and information on which public policy decisions are based. Dr. Straf has developed Committee studies in many areas of national policy, including education, immigration, foreign trade, international finance, defense, agriculture, transportation, energy, health care, occupational safety and health, disability, public health, retirement, poverty, social welfare, and the environment. He has worked to improve programs of many federal statistical agencies, statistics collected on children and the aging population, national income accounting, statistical evidence in the courts, methods for confidentiality and meta-analysis, sharing of research data, and the decennial census. In 1983, Dr. Straf was the study director of the Committee's Advanced Research Seminar on Cognitive Aspects of Survey Methodology.

Judy Tanur

Judith M. Tanur is Distinguished Teaching Professor of Sociology at the State University of New York at Stony Brook, where she has been on the faculty for 30 years. She holds a B.S. in Psychology and an M.A. in Mathematical Statistics from Columbia University and a Ph.D. in Sociology from Stony Brook. Her honors include: Phi Beta Kappa; Fellow, American Statistical Association Fellow, American Association for the Advancement of Science; Elected Member, International Statistical Institute; Stony Brook's President's and SUNY Chancellor's Awards for Excellence in Teaching.

Her major publications include *Statistics: A Guide to the Unknown* (edited with Frederick Mosteller, William Kruskal, and others), *The International Encyclopedia of Statistics* (edited with William Kruskal), *Cognitive Aspects of Surveys: Building a Bridge Between Disciplines* (edited with Thomas Jabine, Miron Straf, and Roger Tourangeau), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys,* and a series of papers with Stephen Fienberg on the parallels in design and analysis between sample surveys and the design of experiments, for a total of 5 books, over 40 published papers, and numerous presentations at conferences. Professor Tanur has served on the Committee on National Statistics of the National Academy of Sciences and on several of its panels, as chair of two sections of the American Statistical Association and two of the American Association for the Advancement of Science, as Book Review Editor of the *Journal of the American Statistical Association*, co-editor of *Chance*, and on numerous other professional committees, panels, and boards.

Roger Tourangeau

Roger Tourangeau, Senior Methodologist at the Gallup Organization, has been a survey researcher for more than 17 years. Before joining Gallup, Tourangeau was a Research Vice President at NORC, where he established the Statistics and Methodology Center and was responsible for the design and selection of NORC's national sampling frame.

Tourangeau is known primarily for the application of theories and methods drawn from cognitive psychology to the reduction of nonsampling errors in surveys. He served as a consultant to the National Academy of Sciences, helping to organize the first seminar on Cognitive Aspects of Survey Methods and co-authoring the report on the seminar. He developed a leading model of the survey response process and, in collaboration with Ken Rasinski, Norman Bradburn, and others, applied this model to the investigation of context effects and forgetting in surveys. He is currently working on a book with Lance Rips and Ken Rasinski on the psychology of surveys.

Dr. Tourangeau received his Ph.D. in Psychology from Yale University in 1978. He has taught courses in psychology, statistics, and survey methods at Yale, Columbia, Northwestern, and Wisconsin. He is currently affiliated with the Joint Program in Survey Methodology at the University of Maryland.

Clyde Tucker

Clyde Tucker is Director of the Behavioral Science Research Center in the Office of Survey Methods Research at BLS. He has worked in the areas of measurement and nonsampling error in surveys for almost twenty years. This experience included several years designing exit polls for CBS News and nine years as a mathematical statistician at BLS in the Office of Prices and Living Conditions. He has an M.S. in Statistics and a Ph.D. in Political Science from the University of Georgia.

Tracy Wellens

Tracy Wellens received her Ph.D. in Social/Personality Psychology from New York University. She then completed a post-doctoral research fellowship at ZUMA, a survey research center in Mannheim, Germany. She joined the Center for Survey Methods Research at the Census Bureau in 1993 where she conducts cognitive research and other methodological studies to investigate nonsampling errors in surveys. She is currently working at the Committee on National Statistics, National Academy of Sciences, under a Interagency Personnel Agreement.

Ed Welniak

Edward J. Welniak began his career at the Census Bureau in 1978 after graduating from Case Western Reserve University with an undergraduate degree in applied statistics. He has worked in the Income Statistics Branch, Housing and Household Economic Statistics Division, U.S. Bureau of the Census for the past 19 years and currently serves as its chief. The income branch is primarily responsible for collecting, processing, analyzing, and publishing income data collected in the Current Population Survey, the Survey of Income and Program Participation, the American Community Survey, and the Decennial Census. In 1995, Ed received the Census Bureau's Bronze medal for his accomplishments in collecting, analyzing, and disseminating income statistics.

Karen Whitaker

Karen Whitaker is the manager for the NCHS Questionnaire Design Research Laboratory. She is responsible for advertising, recruiting, scheduling subjects and coordinating all laboratory testing activities. In addition to managing the lab, she has conducted many cognitive interviews and has collaborated with other NCHS staff on ways to improve the analysis and evaluation of cognitive interview findings. Before joining the NCHS staff, Ms. Whitaker worked as a office/sales manager for a Deaf-owned telecommunications company. She attended Gallaudet University with a major in Interpreting for the Deaf. She is nearing completion of her BS at University of Maryland's University College with a major in Psychology and minor in Gerontology.

Gordon Willis

As a Survey Statistician at the National Center for Health Statistics in Hyattsville, Maryland, Gordon Willis worked in the Questionnaire Design Research Laboratory to test, develop, and conduct research related to health survey questionnaires. Dr. Willis received his bachelor's degree from Oberlin College, and his doctoral degree in experimental psychology at Northwestern University. Before working at NCHS, he was Associate Director for Computing and Research at the Center for Talent Development, Northwestern University, and Research Associate in the Northwestern School of Education. At Northwestern, he worked in the area of early childhood mathematics education research, and developed cognitive interviewing techniques for use in determining the cognitive processes used by children to solve math word problems. His current research focuses on the application of cognitive interviewing techniques in order to improve survey questionnaires, and in particular, the evaluation of the usefulness of these methods. He has also taken primary responsibility for both cognitive laboratory testing and pretesting of large-scale NCHS surveys such as the National Health Interview Survey, and provides consultation and training on a wide range of questionnaire design issues.

Barbara Wilson

Barbara Foley Wilson is a survey statistician at the National Center for Health Statistics. She received her degree in demography from Georgetown University in 1974 and worked in Marriage and Divorce Statistics Section, Division of Vital Statistics, for twenty years before moving to the Questionnaire Design Research Laboratory. Although her roots are in Behaviorism she has enjoyed learning the cognitive approach to research problems and has benefitted from several courses given by the Joint Program in Survey Methodology. She worked with Doug Herrmann in the NCHS Map Design Laboratory and with Gordon Willis, Paul Beatty and Susan Schechter on a variety of surveys that have been tested in the QDRL. Currently she is applying cognitive techniques to testing a user survey for the NCHS Home Page on the World Wide Web. Another project demanding current attention is the Videotaped Oral Histories of the CASM movement.

Ron Wilson

Ronald Wilson is Special Assistant to the Associate Director of Analysis, Epidemiology and Health Promotion at the National Center for Health Statistics, Centers for Disease Control and Prevention (1994 to present.) He also serves as the NCHS HIV/AIDS Coordinator and was responsible for the initial development in 1987 of the AIDS Attitude and Knowledge Supplement to the National Health interview Survey. For 15 years he was the Director of the Division of Epidemiology and Health Promotion at NCHS. One of his major activities during this period was working with the Office of Disease Prevention and Health Promotion on data issues related to monitoring the 1990 Objectives for the Nation and the Year 2000 Healthy People Objectives. His primary interests are issues related to the measurement of health status and the collection of health behavior data and most recently how these issues relate to the growing interest in the development of performance measures to assess program outcomes. He has been with the National Center for Health Statistics since 1965, working the first 10 years with the National Health Interview Survey. He has degrees in sociology, focusing on survey research, from the University of Wisconsin, Lacrosse and the University of Iowa.

Joseph Young

Joseph L. Young is Information Director in the Division of Social, Behavioral, and Economic Sciences at the National Science Foundation. Previously, he spent more than twenty years as NSF's Program Director for Human Cognition and Perception (previously Memory and Cognitive Processes). Prior to joining NSF, Dr. Young was Assistant Director of the Personnel and Training Research Programs at the Office of Naval Research and Assistant Professor of Psychology at the State University of New York at Stony Brook. Dr. Young holds a Ph.D. in Mathematical Psychology from Stanford University and a B.A. in Psychology from Yale University. He is a Fellow of the AAAS and a member of Sigma Xi and a number of general and more specialized psychological societies.