

RESPONSE EFFECTS IN SURVEY MEASURES OF BEHAVIOR: INSIGHTS FROM RESEARCH IN OTHER FIELDS

Dr. Scott C. McDonald, Time Inc.

Abstract

This paper reviews recent developments other fields of possible interest to professionals in readership research. In particular, it considers contributions from cognitive psychologists and sociologists to understand the nature of response effects that undermine the accuracy of reports of past or prospective behaviors. These studies include a number of "record-check" studies comparing actual to reported levels of crime victimization, voting, health maintenance activities, and consumer purchases. The paper focuses on two specific types of response effects: those arising from incomprehension of the meaning of questions, and those arising from problems of memory (including telescoping). In conclusion, it considers a variety of memory enhancing techniques currently being recommended in the various fields under review.

Sample surveys, in their modern form, began with studies of London's poor populations in the late nineteenth century. Since that time, researchers have striven to refine the methods and improve the accuracy of their estimates. Some of these improvements have occurred rather rapidly and have marked the experience of the social scientists working in a given decade. Early improvements mostly addressed problems of sample bias. In the 1930's, government agencies in the United States embraced probability sampling methods, thereby achieving an enormous improvement in the quality and representativeness of public data. In the wake of the disastrous performance of pre-election polls in the 1948 presidential election, most private survey organizations followed suit, with a consequent beneficial effects on polling and opinion data. Indeed, by the 1960s, the field had progressed enough so that most errors and biases related to sampling could be controlled and quantified. Though sampling remains an active area of methodological research, the focus has generally shifted to more arcane sub-specialties, to issues of cost and efficiency, or to techniques for contending with new technologies (such as phone answering machines) that can threaten sample representativeness.

By contrast, more recent decades have seen a marked increase in methodological work on the other major source of errors in surveys -- nonsampling errors. This shift was caused by several factors: the growing availability of high-speed electronic computers facilitated a proliferation in surveys, with attendant decline in compliance rates and increased risk of non-response biases; the advent, in the 1960's and 1970's, of many more large-scale government-sponsored longitudinal studies created greater demands for accuracy by policy-oriented end-users; and, as already indicated, the field's earlier successes in reducing and controlling sampling errors meant that, for scholars interested making the biggest impact on the field, the most important work probably remained in the area of reducing response effects -- those errors introduced by misinterpretations of questions, outright lying, misremembering events, yielding to perceived cues from peers or from interviewers, and so forth.

These concerns about factors that affect responses are not new of course. Survey researchers have long been concerned with the quality of measurement and with the threats to validity posed by response effects. And, indeed, one gets the impression that many of the new crop of researchers in this field could save themselves some trouble by attending a bit more closely to the classic studies from the 40's and 50's by survey sociologists like Hyman, Stouffer, and Lazarsfeld. But what IS new, perhaps, is the systematic attempt to apply the perspectives and methods of cognitive psychology to the problems of reducing response effects in surveys. Spawned in part by a 1980 conference focused on problems in the National Crime Survey and by a subsequent 1983 seminar on Cognitive Aspects of Survey Methodology, sponsored by the National Research Council, researchers in this tradition have come to think of themselves as a "movement" (Tanur, 1992).

My job today is to give you a brief introduction to this emerging cross-disciplinary literature and to suggest some implications that it may hold for readership research. It is not possible, within the constraints of this conference's format, to provide an exhaustive review of the entire literature. Nor, indeed, would that even be necessary, given the recent appearance of several works that either do exactly that (Jobe and Mingay, 1991; Aborn, 1989), or that present compilations of studies in the cognitive tradition (Tanur, 1992; Hippler, Schwarz, and Sudman, 1987). Rather, I hope to provide enough of an orientation to pique your interest and invite you to further inquiry.

I will focus today on the part of the literature that concerns itself with accuracy in reported behavior -- either actual or intended, reported either retrospectively or prospectively -- since this is the area of greatest overlap with issues commonly discussed at these symposia on readership research. By focusing on this literature, we also have the theoretical advantage of attempting to measure some phenomenon that, in principle at least, has a "true" value that can be verified by means external to the survey itself. Thus, this literature has roots in the tradition of "record check" studies in the social sciences, a tradition nearly as old as the social survey itself.

However our decision to concentrate on studies concerning reported behavior necessarily excludes the extensive and fascinating literature on response effects in the measurement of attitudes, some of which could still be relevant to aspects of our work in readership research. For example, Pearson, Ross and Dawes (1991) used longitudinal data to examine attitude change and stability over time. Though they, of course, did not have access to any "true", externally verifiable set of attitudes for their subjects, they were able to compare a respondent's recalled attitudes to the same person's attitudes expressed earlier in time. Their experiments demonstrated that one's current recollection of past attitudes is shaped by one's personal beliefs about the stability of those attitudes. In other words, we tend to reconstruct our pasts to accord with our current view of the past. To my knowledge, no researchers have yet addressed the issue of whether a similar dynamic is at play in the retrospective reconstruction of behavior.

Other recent work by cognitive psychologists (e.g. Krosnick and Abelson, 1991) has amplified the desirability of measuring, not just the direction, but also the strength of attitudes. While this imperative is certainly not new, the more recent studies by cognitive psychologists and sociologists have demonstrated the not entirely unexpected result that the stronger the attitude, the more likely it is to be manifested in an observable behavior (Sample and Warland, 1973; Fazio and Zanna, 1978; Peterson and Dutton, 1975; Krosnick, 1988; Schuman and Presser, 1981; Rokeach and Kliejunas, 1972; Jaccard and Becker, 1985). Indeed, in a recently published meta-analysis, Kim and Hunter (1993) integrated the findings from 138 studies exploring attitude-behavior correlations; their resulting database had a sample size of 90,908 and ranged across 19 different substantive areas, including speech and communication, gerontology, marriage and family studies, consumer research, social psychology, and marketing research. Though correlations computed through such meta-analyses should be viewed with some suspicion, the authors found a consistent and strong correlation across all domains -- again with more intense attitudes associated with more certain behavioral outcomes.

From the perspective of cognitive psychology, attitude strength itself has several distinct dimensions that may be poorly measured by any single question or scale. For example, Krosnick and Abelson (1991) argue that attitude strength has five dimensions -- extremity, intensity, certainty, importance and knowledge -- though they do not present compelling evidence that these dimensions are distinct. For those, like myself, who are sometimes obliged to predict a subscription renewal rate or a new product acceptance rate from survey data on attitudes and intentions, these perspectives on the behavior-intention discrepancy are useful. Or, for those, also like myself, who are merely fascinated by the broad connection (or lack of connection) between attitudes and behaviors, this literature is worth pursuing.

However for our purposes today, let us concentrate on the part of the literature concerned directly with the factors affecting accuracy of recall of behavior -- and here I intend to combine recent contributions by cognitive psychologists with those of other more traditional survey methodologists, sociologists and social psychologists. For convenience (and following Tanur (1992)), I'll divide this literature into two parts: one dealing with issues of meaning and comprehension, the other dealing with the limitations of memory itself.

Issues of Meaning and Comprehension

Survey researchers have long appreciated the powerful effects that question wording and context can have upon responses, and the literature is replete with literally thousands of individual studies demonstrating such effects. However the results have not generally fallen into such neat patterns as to allow the formulation of general principles that can alert investigators to the possible presence of wording or context effects. There are, however, a few areas where generalizations may be ventured.

Sensitive Topics: Bradburn (1978) has demonstrated that threatening questions on sensitive topics can influence responses, including the level of self-reports of behaviors. If respondents admit to participating in a given activity (e.g. drinking, smoking, engaging in sexual activity, or other socially sanctioned activity), the level of that activity is not strongly influenced. However, threatening questions can lead to base underreporting. He recommends asking direct questions about whether "most people" would feel uneasy being asked questions about the given activities, and demonstrates how such opinions are related to underreporting of the given activity. He also demonstrates procedures for making statistical adjustments to correct for underreporting.

Another fairly recent technique for handling threatening questions on behaviors that are likely to be underreported is the randomized response technique. Originally proposed by Warner (1965), it has been the subject of considerable experimentation in criminological and health-related studies. The technique centers around a procedure in which the respondent tosses a coin to determine whether to answer threatening Question A or non-threatening Question B; only the respondent (and not the interviewer) is aware of which question is being answered (see Zdep and Rhodes, 1976, for details on the mechanics of the procedure). The procedure's name is a bit inaccurate, since it really involves randomizing the question; indeed, the technique requires that response categories for both questions be identical (e.g. "yes" or "no"). Experimental results indicate that the procedure yields more truthful reporting of socially sensitive behaviors and, thus, more valid survey estimates of those behaviors (Zdep et. al, 1979).

Question Length: Question length has only recently been recognized as a possible cause of response effects. The conventional wisdom of the past has tended to recommend parsimony -- the shorter the better -- in the formulation of questions. Indeed, most of us have been trained to edit questionnaires ruthlessly and to keep interviews "flowing". However a body of research has emerged since the 1970s that has caused many to question that conventional wisdom. In a series of experiments, Marquis, Cannell and Laurent (1972) modified the National Center for Health Statistics' Health Interview Survey to ascertain how question length affected recall of symptoms, conditions and illnesses. They found that longer questions and interviewer reinforcements resulted in 29% higher reporting levels for most illnesses and symptoms; a similar result was obtained by Blair (1977). However in neither case were record checks performed to verify the authenticity of the claims. By contrast, Sudman and Bradburn (1974) found no effect of question length in face-to-face interviews, but found that short questions elicited sizeable underreporting of behaviors in self-administered questionnaires. Though the literature is not definitive, it at least suggests that brevity may not always be best.

Vague Quantifiers: Though most surveys make liberal use of questions that ask whether certain things are done "often, seldom or never" (using that or some similar quantifying scale), the experimental literature suggests that this is problematic. As early as 1941 (Mosier), researchers had noticed that the exact meaning of these scale points varied considerably among different respondents. Some researchers have tended to emphasize individual variability, and thereby recommend direct magnitude estimations (Hakel 1969). For example, Belson (1981) posed a series of questions about television viewing with vague response categories like "usually" and "few" -- a practice not at all atypical. Respondents answered readily, without complaint. Yet the follow-up probes revealed that respondents had strikingly different interpretations of the "few" in "over the past few years": it meant "no more than two years" to 12% of respondents, "seven or more years" to 32%, and "ten years or more" to 19%. Indeed, Belson found that respondents assign meanings not intended by the researcher 70% of the time, all without any sign of discomfort or incomprehension.

Other researchers have emphasized contextual rather than individual variability in response to quantifying scales, and have recommended anchoring the meaning intended in the question by providing more explicit information (e.g., Pepper and Prytulak 1974). For example, Loftus (1975) ran a series of experiments in which he varied the perspectives provided by the question wording and context. In one experiment, he asked people one of these two questions:

1. Do you get headaches frequently, and if so, how often?
2. Do you get headaches occasionally, and if so, how often?

Each question presupposes a slightly different frequency, and helps respondents assess a normative perspective. Respondents to the first question estimated that they got 2.2 headaches per week, while respondents to the second question only reported .7 headaches per week. While one might be tempted to argue that such experiments provide evidence of the inherent shared meaning of terms like "frequently" and "occasionally", the results also suggest that even magnitude estimation questions can be affected dramatically by subtle contextual perspectives.

Such perspectives are established by the way that we frame our questions, and they can even distort eyewitness accounts, as linguistic experiments in memory have shown. In one such experiment (Loftus and Palmer 1974), respondents were given a common stimulus -- brief videos of car accidents. For one of those videos, they were asked, "How fast were the cars going when they contacted each other?" The experiment varied the final verb in the question for different groups of respondents, with the verbs escalating in violence from "contacted" to "hit", "bumped", "collided", and finally, to "How fast were the cars going when they smashed each other?". Even with the stimulus fixed in a common video, estimates of speed for the eyewitnesses increased in step with the verb used in the question, from 31.8 mph for "contacted" on up to 40.8 mph for "smashed". Moreover, when the respondents were brought back a week later to answer further questions about what they had "seen", those who had been asked the "smashed" question were more likely

than others to report broken glass at the scene of the accident. Reconstruction of events in memory was influenced by the velocity implied in the verbs used in the questioning.

Even such seemingly straightforward quantifying questions as "How many children do you have?" can be undermined by different cultural meanings. While Mexican researchers found no problem when that question was posed in Mexico City, it was discovered that villagers in the Yucatan interpreted the "ninos" in "Cuandos ninos tienen Ustedes?" as including, not just living children, but also unsuccessful pregnancies and children who had died; the resulting count would have been very wrong had not the problem of mutual meaning been discovered (Clark and Schober, 1992).

These problems of question meaning and interpretability are not new. Indeed, one of the best references on them is still Payne's (1951) classic, *The Art of Asking Questions*. What is new is the attention being given to the matters by cognitive psychologists who attempt to apply their perspectives on how we learn, reason, and derive meaning from language to the arts of survey construction. For example, Clark and Schober (1992) argue that respondents in their natural lives discern the meaning of words and of language through a series of cues, many of which represent efforts to find common ground with the other person. Respondents apply the rules of conversation to answering surveys, even though the structure of surveys is nearly the antithesis of a normal, natural conversation. It is this tension that leads to confusion and miscommunication of meaning. From the cognitive/linguistic perspective, respondents attempt to decipher intentions from the language in surveys by (1) trying to reconstruct a sensible meaning, even when other interpretations are available; (2) attempting to find common ground, an activity which, in normal conversation is cumulative; and (3) searching for a normative perspective suggested in the question itself. Without sensitivity to these respondent activities, survey questions are doomed to misunderstanding and survey results are likely to be distorted.

Cognitive psychologists working in this vein tend to argue for more open-ended probes to verify that our meanings really are mutual. They argue that we cannot merely PRESUME that our questions are interpreted as we intend, nor that acquiescent respondents are truly understanding our questions simply because they are uncomplaining. To comply with this suggestion would in many cases impose undue burdens on respondents and make some surveys impractical; but the experimental literature suggests that we should probably include a few more open-ended questions to verify our assumptions and run the occasional "reality check" -- a suggestion made as far back as 1966 by Howard Schuman (Schuman 1966). It also is likely that, at least in the case of response scales that describe some frequency of a behavior, an explicit numeric anchor can be helpful.

However even very explicit numeric response scales can still cue respondents to different normative expectations, as was demonstrated by a recent experiment in Germany. A study by Schwarz et. al. (1985) asked a split sample the same question, "How many hours a day do you spend watching TV?", but varied the response scale. Half of the respondents got a response scale that skewed toward less watching time, with six scale points starting at "up to 1/2 hour" and moving in half-hour increments to a maximum of "more than 2 1/2 hours". The other half of respondents got a response scale that skewed toward more watching time, with the lowest point on the scale starting at "up to 2 1/2 hours", and progressing again in half-hour increments to a maximum of "4 1/2 hours". What was the result? Estimates of TV watching differed sharply between the two groups. Only 16% of those given the low-skewing response set claimed to watch TV more than 2 1/2 hours a day, while 38% of those given the high-skewing response set made that claim. When respondents were later asked the question, "What is the average TV consumption of German citizens?", the low-skew group estimated 2.7 hours, and the high-skew group estimated 3.2 hours. In effect, the different response scales affected perceptions of what was "normal". Indeed, further questioning of these split samples showed that those in the low-skew group rated TV as more important in their lives than did those in the high-skew group, even though they reported watching less TV. This is consistent with a normative interpretation in which the low-skew group assumes that less TV watching is "normal" because of the cues provided in the scale; they thus infer that they must be "above normal" and that, therefore, TV is more important in their lives.

Issues of Memory and Recall

There is also much in the cognitive and survey research literature that bears on matters of recall and memory. Indeed, one need not attend a Readership Research Symposium to enjoy fevered discussions of telescoping. In this review, I encountered record-check studies that found telescoping in police and crime reports (Turner, 1972; Garofalo and Hindenlang, 1977), in studies of household appliance purchases (Sudman and Ferber, 1970), and in response to numerous other questions about shopping, leisure, health events and voting (Bradburn, 1979; Clayton, 1979; Cherry and Rodgers, 1979; Sudman and Bradburn, 1982). However at least one study, by Miller and Groves (1985), criticizes the record-check method and argues that match rates are themselves quite dependent upon our variable definitions of what constitutes a

"match", our instructions to coders, and the differing interpretations of those instructions and definitions adopted by the coders. Since few record-check studies provide details about these procedures, and even fewer attempt to assess inter-rater reliability, the authors call into question the advisability of assuming that records are the unassailable arbiters of survey response quality. Rather than try to mediate that debate, I shall review some of the suggestions for reducing telescoping and enhancing memory that have emerged from the recent body of research.

Ask About Recent Events: Let's begin with the most obvious and consistent finding in the research: memory decays over time. As far as I know, this was first demonstrated in a psychological experiment by Herman Ebbinghaus in 1885 using a sequence of meaningless but pronounceable nonsense syllables (Baddeley, 1979), and the finding has been supported in innumerable studies since then. So the first and most obvious suggestion is to ask about recent behavior.

Use Dual Time Frames to Reduce Overreporting and Telescoping: Crespi and Swinehart (1982) demonstrated a relatively simple procedure for reducing overreporting of past events. They were able to obtain more accurate reporting of socially desirable health behaviors in the previous two months by preceding the question with another asking about the same behaviors in the previous six months. Presumably, this procedure works for three reasons. First, it allows respondents to claim "credit" for a socially desirable behavior through the initial 6-month question. Second, the events that happened 3 or 4 months ago that might be telescoped into the 1-month response instead get absorbed in the 6-month response. Third, the follow-up 1-month question communicates to the respondent a greater demand for precision in answering. The finding has been replicated and extended in a recent record-check study using data from a Seattle-based health maintenance organization (Loftus et. al., 1992). However the use of dual time frames has not been shown to improve the accuracy of self-reports in voting (Abelson, et. al., 1992; indeed, there is evidence that misreporting of voting and vote registration tends to be more respondent-specific than item-specific, with the unhappy consequence that error terms are correlated (Presser, 1984). On several counts, the literature suggests that questions about voting represent something of a special case, with limited applicability to other domains of research on recall (Katosh and Traugott, 1981; Abelson et. al., 1992; Presser, 1990).

Impose Backward Retrieval to Stimulate Memory and Reduce Underreporting: Cognitive psychologists (Loftus et. al., 1992) have recently experimented with asking people to recount events in a chronological or forward order (from most distant to most recent), in a backward order (from most recent to most most distant), or in a free order (whichever they prefer, which is preponderantly the forward order). These experiments are designed to stimulate memory and reduce forgetting (and underreporting) of events. Though there have been some inconsistent results (Jobe et. al., 1990), there are indications that memory is enhanced by the imposition of backward recall (Loftus et. al. 1992; Loftus and Fathi, 1985).

Use Landmarks to Reduce Telescoping: Many of the events that are queried in surveys are not very salient to respondents and, thus, difficult to recall. In the past decade, researchers have experimented with the use of landmarks (or highly salient events) to mark more effectively the beginning of the reference period (Baddeley, 1979; Loftus and Marburger, 1983; Loftus et.al, 1992). The evidence suggests that this procedure significantly reduces forward telescoping and thereby deflates the frequency estimates of a given behavior. Alas, in the case of the Seattle HMO record-check experiments (Loftus et. al. 1992), the use of landmarks actually resulted in poorer net estimates of number of doctor visits. The researchers speculate that, in real life, the forgetting of actual visits is partially compensated by the forward telescoping of earlier visits, thereby resulting in an estimate that contains telescoped errors but that is closer to reality.

To Reduce Telescoping. Reduce Pressure to Respond Quickly: Much of conversational discourse is composed of "adjacency pairs", dyads in conversants exchange repartee. Examples include requests and compliance, offer and acceptance, greeting and greeting, compliment and thanks, and so forth. The cognitive psychologists remind us that, for respondents, the conventions of turn taking in conversation apply to surveys. Even though the nature of a survey is not really conversational at all, its question-and-answer format has a structure that resembles the adjacency pair structure of much conversation. One of the implied rules of conversation calls for the avoidance of long pauses between the two parts of an adjacency pair. The evidence is that the pressure to respond quickly reduces accuracy of recall of behavior (Cannell et. al., 1977; Neter and Waksberg, 1964; Sudman and Bradburn, 1982). Thus, if you are asked something like "How many times did you get take-out food last week?" in conversational mode, you will answer quickly with some answer like "Four", without explicitly counting the number of times. Experiments in the more recent cognitive tradition have shown the benefits to accuracy of an instruction like: "Take a minute to recall and count the number of takeout meals you got last week, then tell me the number". If people have time to calculate rather than estimate, their responses will generally be more accurate. The cognitive survey approach encourages people to "think out loud" or to "count out loud", recognizing that while memory itself may be fallible, it is the pressure to respond quickly prevents people from fully accessing their memories.

Given the growing sophistication of computer-assisted interviewing techniques, this suggests some interesting experiments that can be developed relating the response time to response accuracy. In attitude research, this technique for measuring response time has already been tried in an effort to distinguish between "spontaneous" and "deliberative" attitudes (Dovidio and Fazio, 1992), but the results are, to my mind, inconclusive.

The problem of response time interacts with mode effects. Time pressure is negligible for self-administered questionnaires, and is most acute for telephone surveys, where conversational delaying signals are more limited than in face-to-face (limited, indeed, to utterances like "um"). This higher pressure to respond quickly in telephone surveys probably accounts, in part, for their well documented tendency to go faster than face-to-face surveys (Groves and Kahn, 1979; Williams, 1977). On the telephone, people give shorter answers and allow fewer pauses. Chances are, they do less calculation in responding to retrospective behavior questions, so there is reason to fear that this mode of data collection yields the least reliable estimates of recalled behavior unless explicit steps are taken to encourage people to slow down and reflect on their answers. (Of course, adopting these cognitive interview strategies would also increase the cost of a given telephone survey, since average time per interview would increase).

For Comparative Purposes, Ask for General Estimates of Frequency: The Seattle HMO study that I already mentioned (Loftus et. al., 1992) found that when asked to provide a general estimate of the number of visits made to a doctor in the past year, they report 87% of the actual visits. However when frequency is estimated from the recollection of specific visits, only 39% of the actual visits are remembered. This underreporting occurs, even if lax rules are applied to the dating of visits. If people have a pretty good idea of how many times certain events happened in their lives, even if they cannot specifically recall each instance, then the more general estimates may have value.

To Enhance Memory and Reduce Forgetting, Use Techniques Derived from the "Cognitive Interview": I must admit that this bit of advice is a bit vague, partly because I think the research to date has not yet fully justified it. The "cognitive interview" is a technique that was initially applied to efforts to enhance eyewitness memory for details of a crime (Fisher et. al., 1987). Based initially upon extensive analysis of tape-recorded interviews and laboratory research, it draws upon a theoretical conception of five general principles of cognition and memory retrieval: (1) context reinstatement -- establish the same psychological environment in which the event occurred; (2) focused retrieval -- recall of details requires concentration; (3) extensive retrieval -- the more retrieval attempts one makes, the greater the success; (4) varied retrieval -- memories not activated by one probe may yield to another one; and (5) multiple representations -- memories are stored in different forms, so they may be sometimes accessed indirectly, through analogy or alternative representation. Elements of this schema are present in the other research techniques (including some we have already discussed), but the full-blown "cognitive interview" has still had its fullest exposition in eyewitness testimony work (Fisher et. al., 1993). Fisher and Quigley (1992) made a very limited attempt to apply the techniques to food recall questions in a tainted food investigation, but the study was far too crude to yield any reliable conclusions. Indeed, though the ideas have some appeal and deserve further investigation and experimentation, the five principles of the cognitive interview, as articulated by Fisher and his colleagues, seem to restate in fancier terms the longstanding notion that respondents who try harder will recall more (Cannell, 1977).

Though some of the "new" findings and insights from the cognitive survey "movement" may not be as new as its proponents suggest, I do not really mean to dismiss or belittle it. Many of the ideas and findings are quite intriguing, and they serve to remind us of the numerous ways in which our efforts can be undermined by complacency. They also should encourage us to avoid parochialism in our own research. There are great advantages to be gained from cross-disciplinary perspectives. Just as a memory of an event might be activated by a varied set of questions or a new representation, so too might a methodological problem that has seemed intractable in readership research be jogged into a new light by the insights from another branch of the social sciences. That, at least, is the hope that has animated my efforts in preparing this brief review.

References

- Abelson, R., Loftus, E., and Greenwald, A. (1992) Attempts to improve the accuracy of self-reports of voting. In Tanur, J. (ed.), *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Aborn, M. (1989) Is CASM bridging the chasm? Evaluation of an experiment in cross-disciplinary survey research. Paper presented at the American Statistical Association 1989 Winter Conference, San Diego, CA, January 4-6.
- Baddeley, A. (1979) The limitations of human memory: Implications for the design of retrospective surveys. In Moss, L., and Goldstein, H., (eds.). *The Recall Method in Social Surveys*. 1979. London: University of London, Institute of Education.
- Belson, W. (1981) *The Design and Understanding of Survey Questions*. Aldershot, England: Gower.
- Blair, E., Sudman, S., Bradburn, N., and Stocking, C. (1977) How to ask questions about drinking and sex: Response effects in measuring consumer behavior. *Journal of Marketing Research* 14 (August): 316-321.
- Bradburn, N., and Miles, C. (1979) Vague quantifiers. *Public Opinion Quarterly* 43: 1. Reprinted in Singer, E., and Presser, S. (eds.) *Survey Research Methods*. Chicago: University of Chicago Press, 1989.
- Bradburn, N., Sudman, S., Blair, E., and Stocking, C. (1978). Question threat and response bias. *Public Opinion Quarterly* 42:2. Reprinted in Singer, E., and Presser, S. (eds.) *Survey Research Methods*. Chicago: University of Chicago Press, 1989.
- Bradburn, N., Sudman, S., and Associates. (1979) *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- Cannell, C., Oksenberg, L., and Converse, J. (1977) Striving for response accuracy: Experiments in new interviewing techniques. *Journal of Marketing Research* 14 (3): 306-315.
- Cherry, N., and Rodgers, B. (1979) Using a longitudinal study to assess the quality of retrospective data. In Moss, L., and Goldstein, H., (eds.). *The Recall Method in Social Surveys*. 1979. London: University of London, Institute of Education.
- Clark, H., and Schober, M. (1992) Asking questions and influencing answers. in Tanur, J. (ed.) *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Clayton, D. (1979) Retrospective studies of the aetiology of chronic diseases. In Moss, L., and Goldstein, H., (eds.). *The Recall Method in Social Surveys*. 1979. London: University of London, Institute of Education.
- Crespi, I., and Swinehart, J. (1982) Some effects of sequenced questions using different time intervals on behavioral self-reports: A field experiment. Paper presented at the annual conference of the American Association for Public Opinion Research, May.
- Croyle, R., and Loftus, E. (1992) Improving episodic memory performance of survey respondents. In Tanur, J. (ed.), *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Davidson, A., Yantis, S., Norwood, M., and Montano, D. (1985) Amount of information about the attitude object and attitude-behavior consistency. *Journal of Personality and Social Psychology* 49:1184-1198.
- Dovidio, J., and Fazio, R. (1992) New technologies for the direct and indirect assessment of attitudes. In Tanur, J. (ed.), *Questions About Questions: Enquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Fazio, R. and Zanna, M. (1978) Attitudinal qualities relating to the strength of the attitude-behavior relationship. *Journal of Experimental Psychology* 14: 398-408.
- Fisher, R., and Geiselman, R. (1993) *Memory Enhancement Techniques for Investigative Interviews: Increasing Eyewitness Recall with the Cognitive Interview*. Springfield, IL: Thomas.

- Fisher, R., Geiselman, R., Raymond, D., Jurkevich, L., and Warhaftig, M. (1987) Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science and Administration* 15: 177-185.
- Fisher, R., and Quigley, K. (1992) Applying cognitive theory in public health investigations: Enhancing food recall with the cognitive interview. In Tanur, J. (ed.), *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Garofalo, J., and Hindenlang, M. (1977) *An Introduction to the National Crime Survey*. Washington: U.S. Department of Justice.
- Groves, R., Fultz, N., and Martin, E. (1992) Direct questioning about comprehension in a survey setting. In Tanur, J. (ed.), *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Groves, R., and Kahn, R. (1979) *Surveys By Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.
- Hakel, M. (1969) How often is often? *American Psychologist* 23: 533-34.
- Hippler, H.J., Schwarz, N., and Sudman, S. (1987) *Social Information Processing and Survey Methodology*. New York: Springer-Verlag.
- Jaccard, J., and Becker, M. (1985) Attitudes and behavior: An information integration perspective. *Journal of Experimental Psychology* 21: 440-465.
- Jobe, J., and Mingay, D. (1991) Cognition and survey measurement: History and overview. *Applied Cognitive Psychology* 5: 176-193.
- Jobe, J., White, A., Kelley, C., Mingay, D., Sanchez, M., and Loftus, E. (1990) Recall strategies and memory for health care visits. *Millbank Memorial Fund Quarterly/Health and Society* 68: 171-189.
- Katosh, J., and Traugott, M. (1981) The consequences of validated and self-reported voting measures. *Public Opinion Quarterly* 45:4. Reprinted in Singer, E., and Presser, S. (eds.), *Survey Research Methods: A Reader*. 1989. Chicago: University of Chicago Press.
- Kim, M., and Hunter, J. (1993) Attitude-behavior relations: a meta-analysis of attitudinal relevance and topic. *Journal of Communication* 43 (1): 101-142.
- Krosnick, J. (1988) The role of attitude importance in social evaluation: A study of policy preferences, presidential candidate evaluations, and voting behavior. *Journal of Personality and Social Psychology* 55: 196-210.
- Krosnick J., and Abelson, R. (1992) The case for measuring attitude strength in surveys. In Tanur, J. (ed). *Questions on Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Loftus, E. (1975) Leading questions and the eyewitness report. *Cognitive Psychology* 7: 560-572.
- Loftus, E., and Fathi, D. (1985) Retrieving multiple autobiographical memories. *Social Cognition* 3: 280-295.
- Loftus, E., Klinger, M., Smith, K., and Fiedler, J. (1990) A tale of two questions: Benefits of asking more than one question. *Public Opinion Quarterly* 54: 330-345.
- Loftus, E., and Marburger, W. (1983) Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory and Cognition* 11: 114-120.
- Loftus, E., and Palmer, J. (1974) Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior* 13: 585-589.

- Loftus, E., Smith, K., Klinger, M., and Fiedler, J. (1992) Memory and mismemory for health events. In Tanur, J. (ed). *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Marquis, K., Cannell, C., and Laurent, A. (1972) Reporting for health events in household interviews: Effects of reinforcement, question length and reinterviews. *Vital and Health Statistics*. National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 45. Washington: U.S. Government Printing Office.
- Miller, P., and Groves, R. (1985) Matching survey responses to official records: an exploration of validity in victimization reporting. *Public Opinion Quarterly* 49:3. Reprinted in Singer, E., and Presser, S., (eds.). *Survey Research Methods: A Reader*. 1989. Chicago: University of Chicago Press.
- Mosier, C. (1941) A psychometric study of meaning. *Journal of Social Psychology* 13: 123-40.
- Neter, J., and Waksberg, J. (1964) A study of response errors in the expenditures data from household interviews. *Journal of the American Statistical Association* 59: 18-55.
- Pearson, R., Ross, M., and Dawes, R. (1992) Personal recall and the limits of retrospective questions in surveys. in Tanur, J. (ed.) *Questions on Questions: Inquiries Into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Pepper, S., and Prytulak, L. (1974) Sometimes frequently means seldom: context effects in the interpretation of quantitative expressions. *Journal of Research in Personality* 8: 95-101.
- Petersen, K., and Dutton, J. (1975) Certainty, extremity, intensity: Neglected variables in the research on attitude-behavior consistency. *Social Forces* 54: 393-414.
- Presser, S. (1984) Is inaccuracy on factual survey items item-specific or respondent-specific? *Public Opinion Quarterly* 48: 1B. Reprinted in Singer, E., and Presser, S. (eds.), *Survey Research Methods: A Reader*. 1989. Chicago: University of Chicago Press.
- Presser, S. (1990) Can context changes reduce vote overreporting? *Public Opinion Quarterly* 54: 586-593.
- Rokeach, M., and Kliejunas, P. (1972) Behavior as a function of attitude-toward-object and attitude-toward-situation. *Journal of Personality and Social Psychology* 22:194-201.
- Sample, J., and Warland, R. (1973) Attitudes and the prediction of behavior. *Social Forces* 51: 292-304.
- Schuman, H. (1966) The random probe -- A technique for evaluating the validity of closed questions. *American Sociological Review* 31 (2): 218-222.
- Schuman, H., and Presser, S. (1981) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Schwarz, N., Hippler, H., Deutsch, B., and Strack, F. (1985) Response categories: Effects of behavioral reports and comparative judgments. *Public Opinion Quarterly* 49: 388-395.
- Sudman, S., and Bradburn, N. (1974) *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Sudman, S., and Bradburn, N. (1982) *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
- Sudman, S., and Ferber, R. (1970) *Experiments in Obtaining Consumer Expenditures in Durable Goods by Recall Procedures*. Urbana, IL: Survey Research Laboratory, University of Illinois.
- Tanur, J. (1992) Cognitive aspects of surveys and this volume, in Tanur, J. (ed.). *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys*. New York. Russell Sage Foundation.
- Turner, A. (1972) *The San Jose Methods Test of Known Crime Victims*. Washington: National Criminal Justice Information and Statistical Service.

- Warner, S. (1965) Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60: 63-69.
- Williams, E. (1977) Experimental comparisons of face-to-face and mediated communications: A review. *Psychological Bulletin* 84 (5): 963-976.
- Zdep, S., and Rhodes, I. (1976) Making randomized response techniques work. *Public Opinion Quarterly* 40: 531-37.
- Zdep, S., Rhodes, I., Schwarz, R., and Kilkenny, M. (1979) The validity of randomized response technique. *Public Opinion Quarterly* 43: 4. Reprinted in Singer, E., and Presser, S. (eds.), *Survey Research Methods: A Reader*. Chicago: University of Chicago Press, 1989.