

**TELEPHONE VERSUS FACE-TO-FACE
INTERVIEWING OF NATIONAL
PROBABILITY SAMPLES WITH LONG
QUESTIONNAIRES**
COMPARISONS OF RESPONDENT SATISFICING
AND SOCIAL DESIRABILITY RESPONSE BIAS

ALLYSON L. HOLBROOK
MELANIE C. GREEN
JON A. KROSNICK

Abstract The last 50 years have seen a gradual replacement of face-to-face interviewing with telephone interviewing as the dominant mode of survey data collection in the United States. But some of the most expensive and large-scale nationally funded, long-term survey research projects involving national area-probability samples and long questionnaires retain face-to-face interviewing as their mode. In this article, we propose two ways in which shifting such surveys to random digit dialing (RDD) telephone interviewing might affect the quality of data acquired, and we test these hypotheses using data from three national mode experiments. Random digit dialing telephone respondents were more likely to satisfice (as evidenced by no-opinion responding, nondifferentiation, and acquiescence), to be less cooperative and engaged in the interview, and were more likely to express dissatisfaction with the length of the interview than were face-to-face respondents, despite the fact that the telephone interviews were completed more quickly than the face-to-face interviews. Telephone respondents were also more suspicious about the interview process and more likely to present themselves in socially desirable ways than were face-to-face respondents. These findings shed light on the nature of the survey response process, on the costs and benefits associated with particular survey modes, and on the nature of social interaction generally.

ALLYSON L. HOLBROOK is Assistant Professor of Public Administration and Psychology at the Survey Research Laboratory at the University of Illinois at Chicago, MELANIE C. GREEN is Assistant Professor of Psychology at the University of Pennsylvania, and JON A. KROSNICK is Professor of Psychology and Political Science at The Ohio State University and University Fellow at Resources for the Future. The authors thank Nancy Burns, Donald Kinder, Virginia Sapiro,

Fifty years ago, the vast majority of high-quality surveys in America were conducted via face-to-face interviewing. But following the widespread introduction of the telephone in the United States in the mid-twentieth century, survey researchers began a swift shift to conducting surveys via telephone instead. Today, most local, regional, national, and listed sample surveys are conducted by telephone. When researchers conducting such surveys seek representative general population samples, they most often sample via random digit dialing (RDD).

The appeal of telephone interviewing is multifaceted, because this method has many practical advantages, most notably reduced cost, the possibility of quick turnaround time, and the possibility of closer supervision of interviewers to assure greater standardization of administration. Initially, telephone interviewing had another unique advantage as well: the possibility of computer-driven questionnaire presentation. With the advent of Computer Assisted Personal Interviewing (CAPI), telephone interviewing's edge in this regard is gone, but this mode continues to maintain its other unique advantages and its popularity in practice.

Telephone interviewing has obvious disadvantages, too. For example, show cards, which are often used to present response choices in face-to-face interviews, are more difficult to employ in telephone surveys, requiring advance contact, mailing of cards to respondents, and respondent responsibility for manipulating the cards during the interview. Therefore, telephone surveys routinely forgo the use of show cards (but see Miller [1984] for a discussion of the effects of this omission and Groves and Kahn [1979] for a discussion of possible disadvantages of show cards). As of 1998, about 5 percent of the U.S. population did not have a working telephone in their household, thereby prohibiting these individuals from participating in telephone surveys (Belinfante 1998). And for a variety of reasons, it has always been more difficult to obtain response rates in telephone surveys as high as those obtained in face-to-face surveys (e.g., Groves 1977; Mulry-Liggan 1983; Shanks, Sanchez, and Morton 1983; Weeks et al. 1983). Thus, it is not obvious that data quality in telephone surveys will meet or exceed that obtained from face-to-face surveys.

Kathy Cirkseña, James Lepkowski, Robert Belli, Robert Groves, Robert Kahn, John Van Hoyke, Ashley Grosse, Charles Ellis, Paul Biemer, the members of the National Election Study Ad Hoc Committee on Survey Mode (Norman Bradburn, Charles Franklin, Graham Kalton, Merrill Shanks, and Sidney Verba), and the members of the National Election Study Board of Overseers for their help, encouragement, and advice. We are also grateful to Aldena Rogers and Chris Mesmer for their assistance in collecting data for one of the social desirability validation studies. This research was supported by a grant from the National Science Foundation (SBR-9707741). Correspondence should be addressed to Allyson L. Holbrook at Survey Research Laboratory, 412 S. Peoria St., Sixth Floor, Chicago, IL 60607-7069 (e-mail: allyson@uic.edu), Jon A. Krosnick at the Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210 (e-mail: Krosnick@osu.edu), or Melanie Green at the Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104-6196 (e-mail: mcgreen@psych.upenn.edu).

Perhaps for such reasons, a diverse group of the nation's most expensive, long-term, large-scale, federally funded survey studies of national samples involving long questionnaires have retained the face-to-face method while most other survey research moved to the telephone. For example, the National Election Studies have conducted face-to-face interviews since the 1940s; the General Social Survey has done so since 1972. The National Health Interview Survey (conducted by the U.S. Census Bureau for the Centers for Disease Control and Prevention), the National Crime Victimization Survey (conducted by the U.S. Census Bureau for the Bureau of Justice Statistics), and many other such large survey projects sponsored by government agencies have done so as well.

In this article, we explore whether there are potential benefits of such continued reliance on face-to-face interviewing of national area probability samples for studies with long questionnaires, as compared to the obvious alternative of RDD telephone interviewing.¹ We focus in particular on two possible sources of response error differences. First, we consider the possibility that face-to-face respondents may be more likely to exert the required cognitive effort to answer questions carefully, whereas telephone respondents may be less likely to do so. As a result, the latter individuals may manifest more survey satisficing, thereby compromising response quality. Second, we consider the possibility that face-to-face respondents may differ from telephone respondents in the comfort they have in reporting socially undesirable attitudes, beliefs, or behaviors. As a result, the magnitude of social desirability response bias may differ between the modes.

We begin below by outlining the theoretical rationales underlying these hypotheses. Then, we review the results of many past studies that at first seem to offer evidence regarding these hypotheses. But as we will explain, the designs of these studies render most of them uninformative about the issues of interest here. We therefore proceed to describe the results of new tests of the satisficing and social desirability hypotheses using data from three large-scale experiments that involved long interviews of representative national samples.

Hypotheses

SATISFICING

The last 30 years have seen a blossoming of the literature on response errors in surveys, and many interesting theoretical approaches have been offered and

1. Although it is impossible to specify a precise length to separate short questionnaires from long ones, it is easier to note that in practice most telephone surveys are kept to lengths shorter than 30 minutes on average, whereas many face-to-face surveys involve interviewing that lasts notably longer than that. Our interest in this article is in surveys of this latter type.

developed to characterize and explain such errors. Some work has focused on the impact of misdating or forgetting on reports of behavioral events (e.g., Abelson, Loftus, and Greenwald 1992; Belli et al. 1999; Burton and Blair 1991; Sudman and Bradburn 1974). Other work has examined conversational conventions and norms and the ways in which they govern respondent behavior (e.g., Schwarz 1996; Schwarz et al. 1991). Still other work has focused on how linguistic processing of words in questions is accomplished by respondents (e.g., Tourangeau, Rips, and Rasinski 2000). All of these sorts of perspectives and others as well could be useful in exploring mode effects. In this article, we focus on another one of these theoretical accounts: satisficing theory.

Satisficing theory. Krosnick's (1991, 1999) theory of survey satisficing is based upon the assumption that optimal question answering involves doing a great deal of cognitive work (see also Cannell, Miller, and Oksenberg 1981; Tourangeau 1984). A respondent must interpret the meaning and intent of each question, retrieve all relevant information from memory, integrate that information into a summary judgment, and report that judgment accurately. Many respondents who initially agree to be interviewed are likely to be willing to exert the effort necessary to complete an interview optimally. But many other respondents who agree to be interviewed may become fatigued and may lose their motivation to carry out the required cognitive steps as they progress through a questionnaire, or respondents may be willing to carry out the required cognitive steps but lack the ability to do so. And some respondents who reluctantly agree to be interviewed may do so with no intention of thinking carefully about any of the questions to be asked.

According to the theory, people can shortcut their cognitive processes in one of two ways, via either weak satisficing or strong satisficing. Weak satisficing amounts to a relatively minor cutback in effort: a respondent executes all the cognitive steps involved in optimizing, but less completely and with bias. When a respondent completely loses motivation, he or she is likely to seek to offer responses that will seem reasonable to the interviewer without having to do any memory search or information integration. This is referred to as strong satisficing, which can be done by looking for cues in questions pointing to easy-to-defend answers.

The likelihood that a respondent will satisfice is thought to be a function of three classes of factors: respondent ability, respondent motivation, and task difficulty. People who have more limited abilities to carry out the cognitive processes required for optimizing are more likely to shortcut them. People who have minimal motivation to carry out these processes are likely to shortcut them as well. And people are most likely to shortcut when the cognitive effort required by optimizing is substantial. Respondents' dispositions are thought to interact with situational factors in determining the degree to which any given person will satisfice when answering any given question (see Krosnick 1991, 1999; Krosnick, Narayan, and Smith 1996). That is, satisficing may be most likely when a person is disposed to do so and when circumstances

encourage it. In light of this theoretical perspective, it seems possible that interview mode might affect response quality to factual and opinion questions for a series of reasons, involving nonverbal communication, pace, and multi-tasking, as we describe next.

Nonverbal communication. When an interviewer conducts a face-to-face conversation with a respondent, the interviewer's nonverbal engagement in the process of exchange is likely to be infectious (e.g., Chartrand and Bargh 1999). A respondent whose motivation is flagging or who questions the value of a survey can observe his or her interviewer obviously engaged and enthusiastic about the data collection process. Some interviewers may not exhibit this sort of commitment and enthusiasm nonverbally, but many are likely to do so, and they may thereby motivate their respondents to devote effort to the cognitive processing required for generating optimal answers.

Respondents interviewed by telephone cannot observe all of these same nonverbal cues of commitment to and enthusiasm for the task from an interviewer. Interviewers can certainly convey such commitment and enthusiasm verbally and paralinguistically (Barath and Cannell 1976; Oksenberg, Coleman, and Cannell 1986), but those same messages can and probably are conveyed to respondents in face-to-face interviews. These latter interviews permit additional, nonverbal messages to be sent, and their absence during telephone interviews may leave those respondents less motivated. Furthermore, face-to-face interviewers are uniquely able to observe nonverbal cues exhibited by respondents indicating confusion, uncertainty, or waning motivation, and interviewers can react to those cues in constructive ways, reducing task difficulty and bolstering enthusiasm. Face-to-face interviewers can also better observe events that might distract the respondent (e.g., the presence of another person) and may be able to react to overcome or avoid that distraction (Shuy 2002).

Research in psychology and communication offers compelling indirect support for this notion. This research has shown that observing nonverbal behavior during dyadic bargaining and negotiation interactions favorably affects the outcomes of those interactions. People are less competitive, less contradicting, more empathetic and interested in their partners' perspectives, and more generous to one another when interactions occur face to face instead of by telephone (Morley and Stephenson 1977; Poole, Shannon, and DeSanctis 1992; Siegal et al. 1986; Turoff and Hiltz 1982; Williams 1977).

Furthermore, Drolet and Morris (2000) showed that face-to-face contact (as compared to aural contact only) improved cooperation on complex tasks, and this effect was mediated by rapport: face-to-face contact led participants to feel more "in synch" with each other, which led to improved collaborative task performance. Indeed, Drolet and Morris (2000) showed that such improved performance is due to nonverbal cue exchange, because dyads conversing with one another while standing side by side (and therefore unable to see one another) performed less effectively than dyads conversing facing

one another. This is not surprising, because rapport between conversational partners has been shown to arise in particular from the convergence or synchrony of their nonverbal behaviors (Bernieri et al. 1994; Tickle-Degnen and Rosenthal 1990). If nonverbal communication optimizes cooperative performance in bargaining and negotiation for this reason, it seems likely to do so in survey interviews as well by enhancing respondent effort and reducing the likelihood of satisficing.

Pace. A second key difference between survey modes probably is the pace at which the questions are asked. All interviewers no doubt hope to complete each interview as quickly as possible, but there may be special pressure to move quickly on the phone. Silences during telephone conversations can feel awkward, whereas a few seconds of silence during a face-to-face interview are not likely to be problematic if a respondent can see the interviewer is busy recording an answer, for example. This may lead both interviewers and respondents to proceed through a telephone interview more quickly than a face-to-face interview. Furthermore, break-offs are more of a risk during telephone interviews, partly because it is easier to end a phone interview (by simply hanging up) and because talking on the telephone may be especially fatiguing for some people. Therefore, interviewers may feel pressure to move telephone interviews along more quickly than they conduct face-to-face interviews.

Even if interviewers speak more quickly on the telephone than they do face to face, respondents could in principle take the same amount of time to generate answers thoughtfully in the two modes. But respondents might instead believe that interviewers communicate their desired pace of the conversation by the speed at which they speak, and respondents may be inclined to match such desired speeds. Respondents may also choose to speak quickly on the telephone because they are anxious to finish the interview. Consequently, people may spend less time formulating answers carefully during telephone conversations. Furthermore, asking questions at fast speeds may make it more difficult for respondents to understand the questions being asked (thereby increasing task difficulty), which may lead to more satisficing as well.

Multitasking. Finally, multitasking is a phenomenon that may characterize telephone interviews to a greater extent than face-to-face interviews. A telephone respondent can easily be cooking dinner or paying bills or even watching television while answering survey questions without the interviewer's being aware of it. Therefore, doing such multitasking may not be inhibited by the norm of being polite to the interviewer. Certainly, interviewers have relayed remarkable stories of respondents multitasking during face-to-face interviews as well (such as an instance in which an interviewer saw only the feet of a respondent as he answered questions while repairing his car from underneath; see Converse and Schuman [1974]), but this seems less likely to occur during face-to-face conversations than during telephone interviews. In addition, face-to-face interviewers are more likely to be aware of such mul-

titasking and can adapt to it. Because interviewers cannot observe telephone respondents, it is difficult to collect evidence about whether mode is related to multitasking. However, a recent study examining teleconferencing in businesses provides evidence that many people do engage in multitasking during such telephone interactions (Mark, Grudin, and Poltrok 1999). Respondents in telephone interviews may similarly engage in multitasking. The added distraction caused by multitasking enhances task difficulty and therefore may enhance the likelihood of satisficing.

Conclusion. In sum, telephone interviewing may increase the likelihood of respondent satisficing by decreasing the time and effort respondents devote to generating thoughtful and careful answers and increasing the difficulty of the task. Consequently, response quality may decline. It is possible that some measurements may in fact be improved by minimizing the effort people spend generating them, because rumination might cause people to mislead themselves about their own feelings, beliefs, attitudes, or behavior (Wilson and Schooler 1991). So the shortcutting of cognitive processing might actually improve measurement reliability and validity in some cases. But in most surveys, devoting more careful thought is likely to yield more accurate responses. In the most extreme case, respondents who choose to implement strong satisficing are not answering substantively at all. So if telephone interviewing increases strong satisficing, response quality must, by definition, be decreased.

In this article, we investigate the impact of survey mode on three forms of satisficing. Two are forms of strong satisficing: choosing an explicitly offered no-opinion response option and nondifferentiation (see Krosnick 1991, 1999). These are thought to occur when a respondent chooses not to retrieve any information from memory to answer a question and instead seeks an easy-to-select and easy-to-defend answer from among the options offered. If a “don’t know” option is offered, it is particularly appealing in this regard. And if a battery of questions asks for ratings of multiple objects on the same response scale, selecting a reasonable-appearing point and sticking with it across objects (rather than differentiating the objects from one another) is an effective effort-minimizing approach.

The third response strategy we investigated is a form of weak satisficing: acquiescence response bias, which is the tendency to agree with any assertion, regardless of its content. Acquiescence is thought to occur partly because some respondents think only superficially about an offered statement and do so with a confirmatory bias, yielding an inclination toward agreeing (see Krosnick 1999). If respondents interviewed by telephone satisfice more than respondents interviewed face to face, then we should see more no-opinion responding, more nondifferentiation, and more acquiescence among the former than among the latter.

Satisficing theory suggests that the impact of mode might be strongest among respondents who are most disposed to satisfice. An especially powerful

disposition in this regard appears to be the extent of a person's cognitive skills (for a review, see Krosnick 1991), which is very strongly correlated with years of formal education (see Ceci 1991; Nie, Junn, and Stehlik-Barry 1996) and can therefore be effectively measured in that way. We assessed whether the mode effects on satisficing were stronger among less educated respondents.

We also examined the impact of mode on interview length. If telephone interviewing brings with it pressure on both participants to move quickly, this would make respondents' tasks more difficult and would therefore enhance the likelihood of satisficing. We tested whether this speculation about speed is correct.

SOCIAL DESIRABILITY

The second substantive hypothesis we explored involves social desirability response bias, the tendency of some respondents to intentionally lie to interviewers at times. Theoretical accounts from psychology (Schlenker and Weingold 1989) and sociology (Goffman 1959) assert that inherent in social interactions are people's attempts to construct favorable images of themselves in the eyes of others, sometimes via deceit. And a great deal of evidence documents systematic and intentional misrepresentation in surveys, showing that people are more willing to report socially embarrassing attitudes, beliefs, and behaviors when the reporting circumstances assure anonymity (Himmelfarb and Lickteig 1982; Paulhus 1984; Warner 1965) or when respondents believe researchers have other access to information revealing the truth of their thoughts and actions (e.g., Evans, Hansen, and Mittlemark 1977; Pavlos 1972; Sigall and Page 1971). Taken together, these studies suggest that some people sometimes distort their answers to survey questions in order to present themselves as having more socially desirable or respectable characteristics or behavioral histories (see DeMaio [1984] for a review).

The notion that social desirability response bias might vary depending upon data collection mode seems quite plausible. All of the above evidence suggests that people are more likely to be honest when there is greater "social distance" between themselves and their interviewers. Social distance seems to be minimized when a respondent is being interviewed orally, face to face in his or her own home by another person.² Under such conditions, a respondent knows that he or she could observe frowns of disapproval or other nonverbal signs of disrespect from an interviewer. In contrast, a more remote telephone interviewer has less ability to convey favorable or unfavorable reactions to the

2. Our use of the term "social distance" is closest to that of Aquilino (1994), who used it to describe the physical and psychological proximity of one conversational partner to another. This use of the term differs from other uses, referring to discrepancies in social status (e.g., Dohrenwend, Colombotos, and Dohrenwend 1968) or in the desired degree of intimacy between people in different social groups (e.g., Bogardus 1933).

respondent and may therefore be seen as meriting less concern in this regard. Consequently, more social desirability bias might occur in face-to-face interviews than over the phone.

At the same time, the telephone does not permit respondents and interviewers to develop as comfortable a rapport and as much interpersonal trust as emerges in face-to-face interactions (e.g., Drolet and Morris 2000; see also Groves and Kahn 1979, p. 222). Consequently, respondents may not feel as confident that their interviewers will protect their confidentiality as respondents in face-to-face interviews. Furthermore, the reassurance that face-to-face respondents can get from an interviewer's identification materials and other documentation may increase their comfort with discussing sensitive issues, whereas the greater uncertainty about the interviewer's identity and motives likely to typify telephone respondents may make them less willing to reveal potentially embarrassing facts about themselves. And telephone respondents may be less sure of who will have access to their answers and how they might be used, leading these people to be less honest in discussing potentially embarrassing attitudes or behaviors. If this latter process occurs, social desirability response bias might occur more often in telephone interviews than in face-to-face interviews. Of course, both rapport and social distance may influence social desirability response bias simultaneously and cancel each other out, leading to a null effect of mode.

Available Evidence

At first glance, many past studies appear to be useful for testing these hypotheses, because they compared data collected from face-to-face interviews to data collected from telephone interviews (see Shuy [2002] for a review). However, upon close inspection, the designs of most of these studies make it difficult to draw any inferences with confidence about the hypotheses of interest here. Next, we explain the design features that studies must have to be informative for our present purpose. Then, we describe which studies do and do not meet these criteria.

NECESSARY FEATURES OF A STUDY

Over the years, researchers have employed various different sorts of study designs for investigating differences between face-to-face and telephone surveys. And each study design has value for answering a particular question. For example, studies that began with a single sample of respondents and randomly assigned each person to be interviewed either face to face or by telephone provide a solid basis for making inferences about the impact of interview mode per se. But our interest here is not in isolating the impact of one or more aspects in which RDD telephone and area probability face-to-

face surveys differ. Rather, our goal is to identify the full set of differences that emerge when a survey is moved from the area probability face-to-face approach to the RDD telephone approach, a transition that involves many different sorts of changes in procedures.

In order to identify cleanly differences in satisficing and social desirability response bias between these two types of surveys, a study should have the following eight characteristics. First, one group of respondents should be interviewed face to face, and a different group of people should be interviewed by telephone; if the same people are interviewed first in one mode and then in another, this could produce order and practice effects that alter performance (Smith, Branscombe, and Bormann 1988). Second, the telephone respondents and face-to-face respondents should both be representative samples of the same population. Third, respondents assigned to be interviewed in a particular mode should be interviewed in that mode. In other words, respondents who are difficult to contact or who refuse to be interviewed in one mode should not then be interviewed in another mode, because such reassignment would confound any comparison of modes. Fourth, respondents should be interviewed individually in the face-to-face and telephone surveys, rather than interviewing individual respondents by telephone and groups of respondents simultaneously face to face. For example, if face-to-face interviews are conducted with all available members of a household at once in a group and telephone interviews are conducted with just one household member at a time, then observed differences between the modes could be attributable to differences in the group versus individual interview approach.

Fifth, respondents should not be able to choose whether they will be interviewed face to face or by telephone. Such self-selection could lead other factors to be confounded with mode. Sixth, respondents in both modes should not have been interviewed previously about similar issues, because such prior interviewing could also produce practice effects that would distort comparisons. Seventh, the questions used to gauge satisficing and social desirability response bias should be asked identically in the two sets of interviews, and they should be asked in identical contexts; that is, the number, content, and sequence of prior questions should be the same. And finally, comparisons across modes should be subjected to tests of statistical significance.

IDENTIFYING USEFUL STUDIES

After conducting an exhaustive literature search, we uncovered 48 studies that compared data collected in face-to-face and telephone interviews; these studies are listed down the left side of table 1.³ Some of these studies are potentially

3. Table 1 includes published studies and reports available on the world wide web. These studies were located in a search involving two steps. First, on-line databases of publications and reports were searched for relevant keywords in titles and abstracts. And second, the references of studies found in the first step were used to identify additional books, articles, chapters, and reports to obtain.

useful for our purposes, but many of them are not. The *X*'s in each row of the table indicate features of a study that render it uninformative with regard to the comparison of general population RDD telephone surveying with area probability face-to-face surveying. A study could have *X*'s in multiple columns, indicating multiple such features.

The first column of *X*'s identifies studies in which the same group of respondents was interviewed face to face and over the telephone. Column 2 identifies studies in which face-to-face and telephone respondents were sampled from different populations. Column 3 identifies studies in which some people assigned to one mode were actually interviewed in a different mode. Column 4 identifies studies in which respondents in the face-to-face survey were interviewed in groups and telephone interviews were conducted with individuals one at a time. Column 5 identifies a study in which respondents were given the choice to be interviewed face to face or by telephone (90 percent of people contacted in the study identified in that column chose to be interviewed by telephone). Column 6 identifies studies in which respondents in one or both modes had been interviewed previously. Column 7 identifies studies in which very different questionnaires were used in the face-to-face and telephone surveys. And column 8 identifies studies whose reports did not describe the questionnaires used in sufficient detail to permit adequate evaluation of measurement equivalence across modes.

The 23 studies in the rows in the two lower parts of table 1 were not eliminated by any of the above filters, and these are therefore potentially informative for our present purpose. The seven studies in the bottom third of table 1 compared telephone interviews of national RDD samples to face-to-face interviews of national area probability samples and are therefore of particular interest. However, in order to be useful for addressing the satisficing and social desirability hypotheses of interest in this article, a study must also have assessed the extent of satisficing-related response effects and/or social desirability bias in responding, and many of the 23 studies listed at the bottom of table 1 did not do so.

FINDINGS REGARDING SATISFICING

To gauge satisficing, a questionnaire must include appropriate measures. For example, questions offering explicit no-opinion response options must be asked in order to measure no-opinion responding.⁴ A battery of rating scale questions with identical response options must be asked in order to measure nondifferentiation, and agree/disagree or yes/no opinion questions must be asked in order to measure acquiescence. To gauge social desirability response

4. Volunteering a no-opinion response when it is not explicitly offered does not constitute satisficing, because no cue in the question encourages satisficing in that way. In fact, offering a no-opinion response under these circumstances entails breaking the "rules of the game" (Schuman and Presser 1981) by insisting on going outside the sanctioned set of response options.

Table I. Summary of Previous Studies Comparing Face-to-Face and Telephone Interviewing

Publication	The Same Respondents Were Interviewed Face-to-Face and by Telephone	Telephone and Face-to-Face Samples Were of Different Populations	Some People Assigned to a Mode Were Not Interviewed in That Mode	Face-to-Face Respondents Interviewed in Groups and Telephone Respondents Interviewed Individually	Respondents Were Given the Choice to Be Interviewed Face-to-Face or by Telephone	Respondents in One or Both Modes Were Interviewed Previously	Very Different Questionnaires	Questionnaire Not Described Adequately	Not an RDD Telephone Survey vs. an Area Probability Face-to-Face Survey	Not National Samples
Confounded mode comparisons:										
Henson, Roth, and Cannell 1978	X								X	
Herzog and Rodgers 1988	X								X	
Midanik, Rogers, and Greenfield 2001	X								X	
Rogers 1976	X								X	
Schmiedeskamp 1962	X								X	
Cahalan 1960	X								X	X
Larsen 1952		X							X	X
Siemiatycki 1979			X						X	
Mangione, Hingson, and Barrett 1982			X				X		X	
Hochstim 1962, 1967			X						X	
Herman 1977			X						X	X
Rosenstone, Petrella, and Kinder 1993			X						X	
Thornberry 1987				X						
Cannell, Groves, and Miller 1981				X						
Calsyn, Rodes, and Calsyn 1992					X				X	
Yaffe et al. 1978						X				
Morchio, Sanchez, and Traugott 1985						X			X	
Esaiasson and Granberg 1993						X			X	
Herzog and Rodgers 1999						X			X	
Woltman, Turner, and Bushery 1980						X			X	
Aquilino 1992							X			

bias, questions must be asked about attitudes, beliefs, or behaviors that empirical evidence documents are on sensitive topics and therefore subject to such bias.

Many of the 24 studies in the lower two sections of table 1 did not include appropriate measures to yield useful evidence for gauging mode effects on satisficing, and in the few cases where such measures existed, previous researchers did not examine mode effects on satisficing and report their findings in their publications. Specifically, no published report has examined no-opinion responding in questions that explicitly offered "don't know" response options.⁵ Likewise, none of the reports of these studies examined nondifferentiation in a battery of rating scale questions that was asked identically in the two modes. And none of the relevant publications examined answers to agree/disagree or yes/no questions that were asked identically in the two modes.⁶ In fact, no reports of these past studies compared any indicators of satisficing across modes.

FINDINGS REGARDING SOCIAL DESIRABILITY

Only two of the publications listed in the bottom section of table 1, both reporting analyses of the same data, gauged social desirability response bias in answers to questions with empirically established social desirability connotations.⁷ Colombotos (1965, 1969) asked five social scientists and two physicians to choose the most socially desirable response to a series of questions about the professional conduct of physicians, and some behaviors were thusly identified as respectable or not. New York and New Jersey physicians who were randomly assigned to be interviewed either face to face or via telephone did not give significantly different answers to these questions. However, it is difficult to know whether this result can be generalized to general public samples.

Seven studies of general public samples did not pretest the social desirability connotations of the questions they examined, but some of those questions seem likely to have such connotations. Two of these studies reported a significant mode effect: Weeks et al. (1983) found that telephone respondents were significantly more likely to report that they had visited a dentist during the past

5. A number of past studies examined item nonresponse for questions that did not offer an explicit no-opinion response option (Greenfield, Midanik, and Rogers 2000; Groves and Kahn 1979; Hinkle and King 1978; Jordan, Marcus, and Reeder 1980; Kormendi 1988; Quinn, Gutek, and Walsh 1980; Saris and Kaase 1997). By definition, this behavior is not satisficing, so those studies are not relevant to testing the satisficing hypothesis.

6. Jordan, Marcus, and Reeder (1980) compared agree-disagree questions asked using show cards in face-to-face interviews to comparable items asked without show cards in telephone interviews.

7. Quinn, Gutek, and Walsh (1980) examined questions asking whether or not the respondent or a family member had experienced 28 problems or difficulties, but these questions did not have clear social desirability connotations. Wiseman (1972) and McQueen (1989) examined potentially informative attitudes and behaviors but did not report tests of statistical significance of observed differences between modes.

12 months than were face-to-face respondents, and Aquilino (1994) found that admission of illegal drug use was more likely in face-to-face interviews than in telephone interviews.⁸ In a study of University of Kentucky students, Johnson, Hougland, and Clayton (1989) found reports of more illegal drug use and alcohol use in face-to-face interviews than in telephone interviews, but the social desirability connotations of these behaviors for college students are not necessarily clear. Four other studies found no significant differences between telephone and face-to-face reports of being registered to vote, turning out to vote, being arrested for drunk driving, declaring bankruptcy, and experiencing symptoms associated with depression (Aneshensel et al. 1982; Aquilino 1998; Groves 1977; Groves and Kahn 1979; Locander, Sudman, and Bradburn 1976).

The Present Investigation

In light of how limited the relevant available evidence is regarding mode differences in satisficing and social desirability response bias, it seemed appropriate to explore these issues further, so we did so using three datasets: (1) the 1982 National Election Study Methods Comparison Project (MCP), an experiment designed to compare face-to-face interviewing of a block-listed national sample with RDD telephone interviewing of a national sample and conducted jointly by the University of Michigan's Survey Research Center (SRC) and the Program in Computer-Assisted Interviewing at the University of California, Berkeley, for the National Election Study (NES), (2) a comparable experiment conducted in 1976 by the University of Michigan's Survey Research Center (SRC) for Groves and Kahn (1979), and (3) a comparable experiment conducted as a part of the 2000 National Election Study.

All of these studies met the necessary criteria we outlined above: they all involved essentially identical questionnaires being administered to separate groups of individuals interviewed either face to face or by telephone who had not been previously interviewed and who were selected from their households by the same method. The telephone interviews were conducted with national RDD samples; the face-to-face interviews were conducted with national area probability samples; and the questionnaires were quite lengthy.

In our investigation, we tested our hypotheses controlling for differences between the face-to-face and telephone samples in terms of an array of demographic characteristics, which none of the 18 studies highlighted in the bottom section of table 1 did. There are several reasons to expect that certain demographic groups may be more frequently represented in a sample interviewed

8. Consistent with the notion that frequency of dental check-ups has social desirability connotations, Gordon (1987) reported evidence that having regular dental checkups is socially desirable. Weeks et al. (1983) examined five other health-related behaviors, but these behaviors do not have clear social desirability connotations. Aquilino (1994) found more reports of alcohol consumption in face-to-face interviews than in telephone interviews, but it is not clear that questions about alcohol use have clear social desirability connotations.

by one method than the other. First, as we mentioned, coverage error occurs in RDD telephone samples because about 5 percent of American households are without working telephones (Belinfante 1998). Members of households with telephones are more likely to be highly educated, to have high incomes, and to be women, older, and white than people living in households without phones (Gfroerer and Hughes 1991; Groves and Kahn 1979; Mulry-Liggan 1983; Wolfle 1979).

Second, the two modes may differ in the nature of unit nonresponse error as well if some sorts of people are willing to participate in surveys in one mode but not the other, which seems likely. Contact by a stranger over the telephone always involves a degree of uncertainty, so people who are most socially vulnerable because of a lack of power or resources may feel they have the most to lose by taking the risk of answering and may therefore be reluctant to participate in telephone interviews. Even if survey interviewers' calls are preceded by advance letters, and even if respondents have called a toll-free telephone number to reassure themselves about the identity of their interviewers, respondents cannot be completely sure their interviewers are the people they claim to be and cannot be sure that the questions being asked are truly for their purported purpose.

The same uncertainties exist when an interviewer knocks on a respondent's door, and the same means of reassurance are available. But the doorstep contact offers more: the nonthreatening and professional physical appearance of most interviewers and their equipment, along with their pleasant, friendly, professional, and nonthreatening nonverbal behaviors. All this may help to reassure respondents. Furthermore, the effort expended by the interviewer to travel to the respondent's home communicates a degree of professionalism that may assuage hesitations from reluctant respondents. Consequently, factors such as having limited income, having limited formal education, being female, elderly, and of a racial minority may all make respondents more reluctant to participate in telephone interviews than in face-to-face interviews.

Of course, doorstep contact entails another consideration as well: the risk that the interviewer might be physically threatening or even motivated to rob or otherwise take advantage of the respondent. This might lead some respondents, especially women and the elderly, to be reluctant to let a stranger into their home. Studies comparing respondents in telephone surveys to those who own telephones in face-to-face surveys (a method to eliminate coverage bias when examining nonresponse) suggest that telephone respondents are more likely to be well educated, to have high incomes, and to be male, older, and white (Gfroerer and Hughes 1991; Groves and Kahn 1979; Thornberry 1987; Weeks et al. 1983). This is further reason to expect that socially vulnerable groups will be less well represented in telephone surveys.

Consistent with this expectation, previous studies combining coverage error and unit nonresponse by comparing data collected by face-to-face interviewing of national area probability samples with data collected by national RDD

telephone interviews (the seven studies in the bottom section of table 1) have indeed documented reliable demographic differences between the samples. For example, all of the six studies that compared education levels for the two types of surveys found fewer low-education respondents in the telephone samples than in the face-to-face samples (Greenfield, Midanik, and Rogers 2000; Groves 1977; Groves and Kahn 1979; Klecka and Tuchfarber 1978; Mulry-Liggan 1983; Shanks, Sanchez, and Morton 1983; Weeks et al. 1983). All of the five studies that compared income levels found fewer low-income respondents in the telephone samples than in the face-to-face samples (Greenfield, Midanik, and Rogers 2000; Groves 1977; Groves and Kahn 1979; Klecka and Tuchfarber 1978; Shanks, Sanchez, and Morton 1983; Weeks et al. 1983). Of the six studies that examined age, five found fewer older people in the telephone samples than in the face-to-face samples (Groves 1977; Groves and Kahn 1979; Herzog, Rodgers, and Kulka 1983; Klecka and Tuchfarber 1978; Mulry-Liggan 1983; Weeks et al. 1983). Greenfield, Midanik, and Rogers (2000) found no age differences between a face-to-face sample and a telephone sample. And of the six studies that examined race, five found fewer minority respondents and more white respondents in the telephone samples than in the face-to-face samples (Greenfield, Midanik, and Rogers 2000; Klecka and Tuchfarber 1978; Mulry-Liggan 1983; Shanks, Sanchez, and Morton 1983; Weeks et al. 1983). In the remaining study, there were only slightly more whites interviewed in the telephone survey than in the face-to-face survey (Groves 1977; Groves and Kahn 1979).

Demographic characteristics are sometimes related to the likelihood that a respondent will satisfice (e.g., Narayan and Krosnick 1996) and to the likelihood that a respondent will have performed various sensitive behaviors or will hold certain sensitive attitudes. Therefore, it is important to control for demographic differences in order to isolate the effect of mode on no-opinion responding, nondifferentiation, acquiescence, and social desirability response bias, and we have done so.⁹

The 1982 NES Methods Comparison Project

DATA COLLECTION

The 1982 NES Methods Comparison Project (MCP) involved 998 telephone interviews and 1,418 face-to-face interviews with representative national sam-

9. It is probably impossible to measure all possibly relevant demographic variables, and controlling for demographics in this way requires the assumption that sample members from a particular demographic group adequately represent their population (an assumption routinely made when weighting samples). So our approach here will not completely eliminate all threats due to demographic differences between the samples, but it will help to reduce concern about this alternative explanation for our findings.

ples of noninstitutionalized American adults, conducted during November and December 1982 and January 1983. All of the face-to-face interviews were conducted by the University of Michigan's Survey Research Center and involved their conventional approach to area probability sampling via block-listing. The telephone sample was generated via RDD. Half of the telephone respondents (selected randomly) were interviewed by Michigan's SRC, and the other half were interviewed by the Survey Research Center at the University of California, Berkeley. A respondent was randomly chosen to be interviewed from among all eligible household members. The response rate was 72 percent for the face-to-face sample and 62 percent for the telephone sample (Shanks, Sanchez, and Morton 1983).¹⁰

Essentially identical questionnaires were used for all interviews; show cards that accompanied some questions in the face-to-face interviews were not used during the telephone interviews, but we did not analyze those items. The questionnaire was similar in length and character to those of other National Election Study surveys (which typically last over an hour) and asked about respondents' participation in politics, attitudes toward political candidates and public policies, and much more.

MEASURES

This survey's questionnaire permitted assessment of no-opinion responding, nondifferentiation, and social desirability response bias (for details of the measures and coding procedures, see app. A). No-opinion responding was measured by calculating the percent of questions that offered an explicit no-opinion response option to which each respondent answered "no opinion." Nondifferentiation was measured by counting the number of identical or nearly identical responses each respondent gave in answering two batteries of ratings using the same scale. Social desirability response bias was measured by calculating the proportion of questions with social desirability connotations to which a respondent gave the socially desirable response. We identified these items based upon a pretest designed to determine the extent to which items had social desirability connotations (for details on this survey, see app. B).

All variables were coded to range from 0 to 1, with 0 meaning the least possible no-opinion responses, the least possible nondifferentiation, and the least frequent offering of socially desirable answers, and 1 meaning the most

10. These response rates correspond to AAPOR's response rate 1. These response rates are a bit lower than those observed in other high quality surveys conducted at about the same time. For example, the University of Michigan's Monthly Survey of Consumer Attitudes, a telephone survey, had a response rate of 72 percent in 1982 (AAPOR response rate 2; the numerator included completed and partial interviews, and the denominator included all sampled phone numbers except those known to be ineligible; Curtin, Presser, and Singer 2000), and the General Social Survey, done face-to-face by the National Opinion Research Center, had a response rate of 77.5 percent in 1982 (Smith 1995).

possible no-opinion responses, the most possible nondifferentiation, and the most frequent offering of socially desirable answers.

ASSESSING MODE EFFECTS

We approached the assessment of mode effects in two ways. To gain the maximal statistical power by using the full array of cases, we compared the face-to-face interviews to the full set of telephone interviews. However, this comparison confounds mode with house, because Michigan conducted all the face-to-face interviews, but half the telephone interviews were done by Berkeley. If the standard interviewing practices at these institutions differentially encouraged or discouraged satisficing or socially desirable responses, the confounding of mode with house would yield misleading results regarding mode. To deal with this problem, we also conducted less powerful tests of mode differences comparing only the Michigan telephone respondents to the face-to-face respondents.

All statistical analyses were conducted using Stata, which allowed for proper weighting by the reciprocal of the known probability of selection. In the analyses reported below, the telephone respondents were weighted by the number of adult residents in the household and by the reciprocal of the number of telephone lines in the household, and the face-to-face respondents were weighted by the number of adult residents in the household.¹¹

In order to reduce travel costs for face-to-face interviews, clusters of households were selected for generating the area probability samples. This clustering reduces standard error estimates and makes statistical tests misleadingly liberal. We therefore controlled for clustering at the level of the primary sampling unit in the face-to-face sample.

RESULTS

No-opinion responses. The first two columns of rows 1 and 2 in table 2 display the adjusted mean proportions of no-opinion responses for the face-to-face respondents and the telephone respondents.¹² The first row combines the Michigan and Berkeley telephone respondents, and the second row displays figures using only the Michigan telephone respondents. The remaining columns of the table display the results of ordinary least squares (OLS) regressions predicting the proportion of no-opinion responses using mode (coded

11. The number of telephone lines was not recorded for the Berkeley telephone respondents. Therefore, all Berkeley telephone respondents were assigned a value of 1 for the number of telephone lines.

12. These means were adjusted for demographic differences between the two respondent groups interviewed in the two modes.

Table 2. Regression Coefficients Estimating the Impact of Mode on Satisficing

Response Strategy	House(s)	Adjusted Means		Regression Coefficients										R ²	N
		Face-to-		Mode	Education	Income	Race	Gender	Age	Age ²	Married	Employed			
		Face	Telephone												
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)		
1982 NES MCP:															
No-opinion responding:															
Full sample	Michigan & Berkeley	.17	.24	.07**	-.29**	-.07**	.05*	.02*	-.08	.11 ⁺	.00	.01	.18	2,093	
	Michigan	.17	.26	.08**	-.28**	-.08**	.04	.02	-.06	.10 ⁺	.01	.00	.18	1,680	
Low education	Michigan & Berkeley	.32	.46	.14**	-.37 ⁺	-.12	.01	.08*	.22	-.17	.01	-.02	.10	410	
	Michigan	.32	.53	.22**	-.43 ⁺	-.15 ⁺	-.02	.09**	.19	-.16	.02	-.03	.14	350	
High education	Michigan & Berkeley	.14	.18	.05**	-.22**	-.04 ⁺	.06*	.01	-.17**	.17**	.01	.01	.11	1,683	
	Michigan	.14	.19	.05**	-.21**	-.05*	.05 ⁺	.00	-.16*	.17**	.01	.01	.10	1,330	
Nondifferentiation:															
Full sample	Michigan & Berkeley	.37	.41	.04**	-.02	-.02	-.01	-.01	-.13*	.08 ⁺	.00	.01	.02	2,095	
	Michigan	.37	.41	.03**	-.01	-.02	-.01	-.01 ⁺	-.13*	.08	.00	.02	.02	1,682	
Low education	Michigan & Berkeley	.38	.44	.06**	.04	.05	.00	-.02	-.09	.09	.02	.02	.05	411	
	Michigan	.38	.41	.03	.05	.04	-.01	-.03	.01	.01	.02	.03	.04	351	
High education	Michigan & Berkeley	.37	.41	.04**	.00	-.02	-.01	-.01	-.13*	.07	-.01	.00	.03	1,684	
	Michigan	.37	.40	.03**	.02	-.03	-.01	-.01	-.14*	.07	.00	.01	.03	1,331	

1976 SRC datasets:													
No-opinion responding:													
Full sample	.62	.69	.35**	-2.50**	-.75*	.27	.72**	-1.47	1.35	.16	.10	1,344	
Low education	.80	.86	.42	-2.48	-.34	-.44	.66 ⁺	3.63 ⁺	-4.59 ⁺	-.35	-.22	359	
High education	.55	.63	.34*	-2.46**	-.73 ⁺	.57*	.74**	-2.84*	3.25*	.30 ⁺	.16	985	
Nondifferentiation:													
Full sample	.62	.65	.03*	.00	.16**	-.11**	-.06**	.39**	-.18 ⁺	-.03*	-.03	.09	2,633
Low education	.60	.65	.05*	.10	.20**	-.13**	-.06 ⁺	.41*	-.11	.00	.00	.13	684
High education	.63	.65	.02	-.07 ⁺	.15**	-.09**	-.07**	.44**	-.25 ⁺	-.05**	-.04 ⁺	.08	1,949
Acquiescence:													
Full sample	.33	.37	.03*	-.05	.01	.06*	-.02	.00	.24	.01	-.03	.02	2,485
Low education	.38	.43	.06	-.18	-.23 ⁺	.03	.03	-.23	.36	.08*	.01	.03	605
High education	.32	.34	.03	-.03	.05	.07*	-.03	-.02	.33	.00	-.04	.02	1,880
2000 NES:													
No-opinion responding:													
Full sample	.11	.19	.07**	-.14**	-.05**	.08**	.03**	-.53**	.67**	.01	.03*	.17	1,488
Low education	.14	.22	.08**	-.20**	-.06*	.09**	.04**	-.54**	.67**	.00	.04*	.16	1,029
High education	.06	.12	.06**		-.03	.05*	.01	-.52**	.62*	.01	.01	.07	459
Acquiescence:													
Full sample	.31	.33	.02*	.10**	.00	.02	-.02	.20**	-.19*	.01	-.01	.05	1,488
Low education	.29	.32	.03*	.14**	.01	.01	-.01	.17 ⁺	-.13	.01	-.01	.05	1,029
High education ^a	.34	.35	.01		-.01	.04	-.03*	.30 ⁺	-.34 ⁺	.01	-.02	.03	459
Interview time	70.75	64.96	-5.79**	9.32**	-1.45	.59	-.55	40.82**	-16.65	-.03	-1.99	.09	1,487

NOTE.—All variables were coded to range from 0 to 1. Mode was coded 0 for face-to-face and 1 for telephone. Gender was coded 0 for males and 1 for females. Race was coded 0 for whites and 1 for nonwhites. All coefficients are from OLS regressions except coefficients for no-opinion responding for the 1976 SRC datasets, which are from logistic regressions.

^a Education has no coefficient in this equation because there was no variance in the coding of education within this group of respondents.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.

0 for face-to-face respondents and 1 for telephone respondents) and various demographic control variables.¹³

Higher levels of no-opinion responding occurred in the telephone samples (Michigan and Berkeley adjusted mean = 24 percent, Michigan adjusted mean = 26 percent) than in the face-to-face sample (adjusted mean = 17 percent), consistent with the satisficing hypothesis. The difference between the telephone and face-to-face samples was significant regardless of whether we included or dropped the Berkeley data (b 's = .07 and .08, $p < .01$).¹⁴

To test whether the mode effect varied with respondent education, we repeated these analyses separately for respondents who had not graduated from high school and for respondents with more education (for the rationale for this split, see Narayan and Krosnick [1996]). As shown in rows 3–6 of table 2, the mode effect was larger among the least educated respondents than among more educated respondents. When looking only at the Michigan data, the average proportion of no-opinion responses increased from 32 percent in the face-to-face interviews to 53 percent on the telephone ($b = .22$, $p < .01$). The difference was smaller but nonetheless significant when the Berkeley data were folded in ($b = .14$, $p < .01$). The mode effect was smaller in the highly educated subsample, though it was statistically significant there as well ($b = .05$, $p < .01$).

Nondifferentiation. In rows 7–12 of table 2, we see evidence consistent with the satisficing hypotheses regarding nondifferentiation. There was more nondifferentiation in the telephone samples (adjusted mean = .41) than in the face-to-face sample (adjusted mean = .37). This later rate was significantly lower than the telephone rate, whether we excluded the Berkeley data ($b = .03$, $p < .01$) or included it ($b = .04$, $p < .01$).

When only the Michigan data were considered, the mode effect was no stronger in the least educated group ($b = .03$, N.S.) than in the more educated group ($b = .03$, $p < .01$). But when the Berkeley data were included, the mode effect was stronger in the least educated group ($b = .06$, $p < .01$) than in the more educated group ($b = .04$, $p < .01$), as expected.

Social desirability. Respondents interviewed by telephone gave socially desirable responses more often (Michigan and Berkeley adjusted mean = .46; Michigan adjusted mean = .44) than did respondents interviewed face to face (adjusted mean = .41), regardless of whether the Berkeley respondents were included ($b = .05$, $p < .01$; see row 1 of table 3) or excluded ($b = .03$, $p < .10$; see row 2 of table 3).

13. Ordinary least squares regressions were conducted for all dependent variables with three or more possible values. When a dependent variable had only two possible values, logistic regressions were conducted.

14. Throughout this article, significance tests of directional predictions are one-tailed, and tests of differences for which we did not make directional predictions are two-tailed. When a directional prediction was tested but the observed mean difference was in the opposite direction, a two-tailed test is reported.

Table 3. Regression Coefficients Estimating the Impact of Mode on Reporting Socially Desirable Attitudes and Behaviors and Uneasiness Discussing Such Topics

Response Strategy	House(s)	Adjusted Means		OLS Regression Coefficients									R ²	N
		Face-to-		Mode	Education	Income	Race	Gender	Age	Age ²	Employment			
		Face	Telephone								Married	Status		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)		
Socially desirable responding:														
1982 NES MCP	Michigan & Berkeley	.41	.46	.05**	.33**	.06*	.13**	-.02 ⁺	.34**	.00	.02	.00	.21	2,095
1976 SRC datasets	Michigan	.41	.44	.03 ⁺	.34**	.05 ⁺	.14**	-.03*	.39**	-.08	.02	.00	.21	1,682
2000 NES		.76	.81	.05**	.31**	.05 ⁺	-.12**	.01	.76**	-.56**	.04**	.03*	.20	2,627
Unease discussing sensitive topics:		.41	.44	.03*	.14**	.03	.06**	.00	.48**	-.02	.04**	-.04**	.25	1,488
1976 SRC datasets		.08	.14	.05**	.00	-.00	.01	.01	-.13**	.04	.01	.00	.04	2,630

NOTE.—All variables were coded to range from 0 to 1. Mode was coded 0 for face-to-face and 1 for telephone. Gender was coded 0 for males and 1 for females. Race was coded 0 for whites and 1 for nonwhites.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.

Summary. In sum, the 1982 NES MCP data showed that respondents manifested indications of greater satisficing over the telephone than face to face, and this effect was larger among the least educated respondents. Furthermore, socially desirable attitudes were reported more often by telephone respondents than by face-to-face respondents.

1976 Survey Research Center Datasets

Next, we explored whether these findings would replicate in a second, comparable experiment conducted only by the University of Michigan's Survey Research Center. This survey also involved a comparison of a block-listed national sample interviewed face to face with a national RDD sample interviewed by telephone. The questionnaire allowed us to examine no-opinion responding, nondifferentiation, acquiescence, and social desirability response bias. The questionnaire also allowed us to explore whether respondents were more likely to express unease about discussing sensitive topics over the phone than face to face, which would be consistent with the social desirability findings from the 1982 NES MCP. And we explored whether respondents expressed impatience with telephone interviews more often than they did with face-to-face interviews, which would be consistent with the logic articulated above to justify our suspicions about the tendency to satisfice in telephone surveys.

These data had been analyzed previously by Groves and Kahn (1979), but those investigators did not test most of the hypotheses we explored. Relevant to the social desirability hypothesis, Groves and Kahn (1979) and Groves (1979) reported that respondents expressed more discomfort about discussing sensitive topics (e.g., racial attitudes, political opinions, and voting) over the telephone than face to face, and their telephone sample claimed to have voted in recent elections at higher rates than did their face-to-face sample. And these investigators reported that most respondents said they would prefer to be interviewed face to face rather than by telephone. But none of these differences was tested controlling for the demographic differences between the two modes' samples, and none of the satisficing-related hypotheses articulated above were tested by Groves and Kahn (1979) at all. It therefore seemed worthwhile to revisit these data to conduct more comprehensive analyses of them.

DATA COLLECTION

The face-to-face interviews were conducted during the spring of 1976, with a multistage stratified area probability sample of the coterminous United States. Households were randomly selected from within 74 primary sampling areas, and respondent selection within households was accomplished by the Kish

(1949) method of random selection from a complete household listing. The response rate for the face-to-face mode was 74.3 percent.¹⁵

Two RDD samples were generated for the telephone interviews, which were also conducted during the spring of 1976. The first was a stratified RDD sample, in which working combinations of area codes and three-digit central office codes were selected systematically from a stratified list. A second, clustered sample was generated by selecting among area codes and central office codes for working telephone numbers within the 74 primary sampling units of the Survey Research Center's national sample of dwellings. Respondent selection within households was accomplished by the Kish (1949) method, adapted for telephone administration.¹⁶ The response rate for the telephone mode was 70 percent if we assume that all numbers unanswered after 16 attempts were nonworking, and 59 percent if we assume that none of these numbers were nonworking.¹⁷

The questionnaires used in both modes addressed consumer attitudes and behaviors, economic beliefs, life satisfaction and living conditions, political attitudes, and more. The face-to-face questionnaire was a bit longer than the telephone questionnaire, because some sets of questions asked late in the face-to-face interviews were omitted from the telephone questionnaires. In addition, some questions asked with show cards in the face-to-face interviews were asked without any visual displays during the telephone interviews. Our analyses focus on questions that were asked identically in the two modes, that were asked quite early in the interviews, and that had essentially identical questions preceding them in the two modes.

MEASURES

No-opinion responding, nondifferentiation, and social desirability response bias were assessed as in the 1982 NES MCP. Acquiescence was measured by calculating the percent of the agree/disagree and yes/no questions a respondent was asked to which he or she responded "agree" or "yes." Respondent dissatisfaction with the length of the interview was assessed by a question that explicitly asked respondents whether they thought the interview had taken

15. This response rate is similar to those observed in other high quality surveys being conducted at about the same time. For example, the GSS had a response rate of 73.5 percent in 1978 (Smith 1995).

16. The face-to-face data were weighted by the number of eligible adults in the household, and the telephone data were weighted by the number of eligible adults in the household and by the inverse of the number of residential phone lines in the household. Clustering in primary sampling units was controlled for in the face-to-face data. We were not able to control for clustering in the clustered telephone component of the sample because no variable differentiated the two telephone samples from one another.

17. These response rates were calculated by dividing the number of completed interviews by the total number of eligible households. The higher of these response rates is identical to that of the Survey of Consumer Attitudes (approximately 70 percent) in 1979 (Curtin, Presser, and Singer 2000).

too long and by the interviewer's record of whether the respondent ever spontaneously voiced dissatisfaction that the interview was taking a long time. All measures were coded to range from 0 (meaning the least possible no-opinion responding, nondifferentiation, acquiescence, socially desirable responding, and dissatisfaction with interview length) to 1 (meaning the most possible no-opinion responding, nondifferentiation, acquiescence, socially desirable responding, and dissatisfaction with interview length). Full details about the measures and coding procedures are provided in appendix A.

RESULTS

No-opinion responses. The telephone respondents chose an explicitly offered no-opinion response option more often than did the face-to-face respondents (62 percent for face-to-face respondents and 69 percent for telephone respondents; logistic regression coefficient = .35, $p < .01$; see the first row in the middle panel of table 2). This effect was slightly larger among low-education respondents (logistic regression coefficient = .42, N.S.) than among high-education respondents (logistic regression coefficient = .34, $p < .05$), although the former was not significant and the latter was.¹⁸

Nondifferentiation. Nondifferentiation was more prevalent in the telephone sample than in the face-to-face sample ($b = .03$, $p < .05$; see col. 3 in row 4 of the middle panel of table 2). Furthermore, this mode effect was significant among the least-educated respondents ($b = .05$, $p < .05$; see col. 3 of row 5 in the middle panel of table 2), and smaller and not significant in the high-education group ($b = .02$, N.S.; see col. 3 of row 6 in the middle panel of table 2).

Acquiescence. The telephone respondents were more likely to acquiesce than were the face-to-face respondents ($b = .03$, $p < .05$; see row 7 of the middle panel of table 2). The coefficient estimating this effect was slightly larger among low-education respondents ($b = .06$, N.S.) than among high-education respondents ($b = .03$, N.S.), but neither of those coefficients was significant.

Dissatisfaction with interview length. Respondents interviewed by telephone were significantly more likely than the face-to-face respondents to express dissatisfaction with the interview's length ($b = .06$, $p < .01$; see col. 3 of row 2 in table 4) and to ask how much longer the interview would take (logistic regression coefficient = .98, $p < .01$; see col. 3 of row 1 in table 4).

Social desirability. The telephone respondents were more likely to give socially desirable responses than were the face-to-face respondents ($b =$

18. No-opinion responding was unusually common for this question (greater than 60 percent). Unlike questions that tag a no-opinion response option on at the end of a list of substantive answer choices, this item began by saying: "Not everyone has an opinion on this next question. If you do not have an opinion, just say so." This heavy-handed encouragement of no-opinion responses seems likely to explain their popularity.

Table 4. Regression Coefficients Estimating the Impact of Mode on Respondent Dissatisfaction and Engagement

Respondent Reaction	Adjusted Means		Regression Coefficients									<i>R</i> ²	<i>N</i>
	Face-to-Face (1)	Telephone (2)	Mode (3)	Education (4)	Income (5)	Race (6)	Gender (7)	Age (8)	Age ² (9)	Married (10)	Employed (11)		
1976 SRC datasets:													
Asked how much longer the interview would be	.04	.11	.98**	.41	-.69	.52*	.05	2.10 ⁺	-1.27	.08	-.51*		2,625
Expressed dissatisfaction with interview length	.52	.58	.06**	.03**	-.03 ⁺	.02 ⁺	.00	.03	-.01	.02**	.00	.07	2,570
2000 NES:													
Interviewer rating of respondent interest	.73	.69	-.04*	.21**	.03	.00	-.02 ⁺	.36*	-.27	-.01	-.01	.12	1,487
Interviewer rating of respondent cooperation	.91	.88	-.02*	.08**	.02	-.01	-.01	-.02	.04	.01	.00	.05	1,487
Interviewer rating of respondent suspicion	.08	.12	.04**	-.02	-.01	.05*	.00	.15	-.17	.01	-.01	.02	1,487
Said he or she wanted to stop the interview	.01	.02	.98 ⁺	.35	-1.30	1.21 ⁺	.40	12.39 ⁺	-6.53	.01	.43		1,488
Said interview was too long	.01	.09	2.03**	.18	-.32	.03	.41	2.85	-1.99	-.08	.12		1,488

NOTE.—All variables were coded to range from 0 to 1. Mode was coded 0 for face-to-face and 1 for telephone. Gender was coded 0 for males and 1 for females. Race was coded 0 for whites and 1 for nonwhites. OLS regression coefficients are shown for dissatisfaction with interview length, and respondent suspicion, cooperation and interest, and logistic regression coefficients are shown for asking how much longer the interview would be, complaining about the interview length, and wanting to stop at some point during the interview.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.

.05, $p < .01$; see col. 3 of row 3 of table 3). And respondents interviewed by telephone expressed significantly more unease about discussing sensitive topics than did respondents interviewed face to face ($b = .05$, $p < .01$; see row 5 of table 3).

2000 National Election Study

Finally, we tested the satisficing and social desirability hypotheses using data from a more recent survey conducted by the University of Michigan's Survey Research Center for the 2000 National Election Study. This survey compared an area probability sample of 1,006 people interviewed face to face to an RDD sample of 801 people interviewed by telephone. The questionnaire allowed us to assess the extent of no-opinion responding, acquiescence, social desirability response bias, and respondent dissatisfaction with interview length. We also examined the extent to which respondents expressed suspicion about the trustworthiness of the interview process, which the logic articulated above suggests might be a precursor of social desirability response bias.

In addition, this questionnaire allowed us to examine the effect of mode on interview length. The logic we articulated earlier about the pressure to move quickly through telephone conversations suggests that telephone interviews may be completed more quickly than are face-to-face interviews. And we examined respondents' level of interest in the interview and their cooperativeness, on the assumption that more interest and cooperation were signs of greater engagement in the process and less inclination toward satisficing.

DATA COLLECTION

Face-to-face and telephone interviewing began on September 5, 2000, and ended on November 6, 2000. The population for these surveys was all U.S. citizens of voting age. Within each household, an eligible respondent was randomly chosen to be interviewed.¹⁹ The response rate for the face-to-face interviews was 64.3 percent, and the response rate for the telephone interviews was 56.5 percent.²⁰ The questionnaires addressed political attitudes and behaviors and often focused on the upcoming presidential election.

19. The number of telephone lines in the household was not recorded for the telephone respondents, so the telephone and face-to-face respondents were weighted only by the number of adults in the household. Clustering in primary sampling units was controlled for in the face-to-face data.

20. These response rates were calculated by dividing the number of completed interviews by the total number of potential respondents (and correspond to AAPOR's response rate 1). The response rates for these two surveys were somewhat lower than those observed in other contemporaneous high quality surveys. The Survey of Consumer Attitudes had a response rate of approximately 67 percent in 1996 (Curtin, Presser, and Singer 2000), and a national RDD telephone survey using rigorous methodology conducted in 1997 by the Pew Research Center for the People and the Press had a response rate of 60.6 percent (AAPOR response rate 3; the

MEASURES

No-opinion responding, acquiescence, social desirability response bias, and respondent dissatisfaction with interview length were gauged as in the 1976 SRC datasets. And interviewers recorded spontaneous expressions of respondent suspicion and rated respondent engagement in the interview process. All of these variables were coded to range from 0 (meaning the least possible no-opinion responding, acquiescence, social desirability response bias, respondent dissatisfaction, suspicion, and engagement) to 1 (meaning the most possible no-opinion responding, acquiescence, social desirability response bias, respondent dissatisfaction, suspicion, and engagement). Interview length was recorded in minutes. Details about the measures and coding procedure are provided in appendix A.

RESULTS

No-opinion responses. The telephone respondents were more likely than the face-to-face respondents to choose a no-opinion response option (19 percent for telephone respondents vs. 11 percent for face-to-face respondents; $b = .07, p < .01$; see col. 3, row 1 in the bottom panel of table 2). This effect was slightly stronger among the low-education respondents ($b = .08, p < .01$; see col. 3, row 2 in the bottom panel of table 2) than among the high-education respondents ($b = .06, p < .01$; see col. 3, row 3 in the bottom panel of table 2).

Acquiescence. Respondents interviewed by telephone were significantly more likely to give “agree” and “yes” responses than were respondents interviewed face to face ($b = .02, p < .05$; see col. 3, row 4 in the bottom panel of table 2), and this effect was significant among low-education respondents ($b = .03, p < .05$; see col. 3, row 5 in the bottom panel of table 2), but not among high-education respondents ($b = .01, N.S.$; see col. 3, row 6 in the bottom panel of table 2).

Dissatisfaction with interview length. Respondents interviewed by telephone were significantly more likely than face-to-face respondents to complain that the interview was too long (9 percent of telephone respondents vs. 1 percent of face-to-face respondents; $b = 2.03, p < .01$; see col. 3 of row 5 in the bottom panel of table 4) and to want to stop at some point during the interview (2 percent of telephone respondents vs. 1 percent of face-to-face respondents; $b = .98, p < .10$; see col. 3 of row 4 in the bottom panel of table 4).

Respondent engagement. Respondents interviewed by telephone were rated as less cooperative ($b = -.02, p < .05$; see row 2 in the bottom panel of table

numerator included only completed interviews, and the denominator included all sample numbers known to be eligible and 20 percent of the sample numbers for which eligibility was not known). The GSS had a response rate of 82 percent in 1998 (Smith 1995).

4) and less interested in the survey ($b = -.04, p < .05$; see row 1 in the bottom panel of table 4) than were respondents interviewed face to face.

Interview length. The face-to-face interviews were approximately 6 minutes longer than telephone interviews on average ($b = -5.79, p < .01$; see col. 3, row 7 in the bottom panel of table 2).

Social desirability. The telephone respondents were more likely to give socially desirable answers than were the face-to-face respondents ($b = .03, p < .05$; see col. 3 of row 4 in table 3). And the telephone respondents were more likely to express suspicion about the interview process than were the face-to-face respondents ($b = .04, p < .01$; see col. 3 of row 3 of the bottom panel of table 4).

Meta-analysis of the Education Effect

Satisficing theory anticipates that the mode effect on use of satisficing response strategies may be larger among less-educated respondents. In eight of the nine tests of this hypothesis reported thus far, we saw differences between high- and low-education groups in the expected direction. And in the ninth instance, the mode effect was of equal magnitude in the two education groups. This consistent pattern of differences between the education groups suggests that there may in fact be a meaningful trend here consistent with satisficing theory. However, the difference between the high- and low-education groups was statistically significant only in the case of no-opinion responding in the 1982 NES MCP data and was marginally significant in the case of acquiescence in the 2000 NES (see the last column of table 5).

Meta-analysis was designed precisely to test hypotheses in these sorts of circumstances, where multiple tests point in similar directions (Rosenthal 1984). Therefore, to test whether the moderating effect of education was in fact reliable, we performed a meta-analysis using the data from all three of our studies at once. We compared the difference in the impact of mode between respondents low and high in education across the satisficing indicators using the statistics shown in table 5. At the bottom of table 5 are two meta-analytic test statistics representing the effect of education combined across the 1982 NES MCP, the 1976 Survey Research Center datasets, and the 2000 National Election Study data. One of these tests used the Michigan and Berkeley data from the 1982 NES MCP, in addition to the 1976 and the 2000 data. The other test used only the Michigan data from the 1982 survey, in addition to the 1976 and 2000 data. These statistics were generated by computing a planned contrast between the education groups of the combination of the seven p -values of the mode effect tests within each group.

Both test statistics suggest that the role of education in moderating the effect of mode on satisficing was statistically reliable and in the expected direction (1976 data plus 2000 data plus the 1982 Michigan and Berkeley data: $z =$

Table 5. Meta-analysis of the Moderating Effect of Education on the Impact of Mode

Response Strategy	Low-Education Coefficient	High-Education Coefficient	z-Test of the Education Difference
1982 NES MCP:			
Michigan and Berkeley:			
No-opinion responding	.14** (.03)	.05** (.01)	2.85**
Nondifferentiation	.06** (.02)	.04** (.01)	.89
Michigan:			
No-opinion responding	.22** (.05)	.05** (.01)	3.33**
Nondifferentiation	.03 (.03)	.03** (.01)	.00
1976 SRC datasets:			
No-opinion responding	.42 (.31)	.34** (.15)	.23
Nondifferentiation	.05** (.03)	.02 (.02)	.88
Acquiescence	.06 (.04)	.03 (.02)	.74
2000 NES:			
No-opinion responding	.08** (.02)	.06** (.01)	1.26
Acquiescence	.03** (.01)	.01 (.01)	1.56 ⁺
Combination of significance levels from all three surveys:			
Using the Berkeley and Michigan data from the 1982 NES MCP, the 1976 SRC dataset, and the 2000 NES			3.11**
Using only the Michigan data from the 1982 NES MCP, the 1976 SRC dataset, and the 2000 NES			2.85**

NOTE.—OLS regression coefficients are shown for all effects except no-opinion responding for the 1976 SRC datasets, for which logistic regression coefficients are shown. Standard errors are shown in parentheses.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.

3.11, $p < .01$; 1976 data plus 2000 data plus the 1982 Michigan data only: $z = 2.85$, $p < .01$).

General Discussion

Taken together, this evidence suggests that interview mode can affect response patterns in long interviews with representative national samples. In particular, answers given during telephone interviews of RDD samples appear to have manifested more satisficing and greater social desirability response bias than

did answers given during face-to-face interviews of area probability samples. Furthermore, respondents interviewed by telephone appear to have been more suspicious and less cooperative and less interested in the survey, suggesting they may have had less motivation to generate optimal answers. These differences are consistent with the notion that the rapport probably developed during the lengthy face-to-face interviews may have inspired respondents to work harder at providing high-quality data, even when doing so meant admitting something that may not have been socially admirable.²¹

The magnitudes of the mode effects documented here might appear to be small enough to justify concluding that there is no reason for concern about the telephone mode. And these mode effects on data quality may appear even smaller in light of the large cost savings associated with telephone interviewing relative to face-to-face interviewing. However, we have seen that telephone interviewing is associated with an increase in systematic bias in response patterns, and these effects were sometimes sizable among respondents who were the least educated. Furthermore, it is well established that telephone samples underrepresent low-education respondents, low-income respondents, and minority respondents. Therefore, if one intends survey research to give equally loud voices to all members of society, the biases apparently associated with telephone interviewing may discriminate against population segments that already have limited impact on collective decision making in democracies.

There is reason for concern here even among researchers who do not view surveys as providing vehicles for public influence on public policy and societal deliberation. For example, our findings suggest that basic researchers interested in comparisons across population subgroups may reach different conclusions depending upon which mode they employ. Specifically, many studies have explored the notion that more socially disadvantaged segments of democratic publics are less likely to have opinions on political issues and therefore have less to offer the collective decision-making process (for a review, see Krosnick and Milburn [1990]). One would reach this conclusion more strongly when analyzing telephone survey data than when analyzing face-to-face data.

This perspective suggests potential costs that may be borne by shifting long-term large-scale survey studies that have been done face to face for many years over to the telephone mode in order to cut costs. Because comparisons over time are the lifeblood of these studies, any shift of mode confounding substantive shifts in the phenomena of interest with methodological perturbations may cloud these studies' abilities to make clean historical comparisons.

21. Our findings regarding reports of behaviors and attitudes with social desirability connotations may seem surprising given evidence that people are more likely to disclose potentially embarrassing behaviors and attitudes when their reports are anonymous (Himmelfarb and Lickteig 1982; Paulhus 1984; Warner 1965). However, our findings suggest that any benefit of increased privacy over the telephone for the accuracy of reports of sensitive behaviors and attitudes is less than the advantage of greater rapport developed in face-to-face interviews.

Our results therefore suggest caution regarding mode shifts, pointing to systematic biases that may emerge in the data.

OTHER STUDIES OF SOCIAL DESIRABILITY RESPONSE BIAS AND MODE

Our evidence that telephone respondents manifested more social desirability response bias than face-to-face respondents might appear to some readers to conflict with the findings of some past investigations sometimes viewed as offering evidence on this point (e.g., Aquilino 1992; Aquilino and LoSciuto 1989, 1990; Hochstim 1962; Sykes and Collins 1987, 1988). However, a close look at those past studies shows that their results do not in fact conflict with ours. For example, in all of Aquilino's studies (Aquilino 1992; Aquilino and LoSciuto 1989, 1990), respondents interviewed in the face-to-face mode actually answered the drug use questions (which were used to assess social desirability bias) on self-administered paper-and-pencil questionnaires, not aloud. Hochstim (1962) did not report tests of the statistical significance of observed differences, and Sykes and Collins (1987, 1988) did not report enough details of their quantitative findings to permit a meaningful assessment of the results. Furthermore, neither of those studies documented that the items they presumed had social desirability implications did indeed do so. Therefore, we do not see a basis for concluding that our social desirability findings are inconsistent with those of these other studies.

TRENDS OVER TIME AND ACROSS HOUSES

We have examined new data, from 2000, and what might be considered fairly old data, collected 21 and 27 years ago. Face-to-face and telephone interviewing have certainly changed over the years, and these procedures are no doubt practiced differently today across different "houses" in the United States. So it is difficult to generalize about these methodologies today or in the past from a single study.

In that light, it is reassuring that we saw similar patterns across a series of three studies spread out in time and across different sorts of measures of response quality included in them. Furthermore, there is no notable trend in tables 2–5 indicating that the mode effects we uncovered were any weaker in the older data or in the newer data. Nonetheless, both face-to-face and telephone interviewing could change in the future in ways that make the patterns we have documented here no longer applicable. Indeed, these findings may not even apply directly today to some houses, because the University of Michigan and the University of California at Berkeley's interviewing facilities are relatively expensive, academic operations, which may differ from other academic and nonacademic survey organizations in their procedures and response quality.

IMPLICATIONS FOR REINTERVIEWING

Some previous research yielded evidence that appears consistent with our findings and therefore reinforces confidence in them. Although demographics were not included as controls, Groves (1979) found that respondents interviewed face to face were very satisfied with the process, a majority (78 percent) of them saying they would rather be interviewed face to face than by telephone. In contrast, only 39 percent of respondents interviewed by telephone indicated satisfaction with that method; the majority of these individuals said they would prefer to provide data through face-to-face interviews or self-administered questionnaires. Not surprisingly, people interviewed by telephone said they preferred another mode of interviewing most often because it would allow them more time to think about the questions. This is consistent with the notion that telephone interviewing encourages satisficing even among people who would otherwise prefer to optimize instead.

If, in fact, there are such dramatic differences between the modes in respondent satisfaction, this has at least two interesting implications. First, being interviewed by telephone may be more frustrating than fulfilling for individuals, and these people may be less willing to participate in other surveys in the future because their initial experience was not comfortable and rewarding. Second, individuals who have been interviewed once by telephone may be especially unwilling to participate in a follow-up interview as a part of the same study by telephone, because they can anticipate what the experience will probably be like. Therefore, panel reinterview rates may be lower for telephone surveys than for face-to-face surveys partly because follow-up refusal rates may be higher in the former than the latter.

OTHER ADVANTAGES AND DISADVANTAGES OF FACE-TO-FACE INTERVIEWING

The response quality advantages associated with face-to-face interviewing apparent here are not the only strengths of this method, of course. A significant additional advantage is response rates, which tend to be at least 10 percentage points higher for face-to-face than telephone surveys (Aneshensel et al. 1982; de Leeuw 1992; Henson, Roth, and Cannell 1977; Hox and de Leeuw 1994; Thornberry 1987), as they were in the experiments we analyzed here. As new technologies such as call-blocking make it increasingly difficult to reach potential respondents by telephone, telephone response rates may continue to drop (holding budget constant), while face-to-face response rates may be less susceptible to such declines in participation (Smith 1995). Furthermore, visual aids (show cards) are more difficult to employ in telephone interviews than in face-to-face interviews. Using show cards without in-person assistance may be especially challenging for respondents with more limited literacy.

Another advantage of face-to-face interviewing is the capacity to employ

new measurement technologies. For example, assessments of reaction time with millisecond resolution and subliminal presentation of visual stimuli are core parts of social cognition's new generation of tools for assessing attitudinal dispositions such as racism and various other aspects of cognitive structure and processes (see, e.g., Bargh and Chartrand 2000; Fazio et al. 1995). Enriching the study of public opinion with these methodologies is more difficult with telephone interviewing, though it is readily accomplished if face-to-face interviewers turn their laptops around so that respondents can see the screens and complete tasks on their own. Of course, laptops are routinely used in face-to-face interviewing these days to permit Audio Computer Assisted Self-Interviews (ACASI), so respondents can answer sensitive questions privately, without the involvement of the interviewer.²² Therefore, the use of laptops for unobtrusive measurement of information processing would be a natural and potentially very rich expansion of our toolbox for studying public opinion in the general population.

IMPROVING RESPONSE QUALITY IN TELEPHONE INTERVIEWS

Although the findings reported here suggest that face-to-face interviews provide higher response quality than do telephone interviews, our findings do not pinpoint precisely why these differences appear. We have speculated that the differences are a result of greater trust and rapport and more effective nonverbal communication in face-to-face interviews, as well as less multitasking and more comfort moving slowly through the latter. This logic suggests that telephone interviewing might be improved if it were able to emulate these characteristics of face-to-face interviewing.

Reductions in multitasking and improved nonverbal communication during telephone interviews are not likely to occur until videophones become commonplace in American homes. But steps can be taken now to encourage interviewers to slow the pace of telephone interviews, and enhancing interviewer credibility by sending out advance letters may be at least somewhat effective in shrinking the gap between telephone and face-to-face interviewing by reducing suspicion (see Miller and Cannell [1982] for additional techniques to improve telephone data quality).

Conclusion

The book is far from closed on the relation of interview mode to data quality in national probability sample surveys, and this issue will remain an important

22. Although having respondents answer questions privately on a computer is often done as part of face-to-face interviews, new technology such as telephone audio computer assisted self-interviewing (T-ACASI) may soon make it possible to collect data via similar methods in telephone interviews (Turner et al. 1998).

one for survey researchers. The findings reported here indicate that although telephone interviewing may be particularly appealing to researchers doing such studies because of its affordability, there may be costs associated with this method in terms of response quality. Thus, at least to some extent, we may get what we pay for.

But we must guard against overgeneralizing the findings reported here. Most survey studies conducted around the world today do not involve national American samples and such long questionnaires. Therefore, what we have seen here may not generalize to other, more conventional survey settings. Furthermore, we have focused here only on survey satisficing and social desirability bias, and other types of response errors may not show the same mode-related patterns we have documented here. We hope the findings reported here encourage researchers to continue the investigation of mode effects and to do so in ways driven by theories of information processing and social interaction, so in the long run, we gain a fuller understanding of the trade-offs inherent in mode choices and their impact on the findings of past research.

Appendix A

Measures and Codings

1982 NES Methods Comparison Project

No-opinion responses. Seven questions measuring attitudes explicitly offered respondents no-opinion response options. Five involved 7-point scale ratings of attitudes toward public policies (regarding defense spending, government efforts to improve the social and economic position of minorities, government's role in guaranteeing jobs and a good standard of living, women's rights, and government spending vs. services). The other two asked respondents how the defeat of the Equal Rights Amendment made them feel and their opinion about government regulation of business. For each respondent, we calculated the percent of these questions he or she was asked and answered that were answered "don't know" or "haven't thought much about this." This variable ranged from 0 to 1, with higher numbers meaning more no-opinion responding.

Nondifferentiation. Two batteries of questions asked respondents to make a series of judgments on the same rating scale, which allowed us to assess nondifferentiation. The first battery was a set of seven 101-point feeling thermometer ratings of well-known political figures and groups, such as Ted Kennedy and the Republican Party. For computing nondifferentiation, we divided the 0–100 scale into 10 segments (0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80, 81–90, 91–100). The second battery asked respondents if each of nine personality trait terms described President Ronald Reagan extremely well, quite well, not too well, or not well at all. For each battery, we counted up the maximum number of identical or quasi-identical ratings made by each respondent. These two scores were rescaled to range from 0 to 1 and were averaged to yield a single index of nondifferentiation.

Social desirability. Five items in this questionnaire seemed likely to have widely shared social desirability connotations (see app. B for details about the pretest study in which these items were identified), involving interest in politics, voting in previous elections, and support for government aid to blacks (the latter among Caucasians only). Following government and public affairs “most of the time,” being “very much interested” in political campaigns, having voted in previous elections, and supporting government aid to blacks were considered socially desirable responses. An index of socially desirable responding was created by computing the proportion of these items that a respondent answered by giving the socially desirable response (the question about government aid to blacks was included in this index for white respondents only).

Demographics. Demographic measures included education (coded 0 for respondents who completed eighth grade or less, .33 for respondents who completed between ninth and eleventh grades, .67 for respondents with a high school diploma, and 1 for respondents with more than a high school diploma), income (coded 0 for respondents with incomes less than \$5,000, .14 for incomes between \$5,000 and \$9,999, .29 for incomes between \$10,000 and \$14,999, .43 for incomes between \$15,000 and \$19,999, .57 for incomes between \$20,000 and \$24,999, .71 for incomes between \$25,000 and \$34,999, .86 for incomes between \$35,000 and \$49,999, and 1 for incomes of \$50,000 and above), race (coded 0 for Caucasians and 1 for others), gender (coded 0 for males and 1 for females), age (in years, coded to range from 0 to 1, with 0 meaning the youngest observed age in the sample and 1 meaning the oldest observed age in the sample), married (coded 1 if the respondent was married and 0 otherwise), and employment (coded 1 if the respondent was employed and 0 otherwise).

1976 Survey Research Center Datasets

No-opinion responses. In the 1976 SRC Datasets, a random subsample of respondents was told: “Not everyone has an opinion on the next question. If you do not have an opinion, just say so.” Then they were asked if they agreed or disagreed with the statement “The Arab nations are trying to work for a real peace with Israel.” Respondents who said they had no opinion were coded 1, and those who reported a substantive opinion were coded 0.

Nondifferentiation. Nondifferentiation was assessed using a battery of questions asking whether each of five possible problems that respondents might have had with their house or apartment (e.g., not enough heat, not enough living space, insects) was “a big problem, a small problem, or not a problem at all” for them. The maximum number of identical ratings made by each respondent was computed and rescaled to range from 0 to 1 to measure nondifferentiation.

Acquiescence. Acquiescence was measured using responses to three questions. All respondents were asked about the issues of free speech. A random subsample of respondents were also asked questions about peace in the Middle East and women in politics. We calculated the proportion of these agree/disagree questions that each respondent was asked to which he or she responded “agree.”

Social desirability. Based on the first social desirability pretest study described in appendix B, three items in this survey appeared to have social desirability connotations. Two questions asked about whether the respondent had voted in the 1972 U.S. presidential election and planned to vote in the 1976 U.S. presidential election. Our pretest

suggested that saying one would vote or had voted was socially desirable. The third item asked whether white people should have the right to keep black people out of their neighborhoods or whether black people have a right to live wherever they can afford to. The latter of these answers was found in our pretest to be more socially desirable among white respondents. We calculated the proportion of socially desirable answers given by respondents (the last question was used for Caucasian respondents only).

Unease. Near the ends of the interviews, respondents were asked: "Sometimes, even though a person answers a question, he/she may feel uneasy about discussing the particular subject. I'll mention several types of questions and I would like you to tell me whether or not you felt uneasy about them." Respondents indicated unease about questions on five potentially sensitive topics: their income, racial attitudes, income tax return, voting behavior, and political opinions. We calculated the proportion of topics each respondent felt uneasy discussing.

Dissatisfaction with interview length. The final question in the questionnaire asked respondents whether they felt the interview had been "much too long, too long, about right, too short, or much too short." Responses were coded to range from 0 to 1, with higher numbers indicating greater dissatisfaction with being too long. After the interviews, interviewers recorded whether the respondent had at any time asked how much longer the interview would take. Respondents who asked such a question were coded 1, and those who did not were coded 0.

Demographics. The survey included measures of race, gender, age, marital status, and employment that were coded as in the 1982 NES MCP. The survey also included measures of education (coded 0 for respondents who completed eighth grade or less, .2 for respondents who completed between ninth and eleventh grades, .4 for respondents with a high school diploma, .6 for respondents with some college, .8 for respondents with a college degree, and 1 for respondents with an advanced/graduate degree) and income. Income was measured differently for face-to-face and telephone respondents. Face-to-face respondents were given a show card listing 18 dollar ranges and were asked to indicate in which range their total 1975 family income fell. Telephone respondents were asked directly to report their total family incomes in dollars to the interviewer. We recoded these latter responses into the ranges offered to the face-to-face respondents and then coded the ranges to span from 0 (meaning the lowest income range) to 1 (meaning the highest income range).

Telephone respondents who refused to answer the initial income question were asked to place their income in one of three broad categories: less than \$7,500, between \$7,500–15,000, and more than \$15,000. Individuals who answered this follow-up question were assigned the midpoint of the range they specified (\$3,750 for the lowest, \$11,250 for the middle, and \$24,525 for the highest; this last value was the median of the amounts above \$15,000 reported by people who answered the open-ended initial income question). We then assigned these respondents scores on the income index accordingly.

2000 NES

No-opinion responses. Five questions measuring attitudes explicitly offered all respondents no-opinion response options. Four of these questions measured attitudes

toward public policies (regarding government services, government's role in guaranteeing people jobs and a good standard of living, how much the government should help blacks, and environmental protection), and one asked respondents about their political ideology. Four additional questions explicitly offered a random subset of respondents in each mode no-opinion response options. Three of these questions dealt with attitudes toward public policies (placing limits on foreign imports, protecting the environment, support for school voucher programs), and the fourth question dealt with respondents' beliefs about the roles men and women should play in today's society.²³ We calculated the percent of no-opinion responses a respondent gave to the questions he or she was asked.

Acquiescence. Acquiescence was gauged using answers to eight agree/disagree or yes/no questions asked of all respondents. Some of these questions asked about whether George W. Bush and Al Gore had ever made them feel angry, hopeful, proud, and afraid. Respondents interviewed before September 28 were asked whether Pat Buchanan elicited these emotional responses, and a random subset of respondents were asked whether President Bill Clinton elicited these emotional responses. A random subset of respondents was also asked whether they agreed or disagreed with an isolationist foreign policy, and a different subset of respondents was asked whether they thought companies who have a history of discriminating against blacks should be required to have an affirmative action program. We then calculated the percent of these questions each respondent was asked and answered "agree" or "yes."

Social desirability. The first social desirability experiment described in appendix B suggests that three questions asked in this survey had social desirability connotations: reported voter turnout in the 1996 U.S. presidential election, intentions to vote in the 2000 election, and interest in political campaigns. The NES also conducted a pilot study to test whether other typical NES questions had social desirability connotations (see the description of study 2 in app. B). This investigation identified three other questions with social desirability connotations asked in the 2000 NES: frequency of religious services attendance, watching late afternoon/early evening local television news, and watching late evening local television news. Voting in 1996, planning to vote in 2000, being interested in political campaigns, attending religious services every week, watching late afternoon/evening news every day, and watching late evening news every day were considered socially desirable responses. The proportion of socially desirable responses was calculated for each respondent.

Respondent suspicion. After completing an interview, interviewers rated how suspicious the respondent was about the interview. This variable was coded to range from 0 to 1, with higher numbers indicating greater suspicion.

Dissatisfaction with interview length. After each interview, interviewers recorded whether the respondent complained that the interview was too long or said at some point during the interview that he or she wanted to stop (each of these was coded 1 if a respondent did so and 0 if he or she did not).

Respondent engagement. Interviewers rated how cooperative the respondent was and the respondent's interest in the interview. These variables were coded to range from 0 to 1, with higher numbers indicating greater cooperation and interest, respectively.

23. Some of these questions were presented as rating scales in the face-to-face interviews and as branching questions in the telephone interviews. Previous research indicates that this format difference does not affect rates of "don't know" answering (e.g., Krosnick and Berent 1993).

Demographics. The survey included measures of gender, age, race, marital status, and employment coded as in the first two studies. Measures of education (coded into four categories: 0 for people who did not have a high school diploma, .33 for people with a high school diploma, but no further education, .67 for people with more than a high school degree, but less than a 4-year degree, and 1 for people with at least a 4-year degree) and household income (coded into seven categories: 0 for less than \$15,000, .17 for \$15,000–\$24,999, .33 for \$25,000–\$34,999, .5 for \$35,000–\$49,999, .67 for \$50,000–\$64,000, .83 for \$65,000–\$74,999, and 1 for \$75,000) were also included.

Appendix B

Social Desirability Studies

Study 1

In the 1982 NES MCP, only five items seemed to us likely to have widely shared social desirability connotations, involving interest in politics, voting in previous elections, and support for government aid to blacks (the latter among Caucasians only). Interest and participation in politics are presumably civic virtues in this culture, and the entire 1982 NES MCP interview was on the topic of politics, suggesting that the interviewer and researchers valued political interest. Previous research suggests that Caucasian respondents intentionally underreport animosity toward African Americans, presumably because reporting such feeling is not socially respectable (Pavlos 1972; Sigall and Page 1971). So these items seemed to have sufficient social desirability connotations to allow detection of mode differences in social desirability response bias.

To test our suspicion that these items did evoke social desirability concerns, we asked a sample of 112 adults to answer the same questions, interspersed with filler items. Respondents were 48 males, 63 females, and one person who did not report gender, all attending Ohio State University. Half of the sample (selected randomly) was asked to “fake bad”: give socially undesirable answers, described as those that would “create a negative reaction from society . . . the answers you would least respect or admire from another person answering this questionnaire.” The other half of the sample was asked to “fake good”: provide socially desirable answers, responses that were “most likely to create a positive reaction from society.” If these two groups of respondents gave significantly different answers to the key items, this would indicate that there was a generally agreed-upon desirable answer to each one (e.g., Wiggins 1959, 1962; see also DeMaio 1984).

As expected, significant differences appeared between the “fake good” and “fake bad” respondents on reported voter turnout, $F(1, 110) = 58.79, p < .001$, following government and public affairs, $F(1, 110) = 103.35, p < .001$, interest in political campaigns, $F(1, 110) = 39.16, p < .001$, and government aid to blacks (among white respondents only), $F(1, 84) = 22.37, p < .001$. Not surprisingly, people who “faked good” were more likely to report voting (66 percent of respondents faking good reported voting, compared to 8.9 percent of respondents faking bad), following government and public affairs closely (only 1.8 percent of “fake good” respondents said

that they follow public affairs “hardly at all,” while 73.2 percent of “fake bad” respondents gave this answer), and being interested in political campaigns (only 16.1 percent of “fake good” respondents indicated “not much interest,” while 76.8 percent of “fake bad” respondents selected that choice). Also, “fake good” respondents who were white said that they believed that the government should provide more help to blacks (only 4.7 percent of “fake good” respondents selected “government should not make any special effort to help blacks,” while 46.5 percent of “fake bad” respondents chose that option). These data were collected nearly 20 years after the 1982 NES MCP was conducted, and social desirability connotations of opinions may have shifted during the intervening years. But this evidence is at least reassuring that these items are potentially reasonable diagnostic tools.

Study 2

In order to identify other questions with social desirability connotations, the NES conducted a similar pretest. In that experiment, half of respondents were asked about one set of four questions (attending religious services, following politics, social security spending, and school integration), and the other half of respondents were asked about a different set of four questions (voting, term limits, religion provides guidance, and frequency of watching local television news). Half of the respondents who were asked about each set were asked to say how they would answer questions if they were trying to make the best impression possible on the interviewer (corresponding to the “fake good” condition reported in study 1), and half were asked to say how they would answer the same questions if they were trying to make the worst impression possible on the interviewer (corresponding to the “fake bad” condition in study 1).

Of these eight questions, four were similar to questions asked in the 2000 NES preelection interview, and we focus on those results here. Significant differences appeared between the “fake good” and “fake bad” respondents on all these items (frequency of religious services attendance: $t(211) = 9.09, p < .001$; social security spending: $t(211) = 5.62, p < .001$; reported voter turnout: $t(211) = 9.10, p < .001$; and frequency of watching local television news: $t(211) = 9.09, p < .001$). People who “faked good” were more likely to report attending religious services (42.7 percent of respondents faking good reported attending services every week, and 16.7 percent reported that they never attended services while 9.4 percent of respondents faking bad reported attending services every week, and 76.0 percent reported that they never attended religious services), voting (81.1 percent of “fake good” respondents reported they voted in the 1998 election, and 18.9 percent reported they did not, while 28.0 percent of “fake bad” respondents reported they voted in the 1998 election, and 63.4 percent respondents reported they did not), and watching local television news (58.6 percent of “fake good” respondents said they watched local television news every day, and 5.4 percent reported they never watched local television news; in contrast 27.7 percent of “fake bad” respondents said they watched local television news every day, and 63.4 percent said they never watched local television news). For social security spending, 65.6 percent of “fake good” respondents reported that spending should be increased, and 27 percent reported it should be decreased, but there was little consensus about the socially undesirable response (although 44 percent of “fake bad” respondents

said spending should be decreased, 39 percent said it should be increased). Therefore, we did not analyze the impact of mode on attitudes toward social security spending.

References

- Abelson, Robert P., Elizabeth F. Loftus, and Anthony G. Greenwald. 1992. "Attempts to Improve the Accuracy of Self-Reports of Voting." In *Questions about Questions*, ed. Judith M. Tanur, pp. 138–53. New York: Russell Sage Foundation.
- Aneshensel, Carol S., Ralph R. Frerichs, Virginia A. Clark, and Patricia A. Yokopenic. 1982. "Measuring Depression in the Community: A Comparison of Telephone and Personal Interviews." *Public Opinion Quarterly* 46:110–21.
- Aquilino, William S. 1992. "Telephone versus Face-to-Face Interviewing for Household Drug Use Surveys." *International Journal of Addictions* 27:71–91.
- . 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment." *Public Opinion Quarterly* 58:210–40.
- . 1998. "Effects of Interview Mode on Measuring Depression in Younger Adults." *Journal of Official Statistics* 14:15–29.
- Aquilino, William S., and Leonard A. LoSciuto. 1989. "Effects of Mode of Data Collection on the Validity of Reported Drug Use." In *Conference Proceedings Health Survey Research Methods*, ed. Floyd J. Fowler. Washington, DC: U.S. Department of Health and Human Services.
- . 1990. "Effects of Interview Mode on Self-Reported Drug Use." *Public Opinion Quarterly* 54:362–95.
- Barath, Arpad, and Charles F. Cannell. 1976. "Effect of Interviewer's Voice Intonation." *Public Opinion Quarterly* 40:370–73.
- Bargh, John A., and Tanya L. Chartrand. 2000. "The Mind in the Middle: A Practical Guide to Priming and Automaticity Research." In *Handbook of Research Methods in Social and Personality Psychology*, ed. Harry T. Reis and Charles M. Judd. New York: Cambridge University Press.
- Belinfante, Alexander. 1998. *Telephone Subscribership in the United States*. Washington, DC: Federal Communications Commission.
- Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63:90–108.
- Bernieri, Frank J., Janet M. Davis, Robert Rosenthal, and Robert C. Knee. 1994. "Interactional Synchrony and Rapport: Measuring Synchrony in Displays Devoid of Sound and Facial Affect." *Personality and Social Psychology Bulletin* 20:303–11.
- Biemer, Paul P. 1997. *Dual Frame NHIS/RDD Methodology and Field Test: Analysis Report*. Report prepared for the National Center for Health Statistics, Hyattsville, MD.
- . 2001. "Nonresponse Bias and Measurement Bias in a Comparison of Face-to-Face and Telephone Interviewing." *Journal of Official Statistics* 17:295–320.
- Bogardus, Emory S. 1933. "A Social Distance Scale." *Sociology and Social Research* 17:265–71.
- Burton, Scot, and Edward Blair. 1991. "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys." *Public Opinion Quarterly* 55:50–79.
- Cahalan, Don. 1960. "Measuring Newspaper Readership by Telephone: Two Comparisons with Face-to-Face Interviews." *Journal of Advertising Research* 1:1–6.
- Calsyn, Robert J., Laurie A. Roades, and Dylan S. Calsyn. 1992. "Acquiescence in Needs Assessment Studies of the Elderly." *Gerontologist* 32:246–52.
- Cannell, Charles, Robert M. Groves, and Peter V. Miller. 1981. "The Effects of Mode of Data Collection on Health Survey Data." *Proceedings of the Section on Social Statistics: American Statistical Association*, pp. 1–6. Washington, DC: American Statistical Association.
- Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. "Research on Interviewing Techniques." *Sociological Methodology* 12:389–437.
- Ceci, Stephen J. 1991. "How Much Does Schooling Influence General Intelligence and Its Cognitive Components? A Reassessment of the Evidence." *Developmental Psychology* 27:703–22.

- Chartrand, Tanya L., and John A. Bargh. 1999. "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76:893–910.
- Colombotos, John. 1965. "The Effects of Personal versus Telephone Interviews on Socially Acceptable Responses." *Public Opinion Quarterly* 29:457–58.
- . 1969. "Personal versus Telephone Interviews: Effects on Responses." *Public Health Reports* 84:773–82.
- Converse, Jean M., and Howard Schuman. 1974. *Conversations at Random*. New York: Wiley.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64:413–28.
- de Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: TT-Publikaties.
- de Leeuw, Edith D., and Joop J. Hox. 1993. "Mode Effects in Structural Modeling; A LISREL Multi-Group Comparison of Mail, Telephone, and Face-to-Face Survey Data." In *Methodische Grundlagen und Anwendungen von Strukturgleichungsmodellen*, ed. Jost Reinechke and Gaby Krekler. Mannheim: FRG e.V.
- DeMaio, Theresa J. 1984. "Social Desirability and Survey Measurement: A Review." In *Surveying Subjective Phenomena*, vol. 2, ed. Charles F. Turner and Elizabeth Martin, pp. 257–82. New York: Russell Sage Foundation.
- Dohrenwend, Barbara S., John Colombotos, and Bruce P. Dohrenwend. 1968. "Social Distance and Interviewer Effects." *Public Opinion Quarterly* 32(3):410–22.
- Drolet, Aimee L., and Michael W. Morris. 2000. "Rapport in Conflict Resolution: Accounting for How Face-to-Face Contact Fosters Mutual Cooperation in Mixed-Motive Conflicts." *Journal of Experimental Social Psychology* 36:26–50.
- Esaiaasson, Peter, and Donald Granberg. 1993. "Hidden Negativism: Evaluation of Swedish Parties and Their Leaders under Different Survey Methods." *International Journal of Public Opinion Research* 5:265–77.
- Evans, Richard I., William B. Hansen, and Maurice B. Mittlemark. 1977. "Increasing the Validity of Self-Reports of Smoking Behavior in Children." *Journal of Applied Psychology* 62:521–23.
- Fazio, Russell H., Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams. 1995. "Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?" *Journal of Personality and Social Psychology* 69:1013–27.
- Gfroerer, Joseph C., and Arthur L. Hughes. 1991. "The Feasibility of Collecting Drug Abuse Data by Telephone." *Public Health Reports* 106:384–93.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday/Anchor.
- Gordon, Randall A. 1987. "Social Desirability Bias: A Demonstration and Techniques for Its Reduction." *Teaching of Psychology* 14:40–43.
- Greenfield, Thomas K., Lorraine T. Midanik, and John D. Rogers. 2000. "Effects of Telephone versus Face-to-Face Interview Modes on Reports of Alcohol Consumption." *Addiction* 95: 277–84.
- Groves, Robert M. 1977. "An Experimental Comparison of National Telephone and Personal Interview Surveys." *Proceedings of the Section on Social Statistics: American Statistical Association*, pp. 232–41. Washington, DC: American Statistical Association.
- . 1978. "On the Mode of Administering a Questionnaire and Responses to Open-Ended Items." *Social Science Research* 7:257–71.
- . 1979. "Actors and Questions in Telephone and Personal Interview Surveys." *Public Opinion Quarterly* 43:190–205.
- Groves, Robert M., and Robert L. Kahn. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.
- Hawkins, Del I., Gerald Albaum, and Roger Best. 1974. "Stapel Scale or Semantic Differential in Marketing Research?" *Journal of Marketing Research* 11:318–22.
- Henson, Ramone, Aleda Roth, and Charles F. Cannell. 1977. "Personal versus Telephone Interviews: The Effects of Telephone Re-interviews on Reporting of Psychiatric Symptomatology." In *Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971–1977*, ed. Charles F. Cannell, Lois Oksenberg, and Jean M. Converse, pp. 205–19. Hyattsville, MD: U.S. Department of Health, Education and Welfare, Public Health Service, Health Resources Administration, National Center for Health Services Research.

- Herman, Jeanne B. 1977. "Mixed-Mode Data Collection: Telephone and Personal Interviewing." *Journal of Applied Psychology* 62:399-404.
- Herzog, A. Regula, and Willard L. Rodgers. 1988. "Interviewing Older Adults: Mode Comparison Using Data from a Face-to-Face Survey and a Telephone Re-survey." *Public Opinion Quarterly* 52:84-99.
- . 1999. "Cognitive Performance Measures in Survey Research on Older Adults." In *Cognition, Aging, and Self-Reports*, ed. Norbert Schwarz, D. C. Park, B. Knäuper, and Seymour Sudman. Philadelphia: Psychology Press.
- Herzog, A. Regula, Willard L. Rodgers, and Richard A. Kulka. 1983. "Interviewing Older Adults: A Comparison of Telephone and Face-to-Face Modalities." *Public Opinion Quarterly* 47:405-18.
- Himmelfarb, Samuel, and Carl Lickteig. 1982. "Social Desirability and the Randomized Response Technique." *Journal of Personality and Social Psychology* 43:710-17.
- Hinkle, Andrew L., and Gary D. King. 1978. "A Comparison of Three Survey Methods to Obtain Data for Community Mental Health Program Planning." *American Journal of Community Psychology* 6:389-97.
- Hochstim, Joseph R. 1962. "Comparison of Three Information-Gathering Strategies in a Population Study of Sociomedical Variables." In *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 154-59. Washington, DC: American Statistical Association.
- . 1967. "A Critical Comparison of Three Strategies of Collecting Data from Households." *Journal of the American Statistical Association* 62:976-89.
- Hox, Joop J., and Edith D. de Leeuw. 1994. "A Comparison of Non-response in Mail, Telephone, and Face-to-Face Surveys." *Quality and Quantity* 28:329-44.
- Johnson, Timothy P., James G. Houglan, Jr., and Richard R. Clayton. 1989. "Obtaining Reports of Sensitive Behavior: A Comparison of Substance Use Reports from Telephone and Face-to-Face Interviews." *Social Science Quarterly* 70:174-83.
- Jordan, Lawrence A., Alfred C. Marcus, and Leo G. Reeder. 1980. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly* 44:210-22.
- Kish, Leslie. 1949. "A Procedure for Objective Respondent Selection within Households." *Journal of the American Statistical Association* 44:380-87.
- Klecka, William R., and Alfred J. Tuchfarber. 1978. "Random Digit Dialing: A Comparison to Personal Surveys." *Public Opinion Quarterly* 42:105-14.
- Kormendi, Eszter. 1988. "The Quality of Income Information in Telephone and Face-to-Face Surveys." In *Telephone Survey Methodology*, ed. Robert M. Groves, Paul Biemer, Lars Lyberg, James Massey, William L. Nicholls, and Joseph Waksberg. New York: Wiley.
- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-36.
- . 1999. "Survey Research." *Annual Review of Psychology* 50:537-67.
- Krosnick, Jon A., and Matthew K. Berent. 1993. "Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format." *American Journal of Political Science* 37:941-64.
- Krosnick, Jon A., and Michael A. Milburn. 1990. "Psychological Determinants of Political Opinionation." *Social Cognition* 8:49-72.
- Krosnick, Jon A., Sowmya Narayan, and Wendy R. Smith. 1996. "Satisficing in Surveys: Initial Evidence." *New Directions for Evaluation* 70:29-44.
- Larsen, Otto N. 1952. "The Comparative Validity of Telephone and Face-to-Face Interviews in the Measurement of Message Diffusion by Leaflets." *American Sociological Review* 17:471-76.
- Locander, William, Seymour Sudman, and Norman Bradburn. 1976. "An Investigation of Interview Method, Threat, and Response Distortion." *Journal of the American Statistical Association* 71:269-75.
- Mangione, Thomas W., Ralph Hingson, and Jane Barrett. 1982. "Collecting Sensitive Data: A Comparison of Three Survey Strategies." *Sociological Methods and Research* 10:337-46.
- Mark, Gloria, Jonathan Grudin, and Steven E. Poltrock. 1999. "Meeting at the Desktop: An Empirical Study of Virtually Collocated Teams." In *ECSCW '99: Proceedings of the Sixth European Conference on Computer Supported Cooperative Work, 12-16 September 1999, Copenhagen, Denmark*, ed. Susan Bødker, Morten Kyng, and Kjeld Schmidt. Boston: Kluwer.
- McQueen, David V. 1989. "Comparison of Results of Personal Interview and Telephone Surveys of Behavior Related to Risk of AIDS: Advantages of Telephone Techniques." In *Conference*

- Proceedings Health Survey Research Methods*, ed. Floyd J. Fowler. Washington, DC: U.S. Department of Health and Human Services.
- Midanik, Lorraine T., John D. Rogers, and Thomas K. Greenfield. 2001. "Mode Differences in Reports of Alcohol Consumption and Alcohol-Related Harm." In *Health Survey Research Methods*, ed. M. K. Cynamon and R. A. Kulka. Hyattsville, MD: U.S. Department of Health and Human Services.
- Miller, Peter V. 1984. "Alternative Question Forms for Attitude Scale Questions in Telephone Interviews." *Public Opinion Quarterly* 48:766–78.
- Miller, Peter V., and Charles F. Cannell. 1982. "A Study of Experimental Techniques for Telephone Interviewing." *Public Opinion Quarterly* 46:250–69.
- Morchio, Giovanna, Maria Sanchez, and Santa Traugott. 1985. "Mode Differences: DK Responses in the 1984 Post-Election Study." Unpublished memo submitted to the Board of Overseers of the National Election Study, Ann Arbor, MI.
- Morley, Ian E., and Geoffrey M. Stephenson. 1977. *The Social Psychology of Bargaining*. London: Allen & Unwin.
- Mulry-Liggan, Mary H. 1983. "A Comparison of a Random Digit Dialing Survey and the Current Population Survey." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 214–19. Washington, DC: American Statistical Association.
- Narayan, Sowmya, and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60:58–88.
- Nie, Norman H., Jane Junn, and Kenneth Stehlik-Barry. 1996. *Education and Democratic Citizenship in America*. Chicago: University of Chicago Press.
- Oakes, R. H. 1954. "Differences in Responsiveness in Telephone versus Personal Interviews." *Journal of Marketing* 19:169.
- Oksenberg, Lois, Lerita Coleman, and Charles Cannell. 1986. "Interviewers' Voices and Refusal Rates in Telephone Surveys." *Public Opinion Quarterly* 50:97–111.
- Paulhus, Delroy L. 1984. "Two-Component Models of Socially Desirable Responding." *Journal of Personality and Social Psychology* 46:598–609.
- Pavlos, Andrew J. 1972. "Racial Attitude and Stereotype Change with Bogus Pipeline Paradigm." *Proceedings of the 80th Annual Convention of the American Psychological Association* 7: 291–92.
- Poole, Marshall S., Dale L. Shannon, and Gerardine DeSanctis. 1992. "Communication Media and Negotiation Processes." In *Communication and Negotiation: Sage Annual Reviews of Communication Research*, ed. Linda L. Putnam and Michael E. Roloff, pp. 46–66. Thousand Oaks, CA: Sage.
- Quinn, Robert P., Barbara A. Gutek, and Jeffrey T. Walsh. 1980. "Telephone Interviewing: A Reappraisal and a Field Experiment." *Basic and Applied Social Psychology* 1:127–53.
- Rogers, Theresa F. 1976. "Interviews by Telephone and in Person: Quality of Responses and Field Performance." *Public Opinion Quarterly* 40:51–65.
- Rosenstone, Steven J., Margaret Petrella, and Donald Kinder. 1993. "Excessive Reliance on Telephone Interviews and Short-Form Questionnaires in the 1992 National Election Study: Assessing the Consequences for Data Quality." Unpublished memo submitted to the Board of Overseers of the National Election Study, Ann Arbor, MI.
- Rosenthal, Robert. 1984. *Meta-analytic Procedures for Social Research*. Beverly Hills, CA: Sage.
- Saris, Willem E., and Max Kaase. 1997. *Eurobarometer Measurement Instruments for Opinions in Europe*. Mannheim: ZUMA.
- Schlenker, Barry R., and Michael F. Weingold. 1989. "Goals and the Self-Identification Process: Constructing Desires and Identities." In *Goal Concepts in Personality and Social Psychology*, ed. Lawrence A. Pervin, pp. 243–90. Hillsdale, NJ: Erlbaum.
- Schmiedeskamp, James W. 1962. "Reinterviews by Telephone." *Journal of Marketing* 26:28–34.
- Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- Schwarz, Norbert. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Erlbaum.
- Schwarz, Norbert, Barbel Knauper, Hans-J. Hippler, Elisabeth Noelle-Neumann, and Leslie Clark. 1991. "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55:570–82.
- Shanks, J. Merrill, Maria Sanchez, and Betsy Morton. 1983. "Alternative Approaches to Survey

- Data Collection for the National Election Studies: A Report on the 1982 NES Method Comparison Project." Unpublished memo submitted to the Board of Overseers of the National Election Study, Ann Arbor, MI.
- Shuy, Roger W. 2002. "In-Person versus Telephone Interviewing." In *Handbook of Interview Research Context and Method*, ed. Jaber F. Gubrium and James A. Holstein. Thousand Oaks, CA: Sage.
- Siegel, Jane, Vitaly Dubrovsky, Sara Kiesler, and Timothy W. McGuire. 1986. "Group Processes in Communication." *Organizational Behavior and Human Decision Processes* 37:157–87.
- Siemiatycki, Jack. 1979. "A Comparison of Mail, Telephone, and Home Interview Strategies for Household Health Surveys." *American Journal of Public Health* 69:238–45.
- Sigall, Harold, and Richard Page. 1971. "Current Stereotypes: A Little Fading, a Little Faking." *International Journal of Psychology* 25:1–12.
- Smith, Eliot R., Nyla Branscombe, and Carol Bormann. 1988. "Generality of the Effects of Practice on Social Judgment Tasks." *Journal of Personality and Social Psychology* 54:385–95.
- Smith, Thomas W. 1995. "Trends in Nonresponse Rate." *International Journal of Public Opinion Research* 7:157–71.
- Sudman, Seymour, and Norman M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Sykes, Wendy M., and Martin Collins. 1987. "Comparing Telephone and Face-to-Face Interviewing in the United Kingdom." *Survey Methodology* 13:15–28.
- . 1988. "Effects of Mode of Interview: Experiments in the UK." In *Telephone Survey Methodology*, ed. Robert M. Groves, Paul Biemer, Lars Lyberg, James Massey, William L. Nicholls, and Joseph Waksberg. New York: Wiley.
- Thornberry, Owen T. 1987. *An Experimental Comparison of Telephone and Personal Health Interview Surveys*. Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, National Center for Health Statistics.
- Tickle-Degnen, Linda, and Robert Rosenthal. 1990. "The Nature of Rapport and Its Nonverbal Correlates." *Psychological Inquiry* 1:285–93.
- Tourangeau, Roger. 1984. "Cognitive Sciences and Survey Methods." In *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, ed. Thomas B. Jabine, Miron Straf, Judith Tanur, and Roger Tourangeau, pp. 73–100. Washington, DC: National Academy Press.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
- Turner, Charles F., Barbara H. Forsyth, James M. O'Reilly, Philip C. Cooley, T. K. Smith, S. M. Rogers, and H. G. Miller. 1998. "Automated Self-Interviewing and the Survey Measurement of Sensitive Behaviors." In *Computer Assisted Survey Information Collection*, ed. Mick P. Couper, Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholis II, and James M. O'Reilly, pp. 455–73. New York: Wiley.
- Turoff, Murray, and Starr Roxanne Hiltz. 1982. "Computer Support for Group versus Individual Decisions." *IEEE Transactions on Communication* 30:82–90.
- Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63–69.
- Weeks, Michael F., Richard A. Kulka, Judith T. Lessler, and Roy W. Whitmore. 1983. "Personal versus Telephone Surveys for Collecting Household Health Data at the Local Level." *American Journal of Public Health* 73:1389–94.
- Wiggins, Jerry S. 1959. "Interrelationships among MMPI Measures of Dissimulation under Standard and Social Desirability Instructions." *Journal of Consulting Psychology* 23:419–27.
- . 1962. "Definitions of Social Desirability and Acquiescence in Personality Inventories." In *Measurement in Personality and Cognition*, ed. Samuel Messick and John Ross. New York: Wiley.
- Williams, Ederyn. 1977. "Experimental Comparisons of Face-to-Face and Mediated Communication: A Review." *Psychological Bulletin* 84:963–76.
- Wilson, Timothy D., and Jonathan W. Schooler. 1991. "Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions." *Journal of Personality and Social Psychology* 60:181–92.
- Wiseman, Frederick. 1972. "Methodological Bias in Public Opinion Surveys." *Public Opinion Quarterly* 36:105–8.

- Wolfe, Lee M. 1979. "Characteristics of Persons With and Without Home Telephones." *Journal of Marketing Research* 16:421-25.
- Woltman, Henry F., Anthony G. Turner, and John M. Bushery. 1980. "A Comparison of Three Mixed-Mode Interviewing Procedures in the National Crime Survey." *Journal of the American Statistical Association* 75:534-43.
- Yaffe, Richard, Sam Shapiro, Robert R. Fuchsberg, Charles A. Rohde, and Helen C. Corpeño. 1978. "Medical Economics Survey-Methods Study: Cost Effectiveness of Alternative Survey Strategies." *Medical Care* 16:641-59.