

Self-Reports

How the Questions Shape the Answers

Norbert Schwarz
University of Michigan

Self-reports of behaviors and attitudes are strongly influenced by features of the research instrument, including question wording, format, and context. Recent research has addressed the underlying cognitive and communicative processes, which are systematic and increasingly well-understood. I review what has been learned, focusing on issues of question comprehension, behavioral frequency reports, and the emergence of context effects in attitude measurement. The accumulating knowledge about the processes underlying self-reports promises to improve questionnaire design and data quality.

Self-reports are a primary source of data in psychology and the social sciences. From laboratory experiments to public opinion surveys, researchers rely on the answers that research participants provide to learn about individuals' thoughts, feelings, and behaviors and to monitor societal trends, from the nation's unemployment rate to the development of crime. Unfortunately, self-reports are a fallible source of data, and minor changes in question wording, question format, or question context can result in major changes in the obtained results, as a few examples may illustrate:

- When asked what they consider "the most important thing for children to prepare them for life," 61.5% of a representative sample chose the alternative "To think for themselves" when this alternative was offered on a list. Yet, only 4.6% volunteered an answer that could be assigned to this category when no list was presented (Schuman & Presser, 1981).
- When asked how successful they have been in life, 34% of a representative sample reported high success when the numeric values of the rating scale ranged from -5 to 5, whereas only 13% did so when the numeric values ranged from 0 to 10 (Schwarz, Knäuper, Hippler, Noelle-Neumann, & Clark, 1991).
- When asked how often they experience a variety of physical symptoms, 62% of a sample of psychosomatic patients reported symptom frequencies of more than twice a month when the response scale ranged from "twice a month or less" to "several times a day." Yet, only 39% reported frequencies of more than twice a month when the scale ranged from "never" to "more than twice a month" (Schwarz & Scheuring, 1992).
- Whether we conclude that marital satisfaction is a major or a minor contributor to general life-satisfaction depends on the

order in which both questions are asked, with correlations ranging from .18 to .67 as a function of question order and introduction (Schwarz, Strack, & Mai, 1991).

Although findings of this type often come as unpleasant surprises, the underlying cognitive and communicative processes are systematic and increasingly well-understood. Since the early 1980s, psychologists and survey methodologists developed an interdisciplinary field of research that is devoted to understanding the nature of self-reports and to improving the quality of data collection. Research in this field has addressed a wide range of topics: How do respondents make sense of the questions asked of them? What is the role of autobiographical memory in retrospective reports of behaviors and how can we increase the accuracy of these reports? What are the judgmental processes underlying the emergence of context effects in attitude measurement? Do the processes underlying self-reports of behaviors and attitudes change across the adult life span? Which techniques can we use to determine if a question "works" as intended?

Reflecting the need to bring together researchers from diverse disciplines, interdisciplinary conferences played a pivotal role in the development of this area, resulting in a considerable number of edited volumes (Hippler, Schwarz, & Sudman, 1987; Jabine, Straf, Tanur, & Tourangeau, 1984; Jobe & Loftus, 1991; Schwarz, Park, Knäuper, & Sudman, 1998; Schwarz & Sudman, 1992, 1994, 1996; Sirken et al., 1998; Tanur, 1992). In addition, a recent monograph (Sudman, Bradburn, & Schwarz, 1996) reviewed what has been learned from this research and discussed its implications for questionnaire design.

Although this area of research is typically referred to as "cognitive aspects of survey methodology," the cognitive and communicative processes investigated apply to the question-answering process in all standardized research situations. In this article, I review selected aspects of this work, focusing on basic psychological issues that apply to

Editor's note. Denise C. Park served as action editor for this article.

Author's note. Correspondence concerning this article should be addressed to Norbert Schwarz, Institute for Social Research, University of Michigan, Ann Arbor, MI 48106-1248. Electronic mail may be sent to nschwarz@umich.edu.



Norbert Schwarz

questionnaire studies as well as laboratory experiments. The first section addresses how respondents make sense of the questions asked of them and highlights the role of conversational inference processes in question comprehension. The second section addresses how respondents answer behavioral questions and relates these questions to issues of autobiographical memory and estimation strategies. Finally, the third section addresses attitude questions and reviews the conditions that give rise to context effects in attitude measurement. Throughout, I focus on how features of the research instrument shape the answers that respondents provide and influence the conclusions that we, as researchers, would draw from the obtained reports.

Making Sense of the Question Asked

Not surprisingly, the first task that respondents face is to understand the question asked of them (Strack & Martin, 1987; Tourangeau, 1984). The key issue is whether the respondent's understanding of the question matches what the researcher had in mind: Is the attitude object, or the behavior, that the respondent identifies as the referent of the question the one that the researcher intended? Does the respondent's understanding tap the same facet of the issue and the same evaluative dimension? From a psychological point of view, question comprehension reflects the operation of two intertwined processes (Clark & Clark, 1977; Clark & Schober, 1992).

The first refers to the semantic understanding of the utterance. Comprehending the *literal meaning* of a sentence involves the identification of words, the recall of lexical information from semantic memory, and the construction of a meaning of the utterance, which is constrained by its context (Anderson, 1980). Not surprisingly, textbooks urge researchers to write simple questions and to

avoid unfamiliar or ambiguous terms. However, understanding the words is not sufficient to answer a question. For example, if respondents are asked, "What have you done today?" they are likely to understand the meaning of the words. Yet, they still need to determine what kind of activities the researcher is interested in. Should they report, for example, that they took a shower, or not? Hence, understanding a question in a way that allows an appropriate answer requires not only an understanding of the literal meaning of the question but also involves inferences about the questioner's intention to determine the *pragmatic meaning* of the question.

To infer the pragmatic meaning of a question, respondents rely on the tacit assumptions that govern the conduct of conversation in everyday life. These tacit assumptions were described by Paul Grice (1975), a philosopher of language (see Levinson, 1983, for a detailed introduction). According to Grice's analysis, conversations proceed according to a cooperativeness principle, which can be expressed in the form of four maxims.

First, a *maxim of relation* enjoins speakers to make their contribution relevant to the aims of the ongoing conversation. In research situations, this maxim licenses the use of contextual information in question interpretation and invites respondents to relate the question to the context of the ongoing exchange.

Second, a *maxim of quantity* enjoins speakers to make their contribution as informative as is required, but not more informative than is required. This maxim invites respondents to provide information the questioner seems interested in, rather than other information that may come to mind. Moreover, it discourages the reiteration of information that has already been provided earlier, or that "goes without saying" (such as "taking a shower" in the above example).

Third, a *maxim of manner* holds that the contribution should be clear rather than obscure, ambiguous, or wordy. In research situations, this maxim entails an "interpretability presumption": research participants can assume that the researcher "chose his wording so they can understand what he meant—and can do so quickly" (Clark & Schober, 1992, p. 27). Hence, the most obvious meaning seems likely to be the correct one, and if there is no obvious meaning, respondents may consult the immediate context to determine one. As numerous studies have shown, the researcher's contributions include formal aspects of questionnaire design, such as the response alternatives provided as part of the question, and respondents draw on these features in interpreting the question.

Finally, a *maxim of quality* enjoins speakers not to say anything they believe to be false or lack adequate evidence for. This maxim is often violated in psychological experiments, for example, when the researcher deliberately presents misleading or uninformative material in a context that suggests otherwise. This topic, however, is beyond the scope of the present article (see Bless, Strack, & Schwarz, 1993; Hilton, 1995; Schwarz, 1994, 1996, for reviews).

In summary, speakers should try to be informative, truthful, relevant, and clear, and listeners interpret the

speakers' utterances "on the assumption that they are trying to live up to these ideals" (Clark & Clark, 1977, p. 122). These rules of cooperative conversational conduct are essential for understanding how respondents make sense of the questions asked of them, as the following examples illustrate (see Clark & Schober, 1992; Schober, in press; Schwarz, 1994, 1996; Strack, 1994a, 1994b, for more extended reviews).

Response Alternatives

Open versus closed response formats.

Suppose that respondents are asked in an open-response format, "What have you done today?" To give a meaningful answer, respondents have to determine which activities may be of interest to the researcher. In an attempt to be informative, respondents are likely to omit activities that the researcher is obviously aware of (e.g., "I gave a survey interview") or may take for granted anyway (e.g., "I took a shower"), thus observing the maxim of quantity. If respondents were given a list of activities that included giving an interview and taking a shower, most respondents would endorse them. At the same time, however, such a list would reduce the likelihood that respondents would report activities that are not represented on the list (see Schuman & Presser, 1981; Schwarz & Hippler, 1991, for a review of relevant studies). Both of these question-form effects reflect that response alternatives can clarify the intended meaning of a question, in the present example by specifying the activities the researcher is interested in. In addition, response alternatives may remind respondents of material that they may otherwise not consider.

In combination, these processes can result in pronounced and systematic differences between open- and closed-question formats, as a study on parental values illustrated. When asked what they consider "the most important thing for children to prepare them for life," 61.5% of the respondents picked "To think for themselves" when this alternative was offered as part of a list. Yet, only 4.6% provided an answer that could be assigned to this category in an open-response format (Schuman & Presser, 1981, pp. 105–107). Obviously, the researchers would draw very different conclusions about parental values depending on the question format used.

Frequency scales and reference periods.

Suppose that respondents are asked how frequently they felt "really irritated" recently. To provide an informative answer, respondents have to determine what the researcher means with "really irritated." Does this term refer to major or to minor annoyances? To identify the intended meaning of the question, they may consult the response alternatives provided by the researcher. If the response alternatives present low-frequency categories, for example, ranging from "less than once a year" to "more than once a month," respondents may conclude that the researcher has relatively rare events in mind. Hence, the question cannot refer to minor irritations that are likely to occur more often, so the researcher is probably interested in more severe episodes of irritation. In line with this assumption, Schwarz, Strack, Müller, and Chassein (1988; see also Gaskell,

O'Muircheartaigh, & Wright, 1994) observed that respondents who had to report the frequency of irritating experiences on a low-frequency scale assumed that the question referred to major annoyances, whereas respondents who had to give their report on a high-frequency scale assumed that the question referred to minor annoyances. Thus, respondents identified different experiences as the target of the question, depending on the frequency range of the response alternatives provided to them.

Similarly, Winkielman, Knäuper, and Schwarz (1998) observed that the length of the reference period can profoundly affect question interpretation. In their studies, respondents were asked how frequently they had been angry either "last week" or "last year." Again, they inferred that the researcher is interested in more frequent and less severe episodes of anger when the question pertained to one week rather than one year, and their examples reflected this differential question interpretation.

These findings have important implications for the comparison of concurrent and retrospective reports of behaviors and emotions. Empirically, individuals report more intense emotions (e.g., Parkinson, Briner, Reynolds, & Totterdell, 1995; Thomas & Diener, 1990), and more severe marital disagreements (e.g., McGonagle, Kessler, & Schilling, 1992), in retrospective than in concurrent reports. Whereas findings of this type are typically attributed to the higher memorability of intense experiences, Winkielman et al.'s (1998) results suggest that discrepancies between concurrent and retrospective reports may in part be due to differential question interpretation: Concurrent reports necessarily pertain to a short reference period, with one day typically being the upper limit, whereas retrospective reports cover more extended periods. Hence, the concurrent and retrospective nature of the report is inherently confounded with the length of the reference period. Accordingly, participants who provide a concurrent report may infer from the short reference period used that the researcher is interested in frequent events, whereas the long reference period used under retrospective conditions may suggest an interest in infrequent events. Hence, respondents may deliberately report on different experiences, rendering their reports incomparable.

Rating scales. Similar considerations apply to psychologists' favorite question format, the rating scale. Suppose respondents are asked, "How successful would you say you have been in life?" accompanied by a rating scale that ranges from "not at all successful" to "extremely successful." To answer this question, respondents have to determine what the researcher means by "not at all successful": Does this term refer to the absence of outstanding achievements or to the presence of explicit failures? To do so, they may draw on what is supposedly a purely formal feature of the rating scale, namely its numeric values. Specifically, Schwarz, Knäuper, et al. (1991) presented the success-in-life question with an 11-point rating scale that ranged either from 0 (*not at all successful*) to 10 (*extremely successful*), or from -5 (*not at all successful*) to 5 (*extremely successful*). The results showed a dramatic impact

of the numeric values presented to respondents. Whereas 34% of the respondents endorsed a value between -5 and 0 on the -5 to 5 scale, only 13% endorsed one of the formally equivalent values between 0 and 5 on the $0-10$ scale.

Subsequent experiments indicated that this difference reflects differential interpretations of the term "not at all successful." When this label was combined with the numeric value "0," respondents interpreted it to reflect the absence of outstanding achievements. However, when the same label was combined with the numeric value " -5 ," and the scale offered "0" as the midpoint, they interpreted it to reflect the presence of explicit failures (see also Schwarz, Grayson, & Knäuper, 1998; Schwarz & Hippler, 1995a). In general, a format that ranges from negative to positive numbers conveys that the researcher has a bipolar dimension in mind, where the two poles refer to the presence of opposite attributes. In contrast, a format that uses only positive numbers conveys that the researcher has a unipolar dimension in mind, referring to different degrees of the same attribute.

These and related findings (see Schwarz, 1996, chapter 5, for a review) highlight that respondents draw on apparently formal features of the research instrument to disambiguate the meaning of the questions posed to them. Unless researchers learn to take the informational value of these features into account, we may often be surprised by the answers we obtain.

Question Context

Researcher's epistemic interest. An important but frequently overlooked context variable that may influence respondents' question interpretation is clues provided by the researcher's affiliation. Recall that the norms of conversational conduct (Grice, 1975) require speakers to provide information that the questioner is interested in, which entails inferences about the questioner's likely epistemic interest. A relevant source of information in this regard is the researcher's affiliation. For example, Norenzayan and Schwarz (in press) presented respondents with newspaper accounts of mass murders and asked them to explain why the mass murder occurred. In one condition, the questionnaire was printed on the letterhead of an alleged "Institute for Personality Research," whereas in the other condition it was printed on the letterhead of an "Institute for Social Research." As expected, respondents' explanations showed more attention to personality variables or to social-contextual variables, depending on whether they thought the researcher was a personality psychologist or a social scientist. Apparently, they took the researcher's affiliation into account in determining the kind of information that would be most informative, given the researcher's likely epistemic interest.

Adjacent question. Respondents' interpretation of a question's intended meaning is further influenced by the content of adjacent questions. As an extreme case, consider research in which respondents are asked to report their opinion about a highly obscure, or even completely fictitious, issue, such as the "Agricultural Trade Act of

1978" (e.g., Bishop, Oldendick, & Tuchfarber, 1986; Schuman & Presser, 1981). Public opinion researchers introduced such questions to explore the extent to which respondents are willing to report an opinion in the absence of any knowledge about the topic. In fact, about 30% of any representative sample do offer an opinion on fictitious issues. Yet, their answers may be more meaningful than has typically been assumed.

From a conversational point of view, the sheer fact that a question about some issue is asked presupposes that this issue exists—or else asking a question about it would violate every norm of conversational conduct. But respondents have no reason to assume that the researcher would ask a meaningless question and will hence try to make sense of it. To do so, they are likely to turn to the context of the ambiguous question, much as they would be expected to do in any other conversation. Once they have assigned a particular meaning to the issue, thus transforming the fictitious issue into a subjectively better defined one that makes sense in the context of the questionnaire, they may have no difficulty reporting a subjectively meaningful opinion. Even if they have not given the particular issue much thought, they may identify the broader set of issues to which this particular one apparently belongs, allowing them to derive a meaningful answer.

Supporting this assumption, Strack, Schwarz, and Wänke (1991, Experiment 1) observed that German university students reported different attitudes toward the introduction of a fictitious "educational contribution," depending on the nature of a preceding question. Specifically, some students were asked to estimate the average tuition fees that students have to pay at U.S. universities (in contrast to Germany, where university education is free), whereas others had to estimate the amount of money that the Swedish government pays every student as financial support. As expected, respondents inferred that the fictitious "educational contribution" pertained to students having to pay money when it followed the tuition question, but to students receiving money when it followed the financial support question. Reflecting this differential interpretation, they reported a more favorable attitude toward the introduction of an "educational contribution" in the former than in the latter case—hardly a meaningless response.

Summary

As the preceding examples illustrate, question comprehension is not solely an issue of understanding the literal meaning of an utterance. Rather, question comprehension involves extensive inferences about the speaker's intentions to determine the pragmatic meaning of the question. To make these inferences, respondents draw on the nature of preceding questions as well as the response alternatives. Accordingly, researchers' traditional focus on using the "right words" in questionnaire writing needs to be complemented by a consideration of the conversational processes involved in the question-answering process (see Schwarz, 1996, for a comprehensive review).

To safeguard against unintended question interpretations and related complications, psychologists and survey

methodologists have developed a number of procedures that can be used in questionnaire pretesting (see the contributions in Schwarz & Sudman, 1996, for comprehensive reviews). These procedures include the extensive use of probes and think-aloud protocols (summarily referred to as "cognitive interviewing"; e.g., DeMaio & Rothgeb, 1996), detailed codings of interview transcripts (e.g., Bolton & Bronkhorst, 1996; Fowler & Cannell, 1996), and the use of expert systems that alert researchers to likely problems (see Lessler & Forsyth, 1996). In the domain of survey research, these procedures are routinely applied at the questionnaire pretesting stage, and most major survey centers at government agencies, as well as some centers in the academic and private domains, have established "cognitive laboratories" for this purpose. At present, these techniques are less frequently used in psychological research, where questionnaire development is often of a more ad hoc nature.

Once respondents have determined the intended meaning of the question, they face additional tasks (Strack & Martin, 1987; Tourangeau, 1984). These tasks include the recall of relevant information from memory, the computation of a judgment, and the formatting of these judgments in line with the response alternatives provided by the researcher. Moreover, respondents may want to edit their private judgment before they report it to the researcher, due to reasons of social desirability and self-presentation. However, a caveat needs to be added. Although it is conceptually useful to present respondents' tasks as following the above sequence of question comprehension, recall, judgment, and overt report (Strack & Martin, 1987; Tourangeau, 1984), deviations from this sequence are obviously possible. For example, respondents may revise their initial interpretation of a question when their answer does not fit the response alternatives. Next I address these tasks in the context of behavioral questions and attitude questions.

Reporting on One's Behaviors

Many questions require respondents to report on the frequency with which they engaged in a specific behavior during a specified reference period. Ideally, the researcher wants respondents to identify the intended behavior, to search memory for relevant episodes, to date these episodes

with regard to the reference period, and to count them up to arrive at a numeric answer. Unfortunately, this is the course of action that respondents are least likely to follow. In fact, unless the behavior is rare and of considerable importance, respondents are unlikely to have detailed episodic representations available in memory. Instead, the individual instances of frequent behaviors blend into generic, knowledge-like representations that lack the time and space markers that allow for episodic recall (see Strube, 1987). Accordingly, a "recall-and-count" model does not capture how people answer questions about frequent behaviors or experiences. Rather, their answers are likely to be based on some fragmented recall and the application of inference rules to compute a frequency estimate (see Bradburn, Rips, & Shevell, 1987; Sudman et al., 1996, for reviews).

Drawing on basic research into the structure of autobiographical memory, researchers have developed a number of strategies that are designed to facilitate autobiographical recall (for reviews see Schwarz, 1990; Sudman et al., 1996; and the contributions in Schwarz & Sudman, 1994). These strategies are beyond the scope of the present article, which focuses on contextual influences of questionnaire design. Specifically, I address how the frequency alternatives presented as part of a closed-question format influence respondents' frequency estimates and subsequent related judgments.

Frequency Alternatives

Behavioral reports. In many studies, respondents are asked to report their behavior by checking the appropriate alternative from a list of response alternatives of the type shown in Table 1. In this example, taken from Schwarz, Hippler, Deutsch, and Strack (1985), German respondents were asked how many hours they watch television on a typical day. To provide their answer, they were presented with a frequency scale that offered either high- or low-frequency response alternatives. Although the selected response alternative is assumed to inform the researcher about the respondent's behavior, it is frequently overlooked that a given set of response alternatives may also constitute a source of information for the respondent, as already seen in the section on question comprehension.

Essentially, respondents assume that the researcher

Table 1
Reported Daily TV Consumption as a Function of Response Alternatives

Low-frequency alternatives	Daily consumption	High-frequency alternatives	Daily consumption
Up to 1/2 hour	7.4%	Up to 2 1/2 hours	62.5%
1/2 hour to 1 hour	17.7%	2 1/2 hours to 3 hours	23.4%
1 hour to 1 1/2 hours	26.5%	3 hours to 3 1/2 hours	7.8%
1 1/2 hours to 2 hours	14.7%	3 1/2 hours to 4 hours	4.7%
2 hours to 2 1/2 hours	17.7%	4 hours to 4 1/2 hours	1.6%
More than 2 1/2 hours	16.2%	More than 4 1/2 hours	0.0%

Note. $N = 132$. From "Response Categories: Effects on Behavioral Reports and Comparative Judgments," by N. Schwarz, H. J. Hippler, B. Deutsch, & F. Strack, 1985, *Public Opinion Quarterly*, 49, p. 391. Copyright 1985 by The University of Chicago Press. Adapted with permission.

constructs a meaningful scale, based on his or her knowledge of, or expectations about, the distribution of the behavior in the "real world." Accordingly, respondents assume that the values in the middle range of the scale reflect the "average" or "usual" behavioral frequency, whereas the extremes of the scale correspond to the extremes of the distribution. Given this assumption, respondents can use the range of the response alternatives as a frame of reference in estimating their own behavioral frequency.

This strategy results in higher estimates along scales that present high- rather than low-frequency response alternatives, as shown in Table 1. In this study (Schwarz et al., 1985), only 16.2% of a sample of German respondents reported watching TV for more than two and a half hours a day when the scale presented low-frequency response alternatives, whereas 37.5% reported doing so when the scale presented high-frequency response alternatives. Similar results have been obtained for a wide range of different behaviors, including sexual behaviors (e.g., Schwarz & Scheuring, 1988; Tourangeau & Smith, 1996), consumer behaviors (e.g., Menon, Raghubir, & Schwarz, 1995), and reports of physical symptoms (e.g., Schwarz & Scheuring, 1992; see Schwarz, 1990, 1996, for reviews).

For example, Schwarz and Scheuring (1992) asked 60 patients of a German psychosomatic clinic to report the frequency of 17 symptoms along a low-frequency scale that ranged from "never" to "more than twice a month," or along a high-frequency scale that ranged from "twice a month or less" to "several times a day." Across 17 symptoms, 62% of the respondents reported average frequencies of more than twice a month when presented with the high-frequency scale, whereas only 39% did so when presented with the low-frequency scale, resulting in a mean difference of 22 percentage points. This impact of response alternatives was most pronounced for the ill-defined symptom of "responsiveness to changes in the weather," to which 75% of the patients reported a frequency of more than twice a month along the high-frequency scale, whereas only 21% did so along the low-frequency scale. Conversely, the influence of response alternatives was least pronounced for the better defined symptom "excessive perspiration," with 50% versus 42% of the respondents reporting a frequency of more than twice a month in the high- and low-frequency scale conditions, respectively.

As expected on theoretical grounds, the impact of response alternatives is more pronounced the more poorly the behavior is represented in memory, thus forcing respondents to rely on an estimation strategy. When the behavior is rare and important, and hence well-represented in memory, or when the respondent engages in the behavior with high regularity (e.g., "every Sunday"), the impact of response alternatives is small because no estimation is required (see Menon, 1994; Menon et al., 1995, for a discussion). Moreover, the influence of response alternatives is particularly pronounced when the behavior is ill-defined (as seen above), in which case the response alternatives influence respondents' interpretation of the question (Schwarz et al., 1988) as well as the estimation strategy used. I return to the methodological implications of these findings below.

Subsequent judgments. In addition to affecting respondents' behavioral reports, response alternatives may also affect subsequent judgments. For example, a frequency of "2½ hours a day" constitutes a high response on the low-frequency scale, but a low response on the high-frequency scale shown in Table 1. A respondent who checks this alternative may therefore infer that her own TV consumption is above average in the former case, but below average in the latter. As a result, Schwarz et al. (1985) observed that respondents were less satisfied with the variety of things they do in their leisure time when the low-frequency scale suggested that they watch more TV than most other people (see Schwarz, 1990, for a review). Similarly, the psychosomatic patients in Schwarz and Scheuring's (1992) study reported higher health satisfaction when the high-frequency scale suggested that their own symptom frequency is below average than when the low-frequency scale suggested that it is above average. Note that this higher report of health satisfaction was obtained despite the fact that the former patients reported a higher symptom frequency in the first place, as seen above. Findings of this type reflect that respondents extract comparison information from their own placement on the scale and use this information in making subsequent comparative judgments.

However, not all judgments are comparative in nature. When asked how satisfied we are with our health, we may compare our own symptom frequency with that of others. Yet, when asked how much our symptoms bother us, we may not engage in a social comparison but may instead draw on the absolute frequency of our symptoms. In this case, we may infer that our symptoms bother us more when a high-frequency scale leads us to estimate a high symptom frequency. Accordingly, other patients who reported their symptom frequency on one of the above scales reported that their symptoms bother them more when they received a high- rather than a low-frequency scale (Schwarz, in press). Thus, the same high frequency scale elicited subsequent reports of higher health satisfaction (a comparative judgment) or of higher subjective suffering (a noncomparative judgment), depending on which judgment followed the symptom report.

Users of respondents' reports. Finally, the use of frequency scales as a frame of reference in making comparative judgments is not limited to patients, but has also been observed for their physicians. Schwarz, Bless, Bohner, Harlacher, and Kellenbenz (1991, Experiment 2) asked practicing physicians with an average professional experience of eight and a half years to participate in a study allegedly designed to "test if a standard health survey could be shortened without a decrease in usefulness and reliability." As part of this study, practitioners were presented with vignettes that described a patient who had allegedly reported his or her symptoms along one of the scales shown in Figure 1.

For example, in one vignette, "Mr. Z., 25 years old" checked that he suffered twice a week from "aching loins or back," and in another vignette, "Mrs. K., 41 years old" checked that she suffered from a "lack of energy," also

twice a week. Note that "twice a week" constitutes a high response on the low-frequency scale, but a low response on the high-frequency scale. On the basis of these materials, the physicians rated the severity of the symptoms and the extent to which they saw a need for medical consultation. As expected, suffering from a given symptom "twice a week" was rated as more severe, and as more likely to require consultation, when "twice a week" represented a high- rather than a low-frequency response on the respective scale.

Methodological implications. The reviewed findings have important methodological implications of the assessment of frequency reports (see Schwarz, 1990, for more detailed discussions).

First, the numeric response alternatives presented as part of a frequency question may influence respondents' interpretation of what the question refers to, as seen in the section on question comprehension. Hence, the same question stem in combination with different frequency alternatives may result in the assessment of somewhat different behaviors. This is more likely the less well-defined the behavior is.

Second, respondents' use of the frequency scale as a frame of reference influences the obtained behavioral reports. Aside from calling the interpretation of the absolute values into question, this also implies that reports of the same behavior along different scales are not comparable, often rendering comparisons among different studies difficult.

Third, the impact of response alternatives is more pronounced the less respondents can recall relevant episodes from memory. This implies that reports of behaviors that are poorly represented in memory are more affected than reports of behaviors that are well-represented. When behaviors of differential memorability are assessed, this may either exaggerate or reduce any actual differences in the relative frequency of the behaviors, depending on the specific frequency range of the scale.

Fourth, for the same reason, respondents with poorer memory for the behavior under study are more likely to be influenced by response alternatives than respondents with better memory. Such a differential impact of response alternatives on the reports provided by different groups of respondents can result in misleading conclusions about actual group differences.

Finally, the range of response alternatives may further influence subsequent comparative and noncomparative judgments. Hence, respondents may arrive at evaluative judgments that are highly context-dependent and may not reflect the assessments they would be likely to make in daily life.

To avoid these systematic influences of response alternatives, it is advisable to ask frequency questions in an open-response format, such as, "How many hours a day do you watch TV? ___ hours per day." Note that such an open format needs to specify the relevant units of measurement, for example, "hours per day," to avoid answers like "a few."

As another alternative, researchers are often tempted to use vague quantifiers, such as "sometimes," "frequently," and so on. This, however, is the worst possible choice (see Moxey & Sanford, 1992; Pepper, 1981, for reviews). Most important, the same expression denotes different frequencies in different content domains. Thus, "frequently" suffering from headaches reflects higher absolute frequencies than "frequently" suffering from heart attacks. Moreover, different respondents use the same term to denote different objective frequencies of the same behavior. For example, suffering from headaches "occasionally" denotes a higher frequency for respondents with a medical history of migraines than for respondents without that history. Accordingly, the use of vague quantifiers reflects the objective frequency relative to respondents' subjective standard, rendering vague quantifiers inadequate for the assessment of objective frequencies, despite the popularity of their use.

Reporting the Answer

After having determined a frequency estimate, respondents have to report their estimate to the researcher (Strack & Martin, 1987; Tourangeau, 1984). The estimate they communicate may deviate from their private estimate due to considerations of social desirability and self-presentation (see DeMaio, 1984, for a review). Survey researchers have developed a number of techniques to reduce this "editing" of the communicated response. Most of these techniques emphasize the respondent's anonymity and the confidentiality of the collected data (see Sudman & Bradburn, 1983, for a review and good practical advice).

In the context of the preceding discussion of frequency scales, one may wonder to what extent these scales contribute to response editing: Do respondents hesitate to endorse a frequency that seems "deviant" in the context of the scale? This possibility has been suggested by Bradburn and Danis (1984) in a discussion of higher reports of alcohol consumption in an open- as opposed to a closed-response format, but has received little empirical support. Specifically, this self-presentation hypothesis suggests that the impact of response alternatives should be more pronounced when respondents report about their own behavior

Figure 1
Response Alternatives Used in Medical Judgment Study

Low Frequency Scale

()	()	()	()	(X)	()
less than once a month	about once a month	about once in two weeks	about once a week	about twice a week	more often

High Frequency Scale

()	(X)	()	()	()	()
less than twice a week	about twice a week	about four times a week	about six times a week	about once every 24 hours	more often

Note. From "Response Scales as Frames of Reference: The Impact of Frequency Range on Diagnostic Judgment," by N. Schwarz, H. Bless, G. Bohner, U. Harlacher, and M. Kellenbenz, 1991, *Applied Cognitive Psychology*, 5, p. 43. Copyright 1991 by John Wiley & Sons. Reprinted with permission.

than when they report about the behavior of distant others, reflecting that self-presentation considerations are of less concern in the latter case. In contrast, the estimation hypothesis advanced above predicts that the impact of response alternatives increases with decreasing memory for the behavior. If so, the impact of response alternatives should be more pronounced for reports about others than for reports about self, reflecting that one usually knows more about one's own behavior. The available data support the latter prediction (Schwarz & Bienias, 1990). Nevertheless, it is conceivable that self-presentation concerns elicited by highly threatening questions may be compounded if the respondent discovers that his or her report requires the endorsement of a response alternative that seems extreme in the context of the scale. If so, response alternatives may also affect behavioral reports at the editing stage of the response sequence, although compelling empirical evidence for this possibility has yet to be provided.

Summary

In summary, research into behavioral reports consistently demonstrated that mundane and frequent behaviors are poorly represented in memory, forcing respondents to rely on estimation strategies (Bradburn et al., 1987; Schwarz & Sudman, 1994; Strube, 1987; Sudman et al., 1996). One of these strategies entails the use of frequency response alternatives as a frame of reference, resulting in systematic biases in behavioral reports and subsequent related judgments. Other strategies, which are beyond the scope of this article, include the decomposition of the behavior into easier to estimate parts (e.g., Blair & Burton, 1987) and the use of subjective theories of stability and change over time as a framework for reconstructing one's personal history (Ross, 1989). In combination, the bulk of the work in this area highlights that retrospective behavioral reports are highly fallible and strongly affected by the specifics of the research instrument used.

Reporting on One's Attitudes

Public opinion researchers have long been aware that attitude measurement is highly context-dependent. Numerous studies have demonstrated that preceding questions may influence the responses given to subsequent ones (see Schuman & Presser, 1981; Schwarz & Strack, 1991; Sudman et al., 1996; Tourangeau & Rasinski, 1988; and the contributions in Schwarz & Sudman, 1992, for research examples and reviews). Moreover, when a self-administered questionnaire is used, subsequent questions may also influence preceding ones (e.g., Schwarz & Hippler, 1995b) because self-administered questionnaires allow respondents to go back and forth between questions. In recent years, considerable conceptual and empirical progress has been made in this domain, and several related theoretical models have been offered (Feldman, 1992; Feldman & Lynch, 1988; Schwarz & Bless, 1992a; Schwarz & Strack, 1991; Strack & Martin, 1987; Tourangeau, 1987, 1992; Tourangeau & Rasinski, 1988). Below I draw on Schwarz and Bless's (1992a) inclusion-exclusion model, which specifies the conditions under which question-order effects

emerge and predicts their direction, their size, and their generalization across related issues.

The Construal of Targets and Standards

In a nutshell, the model assumes that individuals who are asked to form a judgment about some target stimulus first need to form some mental representation of it. As numerous studies in social cognition have shown (see Bodenhausen & Wyer, 1987; Higgins, 1996; Schwarz, 1995, for reviews), individuals do not retrieve all knowledge that may potentially bear on the target. Instead, they truncate the search process as soon as enough information has come to mind to form a judgment with sufficient subjective certainty. Accordingly, the judgment is based on the subset of potentially relevant information that is most accessible at the time of judgment. Some of this information may always come to mind when the individual thinks about this topic and is hence called *chronically accessible*. Other information may only come to mind due to contextual influences, for example, because it was addressed in a preceding question. Such information is called *temporarily accessible* (see Higgins, 1996). An example may illustrate the difference. Suppose a respondent who suffers from a severe illness is asked to report on her life-satisfaction. Although she may draw on numerous different aspects of her life, her health problems are likely to be chronically accessible, and she will probably consider them independent of whether a preceding question brought health-related issues to mind or not. On the other hand, she may only draw on the quality of her housing arrangements when this aspect was brought to mind by a preceding question, and healthy respondents may not consider their health unless their attention is drawn to it. The influence of preceding questions on the temporary accessibility of information in memory is a primary source of context effects in self-reports, whereas chronically accessible information is a source of context-independent stability in reports.

To form an evaluative judgment, however, it is not sufficient to have a mental representation of the target. In addition, respondents need a mental representation of a standard against which the target is evaluated. Much as the mental representation of the target, the mental representation of a relevant standard is formed on the spot and is based on chronically or temporarily accessible information that happens to come to mind (see Kahneman & Miller, 1986; Schwarz & Bless, 1992a).

How a given piece of accessible information influences the judgment depends on how it is *used* in forming these different representations. For example, Strack, Schwarz, and Gschneidinger (1985) asked respondents to report either three positive or three negative life-events that recently happened to them. Not surprisingly, respondents who had to recall positive events subsequently reported *higher* happiness and life-satisfaction than respondents who had to recall negative events, as shown in the first row of Table 2. This reflects that they included the recent events in the mental representation of their current lives. Other respondents, however, were asked to report three positive or negative events that happened to them at least five years

ago. These respondents reported *lower* current happiness and life-satisfaction after recalling past positive events than after recalling past negative events, as shown in the second row of Table 2. This contrast effect reflects that the distant events did not directly pertain to respondents' current lives and were hence not included in the representation formed of the target "my-life-now." Instead, the accessible past events were used in forming a mental representation of the standard against which respondents evaluated their current lives—and compared with the fun (or trouble) they had five years ago, life now seemed pretty bland (or rather good, respectively).

In general, information that is included in the representation formed of the target of judgment results in *assimilation effects*, that is, more positive judgments when positive rather than negative information comes to mind. On the other hand, information that is used in constructing a standard of comparison results in *contrast effects*. In this case, positive (negative) information results in a more positive (or negative, respectively) standard of comparison, against which the target is evaluated more negatively (or positively, respectively). Hence, the same information may influence the judgment in opposite directions, depending on whether it is used in forming a mental representation of the target or of the standard against which the target is evaluated (see Schwarz & Bless, 1992a, for more detail).

To further illustrate the basic logic of these mental construal processes, I review a recent experiment that explored the impact of thinking about Colin Powell on evaluations of the Republican party and of Bob Dole (Stapel & Schwarz, 1998).

The Republican Who Did Not Want to Become President: Context Effects in Political Judgment

In November 1995, retired General Colin L. Powell declined to compete in the 1996 Presidential race as a Republican candidate, admitting that his candidacy would require "a passion and commitment that, despite my every effort, I do not have for political life" (New York Times, 1995). Simultaneously, however, he also announced that he had just joined the Republican party. How would this dual

decision affect judgments of the Republican party and of Bob Dole?

When asked to evaluate the Republican party, respondents presumably draw on information that bears on this party. When the highly popular Colin Powell comes to mind and is included in the representation formed of this party, the party should be evaluated more positively. To test this prediction, Stapel and Schwarz (1998) asked some respondents, "General Colin L. Powell, the hero of the Gulf War, recently decided to become a member of a political party. Do you happen to know which party that is?" As expected, these respondents subsequently evaluated the Republican party more positively than respondents who were not asked a question about Powell.

Other respondents were also asked a question about Colin Powell, yet this question was designed to set Powell aside from the Republican party. It read, "General Colin L. Powell, the hero of the Gulf War, has recently been wooed by a political party to run as its candidate in the 1996 Presidential elections. He decided not to run. Do you happen to know which party it was that made Powell an offer he rejected?" As expected, these respondents subsequently evaluated the party more negatively than respondents who were not asked a question about Powell.

Thus, bringing Colin Powell to mind through a preceding question always influenced respondents' judgments of the Republican party. The direction of the influence, however, depended on whether Colin Powell was included in the representation formed of this party or excluded from it. In the former case, the judgment became more positive (an assimilation effect), whereas in the latter case Powell served as a very positive standard of comparison, relative to which the party looked less good (a contrast effect). Hence, we need to consider the accessibility as well as the use of contextual information to understand the emergence and direction of context effects, as already seen in the life-satisfaction example discussed above (Strack et al., 1985).

Next, consider how the above questions about Colin Powell may influence judgments of Bob Dole, who ran as the Republican candidate in the 1996 elections. When asked about Bob Dole, respondents presumably draw on information pertaining to this person. When Colin Powell comes to mind, he is unlikely to be included in the representation formed of Bob Dole: The target category "Bob Dole" has only one member, in contrast to the more inclusive target category "Republican party." Hence, Colin Powell should serve as a standard of comparison whenever he comes to mind, independent of which of the two context questions is asked. Confirming this prediction, respondents evaluated Bob Dole more negatively when either of the context questions brought Colin Powell to mind, compared with a condition in which no question about Powell was presented.

This differential impact of Colin Powell on judgments of Bob Dole and the Republican party reflects the categorical relationship between the contextual information and the target of judgment: When the judgment pertains to a target category that is superordinate (Republican party) to

Table 2
Subjective Well-Being: The Impact of Valence of Event and Time Perspective

Time perspective	Valence of event	
	Positive	Negative
Recent events	8.9	7.1
Past events	7.5	8.5

Note. Shown are mean scores of happiness and life-satisfaction, ranging from 1 to 11, with higher values indicating reports of higher well-being. Adapted from Experiment 1 of Strack, Schwarz, and Gschneidinger (1985).

the context information (Colin Powell), the contextual information can be included in, as well as excluded from, the representation formed of the target. Unless the context question elicits exclusion (as in the above experiment), inclusion is more likely. Hence, assimilation effects are more likely to be obtained than contrast effects when the judgment pertains to a superordinate target. Conversely, target categories that are lateral (Bob Dole) to the context information (Colin Powell) do not allow for the inclusion of the context information. Hence, the context information is used in constructing a standard of comparison, resulting in contrast effects (see Schwarz & Bless, 1992a, for a more detailed discussion).

Scandals and the Trustworthiness of Politicians

The diverging impact of the same context information on evaluations of superordinate and lateral target categories often leads to surprising results. Suppose respondents are asked to evaluate the trustworthiness of American politicians and Richard Nixon and his role in the Watergate scandal come to mind due to a preceding question. Nixon is likely to be included in the representation of the superordinate target "American politicians," resulting in judgments of low trustworthiness. Yet, suppose that the question does not pertain to American politicians as a group but to specific exemplars, such as Newt Gingrich. In this case, Nixon cannot be included in the representation of the lateral target "Newt Gingrich" and will serve as a standard of comparison, relative to which Newt Gingrich will look more trustworthy than would otherwise be the case. A study with German respondents, and the German equivalent of Watergate, confirmed these predictions: Thinking about a politician who was involved in this scandal decreased the trustworthiness of politicians in general, but increased the trustworthiness of all individual politicians assessed (Schwarz & Bless, 1992b; see also Bless & Schwarz, 1998).

Such diverging evaluations of groups and exemplars are often observed in public opinion research. For example, Americans distrust Congress, but trust their own representative (e.g., Erikson, Luttbeg, & Tedin, 1988). Similarly, they are likely to favor capital punishment in general, but much less likely to favor its application in any specific case (e.g., Ellsworth & Gross, 1994). Moreover, women and minorities report considerable discrimination against their group as a whole, yet evaluate their own personal experiences as more benign (e.g., Taylor, Wright, & Porter, 1994). These patterns are to be expected because media coverage renders extreme cases of untrustworthiness, hideous crime, and severe discrimination highly accessible in memory. And these extreme examples can be included in the representation formed of the issue in general, yet they serve as a standard of comparison in evaluating individual instances. As a result, the general and specific judgments diverge in the manner observed above.

Conversational Norms and Information Use: Does Marital Satisfaction Contribute to Life-Satisfaction?

In the preceding examples, the use of temporarily accessible information was determined by the categorical relationship between the context information and the target of judgment. In addition, a host of other variables may influence the use of contextual information (see Schwarz & Bless, 1992a, for a comprehensive review). One of these variables is the norms of conversational conduct discussed in the section on question comprehension. Recall that the maxim of quantity (Grice, 1975) enjoins speakers to provide information that is new to the recipient and to avoid redundancy. Consequently, respondents deliberately exclude information from further consideration when it seems redundant in the context of the ongoing conversation.

A study on marital satisfaction and life-satisfaction illustrates this process (Schwarz, Strack, & Mai, 1991; see also Strack, Martin, & Schwarz, 1988; Tourangeau, Rasinski, & Bradburn, 1991). In one condition, respondents were first asked how satisfied they are with their life as a whole and subsequently how satisfied they are with their marriage. In this case, the two judgments correlated $r = .32$. When the question order was reversed, however, this correlation increased to $r = .67$. This reflects that the marital satisfaction question brought marriage-related information to mind, which could be included in the representation formed of one's life as a whole, resulting in an assimilation effect. In a third condition, the two questions were introduced by a joint lead-in, designed to evoke the norm of nonredundancy. This lead-in informed respondents that they would be asked two questions related to their well-being, namely one about their marriage and one about their life as a whole. In this case, the previously observed correlation of $r = .67$ under the same order condition dropped to a nonsignificant $r = .18$. Apparently, respondents interpreted the general life-satisfaction question as if it were worded, "Aside from your marriage, which you already told us about, how satisfied are you with other aspects of your life?" Confirming this interpretation, a control condition in which the general life-satisfaction question was reworded in this way resulted in a highly similar correlation of $r = .12$.

This impact of question order and conversational norms was also reflected in respondents' mean life-satisfaction judgments, as the reports of unhappily married respondents may illustrate (i.e., the reports of the third of the sample with the lowest marital satisfaction). Compared with the condition where the general life-satisfaction question was asked first ($M = 6.8$ on an 11-point scale, with 11 = *very satisfied*), these respondents reported lower general life-satisfaction ($M = 5.8$) when the preceding question brought their unhappy marriage to mind. Yet, when the joint lead-in induced them to disregard their marriage, the rest of life didn't seem so bad by comparison ($M = 8.0$). The reports of the happily married respondents provided a mirror image of these results.

Quite obviously, with correlations ranging from .18 to .67, we would draw very different conclusions about the contribution of marital satisfaction to overall life-satisfaction depending on the order in which these questions were presented and whether they were introduced by a joint lead-in. Moreover, these contextual influences would profoundly affect the outcome of structural equation models based on these data.

Summary

As the reviewed examples illustrate, attitude measurement is subject to pronounced context effects. In fact, some readers may wonder if the reviewed results suggest that we mostly collect artifacts when we ask attitude questions. I do not think so (see Sudman et al., 1996, chapter 5, for a more detailed discussion). Human judgment is always context-dependent, in research situations as in real life, and attitude judgments are no exception to this rule. We construct these judgments on the spot, when needed (Schwarz & Bless, 1992a; Strack, 1994a, 1994b; Wilson & Hodges, 1992), by drawing on the information that is most accessible at that point in time (Higgins, 1996). Some of this information is chronically accessible, whereas other information may only come to mind because it has been addressed in an earlier question or the recent news. Whereas the chronically accessible information provides for some stability in judgments, the temporarily accessible information is the basis of context effects. The direction of context effects depends on how the accessible information is used—and the same information may result in assimilation as well as contrast effects, reflecting the mental construal processes discussed above. These processes are systematic and apply in “real life” as well as in research situations—and many readers will agree that they see their life differently, depending, for example, on whether they take their marriage into account or not. The “problem” is not the context dependency of human judgment but researchers’ hope that this context dependency may—miraculously—not apply to their own study. Unfortunately, this hope is unwarranted, and any given result may lead us astray when we do not take its contextual nature into account. To be alerted to contextual influences, researchers are well-advised to include context manipulations in the design of their studies, a piece of advice that is more often offered than heeded.

Concluding Remarks

Psychologists and social scientists have long been aware that collecting data by asking questions is an exercise that may yield many surprises (e.g., Cantril, 1944; Payne, 1951)—some obvious ones (as when the researcher is left wondering, “When this is the answer, what was the question?”), and some so nonobvious that they can only be detected when we compare responses to different question wording, format, or order conditions. Whereas the obvious surprises are merely annoying, the nonobvious ones may lead us to erroneous conclusions about the substantive issue under study if we do not become aware of them. Over the last decade, psychologists and survey methodologists have made considerable progress in understanding the cognitive

and communicative processes underlying the question-answering process, rendering some of these “surprises” less surprising than they have been in the past. In fact, we can produce findings like the ones reviewed in this article in a reliable and replicable way when we deliberately write questions to satisfy theoretical criteria. Yet, this does not imply that we can always predict how a given question will “behave” when colleagues ask us for advice: In many cases, the given question is too mushy an operationalization of theoretical variables to allow for predictions (although we typically feel we know what would happen if the question were tinkered with, in one way or another, to bring it in line with theoretical models). Nevertheless, the accumulating insights (reviewed in Sudman et al., 1996) are helpful in guiding researchers to avoid many of the more common pitfalls associated with collecting self-reports. Most important, they alert us to likely problems and help us in identifying questions and question sequences that need systematic experimental testing before they are used in a large-scale study.

The most important lesson that has emerged from this research is rather general in nature (Schwarz, 1996). As researchers, we tend to view our questionnaires as “measurement devices” that elicit information from respondents. What we frequently overlook is that our questionnaires are also a source of information that respondents draw on in order to determine their task and to arrive at a useful and informative answer. Far from reflecting “artifacts” or “shallow responding,” findings of the type reviewed in this article indicate that respondents do their best to be cooperative communicators. Consistent with the assumptions that underlie the conduct of conversation in daily life, they assume that all contributions of the researcher are relevant to the goals of the ongoing exchange, and they take these contributions into account in arriving at an answer. Unfortunately, as researchers we are often not fully aware of the information that our questionnaires—or our experimental procedures (see Hilton, 1995; Schwarz, 1994, 1996)—provide, and hence miss the extent to which the questions we ask determine the answers we receive.

REFERENCES

- Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco: Freeman.
- Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, 50, 240–250.
- Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research*, 14, 280–288.
- Bless, H., & Schwarz, N. (1998). Context effects in political judgment: Assimilation and contrast as a function of categorization processes. *European Journal of Social Psychology*, 28, 159–172.
- Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, 23, 149–165.
- Bodenhausen, G. V., & Wyer, R. S. (1987). Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 6–41). New York: Springer Verlag.
- Bolton, R. N., & Bronkhorst, T. M. (1996). Questionnaire pretesting:

- Computer assisted coding of concurrent protocols. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 37–64). San Francisco: Jossey-Bass.
- Bradburn, N., & Danis, C. (1984). Potential contributions of cognitive research to survey questionnaire design. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 101–129). Washington, DC: National Academy Press.
- Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, *236*, 157–161.
- Cantril, H. (1944). *Gauging public opinion*. Princeton, NJ: Princeton University Press.
- Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt, Brace, Jovanovich.
- Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions* (pp. 15–48). New York: Russell Sage Foundation.
- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257–281). New York: Russell Sage Foundation.
- DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 177–196). San Francisco: Jossey-Bass.
- Ellsworth, P. C., & Gross, S. R. (1994). Hardening of the attitudes: Americans' views on the death penalty. *Journal of Social Issues*, *50*, 19–52.
- Erikson, R. S., Luttbeg, N. R., & Tedin, K. T. (1988). *American public opinion* (3rd ed.). New York: Macmillan.
- Feldman, J. M. (1992). Constructive processes in survey research: Explorations in self-generated validity. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 49–61). New York: Springer Verlag.
- Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, *73*, 421–435.
- Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 15–36). San Francisco: Jossey-Bass.
- Gaskell, G. D., O'Muircheartaigh, C. A., & Wright, D. B. (1994). Survey questions about the frequency of vaguely defined events: The effects of response alternatives. *Public Opinion Quarterly*, *58*, 241–254.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford Press.
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, *118*, 248–271.
- Hippler, H. J., Schwarz, N., & Sudman, S. (Eds.). (1987). *Social information processing and survey methodology*. New York: Springer Verlag.
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.
- Jobe, J., & Loftus, E. (Eds.). (1991). Cognitive aspects of survey methodology [Special issue]. *Applied Cognitive Psychology*, *5*.
- Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136–153.
- Lessler, J. T., & Forsyth, B. H. (1996). A coding system for appraising questionnaires. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 259–292). San Francisco: Jossey-Bass.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge, England: Cambridge University Press.
- McGonagle, K. A., Kessler, R. C., & Schilling, E. A. (1992). The frequency and determinants of marital disagreements in a community sample. *Journal of Social & Personal Relationships*, *9*, 507–524.
- Menon, G. (1994). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 161–172). New York: Springer Verlag.
- Menon, G., Raghurir, P., & Schwarz, N. (1995). Behavioral frequency judgments: An accessibility-diagnostics framework. *Journal of Consumer Research*, *22*, 212–228.
- Moxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their use in attitude research. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 279–296). New York: Springer Verlag.
- New York Times. (1995, November 9). *Excerpts from General Powell's news conference*, p. A11.
- Norenzayan, A., & Schwarz, N. (in press). Telling what they want to know: Participants tailor causal attributions to researchers' interests. *European Journal of Social Psychology*.
- Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, *21*, 331–339.
- Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.
- Pepper, S. C. (1981). Problems in the quantification of frequency expressions. In D. W. Fiske (Ed.), *Problems with language imprecision: New directions for methodology of social and behavior science* (Vol. 9, pp. 25–41). San Francisco: Jossey-Bass.
- Ross, M. (1989). The relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*, 341–357.
- Schober, M. F. (in press). Making sense of questions: An interactional approach. In M. Sirken, D. Hermann, S. Schechter, N. Schwarz, J. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research*. New York: Wiley.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology: Review of personality and social psychology* (Vol. 11, pp. 98–119). Beverly Hills, CA: Sage.
- Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 123–162). San Diego, CA: Academic Press.
- Schwarz, N. (1995). Social cognition: Information accessibility and use in social judgment. In E. E. Smith & D. N. Osherson (Eds.), *Thinking: An invitation to cognitive science* (2nd ed., Vol. 3, pp. 345–376). Cambridge, MA: MIT Press.
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Hillsdale, NJ: Erlbaum.
- Schwarz, N. (in press). Frequency reports of physical symptoms and health behaviors: How the questionnaire determines the results. In D. C. Park, R. W. Morrell, & K. Shifren (Eds.), *Processing medical information in aging patients: Cognitive and human factors perspectives*. Mahaw, NJ: Erlbaum.
- Schwarz, N., & Bienias, J. (1990). What mediates the impact of response alternatives on frequency reports of mundane behaviors? *Applied Cognitive Psychology*, *4*, 61–72.
- Schwarz, N., & Bless, H. (1992a). Constructing reality and its alternatives: Assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Hillsdale, NJ: Erlbaum.
- Schwarz, N., & Bless, H. (1992b). Scandals and the public's trust in politicians: Assimilation and contrast effects. *Personality and Social Psychology Bulletin*, *18*, 574–579.
- Schwarz, N., Bless, H., Bohner, G., Harlacher, U., & Kellenbenz, M.

- (1991). Response scales as frames of reference: The impact of frequency range on diagnostic judgment. *Applied Cognitive Psychology*, 5, 37–50.
- Schwarz, N., Grayson, C. E., & Knäuper, B. (1998). Formal features of rating scales and the interpretation of question meaning. *International Journal of Public Opinion Research*, 10, 177–183.
- Schwarz, N., & Hippler, H. J. (1991). Response alternatives: The impact of their choice and ordering. In P. Biemer, R. Groves, N. Mathiowetz, & S. Sudman (Eds.), *Measurement error in surveys* (pp. 41–56). Chichester, England: Wiley.
- Schwarz, N., & Hippler, H. J. (1995a). The numeric values of rating scales: A comparison of their impact in mail surveys and telephone interviews. *International Journal of Public Opinion Research*, 7, 72–74.
- Schwarz, N., & Hippler, H. J. (1995b). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly*, 59, 93–97.
- Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response categories: Effects on behavioral reports and comparative judgments. *Public Opinion Quarterly*, 49, 388–395.
- Schwarz, N., Knäuper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, F. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55, 570–582.
- Schwarz, N., Park, D., Knäuper, B., & Sudman, S. (Eds.). (1998). *Aging, cognition, and self-reports*. Washington, DC: Psychology Press.
- Schwarz, N., & Scheuring, B. (1988). Judgments of relationship satisfaction: Inter- and intraindividual comparison strategies as a function of questionnaire structure. *European Journal of Social Psychology*, 18, 485–496.
- Schwarz, N., & Scheuring, B. (1992). Selbstberichtete verhaltens- und symptomhäufigkeiten: Was befragte aus anwortvorgaben des fragebogens lernen [Frequency reports of psychosomatic symptoms: What respondents learn from response alternatives]. *Zeitschrift für Klinische Psychologie*, 22, 197–208.
- Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. *European Review of Social Psychology*, 2, 31–50.
- Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, 55, 3–23.
- Schwarz, N., Strack, F., Müller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, 6, 107–117.
- Schwarz, N., & Sudman, S. (Eds.). (1992). *Context effects in social and psychological research*. New York: Springer Verlag.
- Schwarz, N., & Sudman, S. (1994). *Autobiographical memory and the validity of retrospective reports*. New York: Springer Verlag.
- Schwarz, N., & Sudman, S. (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.
- Sirken, M., Hermann, D., Schechter, S., Schwarz, N., Tanur, J., & Tourangeau, R. (Eds.). (in press). *Cognition and survey research*. New York: Wiley.
- Stapel, D. A., & Schwarz, N. (1998). The Republican who did not want to become President: An inclusion/exclusion analysis of Colin Powell's impact on evaluations of the Republican Party and Bob Dole. *Personality and Social Psychology Bulletin*, 24, 690–698.
- Strack, F. (1994a). *Zur psychologie der standardisierten befragung* [The psychology of standardized interviews]. Heidelberg, Germany: Springer Verlag.
- Strack, F. (1994b). Response processes in social judgment. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 287–322). Hillsdale, NJ: Erlbaum.
- Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer Verlag.
- Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology*, 18, 429–442.
- Strack, F., Schwarz, N., & Gschneidinger, E. (1985). Happiness and reminiscing: The role of time perspective, mood, and mode of thinking. *Journal of Personality and Social Psychology*, 49, 1460–1469.
- Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition*, 9, 111–125.
- Strube, G. (1987). Answering survey questions: The role of memory. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 86–101). New York: Springer Verlag.
- Sudman, S., & Bradburn, N. M. (1983). *Asking questions*. San Francisco: Jossey-Bass.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.
- Tanur, J. M. (Ed.). (1992). *Questions about questions*. New York: Russell Sage Foundation.
- Taylor, D. M., Wright, S. C., & Porter, L. E. (1994). Dimensions of perceived discrimination: The personal/group discrimination discrepancy. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 7, pp. 233–255). Hillsdale, NJ: Erlbaum.
- Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, 59, 291–297.
- Tourangeau, R. (1984). Cognitive science and survey methods: A cognitive perspective. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). Washington, DC: National Academy Press.
- Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 149–162). New York: Springer Verlag.
- Tourangeau, R. (1992). Attitudes as memory structures: Belief sampling and context effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 35–47). New York: Springer Verlag.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299–314.
- Tourangeau, R., Rasinski, K. A., & Bradburn, N. (1991). Measuring happiness in surveys: A test of the subtraction hypothesis. *Public Opinion Quarterly*, 55, 255–266.
- Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection, mode, question format, and question context. *Public Opinion Quarterly*, 60, 275–304.
- Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Hillsdale, NJ: Erlbaum.
- Winkielman, P., Knäuper, B., & Schwarz, N. (1998). Looking back at anger: Reference periods change the interpretation of (emotion) frequency questions. *Journal of Personality and Social Psychology*, 75, 719–728.