10

Asking Questions:
Measurement in the Social Sciences

Fritz Strack
University of Würzburg, Germany

Norbert Schwarz
University of Michigan

Psychology is an empirical science. This implies that its validity is rooted in reality, and that reality must have a chance to influence our conceptualizations (e.g., Rosenthal & Rosnow, 1984). In the natural sciences, the link between theory and reality is systematic observation. Often, however, manifestations of reality need to be translated in order to be perceived by our senses. Moreover, to communicate what is perceived, we need a shared reference point, at best, a meter. In this case, measurement affords objectivity in that it does not hinge on subjective experience of the observer (e.g., Wilson, 1992).

The diagnosticity of a datum, however, depends not only on reality but also on the theory on which the measurement is based. If we measure the acidity of a fluid by scaling the color of a litmus paper on a red–blue dimension, the recorded color reflects the acidity to the extent that the theory linking the acidity of the fluid to the color of the litmus paper is correct (e.g., Sydenham & Thorn, 1996).

Observation and measurement have their place in the social sciences, where the reality consists of people and their behavior, as well (see Rosenthal & Rosnow, 1984). Some features of persons can be directly perceived (e.g., gender, mother tongue, race, age), whereas other characteristics are not directly observable, such as traits, attitudes, or motives. For them, a special instrument of measurement is needed to assign numbers to objects.

However, the social sciences seem to have an alternative way of accessing human characteristics: asking questions. Because people are capable of answering questions, their responses serve as data of measurement (Sudman & Bradburn, 1982). As for the natural sciences, however, the validity of the theories that mediate between response and the target of measurement determines the diagnosticity of the assessment. Interestingly, the social sciences have more than one theory linking responses to underlying characteristics. Although these theories are not always explicitly stated, they share a similar terminology but reflect entirely different substantive orientations.

MEASUREMENT BY ASKING QUESTIONS

1. Psychometrics: The Behaviorist Model

One of the most widespread models of measurement-by-asking-questions is that of psychometric testing. It is based on the behaviorist assumption that the answer to a question is simply a response elicited by a stimulus, in this case, the question (for a more complete account of psychometric test theory, see Lord & Novick, 1974). The response consists of two components: (a) a *true-value* component and (b) an *error* component. Psychometric test theory further assumes that the error is randomly determined and that its dispersion around the true value will approximate a normal distribution with increasing number of questions. Because the error (i.e., the deviation from the true value that is associated with one particular question) is considered to be random, psychometricians do not focus on the content or the wording of a particular question. Multiple measurement, many questions tapping the same phenomenon, is the route on which the psychometrician approaches the true value. The validity hinges not on one single question but solely on the covariations of the responses with other behaviors, that is, response behaviors under standardized conditions (for a related discussion, see Abelson, 1984).

Moreover, the respondents do not even have to know their "true value." For example, if a psychometrician wants to find out whether a respondent is extraverted or a type A person, it is not necessary that the re-

spondent has any idea what this concept refers to or where he or she would be located on that dimension. Of interest is primarily the relationship between the responses, usually in form of a summary score, and a criterion variable. Accordingly, psychometricians show little interest in how a question is understood and how an answer is generated.

2. Survey Research: The Introspective Model

A rather different metatheory underlies standardized questioning in survey situations. Although survey researchers use a similar terminology (e.g., the terms *true value* and *error*; see Lessler, 1984), their approach to measurement is quite different. On the surface, this is reflected in the fact that, unlike psychometricians, survey researchers often use one single question to address a particular phenomenon and do care extensively about the content of the question as well as its wording and comprehension (e.g., Belson, 1981; Payne, 1951; Schuman & Presser, 1981; Sudman & Bradburn, 1982). Still, as for psychometricians, it is their goal to capture respondents true values. How is that possible without multiple measurements? What is the rationale behind this logic?

Table 10–1 summarizes the classic meta-theory of survey responding, which dominated survey research through the 1980s. (The characterization of the metatheory of survey responding is based on various parts of the handbook edited by Turner & Martin [1984].) In subsequent years, this metatheory changed in ways we address in the next section. In its classic version, the metatheory of survey responding has four components that refer to features of the respondent, to a psychological process that guarantees validity, and to a possible source of error. We discuss each of them in turn.

The starting point is the assumption that respondents possess certain features. These features are either *objective*, such as a specific age and gender, or *subjective*, such as a certain attitude and belief. The only difference between the two classes is the existence of external criteria for the objective features and the absence of external criteria for the subjective ones. Thus, the true value of respondents age can be checked by inspecting their birth certificates, whereas the true value of a specific attitude cannot be examined by such means.

Such objective validation, however, is not a necessary criterion for survey measurement because its internal validity is guaranteed by the method of accessing the true value. It is assumed that independent of whether external criteria exist (i.e., whether the features are objective or subjective), respondents have immediate access to their true value. Quite succinctly, Martin (1984) summarized this position as follows:

<center>**TABLE 10–1**
The Introspection Theory of Standardized Question Situations</center>

Assumption

Features of respondent

   (a) Objective (age, gender, income …)—objective criterion

   (b) Subjective (beliefs, attitudes, evaluations …)—*no* objective criterion

Claim

Immediate access to true value of features

Method of access

Introspection

Source of error

Respondents lie if goal of reporting true value is less desirable than other goals (e.g., making a good impression)

"There is a fundamental assumption in survey research that respondents can give valid reports of their own subjective states" (p. 298). Just as respondents can report their true age, they can describe their true attitude with candor and accuracy (Campbell, 1981, p. 23).

What is the psychological mechanism that guarantees such a privileged, immediate, and unbiased access to one's own subjective features? It is the method of introspection. (For a recent general discussion of the literature on introspection, see Lyons, 1986; for a more empirically oriented treatment of the topic, see Nisbett & Wilson, 1977, and Ericsson & Simon, 1980.) As Martin (1984) put it: "It might be assumed … that respondents base their report on introspective self-examination" (p. 298) Given the assumption that respondents *can* access the true values of their subjective features, errors are possible in this conceptualization only if respondents do not *want* to communicate their true values if they do not tell the truth (although they know it); that is, if they lie. Thus, if the respondents' *competence* is ruled out as a determinant of error, then it is their *motivation* that must be held responsible for deviations from the truth. Almost exclusively, the influence that affects respondents' motivation not to communicate their true values is assumed to be *social desirability*—that is, the desire to make a positive impression, or at least to avoid a negative one (see DeMaio, 1984).

In this metatheory, the key issue in the collaboration of respondent and researcher is a motivational one: Does the respondent comply with

what the questioner wants her to do; that is, does she tell the truth? Although this is an important insight, this introspective theory limits theorizing in survey methodology to addressing only one aspect of collaboration. This shortcoming makes it difficult for the metatheory of survey research to explain a substantial body of findings bearing on the impact of question wording and question context (e.g., Schuman & Presser, 1981). Specifically, survey researchers found that rather innocuous variations, such as changes in the order in which questions are asked, may have enormous effects on respondents answers (for research examples, see Belson, 1981; Payne, 1951; Schuman & Presser, 1981; Schwarz & Sudman, 1992). In explaining these so-called *response effects*, survey methodologists realized the limits of their metatheory; it proved rather implausible to invoke changes in respondents' motivation to collaborate and tell the truth as the major variable underlying the impact of question wording and question context (see Hippler & Schwarz, 1987; Strack & Martin, 1987).

3. Measurement as Cooperative Communication

Given these limitations, an alternative conceptual framework for response processes seemed warranted and has found increasing acceptance in survey research. This conceptualization recognizes that asking and answering questions is a type of conversation and has properties of a natural discourse in which two (or more) people engage in a purposeful verbal interaction. As Paul Grice (1975), a philosopher of language, put it: "Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are … cooperative efforts; and each participant recognizes in them … a common purpose or set of purposes, or at least a mutually accepted direction" (p. 45).

The conversational nature of measurement-by-asking-questions tends to be overlooked because contributions are typically restrained by the standardized format in which questions are asked and answers are to be provided (Schwarz, 1994, 1996; Strack, 1994; Strack & Schwarz, 1992; see also Clark & Schober, 1992). Examples of standardized question situations include attitude surveys (see Schwarz & Strack, 1991) and experiments in the social and psychological sciences (see Bless, Strack, & Schwarz, 1993), which share the standardization of the researcher contributions (e.g., questions, instructions) and constrain the respondent' answers to a specified format. Because of these restrictions, standardized questioning in the social sciences is often considered equivalent to standardized measurement in the natural sciences.

However, to understand response processes it is useful to recognize the conversational nature of interactions in research situations (Schwarz, 1996; Strack, 1994). As indicated before, communications in natural settings give participants a large degree of freedom to generate messages in a format of their choosing. In most situations, questioners and respondents can decide to be more or less specific, to be elliptical or redundant, or to ask for feedback about an earlier comment. This lack of restriction serves an important function in the conversation process (see Clark & Clark, 1977). Specifically, it has become apparent that to identify the intended meaning of a communication, a collaborative interaction between conversants plays a crucial role.

Quite some time ago, Krauss and Weinheimer (1964, 1966) found that in the course of an interaction, respondents became more accurate and efficient in identifying ambiguous objects that the questioner had selected if the respondent received feedback from the questioner. On the basis of these observations, Clark and his collaborators (e.g., Clark & Wilkes-Gibbs, 1986) developed a collaborative theory of reference (Schober & Clark, 1989) to explain the process of understanding in natural discourse. In this collaborative perspective, speakers and listeners give each other feedback to ensure that a communication' intended meaning is understood. The studies conducted within this perspective (Garrod & Anderson, 1987) have convincingly demonstrated that, to understand what is meant, deciphering the semantic meaning of a particular word or sentence is not sufficient. Instead, the respondent must go beyond the linguistic units to identify the intended meaning of an utterance (i.e., the questioner' communicative intention). In the endeavor, the unrestricted interaction between participants plays a crucial role.

Obviously, standardized questioning lacks this type of unconstrained exchange. Respondents typically do not receive feedback if their interpretation of a question corresponds to what the questioner had in mind. Furthermore, the questioner has no indication of whether a response that is provided in a given format is based on the intended meaning of the question. In such situations, the standardized context of questions and answers may serve as a substitute for the unrestricted feedback that occurs in natural situations. Specifically, respondents are likely to rely on contextual features to a greater degree than participants in natural settings (Bless et al., 1993).

At this point, another difference between natural discourses and standardized situations becomes apparent. In natural communications, the communicative intentions of both the questioner and the respondent are often ambiguous; that is, a person who asks a question may not necessarily request information in natural settings. Instead, questions

may represent indirect speech acts (Searle, 1975, 1976) that express or imply behavioral requests (e.g., "Can you open a window?"), threats (e.g., "Do you want me to lock away your bicycle again?"), assertions (e.g., "Don't you think the play was awful?"), and other actions. Similarly, in natural discourses responses may not be intended merely to inform the questioner. However, in standardized situations respondents can (or at least should) assume that questioners want information. (Of course, in a psychological experiment, a question may be asked to influence cognitive processes; however, it is important that this intention is not recognized by the respondent; see Bless et al., 1993.) This intention can be conveyed by a direct request or a question. In turn, respondents in standardized situations most likely will try to obey this request and provide the desired information. To be sure, in specific situations, respondents may strive for alternative goals, particularly the goal to make a good impression. However, as mentioned before, this goal will be activated only under very specific circumstances. Thus, it can be assumed that respondents can recognize the questioner intention and are motivated to cooperate.

The Cooperative Principle. Determination of the motivation to cooperate is necessary, but not sufficient, to understand response effects. One must also identify the mechanisms of cooperation once the motivation is established. The principles best known and studied as rules for communicating in natural situations are those identified by Grice (1975), whose central postulates were subsumed under a general *cooperative principle*. This principle is composed of four maxims.

A *Maxim of Quantity* requires participants in a discourse to provide the right amount of information; that is, a contribution should convey not more and not less information than is necessary to understand what is meant. A Maxim of Quality demands that the conversants tell the truth, whereas a *Maxim of Relevance* requests that contributions should relate to one another. Finally, a *Maxim of Manner* requires the contributions to be clear and without obscurity (for a more detailed discussion of the Gricean principles, see Levinson, 1983). The assumption that speakers adhere to these rules (Higgins, 1981; McCann & Higgins, 1992) is important for the listener to both infer the intended meaning of an utterance and generate a response that meets the expectations of the speaker.

However, the implementation of these rules can require additional information from the speaker. An example is the application of the Maxim of Quantity. To determine the appropriate amount of information, a respondent may ask the questioner for further specification. Thus, the question "Where do you live?" could be countered with whether the re-

quest for information refers to the country, the city, or the neighborhood. Instead of bothering the questioner, however, the respondent may infer what would be new information to the questioner and what would not. Such an inference may be based on the larger context in which the question is posed. Thus, if the information is requested by a foreign colleague, the response "lower East side" would violate the Maxim of Quantity, whereas "New York" would be appropriate. The reverse would be true if the same question were asked by a colleague at a New York university. Who asks a question and under what circumstances allows inferences about the state of knowledge and what would be new to the questioner. This *given-new contract* (Clark, 1985; Clark & Haviland, 1977), according to which participants in a discourse add information to what they assume the partner already knows, must also be realized by monitoring the course of a conversation. That is, an answer should go beyond the information that already has been provided.

Cooperation Under Natural and Standardized Conditions.   In natural situations, it is the context at large that helps to interpret people's communicative intentions. For example, the question "Can you open a window?" will be interpreted as a request for information only if the respondent's pertinent capability is, in fact, questionable and the Maxim of Quantity is observed. If it is not, the respondent will take it as a request for action. Thus, a child may cooperate by answering "Yes, I can," whereas a cooperative adult may respond with "Just a second."

Although such indirect speech acts rarely occur in standardized situations, this example shows how pragmatic characteristics that are external to the question proper determine the response under natural conditions. In standardized situations, the respondent cannot expect the questioner to take his or her specific situation (e.g., his or her capability) into account. Therefore, contextual cues that help determine the communicative intention of the questioner are sought. The particular response format, the order in which questions are asked, and the wording of questions can provide these cues.

In the following paragraphs, we describe how research participants use different aspects of standardized question situations to determine the intended meaning of a question: What is the information that the researcher wants them to provide?

IMPLICATIONS FOR SOCIAL SCIENCE MEASUREMENT

The literature on response effects in survey measurement offers many examples of the pervasive influence of minor changes in question word-

ing, format and order. Although these observations were long treated as surprising oddities, they are to be expected from a conversational perspective. Specifically, respondents bring the tacit assumptions that govern the conduct of conversation in daily life to the research situation and assume that the researchers are cooperative communicators, whose contributions to the research conversation come with a "guarantee of relevance" (Sperber & Wilson, 1986). What is often overlooked is that these contributions include apparently formal features of the research instrument, which the researcher may have chosen for reasons of technical convenience. Respondents, however, draw on these features to determine the pragmatic meaning of the question asked.

We first address the supposedly most "formal" aspect of questionnaires, namely, the nature of response alternatives. Next, we turn to issues of question wording and, finally, we consider the context in which a question is presented, including the preceding questions, introductions to a study, and the researcher affiliation.

1. Response Formats

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 of which the researcher is obviously aware (e.g., "gave a survey interview") or may take for granted anyway (e.g., "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 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 in which the researcher is interested. 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 illustrates. 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, we 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 by *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," they may conclude that the researcher has relatively rare events in mind. Hence, the question cannot refer to minor irritations, which 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 either asked how frequently they had been angry "last week" or "last year." Again, they inferred that the researcher was interested in more frequent and less severe episodes of anger when the question pertained to 1 week rather than 1 year, and their examples reflected this differential question interpretation.

   These findings have important implications for the interpretation of commonly observed differences in 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

1 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. Accordingly, respondents may deliberately report on different experiences, rendering their reports noncomparable.

On theoretical grounds, we may further expect that formal features, such as the values of a frequency scale or the length of a reference period, seem more relevant when they are unique to the question asked rather than shared by many heterogeneous questions. In the latter case, respondents may conclude that this is the format used for all questions, rendering it less informative for the intended meaning of any given one. Empirically, this is the case. In a replication of Winkielman et al.'s (1998) study, Igou, Bless, and Schwarz (2002) observed that using the same reference period for several substantively unrelated questions attenuated or eliminated its influence on respondents' interpretation of the anger question relative to a condition in which each question was associated with a unique reference period.

Numeric Values of 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, Hippler, Noelle-Neumann, and Clark (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 0 and 5 on the 0-to-10 scale, only 13% endorsed one of the formally equivalent values between and 0 on the –5-to-+5 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, 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, 1995). 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 unipolar dimension in mind, referring to different degrees of the same attribute.

Unfortunately, researchers are typically not aware of the informative functions of formal characteristics of their research instruments and choose them mostly on the basis of technical convenience, as the case of rating scales illustrates. (Our summary is based on a conversation with Charles Cannell, who headed the field department of the Survey Research Center at the University of Michigan during those decades.) Rensis Likert (1932) introduced rating scales with a graphic response format, shown in the first row of Table 10–2. With the introduction of punch cards this format was changed to the numeric format shown in the second row to reduce transcription errors at the data entry stage. This format, however, still required two keystrokes for each entry and was hence changed to the format shown in the third row, thus cutting data entry cost. Along the way, a clearly bipolar presentation format changed into the now-familiar unipolar one—even for questions that are intended to present a bipolar response dimension, which is now merely indicated by the verbal end anchors. Of course, these technical changes were not assumed to affect question interpretation. In light of the above findings, however, one may wonder the extent to which they contaminated time series of attitude data by confounding attitude change over time with changes in the response format.

Range of Targets. Respondents' goal of identifying the intended meaning of a question and its accompanying rating scale can also be reached in other ways. When several stimuli have to be judged along the same response scale, the range of stimuli presented may serve as a conversational cue. Assume, for example, that respondents have to rate how pricy a restaurant is. In one condition, the restaurants to be as-

TABLE 10–2
Different Formats of Rating Scales

– – –	– –	–	+	+ +	+ + +
– 3	– 2	– 1	+ 1	+ 2	+ 3
1	2	3	4	5	6

sessed include Joe' Pizza Parlor and The Golden Goose, a restaurant that has been awarded a Michelin star. In another condition, the targets are confined to restaurants that have the Michelin distinction. The first condition suggests that the questioner refers to restaurants in general; the second condition allows the inference that gourmet restaurants are the topic of discourse. As a consequence, the same target is rated as more expensive in the first case than in the second.

This prediction corresponds to explanations that construe the response scale as a flexible rubber band (Postman & Miller, 1945; Volkmann, 1951), rather than a rigid yardstick. In this view, the respondent anchors the scale so that its endpoint corresponds to the most extreme stimulus in the range. In the restaurant example, the lower anchor would be 'Joe,' and all the gourmet restaurants would be assembled at the upper end of the scale. Therefore, the latter restaurants would be rated as more expensive along the scale than they would if Joe' were not among the set of those considered. In other words, the introduction of the pizza parlor as an anchor would produce a contrast effect on ratings of the other stimuli.

Technically, the rubber-band notion does not imply an identification of the topic of discourse. It merely requires that the most extreme values be identified for use in anchoring the scale. However, this presupposes that all stimuli are simultaneously available at the time of judgment. This is not always the case; that is, the targets are often presented sequentially and have to be assessed in a consecutive manner. Thus, judges have to infer the possible range of the stimuli. Of course, such an inference can be drawn if the topic of discourse is identified. When a scale applies to attitudinal judgments, one stimulus that might be considered in construing the range of values to which the scale is relevant is one's own position (Upshaw, 1965). For example, suppose several persons attitudes toward the legalization of drugs have to be rated on a scale ranging from *liberal* to *conservative*. If the judge favors the legalization of heroin and all of the attitude statements considered are less extreme than this position, then the judge's attitude might be used to anchor the scale. Thus, a statement advocating the legalization of marijuana would be judged as more conservative than it would if the judge's attitude were moderate (i.e., within the range of alternatives considered). In other words, the judge's attitude has a contrast effect on the ratings of others' attitudes. More generally, if a respondent's perspective (Upshaw, 1965; Upshaw & Ostrom, 1984) changes as a function of one's own attitude on an issue, one's judgments of other stimuli on the relevant dimension change as well.

The fact that people include their own attitudes into the range of stimuli has consequences for communication. For example, the way a

friend who is extremely conservative will be described to a third person will depend on the recipient's own political stand. That is, if the recipient is liberal, a description implying a higher degree of conservatism (e.g., very conservative) will be provided than if the recipient leans toward conservatism (e.g., rather conservative). At the expense of being inconsistent by using different categories to describe the same stimulus, respondents are more informative if they take the presumed interpretation of the receivers into account, which is determined by their stand on the issue.

Summary. In combination, the reviewed examples highlight that respondents draw on apparently formal features of the questionnaire as a source of relevant information in determining the pragmatic meaning of the question asked. Little do they know how haphazardly those features may have been chosen, as the example of numeric values of rating scales illustrates. When respondents become aware that the feature may be of questionable relevance to the specific question at hand—for example, because it is used for several heterogeneous questions—they no longer rely on it, eliminating its otherwise observed influence. Throughout, these question form effects undermine the comparability of answers to highly similar questions, only differ only in their presumably "formal" features.

2. Question Wording

It is not surprising that the way a question is worded influences its interpretation. The semantic meaning can obviously vary as a function of the words used and thus influence responses. However, different question wordings may influence responses even under conditions in which the wordings seem semantically equivalent.

For example, semantically, to *forbid* and to *allow* are antonyms, and *not allow* seems equivalent to *forbid*. However, the proportion of survey respondents who answered "yes" when asked if an activity (e.g., smoking marijuana) should be "forbidden" was consistently lower than the proportion who answered "no" when asked if this same activity should be "allowed" (Rugg & Cantril, 1944; Schuman & Presser, 1981). This asymmetry suggested that *not forbidding* was not *allowing*. As Hippler and Schwarz (1986) demonstrated, many respondents considered the possibility that they would not actively oppose the activity but would not support it either. Those respondents answered "no" to the "allow" as well as the "forbid" form of the question, resulting in the observed asymmetry.

The *type* of article is another example of how the wording of a question can affect responses. Most prominently, consequences of the use of

the definite versus indefinite article were investigated by Loftus (1975) in the context of eyewitness testimony. Participants in her studies saw a videotape of a car accident. Some were subsequently asked if they had seen "*the* broken headlight," whereas others were asked if they had seen "*a* broken headlight." This manipulation typically resulted in more affirmative responses when the definite article was used.

The explanation of this phenomenon has been primarily memorial in nature. It is assumed that the presupposition semantically implied by the use of the definite article (i.e., "there *was* a broken headlight") distorted the memory representation of the event, which in turn caused erroneous recall. Despite some dissenting opinions (Lindsay & Johnson, 1989; McCloskey & Zaragoza, 1985; Tversky & Tuchin, 1989), memory mechanisms are still widely held responsible for the phenomenon (Loftus & Hoffman, 1989). However, there is evidence that the wording of the question per se is not sufficient to produce the effect (e.g., Dodd & Bradshaw, 1980; Smith & Ellsworth, 1987). Instead, listeners draw on the information provided by the definitive article only when they can assume that the speaker is a cooperative communicator (the default assumption in psychological experiments), but not otherwise. Hence, the wording has little influence when it is introduced by a defendant lawyer, who is assumed to follow a self-serving agenda (e.g., Dodd & Bradshaw, 1980).

Moreover, Strack and Bless (1994) found that the presupposition implied by the use of the definite article was used as a basis of inference only when other strategies were not applicable. In some of their experimental conditions, participants could base their answers to the question of whether they had previously seen a certain object both on the conversationally conveyed presupposition that the object had been presented ("D id you see *the* screwdriver?" and on their own metacognitive knowledge (i.e., the belief that they *would* have remembered the particular object had it been presented). The applicability of this metacognitive strategy was manipulated by varying the salience of the items in the recognition set. The differential use of judgmental strategies was observed when participants were asked if they had seen an item that had not been presented. Then, the use of the direct versus the indirect article only increased false alarms if the object was not salient. If the object was salient, almost all participants correctly rejected the item as not seen before. These findings suggest that, in the absence of a memory trace, judgmental strategies may come into play, and judges may prefer one strategy over the other. These findings also suggest that the surface structure of a task does not fix the mental mechanisms used to solve it. Thus, a memory task may be solved by inferential strategies that are applicable in a given situation. In this perspective, leading questions

influence responses not by altering what has been encoded about the target but by allowing the respondent to infer what was probably the case. If better alternatives are not available, respondents may use those cues to generate a required response.

3. Preceding Questions

In natural conversations, listeners are expected to draw on the context of an utterance in determining its meaning, and not doing so may be interpreted as a lack of attention or interest. In contrast, researchers often hope that each question is considered in isolation and deplore the emergence of context effects in question comprehension. These context effects take two forms. First, respondents may deliberately draw on the content of preceding questions to determine the meaning of subsequent ones. Second, the answers to preceding questions become part of the common ground, and respondents avoid reiterating information that they have already provided earlier, consistent with the Maxim of Quantity. We address both in turn.

Contextual Information and the Resolution of Ambiguity. 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, 1983; 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 less 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. However, 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 when it followed the financial support question they inferred that it pertained to students receiving money. Reflecting this differential interpretation, they reported a more favorable attitude toward the introduction of an educational contribution in the latter than in the former case—hardly a meaningless response.

Common Ground and Redundancy Avoidance.   Questions that were previously asked and answered provide information about the questioner's current state of knowledge. This information is important, because it allows the respondent to obey the Maxim of Quantity by making his or her answer as informative as required. This is the case if an answer adds to what the recipient already knows. However, the respondent's knowledge changes as a function of the ongoing discourse. As a consequence, the informativeness of a statement depends on communications that have preceded it in the conversation. In other words, a contribution should build on the "common ground" (Clark, 1985) that has been established between participants of the discourse. Syntactically, switching from the indirect to the direct article symbolizes that a target has become a given and allows for new information to be added. Clark and Haviland (1977) described this application of Grice's (1975) Maxim of Quantity to a natural discourse as the *given-new contract*.

The fact that the new value of a contribution is determined by one previous contributions requires participants in a discourse to keep track of what one has said before. In a natural situation, this type of monitoring occurs automatically; a conversant would normally not repeat a previous contribution unless there were reason to assume that the recipient has not understood its content. For example, suppose a person is first asked the question "How is your wife?" followed by "And how is your family?" He is unlikely to take his wife's well-being into consideration in answering the second question, because of his previous answer it would not be informative. Note that this is not the case if the questions had been asked in the reverse order.

The given-new contract should be obeyed in standardized situations when two questions overlap in their content. This is the case if a general question follows a more specific one and their contents are in a subset–superset relation or if their content intersects. In addition, the two (or more) questions must be related to each other. In natural contexts, the speaker guarantees that the rule of relation is observed. In standardized situations, however, this rule is not always obeyed. On the contrary, such a perception is actively avoided by placing related questions at different positions in a questionnaire, separating them by several filler items. Thus, a respondent may or may not see a series of questions as belonging together. More generally, a respondent application of the Maxim of Quantity depends on his or her perception of the relatedness of the items involved (Strack, 1992).

This hypothesis was tested in a study by Strack, Martin, and Schwarz (1988; cf. Tourangeau, Rasinski, & Bradburn, 1991), in which the conversational context was manipulated experimentally. Participants were given a questionnaire that included two questions whose content stood in a subset–superset relationship. The more specific question addressed respondents' happiness with their dating, whereas the more general one concerned their happiness with life as a whole. If the two questions are perceived to belong to the same context of discourse, then the given-new contract should be applied, and the respondents should avoid being redundant. In analogy to the previous example, they should not base the judgments of happiness with life in general on their happiness with dating if they have already reported their dating happiness. However, if the questions are not perceived to belong together, then answering the specific question should render the relevant content more accessible and should increase the probability that the answer to the general question is based on the content of the specific one (see Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1979, 1980). Thus, correlations between the answers should be high in the specific–general order if no conversational context is established. However, under the conversational-context condition, the correlation should be reduced, because the same contents should not be communicated twice.

To establish the conversational context, the two questions were introduced with the following statement: "We are now asking two questions about your life, a) happiness with dating, b) happiness with life in general." No such introduction was used in the no-context condition. Moreover, to further avoid the perception of relatedness, the questions in the latter condition were printed on different pages of the questionnaire. The pattern of correlations corresponded to the predictions. Compared with the control conditions, in which the general question preceded the specific one, the correlation decreased when a

conversational context was introduced ($r = .16$) but increased when it was not ($r = .55$).

The assumption that the decreased correlation under the latter condition was caused by an exclusion of the activated content requires a more diagnostic test. Therefore, a conceptual replication was conducted by Schwarz, Strack, and Mai (1991). German adults who had either a spouse or a partner were asked how satisfied they were with both their current relationship and their lives. Both the order of the questions and the conversational context were varied. Two new conditions were added, in which respondents were explicitly instructed either to include or exclude the redundant content of the specific question when they rated their satisfaction with their lives in general.

The previous pattern of correlation coefficients clearly was replicated; that is, the correlation between the answers decreased if the conversational context was introduced. Moreover, the correlations under conditions where respondents were explicitly instructed to include or exclude the specific content matched exactly the conditions under which the given-new contract was expected to implicitly require respondents to consider the specific information or not. Taken together, this set of findings suggests that respondents in standardized situations comply with the Gricean Maxim of Quantity when they answer questions whose content is related in a part–whole fashion.

In an extension of this logic, Strack et al. (1991) applied the same procedure to questions whose contents were semantically similar. Specifically, they asked participants how happy and satisfied they were with their lives. It was assumed that respondents who observed the given-new contract would be more likely to differentiate between the similar concepts of happiness and satisfaction than would respondents who where not concerned about avoiding redundancy. To foster the perception of relatedness, a box was drawn around the questions "Here are two questions about your life." To prevent such a perception, the two questions were presented as being part of two different questionnaires that used different scales, colors, and typefaces and were described as serving different purposes. "Happiness" was the last item of Survey 1, and "satisfaction" was the opening question of the second questionnaire.

In contrast to many cognitive theories (e.g., Wyer & Srull, 1989), the conversational logic predicted that the correlation between the two answers would be higher if the questions were separated and lower if they were presented as conversationally related. These predictions were borne out by the data. The correlation between the similar dimensions of subjective well-being was almost perfect ($r = .96$) if the questions belonged to different surveys. In contrast, if they were perceived as related, the correlation of the answers dropped dramatically ($r = .65$).

These results provide further evidence that conversational principles are often relevant in standardized question situations. However, this is true only if the standardized exchange has features of natural discourse; that is, the questions must represent an ongoing dialogue in which both the questions and answers to them are perceived as part of the same exchange. This is often ambiguous, however, in standardized situations.

It is not necessary to establish a conversational context explicitly. The immediate sequence of questions may be sufficient to elicit such a perception. This was the case in a study conducted by Ottati, Riggle, Wyer, Schwarz, and Kuklinski (1989); they found that respondents expressed a more positive attitude toward the general topic of free speech if a preceding question about the same issue referred to a specific group that was positively evaluated (e.g., the American Civil Liberties Union) than if it referred to a group that was negatively evaluated (e.g., the American Nazi Party). However, this assimilation effect was found only when the two questions were separated in the questionnaire. If the specific question immediately preceded the general one, a contrast effect was found such that the positive content produced a more negative attitude and vice versa.

Another aspect of informativeness concerns the required accuracy of a response. Respondents are often uncertain as to how exact their answer has to be. This is particularly relevant if they are requested to report past occurrences and their frequencies. For example, suppose participants are asked to report whether or how often they went to see a movie or a doctor during the last 6 months. They may not interpret the interviewer's request as a demand to engage in an exhaustive memory search. Instead, they may infer that their communication goal will be attained by providing an estimate that is only approximate. Given the constraints of most question situations, such an interpretation seems to comply with the cooperative principle.

To make such frequency estimates, participants may first recall the number of instances that occurred during a shorter period of time and extrapolate. Thus, in the previous example, they might recall the number of movies they have seen during the last month and extrapolate from that database to the requested time period (Bradburn, Rips, & Shevell, 1987). This strategy could result in over- or underestimations of the actual frequency.

To induce respondents to provide a more precise answer, Loftus, Klinger, Smith, and Fiedler (1990) suggested a "two-time frame questioning procedure." Specifically, these authors recommended asking for the frequency of the same behavior in different time periods. For example, to increase the accuracy of participants' estimates of how often they had had a physical examination within the last 2 months, they first

might be asked to indicate the number of physicals they had during a different period (e.g., the last 6 months). Loftus et al. compared respondents' medical records with their reports of doctor visits and found more accurate responses under such conditions than under conditions in which the initial question had not been asked.

The effectiveness of this procedure apparently results from an inference that respondents draw about the level of accuracy they are expected to attain; that is, the fact that two questions are asked pertaining to the same content in slightly different temporal frames suggests to respondents that the questioner has a specific interest in possibly different frequencies of occurrence of the event at different points in time, and therefore they make a greater effort to compute the frequency accurately. Thus, as in experimental situations where repeated measures draw participants' attention to what the experimenter wants to know (see Bless et al., 1993), the repeated posing of similar survey questions can be used to communicate this interest (see also Strack et al., 1988).

4. Researcher Affiliation

So far, our discussion focused on the information provided by questions and their context in the questionnaire. Note, however, that additional relevant context information is already provided in the cover letter that accompanies written questionnaires or the opening lines of interviews. One such piece of information is the researcher's affiliation, which respondents consider in determining the researcher's epistemic interest. For example, Norenzayan and Schwarz (1999) 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.

CONCLUSIONS

The program of research that has been outlined in this chapter describes the psychological mechanisms of answering questions in both natural and standardized situations. Moreover, it identifies some crucial

influences researchers need to know if they are asking questions to collect data in the social domain. We emphasize context and conversation as the perhaps most important influences.

In detail, we have contended in this chapter that answers are always generated in a social context. Even without specific evidence, it is safe to assume that answering a question is always influenced by the "actual, imagined or implied presence of others" (Allport, 1954), in this case, the presence of the questioner. This true not only for personal or telephone interviews but also for self-administered questionnaires. If recipients of the response are not present, they are always implied. In particular, it is the questioner's anticipated expectation that determines the generation of the response. To understand the specific influences, it is necessary to understand the rules of natural conversation. As we have demonstrated in this chapter, the Gricean (1975) maxims of conversational cooperation have proved exceptionally useful to understand and predict how various aspects of survey questions affect the generation of responses. Thus, the psychological processes that operate in natural communications may be fruitfully transferred to standardized settings. As a result, asking questions for the purpose of social measurement will become less of an art (Sudman & Bradburn, 1982) and more of a methodological practice that is guided by principles rooted in psychological evidence.

REFERENCES

Abelson, R. P. (1984). Psychological measurement: An introduction to the subjective domain. In C. F. Turner & E. M. Martin (Eds.), *Surveying subjective phenomena* (Vol. 1, pp. 117–125). New York: Russell Sage Foundation.

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot, England: Gower.

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1983). Effects of filter questions in public opinion surveys. *Public Opinion Quarterly, 47*, 528–546.

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.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987, April 10). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*, 157–161.

Campbell, A. (1981). *The sense of well-being in America*. New York: Russell Sage Foundation.

Clark, H. H. (1985). Language and language users. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 179–232). New York: Random House.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedl (Ed.), *Discourse production and comprehension* (pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions. Inquiries into the cognitive bases of surveys* (pp. 15–47). New York: Russell Sage Foundation.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1–39.

DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257–282). New York: Russell Sage Foundation.

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior, 19*, 695–704.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–225.

Garrod, S., & Anderson, A. (1987). Saying what you mean in a dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Gaskell, G. D., O'Muircheartaigh, C. A., & Wright, D. B. (1994). Survey questions about the frequency of vaguely defined events. *Public Opinion Quarterly, 58*, 241–254.

Grice, H. P. (1975). Logic and conservation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic.

Higgins, E. T. (1981). The "communication game": Implications for social cognition and persuasion. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 343–392). Hillsdale, NJ: Lawrence Erlbaum Associates.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141–154.

Hippler, H. J., & Schwarz, N. (1986). Not forbidding isn't allowing: The cognitive basis of the forbid-allow asymmetry. *Public Opinion Quarterly, 50*, 87–96.

Hippler, H. J., & Schwarz, N. (1987). Response effects in surveys. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 102–122). New York: Springer.

Igou, E. R., Bless, H., & Schwarz, N. (2002). Making sense of standardized survey questions: The influence of reference periods and their repetition. *Communication Monographs, 69,* 179–187.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science, 1,* 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4,* 343–346.

Lessler, J. T. (1984). Measurement error in surveys. In C. F. Turner & E. M. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 405–440). New York: Russell Sage Foundation.

Levinson, S. (1983). *Pragmatics*. Cambridge, England: Cambridge University Press.

Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*, 349–358.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*, 560–572.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100–104.

Loftus, E. F., Klinger, M. R., Smith, K. F., & Fiedler, J. (1990). A tale of two questions: Benefits of asking more than one question. *Public Opinion Quarterly, 54*, 330–345.

Lord, F. M., & Novick, M. R. (1974). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lyons, W. (1986). *The disappearance of introspection*. Cambridge, MA: MIT Press.

Martin, E. (1984). The tasks posed by survey questions. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 1, pp. 295–300). New York: Russell Sage Foundation.

McCann, C. D., & Higgins, E. T. (1992). Personal and contextual factors in communication: A review of the "communication game." In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 144–172). London: Sage.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: General, 114*, 1–16.

McGonagle, K. A., Kessler, R. C., & Schilling, E. A. (1992). The frequency and determinants of marital disagreements in a community sample. *Journal of Social and Personal Relationships, 9*, 507–524.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Norenzayan, A., & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers interests. *European Journal of Social Psychology, 29,* 1011–1020.

Ottati, V. C., Riggle, E., Wyer, R. S., Schwarz, N., & Kuklinski, J. (1989). Cognitive and affective bases of opinion survey responses. *Journal of Personality and Social Psychology, 57,* 404–415.

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.

Postman, L., & Miller, G. A. (1945). Anchoring of temporal judgments. *American Journal of Psychology, 58,* 43–53.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Rugg, D., & Cantril, H. (1944). The wording of questions. In H. Cantril (Ed.), *Gauging public opinion* (pp. 23–50). Princeton, NJ: Princeton University Press.

Schober, M. F., & Clark, H. H. (1989). Understanding by addresses and overhearers. *Cognitive Psychology, 21,* 211–232.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. Orlando, FL: Academic.

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). New York: Academic.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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. (1995). The numeric values of rating scales: A comparison of their impact in mail surveys and telephone interviews. *International Journal of Public Opinion Research, 7,* 72–54.

Schwarz, N., Knäuper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55,* 570–582.

Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 31–50). Chichester, England: Wiley.

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.

Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 59–82). New York: Seminar.

Searle, J. R. (1976). The classification of illocutionary acts. *Language in Society, 5,* 1–24.

Smith, V. L., & Ellsworth, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. *Journal of Applied Psychology, 72*, 294–300.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Cambridge University Press.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37,* 1660–1672.

Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology, 38*, 841–856.

Strack, F. (1992). "Order" effects in survey research: Activative and informative functions of preceding questions. In N. Schwarz & S. Sudman (Eds.), *Order effects in survey research* (pp. 23–34). New York: Springer.

Strack, F. (1994). *Kognitive und kommunikative Einflüsse in standardisierten Befragungssituationen*. [Cognitive and communicative influences in standardized question situations]. Heidelberg, Germany: Springer.

Strack, F., & Bless, H. (1994). Memory for non-occurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217.

Strack, F., & Martin, L. 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.

Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology, 18,* 429–442.

Strack, F., & Schwarz, N. (1992). Communicative influences in standardized question situations: The case of implicit collaboration. In G. Semin & K. Fiedler (Eds.), *Language and social cognition* (pp. 173–193). London: Sage.

Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition, 9*, 111–125.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions. A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Sydenham, P. H., & Thorn, R. (1996). *Handbook of measurement science: Vol. 1. Theoretical fundamentals*. New York: Wiley.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59,* 291–297.

Tourangeau, R., Rasinski, K. A., & Bradburn, N. (1991). Measuring happiness in surveys: A test of the subtraction hypothesis. *Public Opinion Quarterly, 55*, 255–266.

Turner, C. F., & Martin, E. (Eds.). (1984). *Surveying subjective phenomena* (Vols. 1 and 2). New York: Russell Sage Foundation.

Tversky, B., & Tuchin, M. (1989). A reconciliation of the evidence on eyewitness testimony: Comments on McCloskey and Zaragoza (1985). *Journal of Experimental Psychology: General, 118,* 86–91.

Upshaw, H. S. (1965). The effect of variable perspectives on judgments of opinion statements for Thurstone scales: Equal-appearing intervals. *Journal of Personality and Social Psychology, 2,* 60–69.

Upshaw, H. S., & Ostrom, T. M. (1984). Psychological perspective in attitude research. In J. R. Eiser (Ed.), *Attitudinal judgment* (pp. 23–41). New York: Springer.

Volkmann, J. (1951). Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads* (pp. 273–294). New York: Harper.

Wilson, M. (Ed.). (1992). *Objective measurement: Theory into practice* (Vol. 1). Norwood, NJ: Ablex.

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.

Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Lawrence Erlbaum Associates.