Why Researchers Should Think "Real-Time":
A Cognitive Rationale

Norbert Schwarz
University of Michigan

Version: 12 Dec 2011

When we want to know what people think, feel, and do, we ask them. This reliance on self-reports is based on the tacit assumption that people know their thoughts, feelings, and behaviors and can report on them "with candor and accuracy", as Angus Campbell (1981), a pioneer of survey research, put it. From this perspective, problems arise when the research situation discourages candor and accuracy, when the questions are ambiguous and difficult to understand, or when the task exceeds participants' knowledge and the limits of memory. A large methodological literature addresses these concerns and what to do about them (for reviews see Bradburn, Sudman, & Wansink, 2004; Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000). The lessons learned from this work highlight that many self-report problems can be attenuated by asking questions in close temporal proximity to the event of interest. Doing so constrains the multiple meanings of questions, reduces memory and estimation problems, and facilitates access to episodic detail, all of which can improve self-report. The real-time or close-in-time measures discussed in this handbook take advantage of this insight.

However, these (largely uncontroversial) methodological issues are only some of the reasons why researchers should think real-time. At a more fundamental level, recent research across many areas of psychological science highlights that every aspect of human cognition, emotion, motivation, and behavior is situated and highly context-sensitive, thwarting attempts to understand it in a decontextualized way (see the contributions in Mesquita, Barrett, & Smith, 2010). As this work progresses, it becomes increasingly clear that our methods should acknowledge this insight. They rarely do. This issue goes beyond the familiar methodological questions of "How to ask about X" and presents a fundamental (and controversial) challenge to bring our empirical operations into line with our theoretical assumptions. Studying psychological phenomena in the context of daily life can make important contributions to this development by shedding new light on the situated and embedded nature of human behavior and experience.

Preview

This chapter elaborates on these themes. The first section summarizes basic insights into how respondents answer questions and sets the stage for later sections. To date, research into the cognitive and communicative processes underlying self-reports has rarely addressed real-time (or close-in-time) measurement, which presents its own set of self-report problems. I draw attention to some of them and offer pertinent conjectures. The second section addresses reports of past behavior and reviews issues of autobiographical memory, highlighting the role of inference strategies and lay theories in determining what "must have been" (Ross, 1989). It pays particular attention to what respondents can, or cannot, report on with some accuracy.

The third section turns to reports of emotions and physical symptoms. It compares prospective reports of expected future feelings and retrospective reports of past feelings with concurrent reports of momentary experience. Of particular interest are systematic

convergences and divergences between these reports. On the one hand, predicted feelings usually converge with remembered feelings and the behavioral choices people make; on the other hand, all of these variables are often poorly related to actual experience as assessed by real-time measures (Schwarz, Kahneman, & Xu, 2009). These dynamics illustrate that feelings are fleeting and poorly represented in memory (Robinson & Clore, 2002); once they dissipated, respondents need to reconstruct what their feeling may have been. Shortly after the experience, episodic reconstruction can result in relatively accurate reports, as indicated by convergence with concurrent assessments (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). But as time passes, respondents resort to general knowledge to infer the past experience, which is also the knowledge used for predicting future feelings; these predictions, in turn, are the basis for intention and choice (Would this be good for me?). Hence, prediction, intention, choice, and later global memories converge because they are based on similar inputs – and this convergence seems to confirm that one's predictions and choices were right all along. Unfortunately, concurrent measures often suggest otherwise but this lesson is missed with the fading feeling (Schwarz et al., 2009). These dynamics impair learning from daily experience and challenge researchers' reliance on the consistency of respondents' reports as an indicator of validity.

The final section turns to reports of attitudes and preferences. It reviews the promises and pitfalls of the traditional conceptualization of attitudes as enduring dispositions and notes the malleable nature of attitude reports. Whereas this malleability is usually considered deplorable measurement error, a situated cognition approach suggests that it may reflect something more laudable and adaptive. If evaluation stands in the service of current action, it is likely to benefit from sensitivity to one's current goals and close attention to the affordances and constraints of one's current context (Schwarz, 2007). From this perspective, the context "dependency" that frustrates observers and researchers, who both want to predict an actor's behavior, reflects an adaptive context "sensitivity" that may serve the actor well. Real-time measurement in situ can shed new light on the underlying dynamics, in particular when it adopts the actor's rather than the observer's perspective.

ANSWERING QUESTIONS: THE PSYCHOLOGY OF SELF-REPORT

Answering a question in a research context requires that respondents (1) interpret the question to understand what the researcher wants to know and (2) retrieve and select relevant information to (3) form an answer. In most cases, they cannot provide an answer in their own words but (4) need to map it onto a set of response alternatives provided by the researcher. Finally, (5) respondents may wish to "edit" their answer before they communicate it for reasons of social desirability and self-presentation. Respondents' performance at each of these steps is context sensitive and profoundly influenced by characteristics of the research setting and instrument. Extensive reviews of these issues are available (Schwarz, Knäuper, Oyserman, &

Stich, 2008; Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000); I summarize key points and draw attention to some implications for real-time measurement.

Question Comprehension

The key issue at the question comprehension stage is whether respondents' understanding of the question matches the meaning the researcher had in mind. As all textbooks note, writing simple questions and avoiding unfamiliar or ambiguous terms helps (see Bradburn, Sudman & Wansink, 2004, for good practical advice). But ensuring that respondents understand the words is not enough. When asked, "What have you done today?" respondents will understand the words – but 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? Providing an informative answer requires inferences about the questioner's intention to determine the *pragmatic meaning* of the question (Clark & Schober, 1992; Schwarz, 1996).

Question context and order

To infer the pragmatic meaning, respondents draw on contextual information, from the purpose of the study and the researcher's affiliation to the content of adjacent questions and the nature of the response alternatives. Their use of this information is licensed by the tacit assumptions that govern the conduct of conversation in daily life (Grice, 1975), which respondents bring to the research situation (for reviews see Schwarz, 1994, 1996). Hence, they interpret a given question in the thematic context of the overall interview and a term like "drugs" acquires different meanings when presented in a survey pertaining to respondents' medical history rather than to crime in the neighborhood. Similarly, they attend to the researchers' affiliation to infer the likely epistemic interest behind their questions. Taking this interest into account, their explanations emphasize personality variables when asked by a personality psychologist, but social context variables when asked by a social scientist (Norenzayan & Schwarz, 1999). Respondents further assume that adjacent questions are meaningfully related to one another, unless otherwise indicated, and interpret their intended meaning accordingly (e.g., Strack, Schwarz, & Wänke, 1991).

When the data collection method enforces a strict sequence, as is the case for personal and telephone interviews and computer administered questionnaires that do not allow respondents to return to earlier questions, preceding questions can influence the interpretation of subsequent questions but not vice versa. In contrast, preceding as well as following questions can exert an influence when respondents can become aware of all questions prior to answering them, as is the case for paper-and-pencil questionnaires and computer programs without strict sequencing (Schwarz & Hippler, 1995). Most real-time studies probably fall into the latter category, given that they repeat a small number of questions with high frequency, thus allowing respondents to know what's coming even when the instrument enforces a strict order.

The maxims of cooperative conversational conduct further ask speakers to provide information the recipient needs and not to reiterate information the recipient already has (Grice, 1975). Respondents observe this norm and hesitate to reiterate information they have already provided in response to an earlier question (for a review see Schwarz, 1996). For example, Strack and colleagues (1991) observed a correlation of $r = .95$ when respondents were asked to report their overall happiness and their overall life-satisfaction in two separate questionnaires, attributed to different researchers. However, the correlation dropped to $r = .75$ when the same two questions were presented in the same questionnaire, attributed to the same researcher. In everyday discourse, the same questioner would not request the same information twice, in somewhat different words; hence, respondents differentiate between similar questions when they are presented by the same researcher. Two different researchers, on the other hand, may very well ask the same thing in different words, so identical answers are appropriate.

Note that the repetition of very similar, if not identical, questions is a key feature of many real-time measurement procedures. At present, we do not know how this affects respondents' question interpretation. Do respondents hesitate to repeat information at 4:05pm that they already provided at 3:40pm? If they hesitate to provide the same answer, does their attempt to provide new information increase meaningful differentiation between episodes or does it foster differentiations that go beyond respondents' actual experience in situ?

Formal characteristics of questions

From a conversational perspective, every contribution is assumed to be related to the ongoing conversation, unless marked otherwise. In research settings, the researcher's contributions include formal characteristics of the question, which respondents use in inferring the question's pragmatic meaning (Schwarz, 1994, 1996). Suppose, for example, that respondents are asked how frequently they felt "really irritated" recently. Does this question refer to major or minor annoyances? The numeric values of the frequency scale provide relevant information. When the scale presents low frequencies respondents infer that the researcher is interested in less frequent events than when it presents high frequencies; as a result they report on major annoyances (which are relatively rare) in the former, but on minor annoyances in the latter case (Schwarz, Strack, Müller, & Chassein, 1988). The same logic applies to the length of reference periods (Winkielman, Knäuper, & Schwarz, 1998). Given that major annoyances are less frequent than minor annoyances, respondents infer that the question pertains to minor irritations when it is presented with a short reference period (e.g., "yesterday"), but to major annoyances when presented with a long reference period (e.g., "last six months"). Accordingly, questions with reference periods of differential length assess substantively different experiences, e.g., "minor" rather than "major" episodes of anger.

This has potentially important implications for real-time measurement, which usually includes very short and recent reference periods. When asked at 4:05pm how often they have

been angry since the last measurement at 3:40pm, respondents may report on very minor episodes, which they would not consider worth mentioning for any longer reference period. Moreover, once they assume that this is what the questioner has in mind, they may evaluate each minor episode relative to other minor episodes. Consistent with this shift in the frame of reference, they may then assign each minor episode a high intensity rating, leading the researcher to conclude that intense anger is very frequent. To date, these possibilities have not been addressed and little is known about the potential impact of high density measurement on question interpretation.

Recall and Judgment

Once respondents determined what the researcher is interested in, they need to recall relevant information to form a judgment. In some cases, they may have direct access to a previously formed relevant judgment that they can offer as an answer. More likely, however, they will need to form a judgment when asked, taking the specifics of the question and the questioner's inferred epistemic interest into account. The processes pertaining to different types of reports are discussed in the sections on behaviors, feelings, and attitudes.

Formatting the Response

Unless the question is asked in an open response format, respondents need to format their answer to fit the response alternatives provided by the researcher (for a review see Schwarz & Hippler, 1991; Sudman et al., 1996). Respondents observe these question constraints and avoid answers that are not explicitly offered. Moreover, their selection of response alternatives is influenced by the order in which they are presented. In most cases, a given response alternative is more likely to be chosen when presented early rather than late on the list under visual presentation conditions, reflecting the sequence of reading. Conversely, a given alternative is more likely to chosen when presented late rather than early on the list under auditory presentation conditions; respondents need to wait for the interviewer to finish reading the list and work backward, beginning with the last alternative heard (Krosnick & Alwin, 1985; Sudman et al., 1996, chapter 6). This suggests that real-time data capture through Interactive Voice Responding, where the response alternatives are presented auditorily, may facilitate the emergence of recency effects, whereas the visual presentation formats typical for ESM and daily diaries may facilitate primacy effects.

Finally, respondents' use of rating scales reflects two regularities familiar from psychophysical research; both have been conceptualized in Parducci's (1965) range-frequency theory (see Daamen & de Bie, 1992, for social science examples). First, respondents use the most extreme stimuli to anchor the endpoints of the scale. Accordingly, they will rate a given episode of anger as less intense when the high end of the scale is anchored by an extreme rather than a moderate anger episode. This has important implications for the comparability of

retrospective and real-time reports. When asked to rate a single past episode, the recalled episode is likely to be compared to other memorable instances – which are often memorable because they were extreme. But when asked to rate multiple episodes over the course of a single day, previously rated moderate episodes may still be highly accessible. Hence, the same episode of anger may be rated as more extreme in real-time than in retrospective reports, reflecting the use of differentially extreme scale anchors and comparison standards.

Second, psychophysical research further shows that respondents attempt to use all categories of the rating scale about equally often when the number of to-be-rated stimuli is large. Hence, two similar stimuli may receive notably different ratings when only a few stimuli are presented, but identical ratings when many stimuli have to be located along the same scale. In many real-time studies, respondents are asked to rate a large number of episodes along identical scales over the course of a few hours, which is likely to elicit similar shifts in ratings. Both of these regularities predict systematic differences between retrospective and concurrent ratings as well as between concurrent ratings assessed with differential frequency. Future research may fruitfully test this possibility.

Editing the Response: Social Desirability

As the final step of the question answering sequence, respondents have to communicate their answer. Due to social desirability and self-presentation concerns they may edit their response (see DeMaio, 1984, for a review). This is more likely in face-to-face interviews than under the more confidential conditions of self-administered questionnaires, with telephone interviews falling in between. This is good news for real-time data capture, which predominantly relies on self-administered formats.

The literature further indicates that influences of social desirability are limited to potentially threatening questions and typically modest in size (DeMaio, 1984). Note, however, that a behavior that may seem only mildly unfavorable when reported once for a single specific episode (e.g., "I don't enjoy being with my spouse right now") may become a major self-presentation concern when the same answer would need to be provided over several episodes. If so, high density measurement in real-time studies may accentuate self-presentation concerns relative to retrospective reporting conditions though the cumulative impact of social desirability concerns over multiple similar episodes. Finally, respondents' self-presentation concerns can be reliably reduced through techniques that ensure the anonymity and confidentiality of the answer (see Bradburn et al., 2004, for detailed advice).

REPORTING ON BEHAVIORS

This section focuses on the recall stage of the question answering process and highlights what respondents can and cannot remember and report. It is organized by the type of information the researcher wants to assess.

Historical Information

Some questions pertain to historical information. Examples include, *Have you ever had an episode of heart burn? In what year did you first have an episode of heart burn?* Respondents' memories are usually the only available source of information and real-time measurement is not feasible. The best a researcher can do is to use interviewing techniques that take the structure of autobiographical memory into account to facilitate recall (for advice see Belli, 1998; Schwarz & Oyserman, 2001; Tourangeau, et al., 2000).

Current models of autobiographical memory conceptualize it as a hierarchical network that includes *extended periods* (e.g., "the years I lived in New York") at the highest level of the hierarchy. Nested within each extended period are lower-level *extended events* (e.g., "my first job" or "the time I was married to Lucy"). Further down the hierarchy are *summarized events*, which take the form of knowledge-like representations that lack episodic detail (e.g., "During that time, I was frequently ill."). *Specific events*, like a particular episode of illness, are represented at the lowest level of the hierarchy; to be represented at this level of specificity, the event has to be unique. As Belli (1998, p. 383) notes, this network, organized by time ("the years in New York") and relatively global themes ("first job;" "first marriage;" "illness"), enables "the retrieval of past events through multiple pathways that work top-down in the hierarchy, sequentially within life themes that unify extended events, and in parallel across life themes that involve contemporaneous and sequential events." Such searches take time and their outcome is somewhat haphazard, depending on the entry point into the network at which the search started. Building on these insights, *Event History Calendars* improve recall by using multiple entry points and forming connections across different periods and themes (see Belli, 1998, for a review and detailed advice).

In the absence of such (costly) efforts, respondents are likely to apply extensive inference strategies to the few bits and pieces they remember to infer what "must have" been (Ross, 1989). Suppose, for example, that respondents are asked how much alcohol they drank five years ago. Having no direct access to this information, they are likely to consider their current alcohol consumption as a benchmark and to make adjustments if they see a need to do so. In most cases, their adjustments are insufficient because people assume an unrealistically high degree of stability in their behavior. This results in retrospective reports that are more similar to the present than is warranted, as observed for reports of income (Withey, 1954), pain (Eich et al., 1985) or tobacco, marijuana, and alcohol consumption (Collins, Graham, Hansen, & Johnson, 1985). However, when respondents have reason to believe things were different in the past, they will "remember" change (Ross, 1998), as discussed next.

Reports of Change, Covariation, and Causation

Some questions go beyond mere retrospective reports and ask respondents to report on change over time (*Do you smoke more or less now than you did when you were 30?)* or to

assess the covariation of their behavior with other variables (*Do you smoke more when you are stressed?*). Respondents can rarely retrieve the information that would be needed to answer such questions and rely on extensive inference and estimation strategies to determine what might have been. Their answers are useful to the extent that the underlying lay theories happen to be correct, which is usually unknown.

Although most people assume an unwarranted amount of stability in their behavior, they will readily detect change when their lay theory suggests that change must have occurred. This is particularly likely --and problematic-- when the context suggests change, as is often the case in medical studies: Believing that things get better with treatment (or why else would one undergo it?), patients are likely to infer that their past condition must have been worse than their present condition (e.g., Linton & Melin, 1982; for a review see Ross, 1989). From a cognitive perspective, asking patients whether they feel better now than before their treatment is the most efficient way to "improve" the success rate of medical interventions, which may explain the recent popularity of "patient reported outcomes". Unfortunately, there is no substitute for appropriate study design. If change over time is of crucial interest, concurrent measures at different points in time are the only reliable way to assess it.

Similar problems arise when respondents are asked to report on covariation (*Under which circumstances…?*) or causation (*Why…?*). To arrive at an observation-based answer to these questions, respondents would need to have an accurate representation of the frequency of their behaviors, the different contexts of these behaviors, and the intensity of related experiences. Respondents are often unable to provide accurate reports on any of these components, making their joint consideration an unrealistically complex task.

Covariation and causation are best assessed with real-time data capture. Experience sampling methods excel at this task by prompting respondents to report on their behavior, experiences, and circumstances, allowing researchers to collect all the data needed for appropriate analyses. However, an important caveat needs attention. While real-time or close-in-time measures improve the accurate assessment of covariation, causation, and change, respondents' own behavioral decisions are based on their own perceptions, which may differ from reality. Hence, erroneous lay theories of covariation are often better predictors of behavior than accurate measures of covariation, as reviewed in the section on feelings.

Frequency Reports

Frequency questions ask respondents to report on the frequency of a behavior or experience during a specified reference period, often last week or last month. Researchers typically hope that respondents will identify the behavior of interest, search the reference period, retrieve all instances that match the target behavior, and finally count these instances to determine the overall frequency of the behavior. However, such a *recall-and-count* strategy is rarely feasible. Respondents usually need to rely on extensive inference and estimation

strategies to arrive at an answer; which strategy they use depends on the frequency, importance, and regularity of the behavior (e.g., Brown, 2002; Menon, 1993, 1994; Sudman et al., 1996).

Questions about rare and important behaviors can be answered on the basis of autobiographical knowledge or a recall-and-count strategy. When asked "*How often did you get divorced?*" most people know the answer without extended memory search. When asked *"How often did you relocate to another city?"* many people will not know immediately, but can compute an appropriate answer by reviewing their educational and job history, following a recall-and-count strategy. Respondents' task is more demanding when the behavior is frequent. High frequency of a behavior makes it unlikely that detailed representations of numerous individual episodes are stored in memory; instead, different instances blend into one global, knowledge-like representation that lacks specific time or location markers (see Linton, 1982; Strube, 1987). Frequent doctor visits, for example, result in a well-developed knowledge structure for the general event, allowing respondents to report in considerable detail on what usually goes on during their doctor visits. But the highly similar individual episodes become indistinguishable and irretrievable, making it difficult to report on any specific one. In these cases, respondents need to resort to estimation strategies to arrive at a plausible frequency report. Which estimation strategy they use depends on the regularity of the behavior and the context in which the frequency question is presented.

When the behavior is highly regular, frequency estimates can be computed on the basis of *rate information* (Menon, 1994; Menon, Raghubir, & Schwarz, 1995). Respondents who go to church every Sunday have little difficulty in arriving at a weekly or monthly estimate. However, exceptions are likely to be missed and the estimates are only accurate when exceptions are rare. A related strategy relies on *extrapolation from partial recall*. When asked how often she took pain medication during the last week, for example, a respondent may reason, "I took pain killers three times today, but this was a bad day. So probably twice a day, times 7 days, makes 14 times a week." The accuracy of this estimate will depend on the accuracy of the underlying assumptions, the regularity of the behavior, and the day that served as input into the chain of inferences.

Other estimation strategies may even bypass any effort to recall specific episodes. For example, respondents may simply rely on information provided by the research instrument itself. As an example, consider the *frequency scales* shown in Table 1. Consistent with the maxims of cooperative conversational conduct (Grice, 1975) respondents assume that the researcher constructed a meaningful scale that is relevant to their task (Schwarz, 1996). Presumably, the range of response alternatives reflects the researcher's knowledge about the distribution of the behavior, with values in the middle range of the scale corresponding to the "usual" or "average" behavior and values at the extremes of the scale corresponding to the extremes of the distribution. Drawing on these assumptions, respondents use the range of the

response alternatives as a frame of reference in estimating their own behavioral frequency. This results in higher frequency estimates when the scale presents high rather than low frequency response alternatives, as Table 1 illustrates.

Table 1 about here

Such scale-based estimation effects have been observed for a wide range of behaviors (for a review see Schwarz, 1996); they are more pronounced, the more poorly the respective behavior is represented in memory (Menon, Raghubir, & Schwarz, 1995). When behaviors of differential memorability are assessed, this can either exaggerate or cloud actual differences in the relative frequency of the behaviors, undermining comparisons across behaviors. Moreover, respondents with poorer memory are more likely to be influenced by frequency scales than respondents with better memory (e.g., Knäuper, Schwarz, & Park, 2004), which can undermine comparisons across groups. Finally, frequency scales also invite systematic underestimates of the variance in behavioral frequencies because all respondents draw on the same frame of reference in computing an estimate, resulting in reports that are more similar than reality warrants.

SELF-REPORTS OF FEELINGS

Feelings are subjective phenomena to which the person who has them has privileged access. While this does not imply that feelings are always easy to identify for the experiencer (see Clore, Conway, & Schwarz, 1994; Ellsworth & Scherer, 2003, for a discussion of different types of feelings and the underlying appraisal processes), most researchers consider the experiencer the final arbiter of what his or her feeling is. Unfortunately, that final arbiter is likely to tell us different things at different points in time and numerous studies documented profound discrepancies between people's concurrent and retrospective reports of emotions (for a comprehensive review see Robinson and Clore, 2002). This section reviews why this is the case, presents some illustrative biases, and highlights distinct patterns of convergence and divergence between prospective, concurrent, and retrospective reports as well as the choices people make (for further discussion of emotion measurement see Larsen & Augustine, this volume).

Accessibility Model of Emotion Report

To conceptualize the processes underlying emotion reports, Robinson and Clore (2002) proposed an accessibility model. When people report on their *current* feelings, the feelings themselves are accessible to introspection, allowing for accurate reports on the basis of experiential information. But affective experiences are fleeting and not available to introspection once the feeling dissipated. Accordingly, the opportunity to collect emotion reports that are based on introspective access is limited to methods of real-time data capture, like experience sampling (Stone et al., 1999; see also Larsen & Augustine, this volume). Once the feeling dissipated, the affective experiences need be reconstructed on the basis of

episodic or semantic information. When the report pertains to a *specific recent episode*, people can draw on episodic memory, retrieving specific moments and details of the recent past. Detailed episodic recall can often re-elicit a similar feeling (and is therefore a popular mood manipulation); it can also provide sufficient material for relatively accurate reconstruction. Hence, episodic reports often recover the actual experience with some accuracy, as indicated by convergence with concurrent reports (e.g., Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004; Robinson & Clore 2002; Stone, Schwartz, Schwarz, Schkade, Krueger, & Kahneman, 2006). One method that facilitates episodic reporting is the Day Reconstruction Method (DRM; Kahneman et al., 2004), discussed below. At present, it remains unclear how far in the past an episode can be to still allow reasonably accurate episodic reconstruction. Most likely the answer depends on the uniqueness and memorability of the episode, paralleling the above discussion of behavioral frequency reports.

In contrast episodic reports, *global* reports of past feelings are based on semantic knowledge. When asked how they "usually" feel during a particular activity, people draw on their general beliefs about the activity and its attributes to arrive at a report. The actual experience does not figure prominently in global reports because the experience itself is no longer accessible to introspection and episodic reconstruction is not used to answer a global question.

Extending this accessibility model of emotion report, Schwarz, Kahneman and Xu (2009; Xu & Schwarz, 2009) noted that the same semantic knowledge serves as a basis for *predicting* future feelings, for which episodic information is not available to begin with. Such predictions are usually more extreme than people's actual experiences (for a review see Wilson & Gilbert, 2003) because the predictor focuses on core attributes of the activity at the expense of other information, resulting in a "focusing illusion" (Schkade & Kahneman, 1997). For example, Midwesterners who predict how happy they would be if they moved to California may focus on the pleasant Californian climate, missing, for example, that they would still have to spend most of the day in an office cubicle. Finally, hedonic predictions play an important role in people's daily lives because they serve as input into *choice* (March, 1978; Mellers & McGraw, 2001) and influence which course of action people will or will not take.

Convergences and Divergences

The above rationale predicts a systematic pattern of convergences and divergences, which results directly from the inputs on which the respective reports are based. First, concurrent reports and retrospective reports pertaining to a specific and recent episode are likely to show good convergence, provided that the episode is sufficiently recent to allow detailed and vivid reinstantiation in episodic memory. Second, retrospective global reports of past feelings and predictions of future feelings are also likely to converge, given that both are based on the same semantic inputs. Third, choices are based on predicted hedonic

consequences, and hence converge with predictions and global memories. One unfortunate side-effect of these convergences is that people's global memories seem to "confirm" the accuracy of their predictions and the wisdom of their choices, thus impairing the opportunity to learn from experience (Schwarz & Xu, in press). However, fourth, concurrent and episodic reports will often diverge from prediction, choice, and global memories. As a result, different measures can paint very different pictures of a person's affective experience with the same situation, as a few examples may illustrate (see Schwarz et al., 2009, for a review).

How enjoyable are vacations?

Not surprisingly, people believe that vacations are very enjoyable and this belief shapes their predictions, choices, and global memories, even when their actual recent experience was less rosy. Assessing prospective, concurrent, and retrospective reports of vacation enjoyment, Mitchell and colleagues (1997) found that prospective reports converged with retrospective reports; however, both the predicted and remembered affect was more positive than the affect reported concurrently during the vacation. In a later study, Wirtz and colleagues (2003) tracked college students before, during, and after their spring-break vacations and compared their predicted, concurrent, and remembered affect. They found that predicted and remembered experiences were more intense (i.e., both more positive and more negative) than reported concurrently during the vacation. However, the (biased) remembered experience predicted the desire to repeat the vacation better than the actual experience, illustrating that we learn from our memories, not from our experiences.

How bad was that colonoscopy?

Particularly memorable examples of learning from memory rather than experience have been reported in the medical domain. For example, Redelmeier and Kahneman (1996) observed that retrospective evaluations of pain are dominated by two moments that may be of particular adaptive relevance (Fredrickson, 2000): the peak ("How bad does it get?") and the end ("How does it end?"). Other aspects, like the overall duration of pain, exert little influence. In fact, extending the duration of a colonoscopy by adding a few moments of discomfort at the end improves the overall evaluation of the episode by adding a better ending. It also improves the likelihood of future compliance, again highlighting how memory beats experience in predicting future behavior (Redelmeier, Katz, & Kahneman, 2003).

How much do parents enjoy spending time with their children?

Several decades ago, Juster and colleagues (1975) asked a representative sample of Americans to rate 28 activities from "dislike very much" (0) to "enjoy a great deal" (10). They found that activities with one's children consistently topped the list (ranks 1-4), whereas grocery shopping and cleaning the house were reported as least enjoyable (rank 27 and 28; Juster, 1985, p.336). In stark contrast to these reports, other studies indicate that parents' marital satisfaction drops when children arrive, reaches a life-time low when the children are teenagers, and recovers after the children leave the house (for a review see Argyle, 1999). Are

the children a more mixed pleasure than global reports of enjoyment convey? Close-in-time measures of affective experience, collected with the Day Reconstruction Method, suggest so. Specifically, 909 employed women in Texas recalled their activities during the preceding day and reported how they felt during each specific episode (Kahneman et al., 2004). In these episodic reports, activities coded as "taking care of my children" ranked just above the least enjoyable activities of the day, namely working, housework, and commuting; data from other samples replicated this pattern.

Several processes contribute to this divergence between global and episodic reports. First, global judgments of enjoyment are based on general beliefs ("I enjoy my kids"), which are often supported by belief-consistent memories of great vividness (like fond memories of shared activities). Yet most mundane episodes of a given day are less enjoyable than the episodes that make for fond memories. Second, activities are organized in memory by their focal features. Attempts to recall memories pertaining to one's interactions with one's children will therefore result in an overrepresentation of child focused activities, at the expense of numerous other episodes of the day in which the children were present. The reconstruction of a whole day in the DRM avoids many of these problems of selective recall and provides a fuller assessment of the affective impact of children throughout the day. Hence, the findings suggest that part of the reason that children seem more enjoyable in general than on any given day is simply that parents do not consider the full range of child-related time use when providing global reports. Finally, global reports are subject to higher social desirability pressures than episodic reports. A parent who reports, "I don't enjoy spending time with my children" is clearly a bad parent; but noting that "They were a pain last night" is perfectly legitimate.

Implications

Several methodological implications are worth emphasizing. Researchers who want to assess peoples' actual hedonic experiences should preferably do so with concurrent reports, using experience sampling methodologies (Stone et al., 1999). If this is not feasible, episodic reporting methods, like the Day Reconstruction Method (Kahneman et al., 2004), provide a less burdensome alternative that can capture the experience with some accuracy, provided that the relevant episodes are recent. In contrast, global reports of affect are theory-driven, not experience-driven. They capture respondents' beliefs about their experience rather than the experience itself and are subject to pronounced focusing effects.

However, people's behavioral choices are based on their *expected* hedonic consequences (March, 1978). These expectations converge with global memories, but often diverge from actual experience. Hence, predictions, choices, and global memories are poor indicators of experience. Yet when people make behavioral decisions, global memories and expectations are likely to figure prominently in the information they consider. Ironically, this turns faulty indicators of experience into good predictors of future choices and behaviors (e.g., Wirtz et al., 2003). It also suggests that optimizing a study for accurate description of what

people do and feel does not optimize it for accurate prediction of what they will do next (and vice versa) – description and prediction are different goals and their optimization requires different strategies.

An Example of Episodic Reconstruction: The Day Reconstruction Method

The Day Reconstruction Method (Kahneman et al., 2004) is designed to collect data that describe a person's time use and affect on a given day through a systematic reconstruction conducted on the following day. In a self-administered questionnaire, respondents first reinstantiate the previous day into working memory by producing a short diary consisting of a sequence of episodes, usually covering the time from when they got up to when they went to bed. The diary's format draws on insights from cognitive research with Event History Calendars (Belli, 1998) and facilitates retrieval from autobiographical memory through multiple pathways. Its episodic reinstantiation format attenuates biases commonly observed in retrospective reports (Robinson & Clore, 2002; Schwarz & Sudman, 1994). Respondents' diary entries are confidential and the diary is not returned to the researcher, which allows respondents to use idiosyncratic notes, including details they may not want to share.

Next, respondents draw on their diary to answer a series of questions about each episode, including (1) when the episode began and ended, thus providing time use data; (2) what they were doing; (3) where they were; (4) whom they were interacting with; and (5) how they felt, assessed on multiple affect dimensions. The details of this response form can be tailored to the specific issues under study; only this form is returned to the researcher for analysis. For methodological reasons, it is important that respondents complete the diary before they are aware of the specific content of the later questions about each episode. Early knowledge of these questions may affect the reconstruction of the previous day and may introduce selection bias. The DRM can be administered individually or in group settings and respondents can report on a complete day in 45 to 75 minutes. DRM reports have been validated against experience sampling data and Krueger, Kahneman, Schkade, Schwarz, and Stone (2009) provide a comprehensive review of the methodology and available findings.

SELF-REPORTS OF ATTITUDES: EVALUATION IN CONTEXT

Another common type of self-report question asks people to report on their likes and dislikes. Psychologists commonly assume that these reports reflect a predisposition to evaluate some object in a favorable or unfavorable manner; this predisposition is referred to as an *attitude* (Eagly & Chaiken, 1993, 2005). Attitudes are hypothetical constructs that cannot be directly observed and need to be inferred from individuals' responses to the attitude object. As Gordon Allport (1935, p. 836) put it, "How does one know that attitudes exist at all? Only by necessary inference. *There must be something* to account for the consistency of conduct" (italics added). From this perspective, it is not surprising that attitude questions are often asked without reference

to any specific context – what makes the construct appealing is exactly the promise of predictive power across contexts. Empirically, attitude research never delivered on this promise. In an early review of attitude-behavior consistency, Wicker (1969, p. 65) concluded that "only rarely can as much as 10% of the variance in overt behavioral measures be accounted for by attitudinal data." Even the attitude reports themselves proved highly malleable and minor variations in question wording, question order or response format can elicit profound shifts in reported attitudes, even on familiar and important topics (for early examples see Cantril, 1944; Payne, 1951; for reviews see Schwarz, 1999; Schwarz, Groves, & Schuman, 1998; Schuman & Presser, 1981). Attempts to overcome these disappointments took one of two general paths; one focused on improving measurement of the attitude itself and the other on improving the predictive power of the attitude measure by taking context variables into account.

Stalking the "True" Attitude

Mirroring Campbell's (1981) convictions, many researchers assumed that context effects on attitude reports and low attitude-behavior consistency can be traced to participants' hesitation to report their true feelings "with candor and accuracy". This focused efforts on attempts to reduce respondents' self-presentation concerns (e.g., techniques that ensure respondents' anonymity and the confidentiality of their answers; see Bradburn et al., 2004, for recommendations) or to convince them that "lying" was futile – thanks to sophisticated machinery, the researcher would learn their "true" attitude anyway (e.g., Jones and Sigall's, 1971, "bogus pipeline"). Empirically, such techniques have been found to increase the frequency of socially undesirable answers. For example, people are more likely to admit that they enjoy pornography when they cannot be identified as the source of the answer (Himmelfarb & Lickteig, 1982) and White participants are more likely to report that they dislike African Americans under bogus pipeline conditions (e.g., Allen, 1975). However, external validation of the reports is not available and the procedures themselves may invite correction of one's spontaneous answer in light of the concern about bias that is clearly conveyed.

Whereas these developments assumed that people know their own attitudes but may not want to report them, later developments considered the possibility that people may sometimes not be aware of their own attitudes or may not even want to "admit" them to themselves. Implicit measures of attitudes address this possibility (for overviews see the contributions in Wittenbrink & Schwarz, 2007). These procedures range from evaluative and conceptual priming techniques (for a review see Wittenbrink, 2007) and response competition procedures (e.g., the IAT; for a review see Lane, Banaji, Nosek, & Greenwald, 2007) to low-tech paper-and-pencil measures (e.g., word completion tasks; for a review see Vargas, Sekaquaptewa, & von Hippel, 2007). To many researchers' disappointment, implicit measures did not deliver the robust, context-independent assessment of attitudes that theorists have long hoped for. To the contrary, implicit measures of attitudes are subject to the same context

effects that have been observed with explicit self-reports (for extensive reviews see Blair, 2002; Ferguson & Bargh, 2007). For example, Dasgupta and Greenwald (2001) found that exposure to pictures of liked African Americans and disliked European Americans resulted in shifts on a subsequent IAT that paralleled previously observed effects of exposure to liked or disliked exemplars on explicit measures of attitudes (e.g., Bodenhausen, Schwarz, Bless, & Wänke, 1995). Similarly, Wittenbrink, Judd, and Park (2001) found that the same Black face primes elicited more negative automatic responses when the faces were presented on the background of an urban street scene rather than a church scene. Moreover, automatic evaluations have also been obtained for novel objects, for which no previously acquired object-attitude links could have been stored in memory (e.g., Duckworth, Bargh, Garcia, & Chaiken, 2002).

Such findings make it unlikely that implicit measures provide a "bona fide pipeline" (Fazio, Jackson, Dunton, & Williams, 1995) to people's true and enduring attitudes, formed on the basis of past experience and stored in memory as object-evaluation associations. However, the findings are fully compatible with an alternative conceptualization of attitudes as evaluations in context (for variants on this theme see Ferguson & Bargh, 2007; Lord & Lepper, 1999; Schwarz, 2007).

Attitude Construal: Evaluation in Context

As William James (1890, p. 333) observed, "My thinking is first and last and always for the sake of my doing." Few psychologists doubt this truism, but even fewer heed its implications. To serve action in a given context, any adaptive system of evaluation should be informed by past experience, but highly sensitive to the specifics of the present. It should overweight recent experience at the expense of more distant experience, and experience from similar situations at the expense of experience from dissimilar situations. In addition, it should take current goals and concerns into account to ensure that the assessment is relevant to what we attempt to do now, in this context. In short, only *context-sensitive* evaluation can guide behavior in adaptive ways by alerting us to problems and opportunities when they exist; by interrupting ongoing processes when needed (but not otherwise); and by rendering information highly accessible that is relevant now, in this situation. From this perspective, it is no coincidence that any list of desirable context sensitivities reads like a list of the conditions that give rise to context effects in attitude judgment (Schwarz, 1999; 2007).

Close attention to context also improves the predictive value of attitude reports as reflected in increased attitude-behavior consistency. This was first highlighted in the seminal work of Fishbein and Ajzen (1975), who considered it a measurement issue, not a conceptual issue. However, the underlying principle follows directly from attitude construal models: an evaluation reported at time 1 will map onto an evaluation or behavioral decision at time 2 to the extent that the person draws on the same inputs at both points in time. This *matching principle* (Lord & Lepper, 1999) offers a coherent conceptualization of the conditions of

stability as well as change in attitude reports and predicts when attitude judgments will or will not relate to later behavioral decisions (for reviews see Lord & Lepper, 1999; Schwarz, 2007). Numerous variables – from the person's current goals to the nature of the context and the frequency and recency of previous exposure – can influence the temporary construal of the attitude object and hence the observed consistencies and inconsistencies across time and contexts.

Taking the Actor's Perspective

Construal models of attitudes are compatible with broader current developments in psychological science, most notably our increasing understanding of the situated and embodied nature of cognition, emotion, and motivation (for recent reviews see Barsalou, 2005; Niedenthal et al., 2006; and the contributions in Mesquita et al., 2010). But much as social psychologists would expect, construal models lack the intuitive appeal of dispositional attitude models. After all, the logic of dispositional models is fully compatible with observers' robust preference for dispositional rather than situational explanations, also known as the "fundamental attribution error" (Ross, 1977). In contrast, construal models emphasize the role of contextual variables, which are usually more attended to by the actor (Jones & Nisbett, 1971), who benefits from the context-sensitivity of evaluation in situ. From this perspective, Allport's (1935) hope that enduring attitudes can account for an actor's "consistency of conduct" in the present is an observer's dream, but an actor's nightmare. After decades of conducting attitude research predominantly from the perspective of an observer who tries to predict an actor's behavior, the increasing interest in how people live and experience their lives on a moment-to-moment basis may contribute to a more systematic exploration of evaluation-in-context from an actor's perspective (see also Mehl & Robbins, this volume).

CODA: QUESTIONS IN SITU

As this selective discussion of the complexities of self-report indicates, retrospective questions often ask respondents for information that they cannot provide with any validity, as discussed in the sections on self-reports of behaviors and feelings. Other questions ask for generic answers that may be incompatible with the contextualized and situated nature of human experience. In the case of attitude measurement, much of the appeal of the enterprise rests on the hope of predicting behavior across contexts, leading researchers to discount the context sensitivity of evaluative judgment as undesirable noise. Methods of real-time or close-in-time measurement attenuate these problems by assessing information in situ, thus allowing (at least potentially) for the simultaneous assessment of contextual and experiential variables, and by posing more realistic tasks in the form of questions about respondents' *current* behavior, experiences, and circumstances. These are promising steps.

At the same time, asking questions in situ raises new self-report issues, which have so far received limited attention. Central to these new issues is the high density of most real-time

data capture procedures, which require respondents to answer the same questions multiple times within a relatively short time. As noted in the section on question comprehension, this introduces conversational issues of nonredundancy (Grice, 1975; Schwarz, 1994) that may invite an emphasize on what is unique and new in each episode at the expense of attention to what is shared across episodes and has therefore already been reported earlier, making its repetition a violation of conversational norms. Similarly, rating many episodes along the same scale invites attention to the frequency principle (Parducci, 1965) of rating scale use, eliciting differentiation in the reports that may exceed differences in experience. Moreover, repeated ratings make it likely that previous related episodes are still accessible and serve as scale anchors or comparison standards. In most cases, these recent anchors would be less extreme than the "memorable" events used to anchor rating scales in one-time ratings. If so, a given episode would be rated as more intense in real-time assessment, where it is evaluated against a less extreme anchor, than in retrospective assessment, where it is evaluated against a more distant "memorable" episode. The cognitive and communicative processes underlying real-time self-reports require the systematic exploration and experimentation that advanced the understanding of self-reports in other domains (Schwarz & Sudman, 1996; Sudman et al., 1996). Without such work, we run the risk of merely replacing known biases with unknown ones.

Finally, advocates of real-time measurement do probably not appreciate the conclusion that accurate assessments of real-time experience are poorer predictors of future behavioral choices than faulty memories of the same experience (e.g., Kahneman et al., 1993; Redelmeier et al, 2003; Wirtz et al., 2003). As one reviewer of this chapter put it, "Why should we even bother measuring experience if global or retrospective assessments are the 'better' predictors of choice?" The answer is simple: there's more to behavioral science than the observer's desire to predict others' choices. A full understanding of the human experience requires attention to the actor's perspective and insight into how people live and experience their lives. Real-time measurement in situ is ideally suited to illuminate the dynamics of human experience from the actor's perspective, balancing decades of research that privileged the observer's goals.

REFERENCES

Allen, B. P. (1975). Social distance and admiration reactions of "unprejudiced" Whites. *Journal of Personality, 43,* 709-726.

Allport, G. W. (1935). Attitudes. In C. Murchison (ed.), *A handbook of social psychology* (pp. 789-994). Worcester, Mass: Clark University Press.

Argyle, M. (1999). Causes and correlates of happiness. In D. Kahneman, E. Diener, and N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 353-373). New York: Russell-Sage.

Barsalou, L. W. (2005). Situated conceptualization. In H. Cohen & C. Lefebvre (eds.), *Handbook of categorization in cognitive science* (pp. 619-650). Amsterdam: Elsevier.

Belli, R. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6,* 383-406.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.

Bodenhausen, G. V., Schwarz, N., Bless, H., & Wänke, M. (1995). Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals? *Journal of Experimental Social Psychology, 31*, 48-63.

Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions* (2nd ed.). San Francisco: Jossey Bass.

Brown, N. R. (2002). Encoding, representing, and estimating event frequencies: Multiple strategy perspective. In P. Sedlmeier & T. Betsch (Eds*.), Frequency processing and cognition* (pp. 37-54) New York: Oxford University Press.

Campbell, A. (1981). *The sense of well-being in America*. New York: McGraw-Hill.

Cantril, H. (1944) *Gauging public opinion*. Princeton, N.J.: Princeton University Press.

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: Russel Sage.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer, & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed.; Vol. 1, pp. 323-418). Hillsdale, NJ: Erlbaum.

Collins, L.M., Graham, J.W., Hansen, W.B., & Johnson, C.A. (1985). Agreement between retrospective accounts of substance use and earlier reported substance use. *Applied Psychological Measurement, 9*, 301 - 309.

Daamen, D. D.,L., & de Bie, S. E. (1992). Serial context effects in survey items. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 97-114). New York: Springer Verlag.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of liked and disliked individuals. *Journal of Personality*

and Social Psychology, 81, 800-814.

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.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 13*, 513 - 519.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth: Harcourt, Brace, Jovanovich.

Eagly, A.H., & Chaiken, S. (2005). Attitude research in the 21st century: The current state of knowledge. In D. Albarracin, B. T. Johnson, & M. P. Zanna (eds.), *The handbook of attitudes* (pp. 743-768). Mahwah, NJ: Erlbaum.

Eich,E., Reeves, J.L., Jaeger, B., & Graff-Radford, S.B. (1985). Memory for pain: Relation between past and present pain intensity. *Pain, 23*, 375-380.

Ellsworth, P.C., & Scherer, K.R. (2003). Appraisal processes in emotion. In R.J. Davidson, K.R. Scherer, & H. H. Goldsmith (eds.), *Handbook of affective sciences* (pp. 572-595). New York: Oxford University Press.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes. A bona fide pipeline? J*ournal of Personality and Social Psychology, 69,* 1013-1027.

Ferguson, M. J., & Bargh, J. A. (2003). The constructive nature of automatic evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp.169-188). NJ: Erlbaum.

Ferguson, M. J., & Bargh, J.A. (2007). Beyond the attitude object: Automatic attitudes spring from object-centered-contexts. In B. Wittenbrink & N. Schwarz (eds.), *Implicit measures of attitudes: Progress and controversies*. New York: Guilford.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fredrickson, B.L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion, 14*, 577-606.

Gibbs, R. W. (2006). *Embodiment and cognitive science.* New York: Cambridge University Press.

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

Himmelfarb, S., & Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology, 43,* 710-717.

James, W. (1890). *The principles of psychology* (Vol. 2). New York: Henry Holt.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349-364.

Juster, F. T. (1985). Preferences for work and leisure. In F. T. Juster & F. P. Stafford (eds.), *Time,*

goods, and well-being (pp. 397-414). Ann Arbor: Institute for Social Research.

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., & Redelmeier, D. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4,* 401-405.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The Day Reconstruction Method (DRM). *Science*, 306, 1776-1780.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *Science, 312,* 1908-1910.

Knäuper, B., Schwarz, N., & Park, D.C. (2004). Frequency reports across age groups: Differential effects of frequency scales. *Journal of Official Statistics, 20,* 91-96.

Krosnick, J.A. & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly, 51,* 201-19.

Krueger, A., Kahneman, D., Schkade, D., Schwarz, N., & Stone, A. (2009). National time accounting: The currency of life. In A. Krueger (ed.), *Measuring the subjective well-being of nations: National accounts of time use and well-being* (pp. 9-86). Chicago: Chicago University Press.

Lane, K.A., Banaji, M.R., Nosek, B.A., & Greenwald, A.G. (2007). Understanding and using the Implicit Association Test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (eds.), *Implicit measures of attitudes: Progress and controversies*. New York: Guilford.

Linton, S. J., & Melin, L. (1982). The accuracy of remembering chronic pain. *Pain, 13*, 281-285.

Loewenstein, G. F. & Adler, D. (1995). A bias in the prediction of tastes. *Economic Journal,* 105, 929-937.

Loewenstein, G., & Schkade, D. A. (1999). Wouldn't it be nice? Predicting future feelings. In D. Kahneman, E. Diener, and N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 85-105). New York: Russell-Sage.

Lord, C.G., & Lepper, M. R. (1999). Attitude representation theory. *Advances in Experimental Social Psychology, 31*, 265-343.

March, J. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics*, 9, 587-608.

Mellers, B. A., & McGraw, A. P. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science, 10,* 210-214.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research, 20,* 431-440.

Menon, G. (1994). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman, S. (Eds.) (1994). *Autobiographical memory and the validity of retrospective reports* (pp. 161- 172). New York: Springer Verlag.

Menon, G., Raghubir, P., & Schwarz, N. (1995). Behavioral frequency judgments: An

accessibility-diagnosticity framework. *Journal of Consumer Research, 22*, 212-228.

Mesquita, B., Barrett, L.F., & Smith, E.R. (Eds.) (2010). *The mind in context*). New York: Guilford

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view." *Journal of Experimental Social Psychology*, 33, 421-448.

Niedenthal, P. M., Barsalou, L.W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2006). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9,* 184-211.

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.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72*, 407-418.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Payne, S. L. (1951). *The art of asking questions*. Princeton: Princeton University Press.

Redelmeier, D., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain, 66*, 3-8.

Redelmeier D.A., Katz, J., and Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain, 104*, 187-194.

Robinson, J. P. (1977). *How Americans use time. A social-psychological analysis of everyday behavior.* New York: Praeger.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*, 934-960.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology, 10*, 173-220.

Ross, M. (1989). The relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341-357.

Schkade, D. A., & Kahneman, D. (1997). Does living in California make people happy?: A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9, 340-346.

Schuman, H. (1983). Survey research and the fundamental attribution error. *Personality and Social Psychology Bulletin, 9,* 103-104.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology, 26*, 123-162.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and*

the logic of conversation. Hillsdale, NJ: Erlbaum.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Schwarz, N. (2002). Situated cognition and the wisdom of feelings: Cognitive tuning. In L. Feldman Barrett & P. Salovey (Eds.), *The wisdom in feelings* (pp.144-166). New York: Guilford.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition, XX*

Schwarz, N., Groves, R., & Schuman, H. (1998). Survey methods. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social   psychology* (4th edition, Vol. 1, pp. 143-179). New York: McGraw-Hill.

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, UK: Wiley.

Schwarz, N., & Hippler, H. J. (1995). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly, 59*, 93-97.

Schwarz, N., Kahneman, D., & Xu, J. (2009). Global and episodic reports of hedonic experience. In R. Belli, D. Alwin, & F. Stafford (eds.), *Using calendar and diary methods in life events research* (pp. 157-174). Newbury Park, CA: Sage.

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., Oyserman, D., & Stich, C. (2008). The psychology of asking questions. In E. de Leeuw, J. Hox, & D. Dillman (eds.), *International handbook of survey methodology* (pp. 18-34). New York: Taylor & Francis.

Schwarz, N. & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication and questionnaire construction. *American Journal of Evaluation, 22*, 127-160.

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.) (1994). *Autobiographical memory and the validity of retrospective reports.* New York, NY: Springer Verlag.

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

Schwarz, N., & Xu, J. (in press). Why don't we learn from poor choices? The consistency of expectation, choice, and memory clouds the lessons of experience. *Journal of Consumer Psychoology.*

Smith, E.R., & Conrey, F.R. (2007). Mental representations are states not things: Implications for

implicit and explicit measurement. In B. Wittenbrink & N. Schwarz (eds.), *Implicit measures of attitudes: Progress and controversies*. New York: Guilford.

Smith, E. R., & Semin, G. R. (2004). Socially situated cognition: Cognition in its social context. *Advances in Experimental Social Psychology, 36*, 53-117.

Stone, A. A., Schwartz, J. E., Schwarz, N., Schkade, D., Krueger, A., & Kahneman, D. (2006). A population approach to the study of emotion: Diurnal rhythms of a working day examined with the Day Reconstruction Method (DRM). *Emotion, 6,* 139-149.

Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). In D. Kahneman, E. Diener, & N. Schwarz (eds.), *Well-being: The foundations of hedonic psychology* (pp. 26-39). New York: Russell-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., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Vargas, P.T., Sekaquaptewa,D., & von Hippel, W. (2007). Armed only with paper and pencil: "Low tech" measures of implicit attitudes. In B. Wittenbrink & N. Schwarz (eds.), *Implicit measures of attitudes: Progress and controversies* (pp. 103-122). New York: Guilford.

Wicker, A.W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25,* 41–78.

Willis, G. B. (2004). *Cognitive interviewing. A tool for improving questionnaire design*. Thousand Oaks: Sage.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345-411). San Diego, CA: Academic Press.

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.

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, 14. 520-524.

Withey, S. B. (1954). Reliability of recall of income. *Public Opinion Quarterly, 18*, 31 - 34.

Wittenbrink. B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (eds.), *Implicit measures of attitudes: Progress and controversies* (pp. 17-58). New York: Guilford.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815-827.

Wittenbrink, B., & Schwarz, N. (eds.) (2007). *Implicit measures of attitudes: Procedures and controversie*s. New York: Guilford.

Xu, J., & Schwarz, N. (2009). Do we really need a reason to indulge? *Journal of Marketing Research, 46,* 25-36*.*

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

Reported Daily TV Consumption

Low Frequency Alternatives		High Frequency Alternatives	
Up to 1/2 h	7.4%	Up to 2 1/2h	62.5%
1/2 h to 1h	17.7%	2 1/2h to 3h	23.4%
1h to 1 1/2h	26.5%	3h to 3 1/2h	7.8%
1 1/2h to 2h	14.7%	3 1/2h to 4h	4.7%
2h to 2 1/2h	17.7%	4h to 4 1/2h	1.6%
More than 2 1/2h	16.2%	More than 4 1/2h	0.0%

Note. N = 132. Adapted from Schwarz, N., Hippler, H.J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly, 49*, 388-395. Reprinted by permission.