COGNITION, AGING, AND SELF-REPORTS

Norbert Schwarz
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106-1248, USA
NSchwarz@umich.edu

Bärbel Knäuper
Natalie O. Rosen
Dept. of Psychology
McGill University
1205 Dr. Penfield Avenue
Montreal, Quebec, H3A 1B1

Revised 7/2005

To appear in
D. Park & N. Schwarz, (Eds.),
Cognitive Aging -- A Primer. 2nd edition

The reported research was supported by grant AG14111 from the National Institute of Aging to N. Schwarz, D. Park, and B. Knäuper.
Much of what we know about age-related differences in individuals' behaviors and opinions is based on self-reports obtained from older and younger respondents. Unfortunately, self-reports are a fallible source of data and minor changes in question wording, question format, or question order can profoundly influence the answers that research participants provide (for reviews see Schwarz, 1999a; Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000). Complicating things further, a growing body of research suggests that older and younger respondents may be differentially affected by features of the research instrument (Schwarz, Park, Knäuper, & Sudman, 1999). Hence, it is often difficult to determine whether age-related differences in self-reports reflect (a) age-related differences in respondents' actual attitudes or behaviors, (b) age-related differences in the emergence of context effects or (c) some unknown combination of both.

This chapter introduces readers to the cognitive and communicative processes underlying self-reports. We highlight how age-related changes in cognitive and communicative functioning influence these processes and illustrate how age-sensitive context effects can invite misleading substantive conclusions.

**RESPONDENTS' TASKS**

From a cognitive perspective, answering a question posed in a research setting requires that respondents perform several tasks. First, they need to interpret the question to understand what is being asked of them. If the question is an opinion question, they may either retrieve a previously formed opinion from memory, or they may form an opinion on the spot. While researchers typically hope for the former, the latter is far more likely. Even when respondents have previously formed a judgment accessible in memory, the previous judgment is unlikely to match the specifics of the question asked, forcing respondents to compute a new judgment. To do so, they need to retrieve
relevant information from memory to form a mental representation of the target that they are to evaluate. In most cases, they will also need to retrieve or construct some standard against which the target is evaluated. Once a "private" judgment is formed in their mind, respondents have to communicate it to the researcher. To do so, they may need to format their judgment to fit the response alternatives provided as part of the question. Moreover, respondents may wish to edit their response before they communicate it, due to influences of social desirability and situational adequacy.

Similar considerations apply to behavioral questions. Again, respondents first need to understand what the question refers to, and which behavior they are supposed to report on. Next, they have to recall or reconstruct relevant instances of this behavior from memory. If the question specifies a reference period, they must determine if the recalled instances occurred during this reference period or not. Similarly, if the question refers to their "usual" behavior, respondents have to determine if the recalled or reconstructed instances are reasonably representative, or if they reflect a deviation from their usual behavior. If they cannot recall or reconstruct specific instances of the behavior, or are not sufficiently motivated to engage in this effort, respondents may rely on their general knowledge or relevant contextual information to compute an estimate. Finally, respondents have to communicate their estimate to the researcher. They may need to map their estimate onto a response scale provided to them, and they may want to edit it for reasons of social desirability.

Accordingly, interpreting the question, generating an opinion or a representation of the relevant behavior, formatting the response, and editing the answer are the main psychological components of a process that starts with respondents' exposure to a survey question and ends with their overt report (Strack & Martin, 1987; Tourangeau, 1984). We first consider issues of question
comprehension and subsequently return to respondents’ other tasks in the context of attitude and behavior questions.

**QUESTION COMPREHENSION**

The key issue at the question comprehension stage is whether the respondent's understanding of the question does or does not match what the researcher had in mind. From a psychological point of view, question comprehension reflects the operation of two intertwined processes. The first refers to the semantic understanding of the utterance. Comprehending the literal meaning of a sentence involves the identification of words, the recall of lexical information from semantic memory, and the construction of meaning from the utterance, which is constrained by its context. Not surprisingly, methodology textbooks urge researchers to write simple questions and to avoid unfamiliar or ambiguous terms (e.g., Sudman & Bradburn, 1983). However, understanding the words is not sufficient to answer a question. For example, when respondents are asked, "What have you done today?", they are likely to understand the meaning of 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? Hence, understanding a question in a way that allows an appropriate answer requires not only an understanding of the literal meaning of the question, but involves inferences about the questioner's intention to determine the pragmatic meaning of the question.

To infer the pragmatic meaning of a question, respondents draw on contextual information, including the content of preceding questions and the nature of the response alternatives presented by the researcher. Their use of this information is licensed by the tacit assumptions that govern the conduct of conversation in daily life, as described in Grice's (1975) logic of conversation. These
tacit assumptions entail that all contributions of a speaker are relevant to the goals of the ongoing conversation, unless indicated otherwise. In a research setting, the researcher's contributions include the content of preceding questions as well as apparently formal features of the questionnaire, rendering them a relevant source of information for respondents (for comprehensive reviews see Clark & Schober, 1992; Schwarz, 1996). Respondents are particularly likely to draw on such contextual information when no interviewer is available to clarify the meaning of an ambiguous question, as is the case when data are collected with a self-administered questionnaire.

**Formal Features of Questionnaires**

**Open vs. closed question formats**

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

**Reference periods**

Suppose respondents are asked how often they have been “angry” during some specified time period. To answer this question, they need to determine what the question refers to: only major
episodes of anger or any minor irritation? One relevant clue is the length of the reference period. As Winkielman, Knäuper, and Schwarz (1998) observed, respondents infer that the question pertains to major anger episodes when an extended reference period (such as “last year”) is presented, but to minor irritations when a short reference period (such as “last week”) is presented. After all, they can’t be expected to remember minor irritations over the course of a full year, whereas major anger is too rare to make a period of one week meaningful. This differential interpretation of the question is reflected in respondents’ examples of “typical” anger experiences as well as in their frequency reports. Most importantly, respondents reported a lower frequency of anger for the one-year period than would be expected on the basis of their reports for a one-week period. Taken by itself, this observation may simply reflect that respondents forgot some distant anger episodes. Yet, the differential extremity of their examples indicates that forgetting is only part of the picture. Instead, respondents deliberately reported on differentially intense and frequent types of anger and this meaning shift contributed to their differential frequency reports.

**Numeric values of rating scales**

Even something as “technical” as the numeric values of rating scales can influence respondents’ question interpretation. For example, Schwarz and Hippler (1995a; see also Schwarz, Knäuper, Hippler, Noelle-Neumann, & Clark, 1991) asked respondents to evaluate politicians along an 11-point rating scale, ranging from "don't think very highly of this politician" (0 or -5) to "think very highly of this politician" (11 or +5). To answer this question, they have to determine the meaning of "don't think very highly of this politician" -- does this imply the absence of positive thoughts or the presence of negative thoughts? In making this judgment, respondents draw on the numeric values of the rating scale, inferring that the label pertains to the presence of negative
thoughts ("I have unfavorable thoughts about him.") when accompanied by the numeric value -5, but to the absence of positive thoughts ("I have no particularly favorable thoughts about him.") when accompanied by the numeric value 0. These differential interpretations of the verbal scale anchor are reflected in markedly different ratings: Whereas only 29.3 percent reported a rating below the midpoint along the -5 to +5 scale, 40.2 percent did so along the 0 to 10 scale.

In combination, the above findings demonstrate that respondents draw on formal features of the questionnaire in interpreting the meaning of a question. In doing so, they proceed on the tacit assumption that every contribution is relevant to the aims of the ongoing conversation. In research situations, the researcher's contributions include apparently “technical” features of questionnaire design, such as the numeric values given on a rating scale. Hence, identically worded questions may acquire different meanings, depending on presumably neutral, technical characteristics of the questionnaire (see Schwarz, 1996, for a more extended discussion).

**Age-related Differences.**

Are these processes likely to be affected by age-related changes in cognitive functioning? On the one hand, we can safely assume that older respondents share the tacit assumptions that underlie the conduct of conversation and are hence likely to draw on formal features of the questionnaire in much the same way as younger respondents. On the other hand, using these features to disambiguate the meaning of a question requires that respondents relate the text presented in the body of the question to the accompanying response alternatives, potentially requiring considerable cognitive resources. Given age-related decline in cognitive resources (see Park, this volume), older respondents may therefore be less likely to arrive at an interpretation that reflects the integration of question wording and response alternatives.
At present, the available data show no age-related differences in the impact of open versus closed question formats, frequency scales and reference periods. We did, however, observe differential attention to the numeric values of rating scales. In a replication of the above experiment on the numeric values of rating scales, younger and older respondents were asked to rate Bob Dole, the Republican candidate in the 1996 U.S. Presidential elections (Schwarz et al., 1998). Replicating the previous findings, younger respondents rated Bob Dole more positively when the verbal label "don't think highly of this politician" was accompanied by the numeric value -5 rather than the numeric value 0. Respondents age 70 and older, however, provided the same ratings independent of the type of numeric values offered. This result suggests that they did not draw on the numeric values in interpreting the meaning of the verbal labels, presumably because their limited cognitive resources did not allow them to relate the wording of the question and the numeric values to one another. Supporting this interpretation, we observed that the impact of numeric values varied as a function of respondents' reading span, a measure of cognitive resource. As expected, respondents high in cognitive resource related the text to the numeric values, whereas respondents low in cognitive resource did not.

Based on these limited findings, we conjecture that older and younger adults are similarly influenced by characteristics of the question stem. Older adults, however, may be less likely to consider contextual features, like the numeric values of rating scales, which do not show up in the question stem. This generalization awaits testing.

**ANSWERING ATTITUDE QUESTIONS**

Attitude questions ask respondents to report an evaluative judgment. In most cases,
Self-reports -- p. 9

evaluative judgments require two mental representations: a representation of the attitude object itself and a representation of a standard, against which the object is evaluated. Both representations are formed on the spot, based on information that is temporarily or chronically accessible. Whereas chronically accessible information lends some stability to attitude judgments, temporarily accessible information is a major source of context effects. How accessible information influences the judgment depends on how it is used. Information used to form a representation of the attitude object results in assimilation effects. That is, the object is evaluated more positively the more positive information comes to mind and is included in its representation. Information used to form a representation of the standard, on the other hand, results in contrast effects. That is, the more positive information is included in the representation formed of the standard, the less positively the attitude object is evaluated relative to this standard. Schwarz and Bless (1992) and Sudman et al. (1996, chapters 4-5) provide detailed reviews of these mental construal processes. Here we focus on two key issues, namely the emergence of question order effects and response order effects, both of which show pronounced age differences.

**Question Order Effects**

As many psychological experiments documented, individuals are unlikely to retrieve all information that may potentially bear on a judgment, but truncate the search process as soon as enough information has come to mind to form a judgment with sufficient subjective certainty (see Bodenhausen & Wyer, 1987). Accordingly, their judgments are profoundly influenced by the information that is most accessible at the time. This is usually information that has been used most recently, e.g. for the purpose of answering a preceding question. For example, Strack, Martin, and Schwarz (1988) asked students to report their general life-satisfaction as well as their dating
frequency in different orders. When the life-satisfaction question preceded the dating question, the answers were uncorrelated, $r = -0.1$, suggesting that dating plays a minor role in students’ well-being. When the question order was reversed, however, the correlation increased to $r = +0.62$, suggesting that dating is a major determinant of life-satisfaction. Findings of this type illustrate the contextual nature of mental construal processes: People can draw on many different aspects in forming a representation of the attitude object “my life” and which aspects they consider depends on what happens to come to mind at the time, which is often a function of haphazard influences.

To guard against such order effects, survey researchers often introduce buffer items to separate related questions, thus attenuating the accessibility of previously used information. By the same token, we may assume that age-related declines in memory attenuate question order effects because they render it less likely that previously used information is still accessible. The available data support this prediction (Knäuper, Schwarz, Park, & Frisch, 2004), even in the case of otherwise highly robust question order effects. One such effect was identified by Schuman and Presser (1981), who asked respondents if a pregnant woman should be able to obtain a legal abortion "if she is married and does not want any more children" (Question A) or "if there is a strong chance of a serious defect in the baby" (Question B). Not surprisingly, respondents are more likely to support legal abortion when there is a risk of “a serious defect in the baby” than when a married woman merely “does not want any more children.” More important, support for Question A decreases when Question B is presented first: This is a case of a contrast effect, where the preceding question about birth defects introduces a standard of comparison relative to which merely "not wanting any more children" appears less legitimate.

Secondary analyses of Schuman and Presser’s (1981) data show that this question order
effect results in a difference of 19.5 percentage points for younger respondents, but decreases with age and is no longer observed for respondents aged 65 and older, as shown in Figure 1. Note that we would arrive at different conclusions about cohort differences, depending on the order in which the questions were asked. When the "no more children" question (A) is asked first, support for abortion decreases with age, suggesting that older respondents hold more conservative attitudes toward abortion. Yet no age difference can be observed when the same question is preceded by the "birth defect" question.

Figure 1

Subsequent laboratory experiments (Knäuper et al., 2004) indicated that the attenuation of question order effects among older respondents is due to age-related declines in working memory capacity, as assessed with a reading span test. Specifically, younger respondents (age 20 to 40 years) as well as older respondents (age 60 to 100 years) with good working memory capacity showed the familiar question order effect, whereas older respondents with poor working memory capacity did not. A related experiment with different questions replicated this pattern.

In combination, these findings suggest that question order effects decrease with age and that this decrease can be traced to age-related declines in memory function, which make it less likely that previously used information remains accessible. In contrast to this conclusion, however, Tourangeau, Singer, and Presser (2003) failed to observe reliable age differences on questions pertaining to confidentiality and privacy concerns. Moreover, the attenuation of question order effects observed above may be limited to the relatively uninvolving questions typical for public opinion surveys. When the questions are of high personal relevance or emotionally involving, on the other hand, older adults’ inhibition problems (see Hasher & Zacks, 1988) may exaggerate question
order effects. This possibility remains to be tested.

**Question Order and Rating Scale Use**

In addition to bringing information to mind that respondents may otherwise not consider, preceding questions can also influence respondents' use of rating scales (e.g., Ostrom & Upshaw, 1968; Parducci, 1983). Specifically, respondents use the most extreme stimuli to anchor the endpoints of a rating scale. Hence, a given stimulus will be rated as less extreme if presented in the context of a more extreme one, than if presented in the context of a less extreme one. This impact of the range of stimuli is referred to as the "range effect". If the number of stimuli to be rated is sufficiently large, respondents will also attempt to use all categories of the rating scale about equally often. Accordingly, the specific ratings given also depend on the frequency distribution of the presented stimuli, an effect that is referred to as the "frequency effect." The observed attenuation of question order effects suggests that older respondents may also be less likely to keep track of numerous previous stimuli presented to them, resulting in an attenuation of range and frequency effects on subsequent ratings. Consistent with this assumption, some findings suggest that very old respondents (age 80 and higher) discriminate less between categories of response scales, which also reduces the predictive power of their ratings (Knäuper & Seibt, 1999).

**Response Order Effects**

Another major source of context effects in attitude measurement is the order in which response alternatives are presented. Response order effects are most reliably obtained when a question presents several plausible response options (see Sudman et al., 1996, chapter 6, for detailed discussions). To understand the underlying processes, suppose you are asked to provide a few reasons why “divorce should be easier to get.” You can easily do so, yet you could just as easily
provide some reasons why “divorce should be more difficult to get.” When such alternatives are juxtaposed (as in “Should divorce be easier to get or more difficult to get?”), the outcome depends on which alternative is considered first. While the researcher hopes that respondents (a) hold the question and all response alternatives in mind, (b) evaluate the implications of each alternative, and (c) finally select the one that they find most agreeable, respondents may not do so. Instead, respondents who first think about “easier” may come up with a good supporting reason and may endorse this answer without engaging in the additional steps.

Importantly, the likelihood that respondents elaborate on a given alternative depends on the order and mode in which the response options are presented (Krosnick & Alwin, 1987). When presented in writing, respondents elaborate on the implications of the response options in the order presented. In this mode, an alternative that elicits supporting thoughts is more likely to be endorsed when presented early rather than late on the list, giving rise to primacy effects. In contrast, when the alternatives are read to respondents, their opportunity to think about the early ones is limited by the need to listen to the later ones. In this case, they are more likely to work backwards, thinking first about the last alternative read to them. When this alternative elicits supporting thoughts, it is likely to be endorsed, giving rise to recency effects. As a result, a given alternative is more likely to be endorsed when presented early rather than late in a visual format (primacy effect), but when presented late rather than early in an auditory format (recency effect).

On theoretical grounds, we may expect that older respondents find it more difficult to keep several response alternatives in mind, elaborating on their respective implications to select the most appropriate answer. This should be particularly true under telephone interview conditions, where the alternatives are read to respondents without external memory support. The available data strongly
support this prediction (see Knäuper, 1999, for a comprehensive review and meta-analysis). For example, Schuman and Presser (1981) asked respondents in a telephone interview, "Should divorce in this country be easier to obtain, more difficult to obtain, or stay as it is now?" Depending on conditions, the response alternative "more difficult" was read to respondents as the second or as the last alternative. Overall, respondents were somewhat more likely to select the response alternative "more difficult" when presented last, a recency effect. However, secondary analyses reported by Knäuper (1999) indicate a dramatic age difference: As shown in Figure 2, the size of the recency effect increased with respondents' age, ranging from a nonsignificant 5% for age 54 and younger to 36.3% for age 70 and older. Note that we would again draw different substantive conclusions about the relationship of age and attitudes towards divorce, depending on the order in which the response alternatives are presented: While attitudes towards divorce seem to become much more conservative with age under one order condition, no reliable age differences are obtained under the other order condition.

Figure 2

The available data suggest that age differences in the emergence of response order effects are limited to the more taxing auditory format and not observed when all response alternatives are presented in writing and remain visible. As an example, consider a study in which younger (age 20 to 40) and older (age 65+) adults were asked, "Which of the following four cities do you find most attractive?" (Schwarz, Park, and Knäuper, unpublished data). Washington, D.C., was presented as either the first or last choice. Table 1 shows the results.

Table 1

When the response alternatives were read to respondents, younger as well as older adults
were more likely to choose Washington, D.C. when presented last rather than first. Replicating the pattern of the divorce findings, this recency effect was more pronounced for older than younger respondents, with differences of 24 vs. 8 percentage points. When the response options were presented in writing, however, older as well as younger respondents were more likely to choose Washington, D.C., when it was presented first rather than last. Moreover, this primacy effect under a visual mode was of comparable size for both age groups.

**Summary**

In sum, the reviewed findings suggest that question order effects are likely to decrease with age, whereas response order effects are likely to increase with age, in particular in telephone interviews. Both of these effects can be traced to age-related declines in cognitive resources, which make it more difficult for older respondents to hold large amounts of relevant information in short term memory. As a result, self-reports of attitudes are not only context dependent, but the size of the emerging context effects is itself age-sensitive, rendering comparisons across age-groups fraught with uncertainty.

**ANSWERING QUESTIONS ABOUT BEHAVIORS**

Many questions about respondents' behavior are frequency questions, pertaining, for example, to how often the respondent has bought something, has seen a doctor, or has missed a day at work during some specified period of time. Researchers who ask these questions would ideally like the respondent to identify the behavior of interest; to scan the reference period; to retrieve all instances that match the target behavior; and to count these instances to determine the overall frequency of the behavior. This, however, is the route that respondents are least likely to take.
In fact, except for rare and very important behaviors, respondents are unlikely to have detailed representations of numerous individual instances of a behavior stored in memory. Rather, the details of various instances of closely related behaviors blend into one global representation (Linton, 1982; Neisser, 1986). Thus, many individual episodes become indistinguishable or irretrievable, due to interference from other similar instances, fostering the generation of knowledge-like representations that "lack specific time or location indicators" (Strube, 1987, p.89). The finding that a single spell of unemployment is more accurately recalled than multiple spells (Mathiowetz & Duncan, 1988), for example, suggests that this phenomenon does not only apply to mundane and unimportant behaviors, but also to repeated experiences that profoundly affect an individual's life. Accordingly, a "recall and count" model does not capture how people answer questions about frequent behaviors or experiences. Rather, their answers are likely to be based on some fragmented recall and the application of inference rules to compute a frequency estimate (see Bradburn, Rips, & Shevell, 1987; Schwarz & Sudman, 1994; Sudman et al., 1996).

The most important estimation strategies involve reliance on subjective theories of stability and change and the use of information provided by the response alternatives (for a more comprehensive review see Schwarz & Oyserman, 2001).

**Subjective Theories of Stability and Change**

One particularly important inference strategy is based on subjective theories of stability and change (see Ross, 1989, for a review). In answering retrospective questions, respondents often use their current behavior or opinion as a bench-mark and invoke an implicit theory of self to assess whether their past behavior or opinion was similar to, or different from, their present behavior or opinion. Assuming, for example, that one's political beliefs become more conservative over the life-
span, older adults may infer that they held more liberal political attitudes as teenagers than they do now (Markus, 1986). The resulting reports of previous opinions and behaviors are correct to the extent that the implicit theory is accurate.

In many domains, individuals assume a rather high degree of stability, resulting in underestimates of the degree of change that has occurred over time. Accordingly, retrospective estimates of income or of tobacco, marijuana, and alcohol consumption were found to be heavily influenced by respondents' income or consumption habits at the time of interview (see Ross, 1989). On the other hand, when respondents have reason to believe in change, they will detect change, even though none has occurred. For example, participants in a study skills training inferred that their skills prior to training were much poorer than after training, even though the training had no measurable effect on actual performance (see Ross, 1989).

As this discussion indicates, retrospective reports of changes across the life-span will crucially depend on respondents' subjective theories. At present, little is known about whether and how these theories are themselves subject to change across the life-span: Which aspects of the self do people perceive as stable vs. variable? And at which life-stage do they expect changes to set in? Moreover, it is likely that salient life-events, like retirement or the loss of a spouse, will give rise to subjective theories of profound change. If so, retrospective reports pertaining to earlier time periods may exaggerate the extent to which life was different prior to the event, conflating real change with biased reconstructions. These issues provide a promising avenue for future research.

**Response Alternatives**

Whereas the preceding estimation strategies draw on respondents’ own knowledge, an alternative estimation strategy is based on information provided by the questionnaire itself. In many
Self-reports -- p. 18

studies, respondents are asked to report their behavior by checking the appropriate alternative from a list of response alternatives of the type shown in Table 2. While the researcher assumes that the scale is merely a “measurement device,” respondents draw on the scale as a relevant source of information. They assume that the values in the middle range of the scale reflect the "average" or "usual" behavioral frequency, whereas the extremes of the scale correspond to the extremes of the distribution. Based on this assumption, they use the range of the response alternatives as a frame of reference in estimating their own behavioral frequency.

Table 2

This strategy results in higher estimates along scales that present high rather than low frequency response alternatives, as shown in Table 2. In this study (Schwarz, Hippler, Deutsch, & Strack, 1985), only 16.2% of a sample of German respondents reported watching TV for more than 2 1/2 hours a day when the scale presented low frequency response alternatives, whereas 37.5% reported doing so when the scale presented high frequency response alternatives. Similar results have been obtained for a wide range of different behaviors, from reports of physical symptoms to consumer behavior (see Schwarz 1999a).

Age-related differences

Not surprisingly, the impact of response alternatives is more pronounced the less well the behavior is represented in memory, thus forcing respondents to rely on an estimation strategy (Menon, Rhagubir, & Schwarz, 1995). Given age-related declines in memory, this suggests that the impact of response alternatives may typically be more pronounced for older than for younger respondents. The available data support this prediction with some qualifications. As shown in Table 3, Knäuper, Schwarz, and Park (2004) observed that older respondents were more affected
than younger respondents by the frequency range of the response scale when asked to report the frequency of mundane events, such as buying a birthday present. On the other hand, older respondents were less affected than younger respondents when the question pertained to the frequency of physical symptoms, which older respondents are more likely to monitor (e.g., Borchert, Gilberg, Horgas, & Geiselmann, 1999). In combination, these findings suggest that respondents of all ages draw on the response alternatives when they need to form an estimate. Yet, whether or not they need to form an estimate depends on how much attention they pay to the respective behavior, which itself is age-dependent.

Table 3

Importantly, we would again draw different conclusions about age-related differences in actual behavior from these reports, depending on the scale format used. We would conclude, for example, that age-differences in red meat consumption (a health-relevant dietary behavior) or the purchase of birthday presents (an indicator of social integration) are minor when a low frequency scale is used, but rather pronounced when a high frequency scale is used. To avoid systematic influences of response alternatives, and age-related differences in their impact, it is advisable to ask frequency questions in an open response format, such as, "How many hours a day do you watch TV? ___ hours per day." Note that such an open format needs to specify the relevant units of measurement, e.g., "hours per day" to avoid answers like "a few." While the reports obtained under an open format are far from error free, they are at least not systematically biased by the instrument (see Schwarz, 1999b, for a discussion).

Summary

The findings reviewed in this section emphasize that retrospective behavioral reports are
rarely based on adequate recall of relevant episodes. Rather, the obtained reports are to a large
degree theory driven: Respondents are likely to begin with some fragmented recall of the behavior
under study and to apply various inference rules to arrive at a reasonable estimate. Moreover, if
quantitative response alternatives are provided, they are likely to use them as a frame of reference,
resulting in systematic biases. Although researchers have developed a number of strategies to
facilitate recall (see Sudman et al., 1996; Schwarz & Sudman, 1994), it is important to keep in mind
that the best we can hope for is a reasonable estimate, unless the behavior is rare and of considerable
importance to respondents.

CONCLUSIONS

As the reviewed examples illustrate, minor differences in question wording, question format,
and question order can greatly influence the obtained results, in representative sample surveys as
well as in the psychological laboratory. Although the underlying processes are increasingly well
understood (see Sudman et al., 1996; Tourangeau et al., 2000), we know little about the impact of
age-related changes in cognitive and communicative functioning on the question-answering process.
The little we do know, however, is cause for considerable concern: As our examples illustrate, age-
related changes in the emergence of context effects can invite misleading conclusions about actual
age differences in attitudes and behavior. To reduce this risk, we need to understand how age-related
changes in cognitive and communicative functioning interact with the features of our research
instruments in shaping respondents' reports. The exploration of these issues provides a challenging
avenue for future interdisciplinary research that promises to advance our theoretical understanding
of human cognition and communication across the life-span and to improve the methodology of
social science research.

REFERENCES


Memory observed: Remembering in natural contexts (pp.77-91). San Francisco: Freeman.


Psychologist, 54, 93-105.


### Table 1. Age, Mode, and Response Order Effects

<table>
<thead>
<tr>
<th>Age</th>
<th>Mode</th>
<th>First</th>
<th>Last</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>Auditory</td>
<td>29%</td>
<td>37%</td>
<td>+8%</td>
</tr>
<tr>
<td>Old</td>
<td>Auditory</td>
<td>17%</td>
<td>41%</td>
<td>+24%</td>
</tr>
<tr>
<td>Young</td>
<td>Visual</td>
<td>36%</td>
<td>28%</td>
<td>-8%</td>
</tr>
<tr>
<td>Old</td>
<td>Visual</td>
<td>22%</td>
<td>13%</td>
<td>-9%</td>
</tr>
</tbody>
</table>

*Note.* Respondents aged 20 to 40 (Young) and 65 to 90 (Old) were asked, "Which of the following cities do you find most attractive?" The alternative “Washington, DC” was presented in the first or last (4th) position.
Table 2

Reported Daily TV Consumption as a Function of Response Alternatives

<table>
<thead>
<tr>
<th>Low Frequency Alternatives</th>
<th>High Frequency Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1/2 h</td>
<td>Up to 2 1/2h</td>
</tr>
<tr>
<td>7.4%</td>
<td>62.5%</td>
</tr>
<tr>
<td>1/2 h to 1 h</td>
<td>2 1/2h to 3h</td>
</tr>
<tr>
<td>17.7%</td>
<td>23.4%</td>
</tr>
<tr>
<td>1 h to 1 1/2h</td>
<td>3h to 3 1/2h</td>
</tr>
<tr>
<td>26.5%</td>
<td>7.8%</td>
</tr>
<tr>
<td>1 1/2h to 2 h</td>
<td>3 1/2h to 4h</td>
</tr>
<tr>
<td>14.7%</td>
<td>4.7%</td>
</tr>
<tr>
<td>2h to 2 1/2h</td>
<td>4h to 4 1/2h</td>
</tr>
<tr>
<td>17.7%</td>
<td>1.6%</td>
</tr>
<tr>
<td>More than 2 1/2h</td>
<td>More than 4 1/2h</td>
</tr>
<tr>
<td>16.2%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 3

The Impact of Response Alternatives on Behavioral Reports

as a Function of Content and Respondents' Age

<table>
<thead>
<tr>
<th>Frequency Scale</th>
<th>Low</th>
<th>High</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mundane Behaviors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eating red meat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>24%</td>
<td>43%</td>
<td>19%</td>
</tr>
<tr>
<td>Old</td>
<td>19%</td>
<td>63%</td>
<td>44%</td>
</tr>
<tr>
<td>Buying birthday presents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>42%</td>
<td>49%</td>
<td>7%</td>
</tr>
<tr>
<td>Old</td>
<td>46%</td>
<td>61%</td>
<td>15%</td>
</tr>
<tr>
<td><strong>Physical Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headaches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>37%</td>
<td>56%</td>
<td>19%</td>
</tr>
<tr>
<td>Old</td>
<td>11%</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>Heartburn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>14%</td>
<td>33%</td>
<td>19%</td>
</tr>
<tr>
<td>Old</td>
<td>24%</td>
<td>31%</td>
<td>7%</td>
</tr>
</tbody>
</table>

*Note.* Younger respondents are aged 29-40, older respondents 60-90. Shown is the percentage of respondents reporting eating red meat more than 10 times a month or more often; buying birthday presents more than 5 times a year or more often; and having headaches or heartburn twice a month or more often. From Knäuper, Schwarz, and Park, (2004). Reprinted by permission
Figure 1. Question Order Effects as a Function of Age

![Bar chart showing the percentage of respondents who support abortion for "a married woman who doesn’t want any more children" when this question precedes (A-B) or follows (B-A) the birth defect question.](chart)

Note. Shown is the percentage of respondents who support abortion for “a married woman who doesn’t want any more children” when this question precedes (A-B) or follows (B-A) the birth defect question.
Figure 2. Response Order Effects as a Function of Age

Note. Shown is the percentage of respondents who endorse that “divorce should be more difficult to get” when this alternative is presented as the second or last of three options.