Eur J Ageing (2016) 13:185–193
DOI 10.1007/s10433-016-0369-0

REVIEW ARTICLE

How aging affects self-reports

Bärbel Knäuper[1] · Kimberly Carrière[1] · Melodie Chamandy[2] · Zhen Xu[1] · Norbert Schwarz[3] · Natalie O. Rosen[4]

Published online: 2 April 2016

Abstract A lot of information used in aging research relies on self-reports. Surveys or questionnaires are used to assess quality of life, attitudes toward aging, experiences of aging, subjective well-being, symptomatology, health behaviors, financial information, medication adherence, etc. Growing evidence suggests that older and younger respondents are differentially affected by questionnaire features and the cognitive tasks that question answering pose. This research has shown that age-related changes in cognitive and communicative functioning can lead to age-related differences in self-reports that are erroneously interpreted as real age differences in attitudes and behaviors. The current review highlights how the processes underlying respondents' self-report change as a function of respondents' age; it updates our previous reviews of this literature.

Keywords Aging · Self-reports · Surveys · Questionnaires · Questionnaire features · Survey methodology

Responsible editor: D. J. H. Deeg.

✉ Bärbel Knäuper
barbel.knauper@mcgill.ca

[1] Department of Psychology, McGill University, Montreal, Canada

[2] School of Psychology, University of Ottawa, Ottawa, Canada

[3] Department of Psychology, University of Southern California, Los Angeles, CA, USA

[4] Department of Psychology and Neuroscience, Dalhousie University, Halifax, Canada

A lot of information used in aging research relies on self-reports. Surveys or questionnaires are used to assess quality of life, attitudes toward aging, experiences of aging, subjective well-being, symptomatology, health behaviors, financial information, medication adherence, etc. Over the past 50 years, a strong body of research has uncovered many potential sources of response biases in surveys and has shown that apparently minor variations in question format, question order, or the position of response options can systematically bias responses and lead to false reports of attitudes and behaviors (for reviews see Schwarz 1999a; Sudman et al. 1996; Tourangeau et al. 2000; Abeele et al. 2013; Krosnick et al. 2006). Growing evidence suggests that older and younger respondents are differentially affected by questionnaire features (Schwarz et al. 1999; Yoon et al. 2010). This research has shown that age-related changes in cognitive and communicative functioning can lead to age-related differences in self-reports that are erroneously interpreted as real age differences in attitudes and behaviors. The current review highlights how the processes underlying respondents' self-report change as a function of respondents' age; it updates our previous reviews of this literature (Schwarz 2005; Schwarz et al. 1999; Schwarz and Knäuper 2000). Notably, this review concludes that age-related differences in question or response option comprehension and memory retrieval can lead to erroneous conclusions about age differences in opinions or behaviors. Yet, the examples identified over the last 20 years have not yet led to substantial changes in survey research that systematically address these problems. Also, the available research is still limited in quantity and quality, as detailed below.

What are the cognitive tasks involved in self-reports?

Providing optimal answers to survey questions requires the performance of a number of cognitive tasks (e.g., Strack and Martin 1987; Tourangeau 1984). First, the question's meaning has to be understood. Then, a memory search has to be conducted to retrieve information relevant for answering the question. Once the relevant information is retrieved, the information may need to be integrated to construct a summary judgment. Finally, the retrieved or constructed answer has to be formatted onto a given response scale. The difficulty of these cognitive tasks varies depending on the content or topic of the question, the time frame it covers, and the response options it provides (for an overview see Belson 1981; Clark and Schober 1992). Importantly, the difficulty also varies as a function of respondent characteristics, including their memory capacity. In the current review, we will describe how age differences in question comprehension and interpretation, as well as memory retrieval and decision-making have been shown to affect self-reports.

Age differences in question comprehension and interpretation

Early studies looked at the overall impact of age on survey data quality (Andrews and Herzog 1986; Colsher and Wallace 1989; Gergen and Beck 1966; Herzog and Dielman 1985; Rodgers et al. 1988). Andrews and Herzog (1986) found that the percentage of true score variance in survey measures tends to decline as respondent age increases and that the percentage of random and correlated error variance increases. Subsequently, studies started looking at how more specific age-related changes in memory and cognition affect more specific question comprehension and interpretation issues. These studies were motivated by the empirical observation that, generally, cognitive aging is associated with decreasing working memory capacity (Baddeley 2007). Working memory capacity reflects how much information a person can hold and manipulate at a given moment. Survey cognition researchers thus began to examine whether age differences in question comprehension and interpretation occur as a function of the difficulty of questions and response options and on the order in which response options are presented.

Question and response option difficulty

It had long been known that the accuracy and completeness of self-reports decrease with increasing question difficulty (e.g., Schwarz 1999a). Krosnick (1991) pointed out that difficult questions may lead respondents to provide a *satisfactory* answer instead of an *optimal* answer, a process that he terms "satisficing." He assumes that satisficing can present itself in either incomplete or biased information reports or in no report at all (e.g., answering "don't know"). Krosnick suggested that satisficing is a function of three factors and their interactions: (a) the difficulty of the task (question difficulty), (b) the ability of performing the task, and (c) the motivation to perform the task. We focus in this review on the first two factors and their interaction because these are the factors for which age-related changes in cognitive and communicative functioning are relevant. In a similar vein, Roedder John and Cole (1986) pointed out that information processing deficits in older persons are pronounced in situations where large amounts of information are presented, the information is not presented in an optimal format, and the task requires the use of difficult response formats. Applying these notions to the process of answering questions, and as pointed out in our previous reviews, respondents with lower working memory capacity should be more likely than those with a higher capacity to give incomplete, biased, or no responses to particularly difficult questions (Schwarz 2005; Schwarz et al. 1999; Schwarz and Knäuper 2000; see also Kaldenberg et al. 1994). Importantly, survey or questionnaire results can be systematically biased if respondents with lower working memory capacity do not respond to difficult questions or if they answer them in a systematically biased way.

In support of these arguments, Knäuper et al. (1997) found that the self-reports of older respondents (70+ years of age) with reduced working memory capacity (assessed with measures of cognitive reasoning, attention, and memory) were particularly affected by increased question difficulty. They found that respondents with lower working memory capacity were more likely to answer "don't know" to difficult questions than respondents with higher cognitive abilities.

Adding to these findings more recently, Fox et al. (2007) found through cognitive interviews (an interviewing approach that uses unstructured questions to help the interviewee to reliably recall information from memory) that items of psychometric scales that had complex content (i.e., contain large amounts of information) and items that were worded negatively (e.g., "Physically I feel I am in a bad condition") were associated with comprehension difficulties among older adults (ages 65–96). To reduce the cognitive burden of the items, the authors administered shorter and less complex items to respondents by breaking up the items into multiple smaller, simpler ones. When presented with the simplified versions of the items, participants found nothing in the items or response options unclear, i.e., they no longer had comprehension difficulties. These findings demonstrate that item complexity presents a

cognitive challenge for older adults and that this problem can be rectified through simplifying the items. Unfortunately, no comparisons were made in this study with younger adults.

Fox et al. (2007) further proposed that closed-ended questions with numeric response formats represent obstacles for older adults. Supporting their hypothesis, the authors found that bidirectional response formats that offered distinct numerical categories ranging from 0 (yes, that is true) to 4 (no, that is not true) as response options for questions such as "I can concentrate well" posed particular difficulty. These types of response categories are more cognitively taxing for older respondents, especially for those with cognitive decline, when compared to unidirectional response categories (e.g., from "not at all" to "a lot"). Finally, Fox et al.' participants expressed difficulty in understanding numeric response options. Participants were more comfortable elaborating upon their experiences verbally rather than having to choose numbers, demonstrating their preference for open-ended questions. It is an open question, though, whether the open responses would be more reliable than the responses using numeric response options, an issue we will revisit later in this review.

Adding to Fox et al.'s findings, Krestar et al. (2012) examined interview-based self-reports in relation to cognitive functioning among memory-impaired older adults between the ages of 50 and 95 (mini mental state examination (MMSE) scores of 12 or higher). The authors examined answer frequencies of 4-point unidirectional (e.g., from "none" to "a lot") and 4-point bidirectional response scales (e.g., from "strongly disagree" to "strongly agree"). Interviewers noted down when participants were not able to use the four points of the response scale and rather answered only "yes" or "no" to answer the question. They found that lower MMSE scores were correlated with a higher tendency to answer the questions dichotomously (i.e., yes/no) for unidirectional and bidirectional response scales, with larger effects for bidirectional scales. The authors interpret this to suggest that bidirectional scales pose a higher cognitive load than unidirectional scales because they require participants to switch between two directions. The authors recommend using unidirectional response scales in order to reduce cognitive load and improve response accuracy among individuals with mild to moderate cognitive impairment.

Relatedly, experiments in the area of decision research suggest that older respondents prefer to be given fewer options to choose from when making a decision. Reed et al. (2013) asked younger (18–37 years of age) and older (60–89 years of age) respondents to indicate how many options they would prefer, from 2 to 30 options, when making decisions in 12 different life domains (e.g., apartments, vacations, restaurants, cars). Older adults desired fewer choice options, preferring one-third of the total number of options compared with half for younger adults. In a separate computerized experiment, Reed et al. (2013) administered a behavioral decision task in which participants were asked to make decisions among twenty cars using a standard decision grid. The cars were labeled with names of rare birds (e.g., "Pipit," "Turaco," and "Xenops"). The information grid contained information on the following six attributes: gas mileage, horsepower, turning radius, safety rating, comfort, and dependability. When respondents were given the opportunity to click on numerous cells to reveal information about a specific car model, older adults viewed fewer choice options, as indicated by the amount of information searched based on the proportion of cells viewed in the grid. More specifically, older adults viewed less than half of all available information compared to young adults who viewed two-thirds of the information grid. Together, these findings suggest that preference for the amount of choices and extent of information seeking are negatively correlated with age. Although we assume that this is due to a decline in cognitive capacity, the mechanisms underlying these age-related differences have not yet been directly examined. Longitudinal studies may be of use to better understand to which extent and in which way a decline in cognitive capacity contributes to these findings.

Age-related differences as a function of question difficulty have been equally shown to persist in situations when respondents are offered clarification about question's meaning. In a study assessing the need for clarification in web surveys (Coiner et al. 2002), older respondents, compared to younger respondents, were less likely to seek clarification when confronted with the meaning of an ambiguous question. Such findings suggest that a lack of clarification requests by older respondents may contribute to a decrease in self-report accuracy. This is especially problematic when considering earlier studies using telephone interviews that showed response accuracy to increase relatively to the degree to which respondents received clarification of a question's meaning (Conrad and Schober 2000).

Interviewers are sometimes trying to compensate for perceived question and comprehension difficulties of their respondents by deviating from the questionnaire script and trying to make questions easier to understand. In a survey about physical activities and fall incidents among older adults (van der Zouwen et al. 2005), half of the interviewers deviated from the questionnaire script. Rather than asking older respondents to report on the frequency and the duration of household tasks simultaneously ("integral strategy"), interviewers rephrased the questionnaire and posed questions about the frequency and duration of household tasks separately ("partial strategy"). Using this partial strategy, interviewers decomposed the survey's

broad questions into smaller questions as a means to increase question comprehension and response accuracy in older respondents. Results showed that respondents for whom the interviewer used the partial strategy were less likely to round off and less likely to overestimate the average time spent on household duties than respondents for whom the interviewer used the integral strategy. However, when correlating the respondents estimated times spent on household duties with the actual times they spent on household duties (recorded using a daily dairy), response accuracy was found to be lower in the partial strategy group than the integral strategy group. The authors suggest based on these findings that the integral strategy leads to more accurate responses. They also point out that it is the more efficient process of data collection, as compared to its counterpart, because it requires less than half of the time to collect the same number of answers. Note, however, that this study did not use an experimental design and it is possible that respondents for whom the interviewer stuck with the prescribed integral strategy had a higher cognitive capacity and thus a better ability to give accurate accounts of their time use. Experimental studies on this issue are warranted.

Using response options to understand and interpret questions

Many studies in survey methodology have shown that respondents rely on formal features of the questionnaire (e.g., the response options) to interpret the question's meaning or to estimate how frequently they engage in certain behaviors (for a review, see Schwarz 1996). In fact, most research on age differences in self-reports has been done on this issue, as we previously reviewed (Schwarz, 2005; Schwarz et al. 1999; Schwarz and Knäuper 2000). Here, we recap the main points and present previous as well as recent empirical evidence for age differences.

Survey methodology research has documented that respondents assume that the survey researcher has chosen the response options with care. Respondents assume that the survey researcher placed typical or average behavior frequencies in the middle of the response scale and expects respondents to use the offered behavior frequencies to interpret what is asked of them. For instance, respondents may use the range of response options provided for a question on weekly TV watching to infer what the average TV watching amount is, and then use this information to estimate how much they themselves watch TV as compared to the average. If the researcher happened to choose "25 h" as the middle response options, then this may lead to higher self-reports than if the researcher happened to choose "6 h" as the middle response option because respondents will interpret the middle response option as the average behavior and relate their own behavior to it accordingly. In other words, respondents use the tacit assumptions that underlie the conduct of conversation in everyday life (Grice 1975) also in surveys and draw on formal features of the questionnaire to infer what the researchers' assumptions were when asking the question (cf. Schwarz 1995). Schwarz and colleagues (Schwarz et al. 1985) have shown that this inference strategy results in higher estimates along scales that present high rather than low frequency response alternatives: Only 16.2 % of a sample of German respondents reported watching TV for more than 2½ h 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).

Not surprisingly, the less well the behavior is represented in memory, the more question interpretation is influenced by the response alternatives. If the behavior is not well represented in memory, respondents cannot directly recall how often they engaged in the behaviour and instead have to use estimation strategies (Menon et al. 1995). Given age-related declines in memory, this suggests that the impact of response alternatives may be more pronounced for older than for younger respondents. The available data support this prediction with some qualifications. Knäuper et al. (2004) observed that, on the one hand, 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 found to be less affected than younger respondents when the question pertained to the frequency of physical symptoms, which presumably are more relevant for older people to monitor.

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. For example, older adults pay more attention to physical symptoms because of an increased salience of their mortality (Maxfield et al. 2007). In line with Borchelt et al.'s (1999) finding that older adults' responses were less affected by experimental variations of response choices when responding on personally relevant information (frequency of physical symptoms), Schmitz et al. (2002) found that there was more than 85 % agreement between self-reports and physician reports for the recall of physician visits. This further underlines that the accuracy of self-reports across

the life span does not only depend on cognitive capacity but also on the self-relevance of the subject matter.

To avoid systematic influences of response alternatives on frequency reports 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 frequency reports obtained in an open response format are far from being without error, they are at least not systematically biased by the survey instrument (see Schwarz 1999b, for a discussion).

In summary, we conjecture that older and younger adults are differentially influenced in question comprehension and interpretation by characteristics of the question stem, specifically with regards to negative statements, bidirectional versus unidirectional responses formats, and open versus closed response formats. Older respondents also seem less likely to consider contextual features, such as the numeric values of rating scales, when finding their response and prefer fewer and less complex choices.

Retrieving relevant information from memory and making a judgment

Retrospective behavioral reports

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. Ideally, researchers who ask these questions want the respondent to identify the behavior of interest, scan the reference period, retrieve all instances that match the target behavior, and to count these instances to determine the overall frequency of the behavior. This is, however, not the way respondents usually answer frequency questions.

In fact, except for rare and very important behaviors, it is unlikely that respondents have detailed representations of numerous individual instances of the 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 and Duncan 1988) suggests that this does not only apply to mundane and unimportant behaviors, but also to repeated experiences that are personally very relevant and significant. In sum, a "recall and count"—model does not capture how people answer questions about frequent behaviors or experiences. Rather, their answers tend to be based on some fragmented recall and the application of inference rules to compute a frequency estimate (see Bradburn et al. 1987; Schwarz and Sudman 1994; Sudman et al. 1996).

As one example of age differences in retrospective reports, Rossi and Isaacowitz (2006) found age differences in response to questions asked in open response formats when asking respondents to report on the life domains that are currently considered to be important in one's life. Four assessment techniques were used; two consisting of open-ended response formats and two consisting of closed-ended response formats (a comprehensive checklist of possible life domains and an incidental memory task). When open-ended response formats were used, age differences in the number of selected domains were found. Specifically, older respondents reported fewer important areas of life than younger respondents when using open response formats. No age differences were found for closed response formats. This is in line with previous work on memory and aging that showed that recall of information is more impaired with increasing age than recognition of information (e.g., Zelinski and Burnight 1997). Open-ended questions are similar to memory recall tasks ("What did you eat yesterday?") and closed questions are similar to recognition type memory tasks ("Did you eat X yesterday?"). In sum, these findings suggest that closed-ended questions may be less cognitively taxing when compared to open-ended formats and thus may lead to less bias when comparing the responses of younger and older people.

Age differences in mapping have also been found (Holbrook et al. 2006). Mapping requires respondents to simultaneously remember the question they are answering and provide their response in the correct response format (e.g., choosing a number from a provided rating scale), i.e., requires considerable working memory capacity (Olton et al. 1980). Holbrook et al. found that adults between the ages of 30 to 50, compared with adults between the ages of 18 to 29, were more likely to experience cognitive difficulties with translating mental representations of their judgment onto a given response format. That is, even if questions were well understood, the 30–50 year old respondents showed greater difficulty with keeping the question in mind while forming a relevant response. This finding is in line with recent systematic investigations of the age of peak performance across different cognitive tasks: Hartshorne and Germine (2015) found in a series of experiments that processing speed peaks around age 20, working memory around age 30, and comprehension/vocabulary around age 50.

Thus, what people report is affected by what information they can remember and retrieve from memory and how easy or difficult the question format renders this cognitive task. For questions asking about attitudes, it is particularly relevant what information happens to be temporarily or chronically accessible in memory because such questions ask for an evaluative judgment rather than for objective information. While chronically accessible information contributes stability to attitude judgments, temporarily accessible information is a major source of context effects. Here, we focus on two key issues in attitude measurement, namely the emergence of question and response order effects, both of which show pronounced age differences.

Question order

When asked for their attitude, respondents are unlikely to retrieve all information from memory that is potentially relevant. This has been documented in many psychological experiments. Rather, respondents truncate the search process as soon as enough information has come to mind to form a judgment with sufficient subjective certainty (see Bodenhausen and Wyer 1987), i.e., they are satisficing. Accordingly, their judgments are profoundly influenced by the information that is most accessible at the time. This is usually the information that has been used most recently, e.g., for the purpose of answering a preceding question.

To guard against such question order effects, survey researchers often include buffer questions in between related questions as an attempt to attenuate the accessibility of previously used information. Because of age-related declines in memory, one could expect attenuated question order effects because the earlier questions are more likely to be forgotten. The available data show that this is indeed the case (Holbrook et al. 2006; Knäuper et al. 2007; Schwarz 2005). Knäuper et al. (2007) demonstrated this in secondary analyses of Schuman and Presser's (1981) classic abortion question order experiment. In this experiment, a question that provided a strong legitimization for an abortion was asked either before or after a "weaker" legitimization question. Specifically, the following two questions were asked with the order of the questions reversed in two versions of the questionnaire: Question A: "Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of serious defect in the baby?" (strong legitimization); Question B: "Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?" (weak legitimization). Results showed that the weaker abortion question (Question B) received 13 % more support when asked first than when asked after the stronger child-defect item (Question A). The stronger item, on the other hand, was unaffected by question order. Thus, a contrast effect occurred with the difference in the marginals between the two questions being larger when asked in the strong–weak order than when the questions were each asked first. This reflects that "not wanting any more children" appears as a less legitimate reason for an abortion when contrasted with "serious defects in the baby." Knäuper et al. (2007), however, found that the abortion question order effect is age-dependent: It is sized at 19.5 percentage points for younger respondents, but decreases with age and is no longer observed for respondents aged 65 and older. Importantly, one arrives at different conclusions about age differences in attitudes toward abortion, depending on the order in which the questions were asked. Furthermore, subsequent laboratory experiments (Knäuper et al. 2007) showed 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 20–40 years) as well as the subset of older respondents (age 60–100 years) with good working memory capacity showed the familiar question order effect. In contrast, the responses of older respondents with poor working memory capacity were not affected by the order in which the questions were asked.

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. It should be noted, though, that at least one study did not find decreasing question order effects with increasing age (Tourangeau et al. 2003). The attenuation of question order effects in older age may be limited to relatively uninvolving questions typical of public opinion surveys. When, however, the questions are of high personal or emotional relevance, older adults may show question order effects to the same extent as younger adults. This possibility remains to be tested.

Response order

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 equally plausible response options (see Sudman et al. 1996, chapter 6, for detailed discussions). To understand the underlying processes, suppose respondents are asked to provide a few reasons why "divorce should be easier to get." Some reasons will probably come to mind easily. Similarly, if they were asked why "divorce should be more difficult to get," some reasons will probably also come to mind easily.

When such alternatives are juxtaposed (as in "Should divorce be easier to get or more difficult to get?"), the outcome depends on which of the two alternatives is considered first. While the researcher hopes that respondents (a) hold the question and all response alternatives in mind, (b) consider all the reasons that come to mind for each of the alternatives, and (c) finally select the alternative for which most reasons came to mind, respondents rarely 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 thoughts of why it should be more difficult to get.

Importantly, the likelihood that respondents elaborate on a given alternative depends on the order and mode (visual, auditory) in which the response options are presented (Krosnick and Alwin 1987; Schwarz 2005). 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 (e.g., in a telephone interview), 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, as outlined in our earlier reviews, we expect that older respondents find it more difficult to keep several response alternatives in mind while 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: The size of the recency effect increased with respondents' age, ranging from a statistically not significant 5 % for ages 54 and younger to 36.3 % for ages 70 and older. Note that we would draw different substantive conclusions about the relationship of age and attitudes toward divorce depending on the order in which the response alternatives are presented: While attitudes toward divorce seem to become much more conservative with increasing age under one order condition, no reliable age differences are obtained under the other order condition. The available data suggest that age differences in response order effects are limited to the cognitively more taxing auditory format and are not observed when all response alternatives are presented in writing and remain visible (Knäuper 1999).

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 might be due 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. More research directly testing the contribution of cognitive resources to the emergence of response order effects and decrease of question order effects is needed. As a result, self-reports of attitudes are not only context-dependent but the size of the context effects is itself age-sensitive, rendering comparisons across age-groups fraught with uncertainty.

Conclusions

As the reviewed examples illustrate, minor differences in question wording, question format, response options, 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), they have rarely been examined in older respondents. There might be age-related changes in cognitive and communicative functioning that have specific influences on the question answering process that have so far been overlooked. Age-related differences in question or response option comprehension can lead to erroneous conclusions about age differences in opinions or behaviors. Yet, the examples identified over the last twenty years have not led to substantial changes in survey research that systematically addresses these problems, and the available research is still limited. The little we do know, however, is cause for considerable concern: As the findings of existing studies 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 lifespan and to improve the methodology of aging research that relies on self-reports.

Practical Implications

There is no perfect questionnaire. Designing a questionnaire involves many interdependent decisions and while optimizing one feature of the questionnaire, one might have to compromise on another. Making the best trade-off decisions requires that people who conduct research on aging are knowledgeable about the cognitive processes involved in answering questions and are aware of the possible issues that we outlined in this paper in order to make decisions that minimize potential bias (for guidelines see Sudman et al. 1996; Schwarz and Oyserman 2001). Whenever possible, questionnaires should be pretested using the methods developed for this purpose (see the papers in Sudman and Schwarz 1996) in order to avoid age differences in question interpretation, and in question and response order effects. Given the described age differences in context effects, particular attention should also be paid to what comes before the target questions in the survey instrument as it might bias the responses to the target questions. Part of the pretesting process could involve varying the order of questions or response alternatives to uncover potential age differences in order effects.

While pretesting requires time and resources, the advice for numeric frequency scales, in contrast, is simple: Frequency questions should be asked in an open response format, rather than providing frequency response categories to choose from, to avoid systematic bias by the age of the respondents.

Given that not all issues can be resolved at the same time in a survey instrument, it is important that researchers report in publications exactly how questions and response options were worded, and in which order they were administered. This will allow users of the information to take the context in which the self-reports were received into account when interpreting age differences.

References

Abeele MV, Beullens K, Roe K (2013) Measuring mobile phone use: gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use. Mob Med Comm 1:213–236

Andrews FM, Herzog AR (1986) The quality of survey data as related to age of respondent. J Am Stat Assoc 81:403–410

Baddeley AD (2007) Working memory, thought and action. Oxford University Press, Oxford

Belson WA (1981) The design and understanding of survey questions. Gower, London

Bodenhausen GV, Wyer RS (1987) Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In: Sudman S (ed) Hippler HJ, Schwarz M. Social information processing and survey methodology. Springer Verlag, New York, pp 6–41

Borchelt M, Gilberg R, Horgas AL, Geiselmann B (1999) On the significance of morbidity and disability in old age. In: Baltes PB, Mayer KU (eds) The Berlin aging study: aging from 70 to 100. Cambridge University Press, New York, pp 403–429

Bradburn NM, Rips LJ, Shevell SK (1987) Answering autobiographical questions: the impact of memory and inference on surveys. Science 236:157–161

Clark HH, Schober MF (1992) Asking questions and influencing answers. In: Tanur JM (ed) Questions about questions. Russel Sage, New York, pp 15–48

Coiner TF, Schober MF, Conrad FG, Ehlen P (2002) Assessing respondents' need for clarification in web surveys using age-based user modeling. Proceedings of the American Statistical Association. American Statistical Association, Alexandria

Colsher PL, Wallace RB (1989) Data quality and age: health and psychobehavioral correlates of item nonresponse and inconsistent responses. J Geront 44:P45–P52

Conrad FG, Schober MF (2000) Clarifying question meaning in a household telephone survey. Public Opin Quart 64:1–28

Fox MT, Sidani S, Streiner D (2007) Using standardized survey items with older adults hospitalized for chronic illness. Res Nurs Health 30:468–481

Gergen KJ, Beck KW (1966) Communication in the interview and the disengaged respondent. Public Opin Quart 30:385–398

Grice P (1975) Logic and conversation. In: Cole P, Morgan JL (eds) Syntax and semantics, vol 3., Speech actsAcademic Press, New York, pp 41–58

Hartshorne JK, Germine LT (2015) When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. Psychol Sci 26:433–443

Herzog AR, Dielman L (1985) Age differences in response accuracy for factual survey questions. J Geront 40:350–357

Holbrook A, Cho YI, Johnson T (2006) The impact of question and respondent characteristics on comprehension and mapping difficulties. Public Opin Quart 70:565–595

Kaldenberg DO, Koenig HF, Becker BW (1994) Mail survey response rate patterns in a population of the elderly: does response deteriorate with age? Public Opin Quart 58:68–76

Knäuper B (1999) The impact of age and education on response order effects in attitude measurement. Public Opin Quart 63:347–370

Knäuper B, Belli RF, Hill DH, Herzog AR (1997) Question difficulty and respondents' cognitive ability: the impact on data quality. J Off Stat 13:181–199

Knäuper B, Schwarz N, Park DC (2004) Frequency reports across age groups. J Off Stat 20:91–96

Knäuper B, Schwarz N, Park D, Fritsch A (2007) The perils of interpreting age differences in attitude reports: question order effects decrease with age. J Off Stat 23:1–14

Krestar M, Looman W, Powers S, Dawson N, Judge K (2012) Including individuals with memory impairment in the research process: the importance of scales and response categories used in surveys. J Empir Res Hum Res Ethics 7:70–79

Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cognit Psychol 5:213–236

Krosnick JA, Alwin DF (1987) An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opin Quart 51:201–219

Krosnick JA, Holbrook AL, Visser PS (2006) Optimizing brief assessments in research on the psychology of aging: a pragmatic approach to self-report measurement. In: Carstensen LL, Hartel CR (eds) When I'm 64. National Academies Press, Washington DC, p 231

Linton M (1982) Transformations of memory in everyday life. In: Neisser U (ed) Memory observed: remembering in natural contexts. Freeman, San Francisco, pp 77–91

Mathiowetz NA, Duncan GJ (1988) Out of work, out of mind: response errors in retrospective reports of unemployment. J Bus Econ Stat 6:221–229

Maxfield M, Kluck B, Greenberg J, Pyszczynski T, Cox CR, Solomon S, Weise D (2007) Age-related differences in responses to thoughts of one's own death: mortality salience and judgments of moral transgressions. Psychol Aging 22:341–353

Menon G, Raghubir P, Schwarz N (1995) Behavioral frequency judgments: an accessibility-diagnosticity framework. J Cons Res 22:212–228

Neisser U (1986) Nested structure in autobiographical memory. In: Rubin DC (ed) Autobiogra-phical memory. Cambridge University Press, Cambridge, pp 71–88

Olton DS, Becker JT, Handelmann GE (1980) Hippocampal function: working memory or cognitive mapping? Phys Psychol 8:239–246

Reed AE, Mikels JA, Löckenhoff CE (2013) Preferences for choice across adulthood: age trajectories and potential mechanisms. Psychol Aging 28:625–632

Rodgers WL, Herzog R, Andrews FM (1988) Interviewing older adults: validity of self-reports of satisfaction. Psychol Aging 3:264–272

Roedder John D, Cole CA (1986) Age differences in information processing: understanding deficits in young and elderly consumers. J Cons Res 13:297–315

Rossi N, Isaacowitz DM (2006) What is important to me right now? Age differences in domain selectivity depend on the measure. Ageing Int 31:24–43

Schmitz MF, Russell DW, Cutrona CE (2002) The validity of self-reports of physician use among the older population. J Appl Geront 21:203–219

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

Schwarz N (1995) What respondents learn from questionnaires: the survey interview and the logic of conversation. Int Stat Rev 63:153–168

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

Schwarz N (1999a) Self-reports: how the questions shape the answers. Am Psychol 54:93–105

Schwarz N (1999b) Frequency reports of physical symptoms and health behaviors: How the questionnaire determines the results. In: Park DC, Morrell RW, Shifren K (eds) Processing medical information in aging patients: Cognitive and human factors perspectives. Erlbaum, Mahaw, pp 93–108

Schwarz N (2005) Measurement: Aging and the psychology of self-report. In: Carstensen LL, Hartel CR (eds) When I'm 64: Aging frontiers in social psychology, personality, and adult developmental psychology. National Academic Press, Washington DC, pp 219–230

Schwarz N, Knäuper B (2000) Cognition, aging, and self-reports. In: Park D, Schwarz N (eds) Cognitive aging: a primer. Psychology Press, Philadelphia, pp 233–252

Schwarz N, Oyserman D (2001) Asking questions about behavior: cognition, communication and questionnaire construction. Am J Eval 22:127–160

Schwarz N, Sudman S (1994) Autobiographical memory and the validity of retrospective reports. Springer Verlag, New York

Schwarz N, Hippler HJ, Deutsch B, Strack F (1985) Response categories: effects on behavioral reports and comparative judgments. Public Opin Quart 49:388–395

Schwarz N, Park D, Knäuper B, Sudman S (eds) (1999) Cognition, aging, and self-reports. Psychology Press, Philadelphia

Strack F, Martin L (1987) Thinking, judging, and communicating: A process account of context effects in attitude surveys. In: Hippler HJ, Schwarz N, Sudman S (eds) Social information processing and survey methodology. Springer Verlag, New York, pp 123–148s

Strube G (1987) Answering survey questions: The role of memory. In: Hippler HJ, Schwarz N, Sudman S (eds) Social information processing and survey methodology. Springer Verlag, New York, pp 86–101

Sudman S, Bradburn N, Schwarz N (1996) Thinking about answers: The application of cognitive processes to survey methodology. Jossey-Bass, San Francisco

Tourangeau R (1984) Cognitive science and survey methods: a cognitive perspective. In: Jabine T, Straf M, Tanur J, Tourangeau R (eds) Cognitive aspects of survey methodology: Building a bridge between disciplines. National Academy Press, Washington DC, pp 73–100

Tourangeau R, Rips LJ, Rasinski K (2000) The psychology of survey response. Cambridge University Press, New York

Tourangeau R, Singer E, Presser S (2003) Context effects in attitude surveys: effects on remote items and impact on predictive validity. Soc Methods Res 31:486–513

van der Zouwen J, Smit JH, van der Horst MHL (2005): Reporting the frequency and duration of household tasks by elderly respondents: The effect of different interview strategies on data quality. In: Proceedings of the 60th AAPOR annual conference, May 12–15, 2005, American Statistical Association, Alexandria, pp 4007–4014

Yoon C, Feinberg F, Schwarz N (2010) Why do older consumers tell us they are more satisfied? In: Drolet A, Schwarz N, Yoon C (eds) The aging consumer: Perspectives from psychology and economics. Routledge/Taylor & Francis Group, New York, pp 209–228

Zelinski EM, Burnight KP (1997) Sixteen-year longitudinal and time-lag changes in memory and cognition in older adults. Psychol Aging 12:503–523