Relying on External Information Sources When Answering Knowledge Questions in Web Surveys

Gummer, Tobias; Kunz, Tanja

Article

Sociological Methods & Research
2022, Vol. 51(2) 816–836

DOI: 10.1177/0049124119882470
journals.sagepub.com/home/smr

Relying on External Information Sources When Answering Knowledge Questions in Web Surveys

Tobias Gummer[1] and Tanja Kunz[1]

Abstract

Knowledge questions frequently are used in survey research to measure respondents' topic-related cognitive ability and memory. However, in self-administered surveys, respondents can search external sources for additional information to answer a knowledge question correctly. In this case, the knowledge question measures accessible and procedural memory. Depending on what the knowledge question aims at, the validity of this measure is limited. Thus, in this study, we conducted three experiments using a web survey to investigate the effects of task difficulty, respondents' ability, and respondents' motivation on the likelihood of searching external sources for additional information as a form of over-optimizing response behavior when answering knowledge questions. We found that the respondents who are highly educated and more interested in a survey are more likely to invest additional efforts to answer knowledge questions correctly. Most importantly, our data showed that for these respondents, a more difficult question design further increases the likelihood of over-optimizing response behavior.

[1] GESIS—Leibniz Institute for the Social Sciences, Mannheim, Baden-Wuerttemberg, Germany

Corresponding Author:
Tobias Gummer, GESIS—Leibniz Institute for the Social Sciences, B2, 1, Mannheim, Baden-Wuerttemberg 68159, Germany.
Email: tobias.gummer@gesis.org

Keywords
paradata, knowledge questions, web surveys, measurement error, satisficing

Knowledge questions frequently are used in survey research to measure respondents' topic-related cognitive ability and memory. Respondents are asked to provide an answer for which the true value is known, and thus, correct and false answers can be determined. Sometimes, it might even be interesting to calculate the distance between a respondent's answer and the true value, for instance, if a question involves guessing. In a variety of research fields, the measures of knowledge questions serve as important variables of interest. For instance, Clifford and Jerit (2016:859) describe political knowledge to be "a central construct in political science, communications, and related fields." Accordingly, important political science studies repeatedly have reported relationships between topic-specific knowledge and turnout (e.g., Prior 2005), political participation (e.g., Galston 2001), and the stability of political attitudes (e.g., Delli Carpini and Keeter 1996). With the Programme for International Student Assessment (PISA) and the Programme for the International Assessment of Adult Competencies (PIAAC), large-scale cross-national survey programs exist that are dedicated entirely to measuring students' and adults' knowledge and skills.

In self-administered surveys, knowledge questions can be designed as either closed-ended or open-ended questions (for an overview of different question types and response formats, see Dillman, Smyth, and Christian 2014). Closed-ended questions provide respondents with a fixed set of response options from which they can select their answer(s). These response options give respondents cues about which answers researchers might expect and, thus, ease the answering process and lower their response burden (Tourangeau, Rips, and Rasinski 2000). However, predefined response options limit the variety of answers and degree of detail that respondents are able to provide in their answers (e.g., Reja et al. 2003). Open-ended questions are one possibility to remedy this shortcoming of closed-ended questions by enabling an increased variability and detailedness of answers up to and including narrative responses. The consequences of this more laborious answering process and omission of cues compared to closed-ended questions can be an increased response burden and respondent fatigue that might endanger response quality (Reja et al. 2003; Smyth et al. 2009). In addition, responses to open-ended questions need coding to be analyzable and, thus, demand more work and resources by researchers. In the case of knowledge questions, open-ended questions are considered more suitable than closed-ended questions because answers to open-ended questions better reflect respondents' actual knowledge rather than increases in knowledge due to the possibly "lucky guesses" fostered by the response options in a closed-ended question. However, this advantage comes at the expense of more "don't know" responses in the case of respondents who are unsure about the correct answer or cannot immediately recall the correct answer (Krosnick and Presser 2010; Robison 2015).

The answering of knowledge questions is known to be more than merely a retrieval of requested information from memory. Respondents who know the correct answer or at least think they know it retrieve that answer from their memory. By contrast, respondents who do not know or who are unsure about the correct answer are likely to make inferences or use cues to answer knowledge questions (Nadeau and Niemi 1995); or, they may refer to external sources and look up the correct answer. In self-administered surveys such as web surveys, respondents are free to draw on external sources to look up relevant information for knowledge questions, if they feel the need to search for additional information to provide an informed and correct answer (Clifford and Jerit 2016). This kind of search behavior contradicts a survey researcher's intention for asking such questions, if the knowledge questions are supposed to measure what information is accessible to a respondent when answering a knowledge question rather than a respondent's ability to read up on a specific topic (Jensen and Thomsen 2014). If respondents can perform search activities, the measure of accessible memory is confounded with procedural memory (Prior and Lupia 2008). Consequently, if respondents look up information in external sources to answer a knowledge question, the measure of their accessible memory will be erroneous, and the validity of these knowledge questions needs to be challenged.

Previous research on this specific issue is sparse, primarily because the tools to directly measure whether respondents leave a web survey and switch to other websites or activities have become available only recently (Diedenhofen and Musch 2017; Schlosser and Höhne 2018; Sendelbah et al. 2016). To measure switching behavior when answering knowledge questions, prior studies frequently have relied on the self-reported use of external sources (Clifford and Jerit 2016; Jensen and Thomsen 2014). These reactive measures only serve as vague proxies for leaving a web survey to gather external information to answer survey questions. Recent advances in survey software development enable a tracking of window switching behavior, which can detect whether respondents leave a web survey. Along these lines, Sendelbah et al. (2016) have proposed using nonreactive paradata measures of when a respondent leaves a web survey to detect the respondent's multitasking behavior. Accordingly, Höhne and Schlosser (2018) report that respondents' loss of focus on doing a web survey is associated positively with longer response times. Drawing on similar paradata, Diedenhofen and Musch (2017) have provided evidence that to answer questions correctly, respondents leave the web survey to seek information from external sources. However, nothing is known about what motivates respondents to expend this additional effort to look up information. Höhne and Schlosser (2018:376) have reached the same conclusion and have stated that "it would be desirable if future research takes a closer look at the reasons why respondents leave the web-survey page." In the present study, we address this research gap by focusing on how task difficulty influences respondents' decisions to search for external information when answering knowledge questions. Moreover, we investigate respondent characteristics that are likely to promote this kind of search behavior.

The cognitive model of survey response (Tourangeau et al. 2000) suggests that the cognitive process underlying the answering of survey questions is a four-step process that includes comprehending the question, retrieving relevant information from memory, forming a judgment, and reporting a response. Optimizing in terms of carefully passing through all four steps requires a certain amount of respondent effort. Otherwise, skipping one or more steps of this cognitive process can result in suboptimal answers to survey questions, for instance, if the meaning of the question is not comprehended properly and irrelevant information is retrieved. In line with this reasoning, satisficing theory (Krosnick 1991, 1999) suggests that respondents may skip parts of the four-step process to reduce their response burden; consequently, forms of weak or strong satisficing response behavior are likely to be observed, depending on whether one or several steps are skipped. Both satisficing response strategies involve selecting an answer that the respondents deem adequate with respect to their individual response burden. Satisficing theory further assumes that the likelihood of showing satisficing response behavior is a function of task difficulty, respondent ability, and respondent motivation. Previous studies have routinely reported higher rates of satisficing when more demanding question formats are used (Mavletova and Couper 2016; Roßmann, Gummer, and Silber 2017), when the answering process is more taxing (Couper and Peterson 2017), or when comprehending the question is harder due to vague and ambiguous language (Lenzner, Kaczmirek, and Galesic 2011; Lenzner, Kaczmirek, and Lenzner 2010). Response behavior that is commonly associated with satisficing also has been shown to be more likely for respondents with lower cognitive ability (e.g., Narayan and Krosnick 1996; Toepoel, Das, and Van Soest 2009; Yan and Tourangeau 2008) and lower motivation (e.g., Kleiner, Lipps, and Ferrez 2015; Lenzner 2012; Roßmann et al. 2017). In addition, respondent ability, respondent motivation, and task difficulty have been reported to interact (Lenzner 2012; Roßmann et al. 2017). For instance, Roßmann, Gummer, and Silber (2017) found that more capable and higher motivated respondents are less prone to satisficing response behavior, although these effects are moderated by the difficulty of the question format (i.e., whether questions were presented in an item-by-item or grid design). Previous studies on knowledge questions also found that respondents with high ability and motivation are more likely to provide a substantive answer at all and to answer correctly (Nadeau and Niemi 1995). Especially, when answering knowledge questions based on an open-ended response format, respondents with higher interest and education have been found to try harder to come up with a substantive response, whereas respondents with less interest and education were more likely to satisfice and indicate that they did not know the answer (Robison 2015).

Leaving a web survey to look up external information sources to provide a correct answer to knowledge questions is not satisficing—in this case, respondents do not simply rely on guessing and select the first response option they deem satisfactory to minimize their response burden. On the contrary, respondents who look up information are going the extra mile by increasing their response burden. They perform an additional search operation that requires them to proceed through several additional steps of information gathering and judgment. These extra steps may include deciding on an external information source, formulating a search inquiry, evaluating the relevance of the information found with respect to its credibility and quality, translating the gathered information back into the context of the survey question, and other steps depending on the nature of the search inquiry. Accordingly, the answering process for knowledge questions in self-administered surveys may comprise more than four steps if external information sources are consulted.

Previous studies that have examined the cognitive answering process related to knowledge questions have assumed that when respondents are asked to answer a knowledge question, they either know the answer or at least think they know it, retrieve the answer, and give a substantive response; or that respondents are unsure about their answer or do not know the answer at all. These latter respondents may provide a nonsubstantive response (a "don't know" response or leave the answer blank), or they may provide a substantive response based on a random guess or by drawing on cues, heuristics, beliefs, or feelings from which they make an educated guess (Mondak and Davis 2001; Nadeau and Niemi 1995). The present study has expanded this model to include external information sources accessed by those respondents who are unsure about or do not know the correct answer.

Concerning satisficing theory, we argue that processing all the four steps proposed by Tourangeau, Rips, and Rasinski (2000), and also performing the additional steps of searching for and processing external information, exceeds what is commonly termed *optimizing*. With respect to the extended model of survey response that we have outlined for knowledge questions, we argue that progressing through all the aforementioned four steps and performing the additional steps of searching external information sources can be referred to as *over-optimizing*. In this regard, we assume that highly capable respondents are more likely to temporarily leave a web survey to search external information sources, similar to the respondents who are more interested and involved in a survey topic. This reasoning is based on the assumption that motivated respondents usually have a high degree of self-confidence and thus "are likely to feel that they should know the answer" (Nadeau and Niemi 1995:326). This notion corresponds to the concept of *self-deception*, a component of socially desirable response behavior that aims at enhancing a respondent's self-perception (Paulhus 2002; Tourangeau and Yan 2007). With respect to knowledge questions in self-administered surveys, a search of external information sources is a viable way for respondents who are motivated to meet their self-perceived knowledge level. Moreover, respondents with a higher cognitive ability are presumably better able to, first, identify appropriate external sources, and second, find or deduce the correct answer with reasonable efforts. We also assume that the more difficult and complex a knowledge question is to process and answer, the more likely respondents are to draw on external sources to answer that question. The increasing difficulty of a question will make it harder for the respondent to give the correct answer without using external sources and, thus, should increase the likelihood of those respondents who are willing to invest the extra effort to temporarily leave the survey to search for external information.

To investigate our research questions concerning the factors that drive the search behavior with respect to answering knowledge questions and to address the respective research gap, we performed three experiments in a web survey. The next section introduces our data and method. Then, we present our findings and after provide concluding remarks and implications for survey practitioners and future research.

Data and Method

To investigate our research questions, we performed three experiments in a web survey on "politics and work" in Germany. The survey was fielded in November 2017. Respondents were quota-sampled from a large German online access panel. Of the 3,030 invited panelists, 498 were screened out. Overall, 2,247 respondents completed the survey with a break-off rate of 7% (Callegaro and DiSogra 2008). The break-off was slightly lower in our survey compared to other web surveys done in Germany of similar length, samples, and topics (Gummer, Quoß, and Roßmann 2019). Approximately 21% of the respondents chose to complete the survey on a smartphone. On average, the questionnaire took 32.9 minutes to complete (*Mdn* = 29.3).

We employed the Embedded Client Side Paradata (ECSP) script (Schlosser and Höhne 2018) to collect client-side paradata on whether respondents temporarily left the browser window that hosted the web survey and switched to other browser windows, programs, or applications. For analytical purposes, we dichotomized the information on window switching (0 = *no switch*, 1 = *switch*). This measure indicated that changes in the active browser window occurred; however, it did not provide any specific information regarding the web sites, programs, or applications to which the respondents switched. Furthermore, the ECSP script enabled us to track the absence times that described how long in milliseconds the respondents had left the browser window hosting the web survey. The ECSP script also provided the response times for each experimental question by indicating the time in milliseconds from loading the webpage with the particular experimental question to submitting the webpage by clicking on the "next button." For better readability, we reported all the time data in seconds.

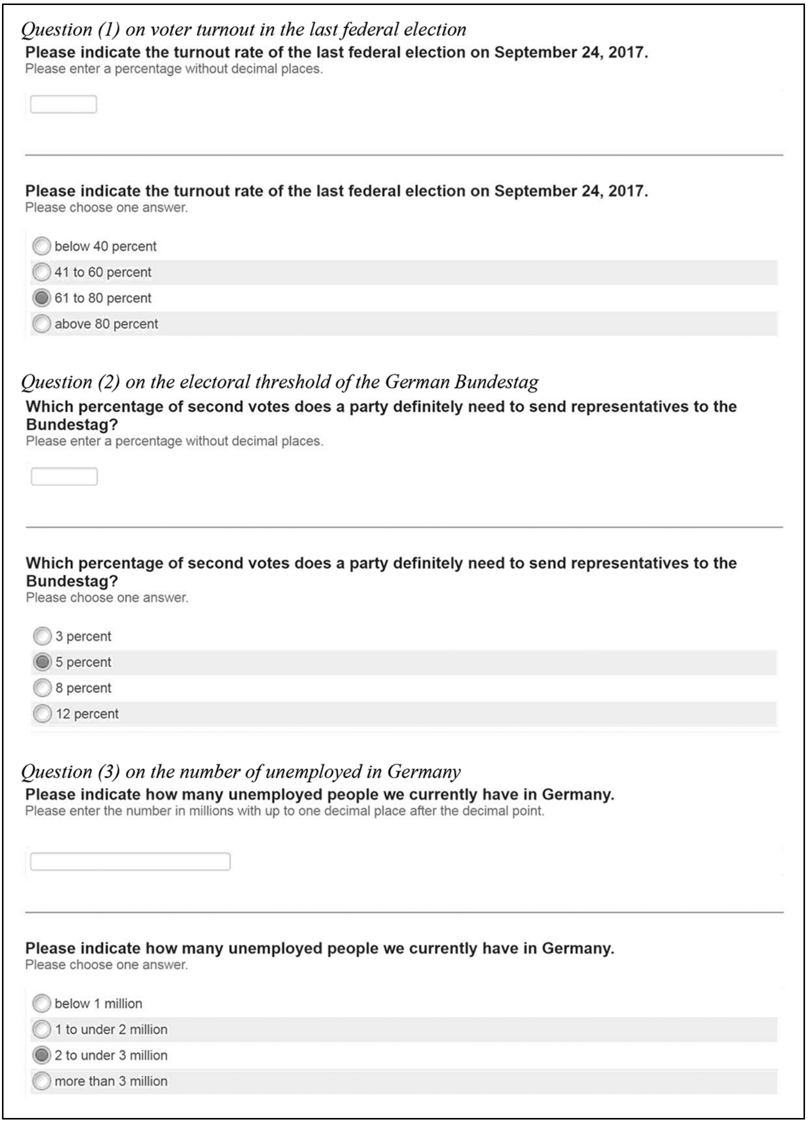
Our experiments were based on three knowledge questions located close to the beginning, the middle, and end of the questionnaire. These experimental questions asked about (1) voter turnout in the last federal election, (2) the electoral threshold of the German Bundestag, and (3) the number of unemployed in Germany. We had 12 survey methodologists from Germany rank the questions regarding the complexity of their content. Based on these expert ratings, question 3 was considered the most complex, followed by question 1, whereas question 2 was deemed quite easy to answer with respect to its content.

The response formats of the experimental questions, and thus the task difficulty, were experimentally varied by independently randomizing the three questions. Using a between-subjects design, we randomly assigned the respondents to one of two experimental groups to investigate the role of task difficulty on switching behavior. We provided the first group with closed-ended response options, and the second group with an open-ended answer field. The closed-ended questions included a set of predefined response options from which the respondents could select from and thus provided additional guidance to respondents regarding the retrieval of relevant information and judgment. Moreover, the closed-ended questions did not require respondents to formulate a response in their own words. Thus, we considered the experimental questions with closed-ended response formats to be less difficult to process and answer compared to the questions with open-ended response formats. In the latter formats, the respondents were required to type their responses in a blank answer field without any suggestions on possible responses. Both the closed-ended and open-ended response formats did not include an explicit "don't know" response option. However, answers to all the survey questions—including the experimental—were voluntary, so respondents could simply skip a question without being prompted if they were not able or willing to provide a substantive answer. Figure 1 provides the English-language versions of the three experimental questions and depicts the response formats for each experimental group. The correct answers are marked for the closed-ended questions. We coded as correct all those answers to the open-ended questions that were within the same range of answers as the correct answer in the closed-ended questions. For instance, regarding the first experimental question that asked for voter turnout in the last German federal election, we coded as correct all the open-ended answers within the range of 61–80 percent. Accordingly, we created three dichotomous variables indicating the correct answers to each of the knowledge questions (0 = *no*, 1 = *yes*).

To analyze the relationship between respondent characteristics and response behavior, we created a set of variables to measure cognitive ability and motivation. Therefore, we created a measure for the respondents' educational level (low, intermediate, and high); and survey evaluation—assessed as interesting, diverse, important for science, long, difficult, and too personal—on fully labeled verbal five-point scales ranging from *very* (4) to *not at all* (0).

Results

For our first step, we looked at the frequency of window switching behavior depending on varying task difficulty. Table 1 details, in line with our expectations, that respondents were more likely to switch away from the web survey in experiments 1 and 3, if an open-ended response format was



Question (1) on voter turnout in the last federal election

Please indicate the turnout rate of the last federal election on September 24, 2017.
Please enter a percentage without decimal places.

Please indicate the turnout rate of the last federal election on September 24, 2017.
Please choose one answer.

- below 40 percent
- 41 to 60 percent
- 61 to 80 percent
- above 80 percent

Question (2) on the electoral threshold of the German Bundestag

Which percentage of second votes does a party definitely need to send representatives to the Bundestag?
Please enter a percentage without decimal places.

Which percentage of second votes does a party definitely need to send representatives to the Bundestag?
Please choose one answer.

- 3 percent
- 5 percent
- 8 percent
- 12 percent

Question (3) on the number of unemployed in Germany

Please indicate how many unemployed people we currently have in Germany.
Please enter the number in millions with up to one decimal place after the decimal point.

Please indicate how many unemployed people we currently have in Germany.
Please choose one answer.

- below 1 million
- 1 to under 2 million
- 2 to under 3 million
- more than 3 million

Figure 1. Three experimental questions with an open-ended (top) and closed-ended response format (bottom; correct answer marked).

Table 1. Window Switching Behavior in Three Web Survey Experiments.

	Response Format			
	Closed-Ended	Open-Ended		
Experiment	Switching (%)	Switching (%)	Cramer's V	χ^2 Test
1	3.13	7.88	.104	24.23 (1)***
2	3.27	4.45	.031	2.13 (1)
3	5.83	9.03	.061	8.29 (1)**

Note: $N = 2{,}247$. χ^2 tests reported as test statistic with *df* in brackets.
*$p < .05$. **$p < .01$. ***$p < .001$.

provided compared to a closed-ended response format that was deemed easier to answer. In experiment 2, we did not find a significant difference in switching behavior between the respondents who answered an open-ended versus closed-ended response format. We assumed this finding was due to the fact that—irrespective of the response format—this question was relatively easy to answer for most of the respondents as our expert ratings also suggested.

In all three experiments, respondents who used a smartphone were less likely to switch windows compared to PC users (all differences significant, $p < .001$). This finding is in line with a study by Berens, Schlosser, and Höhne (2018) who reported less on-device multitasking when completing a survey via a smartphone compared to a PC. Besides the fact that incidence rates for window switching were rather low (see Table 1), only 21% of the respondents used a smartphone. Thus, because we did not experimentally vary the devices that respondents used in this survey, we have to leave it to future studies to analyze in more detail device effects on switching and multitasking behavior.

To investigate respondent characteristics and their effects on switching behavior, we ran logistic regressions on the likelihood of switching away from the web survey. Table 2 details the results of the models for each of the experiments, and a fourth "pooled" model with switching in at least one of the experiments as the dependent variable. In line with our expectations on respondent ability, in experiments 1 and 3 as well as in the pooled model, we found that respondents with a higher educational level were more likely to temporarily switch from the survey compared to low-educated respondents.[1] With respect to experiment 2 with its overall small percentage of switchers, we did not find this effect, which supports our previous notion that the content of this knowledge question made it easy to answer for most of the

Table 2. Logistic Regressions on Window Switching Behavior in Three Web Survey Experiments.

	Experiment			Pooled[a]
	1	2	3	
Independent Variables	Coefficient (*SE*)	Coefficient (*SE*)	Coefficient (*SE*)	Coefficient (*SE*)
Educational level (ref. low)				
Intermediate	.651* (.266)	−.070 (.269)	.828*** (.228)	.473** (.178)
High	1.241*** (.267)	.217 (.283)	1.132*** (.237)	.969*** (.183)
Survey evaluation				
Interesting	.691*** (.182)	.215 (.194)	.456** (.151)	.360** (.124)
Diverse	−.331* (.153)	−.227 (.177)	−.252 (.133)	−.229* (.112)
Important for science	−.082 (.122)	−.145 (.137)	−.144 (.104)	−.068 (.087)
Long	.077 (.092)	.123 (.106)	.065 (.079)	.064 (.066)
Difficult	.243* (.109)	.052 (.135)	.082 (.100)	.119 (.082)
Too personal	−.267* (.115)	−.320* (.138)	−.297** (.101)	−.275*** (.083)
Response format (ref. closed-ended)	1.004*** (.212)	.334 (.227)	.451** (.170)	
Intercept	−5.298*** (.653)	−2.994*** (.660)	−3.681*** (.535)	−2.784*** (.424)
N	2,187	2,187	2,187	2,187
Log. likelihood	−431.256	−349.850	−547.878	−740.037
Pseudo R^2	.078	.018	.047	.033

[a] Dependent variable in pooled model: switching windows in at least one of the three experiments.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3. Relationship between Window Switching Behavior and Providing Correct Answers in Three Web Survey Experiments.

	Correct Answers (%)			
Experiment	No Switch	Switch	Cramer's *V*	χ^2 Test
1	59.38	85.48	.123	33.48 (1)***
2	66.48	77.01	.043	4.19 (1)*
3	37.62	75.90	.205	93.46 (1)***

Note: $N = 2{,}247$. χ^2 tests reported as test statistic with *df* in brackets.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

respondents—despite our experimental variation of task difficulty. Concerning respondent motivation measured by interest in the survey topic, we found that the more interesting respondents perceived the questionnaire, the more likely they were to switch. Again, we found this effect to be significant in experiments 1 and 3 and in the pooled model. Moreover, in all four models, our findings indicated the important role of attitudes toward the survey. If respondents perceived the questionnaire as too personal, they were less likely to switch windows. Furthermore, in experiments 1 and 3, even after controlling for respondent ability and motivation, we found significant effects of the response format (i.e., task difficulty) on the likelihood of switching windows. Those respondents who had to answer questions in an open-ended response format were more likely to switch compared to those who answered questions in a closed-ended response format. This finding is in line with our reasoning that task difficulty affects respondents' likelihood to show over-optimizing response behavior.

Drawing on the fact that each respondent received three knowledge questions, we also investigated within-respondent window switching behavior across the experimental questions. In these analyses, we found window switching in experiment 1 to be strongly correlated with switching again in experiment 2 (Cramer's $V = .31$; $\chi^2 = 209.15$, $p < .001$) and experiment 3 (Cramer's $V = .49$; $\chi^2 = 222.08$, $p < .001$), respectively. Overall, 1.2% of the respondents switched in all three instances, 10.1% switched in one or two instances, and 88.8% never switched. These findings are in line with our previous analyses indicating that a specific subgroup of respondents is more likely to switch windows while answering knowledge questions. Given the characteristics of these respondents, they usually are referred to as *optimizers*.

In the previous discussion, we assumed that window switching behavior in knowledge questions was driven by respondents searching for information to correctly answer the questions. Table 3 shows the relationship between window switching behavior and the proportion of correct answers to the knowledge questions. For all three experiments, we found window switching to be correlated with correctly answering the knowledge question. The differences in the proportion of correct answers among the respondents who switched compared to those who did not switch was 26.1 percentage points in experiment 1, 10.5 percentage points in experiment 2, and 38.3 percentage points in experiment 3 (all differences significant with $p < .05$).[2] These results can be seen as evidence that respondents switch away from surveys to look up relevant information to correctly answer a knowledge question. However, this relationship is not deterministic because not every respondent who switched actually answered correctly. This finding might be an indication that respondents are failing to look up the correct answer or because they are engaged in other multitasking activities such as checking e-mails or updating their social networking status.

Finally, we examined the response times that reflected the overall time respondents took to answer the knowledge questions. In line with the previous research by Höhne and Schlosser (2018), we found that respondents who switched windows had longer response times (in seconds) compared to those who did not switch (experiment 1: $t = -19.1$, $p < .001$; experiment 2: $t = -14.1$, $p < .001$; and experiment 3: $t = -11.1$, $p < .001$). On average, the mean response times in each experiment were between four and six times higher for respondents who switched than for those who did not. This result seems straightforward because the response time measures also comprised the time that respondents spent away from the browser window hosting the web survey, for instance, for the purpose of acquiring information to answer a knowledge question. Accordingly, *absence times* as the time respondents who switched spent away from the survey—on average—accounted for 55.9% ($SD = 22.31$) of their overall response times in experiment 1, 52.5% ($SD = 24.94$) in experiment 2, and 59.4% ($SD = 23.47$) in experiment 3. However, considering *focus times* (i.e., subtracting the absence times from the overall response times) and, thus, looking at the actual time the respondents stayed focused on the survey, the differences between those respondents who switched away and those who did not still remained significant (experiment 1: $t = -4.3$, $p < .001$; experiment 2: $t = -3.1$, $p < .01$; and experiment 3: $t = -5.4$, $p < .001$). Although the differences in focus times were smaller than when using uncorrected response times, with the mean focus times of the switchers 1.5–2.3 times higher than for the non-switchers, the differences were still pronounced. This finding supports our previous reasoning that those respondents who switch windows devote more effort (i.e., time) to thoroughly answering the knowledge questions and try to provide optimal responses.

Conclusion

The present study investigates what motivates respondents who answer knowledge questions in web surveys to invest additional effort in searching external information sources to correctly answer these questions. We conducted three experiments to examine whether task difficulty stimulates respondents' information search behavior. To describe the implications on the answering process of searching external sources for additional information and the response burden when answering knowledge questions, we introduced an extension of the cognitive model of survey response developed by Tourangeau et al. (2000). Also, we drew on satisficing theory (Krosnick 1991, 1999) and differentiated between respondents who progressed through all four steps of the cognitive answering process—which is well-known as *optimizing*—and those respondents who went the extra mile and performed additional cognitive steps by searching external sources for additional information, which is a response strategy we refer to as *over-optimizing*. We found that the knowledge questions that were more difficult to answer and process were more likely to induce window switching behavior. Furthermore, we found respondents who had a higher educational level, who perceived the questionnaire as more interesting, and who were more involved in the survey to be more likely to switch away from the web survey to search for external information to correctly answer a knowledge question. In this regard, we found that these respondents were more likely to over-optimize when the question design was more demanding. This finding supports previous studies on the satisficing theory that respondents' ability and motivation, and task difficulty interact in complex ways (Lenzner 2012; Roßmann et al. 2017). However, we found that an increase in task difficulty due to the more demanding response formats of knowledge questions (i.e., open-ended instead of closed-ended response formats) did not result in more satisficing as predicted by satisficing theory; instead, higher task difficulty led to over-optimizing response behavior (i.e., search for external information). In our view, future studies on satisficing theory should advance this line of research and investigate these interaction effects in more detail. Moreover, we believe that it is important to not only focus on forms of weak and strong satisficing but also on different forms of (over-)optimizing. If we think of attitude and behavioral questions, for example, other response behaviors are conceivable that can be regarded as over-optimizing such as looking up or even modifying previously given answers to be consistent with later ones. When investigating this, future research should systematically extend the cognitive model of survey response proposed by Tourangeau et al. (2000) by the additional steps that may be involved in over-optimizing response behaviors.

Over-optimizing may mean that even if respondents try to give an optimal answer by investing additional effort, this does not necessarily lead to better answers from the researcher's point of view. Instead, depending on what the knowledge question is intended to measure—accessible or procedural memory—the respondents' search for additional information may introduce measurement error. Paradata on window switching can be used to identify those respondents for whom both memory components are measured (i.e., those who switch). The increasing availability of paradata scripts to measure this kind of window switching behavior has provided survey practitioners with the possibility to routinely capture these behaviors and use them for analyses. In addition, it might not be the focus of this study, but we believe that these paradata can be part of a solution to disentangle the measurement of the procedural memory and accessible memory involved in answering knowledge questions.

The present study has several practical implications for survey researchers. Most importantly, our findings suggest that researchers need to be clear on what they want to measure when designing and implementing knowledge questions: accessible memory, procedural memory, or the sum of both. When knowledge questions are supposed to solely measure accessible memory, one needs to take into account that although open-ended questions usually are considered to prevent random guessing and thus better reflect a respondent's actual knowledge, they also are more likely to induce window switching behavior and additional information searching in web surveys. Potentially, respondents may be explicitly instructed not to switch windows to search external sources for additional information. After data collection is completed, the question arises as to what to do with the respondents who switched windows during answering, since their answers are likely to be confounded with procedural memory. The first step to tackle this issue is to investigate whether the data show a relationship between switching behavior and correct answers to the knowledge question. If this relationship is present, researchers may incorporate dummy variables for switching in their substantive models to account for this effect or—a more invasive approach—omit the respective cases from analyses. With respect to the latter, however, it should be noted that studies on speeding (Greszki, Meyer, and Schoen 2015) and inattentive respondents (Gummer, Roßmann, and Silber 2018) have shown that deleting cases did not change the substantive conclusions drawn from multivariate models. The number of switchers and the effect of different correction methods will vary across surveys. Therefore, whenever implementing knowledge questions, we recommend that survey researchers at least examine the relationship between window switching behavior and the correct answering of knowledge questions to assess the risk of potentially confounded measures.

The findings of this study further provide novel insights into how to use paradata to better understand the process of answering knowledge questions. When talking about multitasking, a distinction can be made between *off-device (or nonmedia) multitasking* primarily measured by self-reports and *on-device (or media) multitasking* measured by paradata (Berens, Schlosser, and Höhne 2018; Sendelbah et al. 2016; Zwarun and Hall 2014). Paradata on window switching behaviors in the specific context of knowledge questions may help to subdivide on-device multitasking into off-topic and on-topic activities. Off-topic multitasking behaviors can include a variety of activities, such as reading text messages or e-mails; posting photos, videos, or other media; and other activities that have nothing to do with the web survey. Our study points out that paradata on window switching behaviors also can detect on-topic multitasking behaviors in the specific context of knowledge questions—in our case, searching for additional information to answer the question. In this regard, such kind of over-optimizing response behavior that especially occurs among a specific subgroup of respondents might have adverse effects on the validity of knowledge questions. To follow up on this, it seems worthwhile to use paradata to explore off-topic and on-topic multitasking behaviors in the context of attitudinal and behavioral questions and their implications for data quality. Most importantly, these data could be used to investigate how respondents who switched and answered correctly differed from those who switched and answered incorrectly. Web tracking data are a potential source for in-depth analyses of what respondents do after switching away from a web survey. Combining data from multiple sources by linking web tracking data with survey data and paradata enables researchers to answer important questions about the information sources respondents rely on to answer questions, and how they perform their search. The latter especially can provide an even more detailed measure of procedural memory. Collecting web tracking data often is not as straightforward as setting up a survey because it requires access to suitable tracking software, obtaining the consent of respondents, and may include laborious data management tasks. However, the topic of web tracking has attracted the increasing attention of social science research (e.g., De Vreese et al. 2017; Scharkow 2016) that eventually could result in solutions to these challenges.

The present study is not without limitations that may be addressed by future research projects. First, we focused only on three knowledge questions that asked about key figures on political issues in Germany. We selected these question topics because political knowledge is an essential construct in the social sciences and political science in particular, which may impair the generalizability of our findings with respect to knowledge questions in other fields or subfields. Therefore, we encourage the replication of our study with respect to other topics and with different kinds of knowledge questions. Second, we manipulated the task difficulty by changing the response format of the knowledge questions. Although we have found support that our manipulation was successful, it might be interesting to test other ways to increase task difficulty for the sake of replication. Thus, it might be interesting for future research to experimentally investigate other dimensions of task difficulty by systematically varying the content and position of the question in the questionnaire. The present study serves as a blueprint on how to design such an experimental survey. Third, previous studies have reported an increasing use of smartphones to complete web surveys (e.g., Gummer et al. 2019). Although the experiments of the present study were not designed to investigate device effects—in line with the previous findings on on-device multitasking by Berens et al. (2018)—we found smartphone users to be less prone to switch away from a web survey. In our opinion, further research on device effects is needed but will have to meet the challenges of sample size and self-selection. Fourth and finally, we focused on web surveys as one kind of self-administered surveys. Although web surveys have become increasingly popular and enable the collecting of rich sets of paradata, mail surveys remain an alternative and important mode for conducting self-administered surveys. Similar to web-based questionnaires, respondents can search for external information when completing a paper-based questionnaire. Accordingly, we believe that additional studies on the same issue with respect to mail surveys would be a novel and welcome contribution to this line of research. Observing the response behavior in a mail survey is more challenging and would involve experimentation in a laboratory setting, which has its own problems such as generalizability to other situations and the gathering of a sufficiently large sample to perform the analyses.

Acknowledgments

We would like to thank the anonymous reviewers and editors at *Sociological Methods & Research* as well as the participants of the General Online Research Conference 2018 in Cologne, Germany, for their thoughtful comments and suggestions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tobias Gummer https://orcid.org/0000-0001-6469-7802

Supplemental Material

Supplementary material for this article is available online.

Notes

1. Previous studies (e.g., Holbrook et al. 2007; Knäuper 1999) introduced age as an indicator for respondents' general cognitive abilities. In our study, we primarily referred to task-related ability (i.e., the ability to search for and evaluate information), so for our set of knowledge questions, we assumed education was a better proxy measure than age. Nevertheless, as a robustness check, we recalculated the models predicting window switching behavior with age, sex, and region of residence as control variables. Including these variables did not change the substantive conclusions drawn from the models. Moreover, none of the effects of these variables were significant, except for sex in experiment 2.
2. To test the robustness of our findings, we fitted logistic regressions with answering a knowledge question correctly as dependent variables (see Table A1 in the Online Appendix). Independent variables were the same as in the models presented in Table 2. In addition, we introduced the variables indicating window switching. The results of the robustness checks confirmed that temporarily switching away from the web survey was positively related to correctly answering a knowledge question.

References

Berens, Florian, Stephan Schlosser, and Jan K. Höhne. 2018. "On-device and Off-device Multitasking and Their Effects on Web and Mobile Web Survey Completion." Paper presented at the General Online Research Conference (GOR 18), March 01, 2018, Cologne, Germany.

Callegaro, Mario and Charles DiSogra. 2008. "Computing Response Metrics for Online Panels." *Public Opinion Quarterly* 72:1008-32.

Clifford, Scott and Jennifer Jerit. 2016. “Cheating on Political Knowledge Questions in Online Surveys: An Assessment of the Problem and Solutions.” *Public Opinion Quarterly* 80:858-87.

Couper, Mick P. and Gregg J. Peterson. 2017. “Why Do Web Surveys Take Longer on Smartphones?” *Social Science Computer Review* 35:357-77.

De Vreese, Claes H., Mark Boukes, Andreas Schuck, Rens Vliegenthart, Linda Bos, and Yph Lelkes. 2017. “Linking Survey and Media Content Data: Opportunities, Considerations, and Pitfalls.” *Communication Methods and Measures* 11:221-44.

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, MA: Yale University Press.

Diedenhofen, Birk and Jochen Musch. 2017. “Pagefocus: Using Paradata to Detect and Prevent Cheating on Online Achievement Tests.” *Behavior Research Methods* 49:1444-59.

Dillman, Don A., Jolene D. Smyth, and Leah M. Christian. 2014. *Internet, Phone, Mail, and Mixed-method Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley.

Galston, William A. 2001. “Political Knowledge, Political Engagement, and Civic Education.” *Annual Review of Political Science* 4:217-34.

Greszki, Robert, Marco Meyer, and Harald Schoen. 2015. “Exploring the Effects of Removing ‘Too Fast.’ Responses and Respondents from Web Surveys.” *Public Opinion Quarterly* 79:471-503.

Gummer, Tobias, Franziska Quoß, and Joss Roßmann. 2019. “Does Increasing Mobile Device Coverage Reduce Heterogeneity in Completing Web Surveys on Smartphones?” *Social Science Computer Review* 37:371-84.

Gummer, Tobias, Joss Roßmann, and Henning Silber. 2018. “Using Instructed Response Items as Attention Checks in Web Surveys: Properties and Implementation.” *Sociological Methods & Research*. https://doi.org/10.1177/0049124118769083.

Höhne, Jan K. and Stephan Schlosser. 2018. “Investigating the Adequacy of Response Time Outlier Definitions in Computer-based Web Surveys Using Paradata Surveyfocus.” *Social Science Computer Review* 36:369-78.

Holbrook, Allyson L., Jon A. Krosnick, David Moore, and Roger Tourangeau. 2007. “Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes.” *Public Opinion Quarterly* 71:325-48.

Jensen, Carsten and Jens P. F. Thomsen. 2014. “Self-reported Cheating in Web Surveys on Political Knowledge.” *Quality & Quantity* 48:3343-54.

Kleiner, Brian, Oliver Lipps, and Eliane Ferrez. 2015. “Language Ability and Motivation among Foreigners in Survey Responding.” *Journal of Survey Statistics and Methodology* 3:339-60.

Knäuper, Bärbel. 1999. "The Impact of Age and Education on Response Order Effects in Attitude Measurement." *Public Opinion Quarterly* 63:347-70.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-36.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50:537-67.

Krosnick, Jon A. and Stanley Presser. 2010. "Question and Questionnaire Design." Pp. 263-313 in *Handbook of Survey Research*, edited by P. V. Marsden and J. D. Wright. Bingley, England: Emerald Group.

Lenzner, Timo. 2012. "Effects of Survey Question Comprehensibility on Response Quality." *Field Methods* 24:409-28.

Lenzner, Timo, Lars Kaczmirek, and Mirta Galesic. 2011. "Seeing through the Eyes of the Respondent: An Eye-tracking Study on Survey Question Comprehension." *International Journal of Public Opinion Research* 23:361-73.

Lenzner, Timo, Lars Kaczmirek, and Alwine Lenzner. 2010. "Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment." *Applied Cognitive Psychology* 24:1003-20.

Mavletova, Aigul and Mick P. Couper. 2016. "Grouping of Items in Mobile Web Questionnaires." *Field Methods* 28:170-93.

Mondak, Jeffery J. and Belinda C. Davis. 2001. "Asked and Answered: Knowledge Levels When We Will Not Take "Don't Know" for an Answer." *Political Behavior* 23:199-224.

Nadeau, Richard and Richard G. Niemi. 1995. "Educated Guesses: The Process of Answering Factual Knowledge Questions in Surveys." *Public Opinion Quarterly* 59:323-46.

Narayan, Sowmya and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60:58-88.

Paulhus, Delroy L. 2002. "Socially Desirable Responding: The Evolution of a Construct." Pp. 49-69 in *The Role of Constructs in Psychological and Educational Measurement*, edited by H. I. Braun and D. N. Jackson. Mahwah, NJ: Erlbaum.

Prior, Markus. 2005. "News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout." *American Journal of Political Science* 49:577-92.

Prior, Markus and Arthur Lupia. 2008. "Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills." *American Journal of Political Science* 52:169-83.

Reja, Urša, Katja Lozar Manfreda, Valentina Hlebec, and Vasja Vehovar. 2003. "Open-ended vs. Close-ended Questions in Web Questionnaires." *Developments in Applied Statistics* 19:159-77.

Robison, Joshua. 2015. "Who Knows? Question Format and Political Knowledge." *International Journal of Public Opinion Research* 27:1-21. doi:10.1093/ijpor/edu019

Roßmann, Joss, Tobias Gummer, and Henning Silber. 2017. "Mitigating Satisficing in Cognitively Demanding Grid Questions: Evidence from Two Web-based Experiments." *Journal of Survey Statistics and Methodology* 6:376-400.

Scharkow, Michael. 2016. "The Accuracy of Self-reported Internet Use—A Validation Study Using Client Log Data." *Communication Methods and Measures* 10: 13-27.

Schlosser, Stephan and Jan K. Höhne. 2018. *ECSP—Embedded Client Side Paradata*. (http://doi.org/10.5281/zenodo.1218941).

Sendelbah, Anže, Vasja Vehovar, Ana Slavec, and Petrovčič. Andraž. 2016. "Investigating Respondent Multitasking in Web Surveys Using Paradata." *Computers in Human Behavior* 55:777-87.

Smyth, Jolene D., Don A. Dillman, Leah M. Christian, and Mallory McBride. 2009. "Open-ended Questions in Web Surveys: Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality?" *Public Opinion Quarterly* 73:325-37.

Toepoel, Vera, Marcel Das, and Arthur Van Soest. 2009. "Design of Web Questionnaires: The Effects of the Number of Items Per Screen." *Field Methods* 21:200-13.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133:859-83.

Yan, Ting and Roger Tourangeau. 2008. "Fast Times and Easy Questions: The Effects of Age, Experience, and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology* 22:51-68.

Zwarun, Lara and Alice Hall. 2014. "What's Going On? Age, Distraction, and Multitasking During Online Survey Taking." *Computers in Human Behavior* 41: 236-44.

Author Biographies

Tobias Gummer is a senior researcher and team leader at the GESIS—Leibniz Institute for the Social Sciences. His methodological research interests include survey design, data quality, nonresponse, prevention and correction methods for biases, and analysis of social and individual change. E-mail: tobias.gummer@gesis.org

Tanja Kunz is a senior researcher at the GESIS—Leibniz Institute for the Social Sciences. Her current research interests include issues of visual design and data quality in web surveys, paradata, and questionnaire design. E-mail: tanja.kunz@gesis.org