Survey Research Methods (2023)
Vol. 17, No. 3, pp. 269-300
doi:10.18148/srm/2023.v17i3.7986
European Survey Research Association

Satisficing Response Behavior Across Time: Assessing Negative Panel Conditioning Using an Experimental Design with Six Repetitions

Fabienne Kraemer[1], Henning Silber[1], Bella Struminskaya[2], Matthias Sand[1], Michael Bosnjak[3], Joanna Koßmann[4], and Bernd Weiß[1]

[1]GESIS—Leibniz Institute for the Social Sciences
[2]Utrecht University
[3]Trier University
[4]ZPID—Leibniz Institute for Psychology

Satisficing response behavior can be a threat to the quality of survey responses. Past research has provided broad empirical evidence on the existence of satisficing and its consequences on data quality, however, relatively little is known about the extent of satisficing over the course of a panel study and its impact on response quality in later waves. Drawing on panel conditioning research, we use question design experiments to investigate whether learning effects across waves of a panel study cause changes in the extent of satisficing and if so, whether general survey experience (learning of the survey process) or familiarity with specific question contents (learning of the questionnaire content) accounts for those changes. We use data from a longitudinal survey experiment comprising six panel waves administered within a German non-probability online access panel. To investigate the underlying mechanism of possible learning effects, the experimental study randomly assigned respondents to different frequencies of receiving identical question contents over the six panel waves. Our results show the existence of satisficing in every panel wave, which is in its magnitude similar to the extent of satisficing in the probability-based GESIS Panel that we use as a benchmark study. However, we did not find changes in the extent of satisficing across panel waves, nor did we find moderation effects of the interval between the waves, respondents' cognitive ability, or motivation. Additional validity analyses showed that satisficing does not only affect the distribution of individual estimates by 15 percent or more but also can have an effect on associations between variables.

Keywords: satisficing; learning effects; non-probability panel; response quality; panel conditioning; form-resistant correlation hypothesis

1 Background

Learning effects resulting from repeated interviewing are known as early as 1940 (Lazarsfeld, 1940). In panel studies, these learning effects—also referred to as panel conditioning (Kalton et al., 1989)—can influence response quality in later waves. The effects on response quality can be caused by two different types of learning (Struminskaya, 2016): (a) respondents gain general experience with a survey and with the types of questions asked (i.e., *learning of the survey process*) and (b) respondents become familiar with specific question contents (i.e., *learning of the questionnaire content*). Learning processes in preceding panel waves can cause changes in different response behaviors (Struminskaya & Bosnjak, 2021), such as *satisficing*. Satisficing describes shortcuts in respondents' cognitive response process with the aim to reduce survey burden and cognitive effort (Krosnick, 1991). Past research has demonstrated that various forms of satisficing exist and have negative consequences on survey responses (Roberts et al., 2019). However, the extent of satisficing in panel studies and its consequences on response quality in later waves is much less explored. The few existing longitudinal studies provide mixed evidence on selected indicators of satisficing response behavior: increases or non-significant differences for both non-differentiation (Schonlau & Toepoel, 2015; Sun et al., 2019) and misreporting to filter questions (Bach & Eckman, 2018, 2020; Silber et al., 2019). However, previous studies are mostly non-experimental and leave several other satisficing indicators unaddressed (e.g., saying "don't know", selecting first or middle response options, acquiescence). It remains unclear to which extent previous results can be generalized and whether other indicators produce similar findings.

Our research aims to fill this gap by examining satisficing and its underlying mechanisms in a longitudinal con-

Fabienne Kraemer, GESIS—Leibniz Institute for the Social Sciences, B6 4,5, 68159 Mannheim (E-Mail fabienne.kraemer@gesis.org).

text. Specifically, we investigate the following research questions: (1) Does satisficing response behavior increase or decrease over the course of a panel study? (2) Does learning of the questionnaire contents account for change in satisficing across waves of a panel study? (3) Do different panel intervals, respondents' cognitive ability, and respondents' motivation have an impact on the change in satisficing across panel waves?

To examine our research questions, we conducted a randomized experiment comprising six panel waves that were administered within a German non-probability online access panel. To assess the extent of satisficing response behavior in the panel study, we focus on three satisficing indicators (Krosnick, 1991): choosing the first response option (i.e., primacy effect), saying "don't know" instead of providing substantial answers, and agreeing to given statements in the agree/disagree response format (i.e., acquiescence). The experimental design of the study allows us to assess (a) changes in satisficing over time by conducting within-person comparisons and to examine (b) the underlying mechanism causing the change in satisficing by comparing respondents who received the same (conditioned group) or different question contents (unconditioned group).

We extend previous research by investigating a different set of satisficing indicators over time and, more importantly, by providing experimental evidence on which learning mechanism is responsible for changes in satisficing across waves of a panel study. Our study uses question design experiments with different question versions to measure the extent of satisficing, a method that allows quantifying the amount of satisficing in each wave and has not been used in the context of panel conditioning previously. As a benchmark for the extent of satisficing to compare to our online panel data, we use data of a German probability-based panel study.

2 Satisficing within longitudinal studies

Measurement error is one of the major concerns in survey research. When analyzing survey data, many researchers follow the assumption that respondents provide the best possible answers to survey questions. Ideally, respondents undergo four different stages of cognitive processing when generating an answer to any type of survey question (Tourangeau et al., 2000): (1) comprehension of the question, (2) retrieval of relevant information from memory, (3) formation of a judgment, and eventually, (4) reporting the final answer to the question. If respondents carry out all four stages carefully, they provide *optimized* responses to a survey question (Krosnick, 1991).

However, respondents can take mental shortcuts in the response generation process if they find answering to a survey tiring and burdensome and want to reduce cognitive effort. Such shortcuts are referred to as *survey satisficing* (Krosnick, 1991; Krosnick & Alwin, 1987). Survey satisficing can manifest itself in various response strategies that produce poor quality answers which are completely detached from respondents' true attitudes and values. Common response strategies, which bypass tedious cognitive processing to provide an answer to a question, are (a) choosing first response options irrespective of their content (i.e., primacy effect), (b) selecting mid-responses, (c) simply agreeing with given statements (i.e., acquiescence), (d) non-differentiation of answers in a grid question (i.e., straightlining), or (e) saying "don't know" instead of giving a substantial answer (Krosnick, 1991). Our study focuses on primacy effects, saying "don't know", and acquiescence in detail.

In panel studies, satisficing response behavior can change over time due to learning effects, which result from respondents' participation in previous survey waves (Struminskaya & Bosnjak, 2021). Two different learning mechanisms can cause changes in response behavior in subsequent waves of a study: learning of the survey process and learning of the questionnaire content (Struminskaya, 2016). In the case of learning the survey process, respondents gain general experience with the overall procedure of a survey as they learn about the structure of a questionnaire or how to respond to different types of questions. Learning the questionnaire content refers to respondents becoming familiar with the specific contents of questions in the survey. Both learning mechanisms can result in adapted response strategies that either positively affect the quality of survey responses or have negative consequences for overall response quality in later panel waves.

Past research has mainly documented negative learning effects within panel studies concerning the use of satisficing response strategies. The research suggests that respondents use their general experience of the survey process as well as the acquired knowledge of the questionnaire structure to strategically misreport to questions that would otherwise trigger several follow-up questions (Bach & Eckman, 2018; Kreuter et al., 2011). This response behavior is referred to as *motivated misreporting* (Eckman & Kreuter, 2018; Eckman et al., 2014) and can be considered a special form of satisficing since respondents aim for reducing survey burden by triggering fewer follow-up questions. For example, learning in previous waves fosters the use of this form of satisficing with experienced respondents reporting higher rates of unemployment (Bailar, 1975; Halpern-Manners & Warren, 2012), having fewer household members or being less likely to be a member of a political party (Warren & Halpern-Manners, 2012). Other studies, however, found no change in motivated misreporting over time (Bach & Eckman, 2018, 2020).

Previous studies that investigate further satisficing indicators do not show significant changes in the extent of acquiescence among the same respondents over a 4-year period (Billiet & Davidov, 2008) and provide mixed findings regarding

changes in the extent of straightlining across panel waves: Whereas Sun and colleagues (2019) found no evidence for learning effects causing changes in straightlining over the course of a study, Schonlau and Toepoel (2015) identified increases in the prevalence of straightlining among experienced panelists of the LISS panel.

Schonlau and Toepoel's study suggested that the observed changes in straightlining among experienced respondents are the result of respondents becoming increasingly fatigued and disinterested over the course of a panel study. According to Krosnick (1991), respondents' motivation is one of the central factors that influence the occurrence and extent of satisficing response behavior, and past research has shown that satisficing is indeed more likely to occur if respondents lack the motivation to engage in cognitive processing necessary for answering a question (Kaminska et al., 2010). For panel studies, the assumption of panel fatigue is a central one (Lundmark & Gilljam, 2013): It has been argued that asking respondents identical questions over multiple waves can lead to frustration and lower motivation, increasing the likelihood to satisfice in later waves of the study (Scherpenzeel & Saris, 2017).

Conversely, it can also be argued that repeatedly answering questions with identical content helps with question complexity: Repeated exposure to identical survey content in a panel study, increases respondents' familiarity with specific questions. Respondents become knowledgeable about the issues covered by survey questions as they have repeatedly reflected on the topic or searched for further information. Consequently, we can assume that the perceived difficulty of a response task decreases in subsequent survey waves. The difficulty of tasks also determines the existence and extent of satisficing (Krosnick, 1991) as task difficulty directly influences the ease with which respondents progress through the stages of response processing. It has been shown that satisficing is likely to occur in complex questions, for example, those with wordy question text, uncommon or ambiguous terms, or long and difficult response scales (Alwin & Krosnick, 1991; Krosnick, 1999; Krosnick & Presser, 2010).

Due to their familiarity with the question contents, respondents might therefore, on the other hand, be less susceptible to satisficing response strategies in later waves of a study. Repeated survey participation could thus, lead to either increases or decreases in the use of satisficing response strategies in later waves of a panel study. However, both assumptions on panel fatigue and decreased question complexity with repeated interviewing suggest that respondents show some change in their satisficing response behavior over the course of a study. We hypothesize:

H1 When answering identical questions multiple times, respondents show changes in their level of satisficing across panel waves.

Both cases of change in satisficing in a panel study imply that respondents' repeated exposure to identical question contents causes a shift in response behavior—either toward an increase or a decrease in satisficing. In both cases, we assume that becoming familiar with specific question contents is the underlying mechanism of a change in satisficing across the waves of a study (either by affecting respondent motivation or perceived task difficulty). We therefore assume:

H2 Changes in satisficing across panel waves are caused by learning of the question contents.

Learning effects that cause changes in response behavior are more likely to occur when survey waves are administered closely in time. This is because when waves are only separated by short time intervals, respondents' memories of the survey and question contents are still relatively fresh when they participate in the subsequent survey wave, increasing the likelihood of a change in response behavior. In their research on panel conditioning in longitudinal social science surveys, Halpern-Manners and Warren (2012) report that learning effects are often observed when survey waves are only separated by a month or less. While studies with waves separated by more than one month only show mixed findings, panel conditioning is rarely observed when survey waves are more than one year apart. In line with this, we expect changes in satisficing in this study to be less prevalent when respondents are surveyed within greater panel intervals:

H3 A longer panel interval decreases the change in satisficing across panel waves caused by repeatedly answering identical questions.

The theory of survey satisficing (Krosnick, 1991) assumes that cognitive ability and motivation influence the occurrence and extent of satisficing. Whereas cognitive abilities can facilitate the thorough completion of the response process, respondent motivation reflects the willingness to engage in cognitive processing in the first place. Multiple studies have provided evidence for this assumption showing that respondents with higher cognitive abilities and higher motivation are less likely to satisfice (Holbrook et al., 2007; Kaminska et al., 2010; Narayan & Krosnick, 1996). Based on this broad empirical evidence, we assume that respondents' cognitive ability exerts a similar effect on the occurrence and extent of satisficing response behavior over time in a panel study:

H4 Higher cognitive abilities decrease the change of satisficing across panel waves caused by repeatedly answering identical questions.

Similarly, we expect respondents with higher motivation to be less susceptible to any change in satisficing induced by repeatedly answering similar question contents over several waves:

H5 Higher motivation decreases the change in satisficing across panel waves caused by repeatedly answering identical questions.

2.1 Form-resistant correlation hypothesis

Previous research has explored whether satisficing affects associations of variables of interest with other variables (e.g., Krosnick & Alwin, 1988; Schuman & Presser, 1996). This was often done using the advantages of question design experiments by comparing whether the association of one question form (e.g., the version with agree/disagree response categories) with another variable is significantly different to the association of another question form (e.g., the version with construct specific response categories). In this case, the form-resistant correlation hypothesis would suggest that the association between the variable of interest and another content-related variable is invariant to the changes in the response categories.

3 Methods

3.1 Experimental design

To test our hypotheses, we use data from a survey experiment comprising six panel waves that were administered within a German non-probability online access panel.

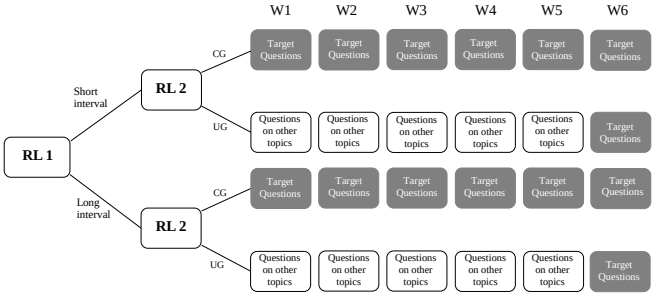
The experimental study involved three randomization levels (see Figure 1 for an illustration of the experimental design). The first randomization level (see "RL 1" in Figure 1, first panel) manipulated the time interval between panel waves to test whether learning effects are more prevalent when the interval between panel waves is short. Respondents were randomly allocated to being either surveyed within a short interval (i.e., monthly; $n = 1,294$) or a longer interval between waves (i.e., about every three months; $n = 1,295$). In the second randomization stage (see " RL 2" in Figure 1, first panel), we manipulated the frequency of receiving identical question contents throughout the study (i.e., conditioning frequency) to enable a differentiation between the two central learning mechanisms that can take place within panel studies (i.e., learning of the questionnaire content vs. learning of the survey process). Specifically, we randomly assigned respondents to one of two groups of which the experimental group received identical target questions (i.e., questions of interest) in each of the six panel waves (i.e., conditioned group), whereas the control group received the target questions only in the last wave of the study (i.e., unconditioned group). Consequently, respondents of the conditioned group were able to learn the content of the recurring questions, while respondents of the unconditioned group were only able to learn the general process of the survey as they received different questions with each survey wave. Conditioned respondents either answered identical question contents every month ($n_{\text{conditioned}} = 441$) or about every three months ($n_{\text{conditioned}} = 442$), whereas unconditioned respondents received the target question contents in the last wave being surveyed either monthly ($n_{\text{unconditioned}} = 441$) or about every three months ($n_{\text{unconditioned}} = 439$) (see Online Appendix A, Table A1 for an overview of the experimental groups).

The third randomization level manipulated the design of the administered target questions to measure the extent of satisficing (see "RL 3" in Figure 1, second panel). In each wave, respondents who received the target questions, were randomly assigned to be asked one of two different versions of six questions related to social attitudes, such as housing or causes of crime. As a result, respondents of the conditioned group did not necessarily receive the same question version throughout the study but randomly received the two different versions of a question with the same content over the course of the six panel waves.

3.2 Data

The fieldwork of the experimental study took place from October 2020 to December 2021. The recruitment of respondents targeted persons aged 18 years and older who were at the time of the study residing in Germany and were registered as a member of a commercial online access panel, which was used as the sampling frame for all study respondents. For the implementation of the study, a quota sample was drawn based on the demographic distribution of the German Microcensus (Mikrozensus). Sample members were selected using cross-quotas on gender (female; male), age (18–35 years; 36–65 years; 66 years and older), and education measured as the highest general school leaving certificate (lower secondary school leaving certificate and lower; secondary school certificate; higher education entrance qualification). Data was collected exclusively online via web-based surveys with each survey wave taking about 15 minutes to complete, except for the first wave of the study that took about 35 minutes to complete due to the inclusion of additional background measures on socio-demographics, political attitudes, and personality.

Selected panel members were explicitly invited to participate in the study via a personalized e-mail by the panel provider. However, every panel member could also participate via the panel's online user interface. Due to this additional option of participating, we do not know how many of the respondents participated in the first wave via a personalized e-mail invitation or via the user interface. Only respondents who completed the first wave, however, were able to participate in the following survey wave. The invited panel members received up to four e-mail reminders. Each reminder was sent after three to four days following the invitation or the preceding reminder e-mail. Respondents received a postpaid incentive after each completed survey wave. The amount of the paid incentive was based on the



median response time of the completed survey, varying from €0.60 to €2.50. Furthermore, respondents received an additional bonus payment of €2 if they successfully participated in all six panel waves.

Across the six waves of the panel study, the completion rates (COMR) (see Callegaro & DiSogra, 2008) varied from about 77% to roughly 97%. Up to the last wave of the study, between 36% and 47% of respondents dropped out of the panel (see Appendix D, Tables D1 and D2 for detailed completion and attrition rates for each wave separated by the different respondent groups; Callegaro and DiSogra, 2008).

Our main analyses to assess changes in satisficing as a result of learning effects in a panel study focus on respondents who were surveyed within the short panel interval (i.e., monthly) and who successfully participated in every wave of the study (analytic sample: $n = 552$), because learning effects are most likely to occur when the time interval between waves is short. Moreover, for the robustness checks that investigate the impact of panel attrition on the extent of satisficing, we draw on the full sample, including respondents who participated in all six panel waves and those who dropped out before the last wave of the study ($n = 1,294$). Lastly, to explore whether the time between waves has an impact on the extent of satisficing over the course of the study, we included data of the respondents who were surveyed within the longer panel interval (i.e., about every three months) and who completed every of the six panel waves ($n = 479$).

Data for benchmark analyses

For benchmark analyses of the extent of satisficing within a probability-based panel study, we additionally used data from the GESIS Panel—a German probability-based mixed-mode panel comprising about 5,000 respondents (GESIS, 2021). The GESIS Panel was initially recruited in 2013, based on a random sample drawn from municipalities' population registers targeting German-speaking persons aged 18 to 70 years who were permanently residing in private households in Germany.

For the benchmark analyses, we used the first survey wave of the GESIS Panel ($n = 4,298$) administered from February to March 2014 in which a study on question design and corresponding response effects was implemented using the identical measures of satisficing indicators as our study (Silber et al., 2018). In this survey wave, the completion rate (COMR) was 88%, whereas the cumulative response rate (CUMR) was 20%. Up to the first wave of the GESIS Panel, 1% of the respondents already dropped out of the panel following the welcome survey (Schaurer et al., 2014).

3.3 Measures

Satisficing response behavior

To capture the extent of satisficing throughout our panel study, we draw on three prominent satisficing response strategies, that is (a) choosing first response options irrespective of the content (i.e., primacy effect), (b) simply agreeing to given statements (i.e., acquiescence), and (c) saying "don't know" instead of providing a substantial answer. To measure these forms of satisficing, we used six between-subject experiments which manipulated different aspects of question design (Silber et al., 2018). Specifically, two question design experiments each manipulated 1) the *order of the response categories*, (2) whether a question was displayed in the *agree/disagree or the construct-specific* response format, or (3) whether a question included a *'don't know' option* or not (see Appendix B, for a detailed overview of the single question design experiments).

Repeatedly answering identical question contents

To investigate whether repeatedly answering identical question contents (i.e., learning of the questionnaire content) causes any changes in satisficing behavior across waves, we used the study's experimental design that manipulated the frequency with which respondents had to answer identical questions over the course of the study. We compared respondents who received identical question contents in every panel wave with respondents who only received the respective question contents in the last wave of the study (see Figure 1).

Panel interval

To explore whether different time intervals between the panel waves influence the extent of satisficing caused by repeatedly answering identical question contents, we compared two groups of respondents who either participated in the single survey waves every month or answered to the survey within a greater interval of about every three months.

Cognitive ability

We used two different measures of cognitive ability. To follow prior research (Holbrook et al., 2007; Narayan & Krosnick, 1996), we first drew on formal education. Based on a respondent's highest general school leaving certificate, we created a variable indicating low, medium, or high education.

In addition, we used a cognitive reflection test comprising three questions to directly measure respondents' actual cognitive performance (Frederick, 2005). Specifically, respondents received three different calculation tasks to which they were asked to provide the correct answer (see Appendix C for detailed question wording and implementation). We constructed a sum score indicating the number of correct answers a respondent gave, ranging from zero (corresponding to low cognitive abilities) to three correct answers (corresponding to high cognitive abilities). As respondents' cognitive reflection test scores were only weakly correlated with their formal education ($r = 0.32$, $p < 0.001$), we included both alternative measures into our analyses of an impact of cognitive ability on the extent of satisficing caused by learning the question contents.

Respondent motivation

We used the NFC-K (Beißert et al., 2014), a German short scale measuring the construct "need for cognition" with four items as a proxy of respondent motivation (Lenzner, 2012; Silber et al., 2019). Respondents were presented with four different statements and asked to indicate to which extent each of the statements applies to them personally on a scale ranging from 1 "Does not apply at all" to 7 "Applies completely" (see Appendix C for detailed question wording). Two of the four items captured respondents' engagement in cognitive tasks whereas the remaining two items measured respondents' enjoyment of cognitive tasks in general. We argue that the extent of a respondent's need for cognition represents their individual motivation to engage in cognitively demanding tasks such as participating in a survey (Kaminska et al., 2010). We constructed a mean score of respondents' need for cognition across the four items ranging from 1 to 7 with higher values representing a higher motivation to properly process and think about survey questions to provide optimal responses.

Content-related variables

To investigate the impact of satisficing on substantive results obtained from the study's data (i.e., associations between variables of interest; see paragraphs on the validity analysis for more information), we drew on several constructs which were substantially related to the issues addressed by the question design experiments (see Appendix C for detailed question wording).

3.4 Analytic strategy

To test *Hypothesis 1*, which predicts a change in satisficing across waves, we solely used data of the conditioned respondent group (CG-short; see Appendix A, Table A1), which received the manipulation of question design in each survey wave. We initially examined the existence and extent of satisficing in each single wave of the panel study by comparing response distributions between the two versions of each question design experiment. To assess the significance of the differences in the responses, we conducted χ^2-tests as well as Fisher's exact tests when the number of observations in individual cells was not sufficiently large. For the main analyses regarding Hypothesis 1, we conducted logistic random-effects panel models with the response distribution of each experiment as the dependent variable of a model (see Appendix E "Analytic strategy" for detailed information on the underlying data structure). We chose random-effects models instead of fixed-effects models due to a considerable number of respondents showing no variation in the dependent variable over time. Hence, to attribute any changes in our latent variable to non-time-constant effects, we used random-effects models. Within each logistic panel model, we estimated the interaction effects of the question design and the separate dummy variables for each panel wave. By including the several interaction effects of question design and the variables for each panel wave, we could assess how response effects change in each following wave compared to Wave 1 (see Table 1 for an overview of the estimated models and interaction terms).

To examine whether learning question contents causes changes in satisficing response behavior across waves (*Hypothesis 2*), we exclusively used data from the last wave of our study. We compared the conditioned respondent group (CG-short), which received identical question contents in each wave of the study, with the unconditioned respondent group (UG-short), which only received the respective question contents in the last wave. For this analysis, we conducted a cross-sectional multilevel logistic regression model, modeling the six question design experiments as different satisficing measurements nested within respondents to estimate the effect of learning the question contents on satisficing response behavior overall (for detailed information on the underlying data structure, see Appendix E "Analytic strategy").

Table 1

Overview of estimated models and interaction effects for each hypothesis

Hypothesis	Database	Estimated model	Estimated interaction effect(s)	Interpretation
H1	Waves 1-6	Logistic random-effects panel model	Question design × wave	Response effects in each wave compared to Wave 1
H2	Wave 6	Cross-sectional multilevel logistic regression model	Question design × conditioning frequency	Response effect for conditioned respondents compared to unconditioned respondents
H3	Wave 6	Cross-sectional multilevel logistic regression model	Question design × conditioning frequency × panel interval	Impact of panel interval on difference in response effects between conditioned and unconditioned respondents
H4	Wave 6	Cross-sectional multilevel logistic regression model	Question design × conditioning frequency × cognitive abilities	Impact of respondents' cognitive ability on difference in response effects between conditioned and unconditioned respondents
H5	Wave 6	Cross-sectional multilevel logistic regression model	Question design × conditioning frequency × motivation	Impact of respondents' motivation on difference in response effects between conditioned and unconditioned respondents

However, the variance measured at the group level is zero likely due to the six experiments being randomly assigned to each person individually. Hence, there is no determinable cluster effect (as tested by calculating the ICC) and therefore no determinable random effect (Donner, 1986). However, we initially conducted cross-sectional multilevel logistic regression models to gain insight into the different variance components and decided against replacing the multilevel models with robust generalized linear models both for reasons of transparency and because multilevel models better reflect the research design (experiments nested within respondents). They enable us to generalize across experiments and types of satisficing instead of conducting multiple individual tests for each experiment. The multilevel model included an interaction effect of question design and conditioning frequency to assess whether response effects are different between conditioned and unconditioned respondents (see Table 1).

To test whether different intervals between panel waves affect the extent of satisficing caused by learning the question contents (*Hypothesis 3*), we again drew on the last wave of the study but additionally included respondents who were surveyed at a greater interval between the panel waves (CG-long and UG-long). By including a dummy variable indicating the panel interval (short vs. long) into our cross-sectional multilevel logistic regression model, we estimated a three-way interaction effect of question design, conditioning frequency, and panel interval. By estimating the three-way interaction, we aimed to assess how different panel intervals affect the difference in response effects between conditioned and unconditioned respondents (see Table 1; for detailed information, see Appendix E "Analytic strategy").

To test our *Hypothesis 4* predicting a negative effect of respondents' cognitive ability on the extent of satisficing caused by repeatedly answering identical question contents, we again estimated a cross-sectional multilevel logistic regression model based on data of the last panel wave and including a three-way interaction of question design, conditioning frequency, and cognitive ability. By doing so, we can assess the moderating effect of respondents' cognitive ability on the differences in response effects between conditioned and unconditioned respondents (see Table 1; for detailed information, see Appendix E "Analytic strategy").

Lastly, to examine the moderating effect of respondent motivation on the extent of satisficing caused by repeatedly answering identical question contents (*Hypothesis 5*), we estimated the equivalent cross-sectional multilevel logistic regression model with a three-way interaction of question design, conditioning frequency, and respondent motivation (see Table 1; for detailed information, see Appendix E "Analytic strategy").

Validity analysis

To assess whether the extent of satisficing in the study affects substantive results, we additionally conducted a va-

lidity analysis. For this, we first used exploratory correlational analyses to identify content-related variables that significantly correlate with the issues addressed in each question design experiment. We set the minimal threshold for correlation at $r = 0.2$ to ensure that we only included content-related variables sufficiently correlating with each respective experiment before estimating the impact of satisficing on their substantive relationship. As the next step, we calculated logistic regression models for each experiment based on the data of Wave 6, including an interaction effect of the selected content-variable and question design to initially assess whether satisficing itself has an impact on the substantial correlations between two variables. Finally, we conducted logistic regression models for each experiment including a three-way interaction effect of question design, the selected content-variable, and conditioning frequency. By that, we assessed whether the impact of the question version on substantial correlations between two variables is different for conditioned or unconditioned respondents. In other words, the estimated interaction effects indicate whether learning the question contents impacted substantial correlations.

Robustness checks

To examine whether respondents who dropped out of the panel systematically differed in their level of satisficing from respondents who did not drop out, we conducted robustness checks comparing the prevalence of satisficing between only those respondents who stayed and the full sample, including respondents who dropped out of the panel before the last wave. We used logistic regression models for every experiment in every wave, including an interaction effect of question design and a dummy variable indicating attrition before the last wave of the study. In case of significant interaction effects, we can assume that respondents who dropped out of the panel showed a different degree of satisficing behavior from those who stayed in the panel.

4 Results

4.1 Extent of satisficing in each wave

Primacy effect

For the first response order experiment on adequate housing in Wave 1, we found a statistically significant primacy effect ($\chi^2(1) = 22.74$; $p < 0.001$) with 78% of respondents indicating that the state and not the individual should be responsible for the provision of housing when this statement appeared first, whereas only 50% indicated that the state should be responsible for housing when it was shown as the last response category. Compared to this response effect of 28.0 percentage points (p.p.), our benchmark analyses on satisficing response behavior within the GESIS Panel showed a weaker but also statistically significant primacy effect of 15.9 p.p. ($\chi^2(1) = 52.77$; $p < 0.001$) (see Table 2). When we extended our analyses across waves (Wave 1 to 6), we found statistically significant and similarly prevalent primacy effects in each wave (ranging from 14.1 to 21.4 p.p.; see Table 2; for more detailed results, see Appendix, Table B1).

However, results of the analysis on changes in satisficing across panel waves did not provide evidence for a change in primacy effect over the course of the panel study (*Hypothesis 1*). Although results of the logistic panel regression showed decreased primacy effects in all waves compared to the first wave, most of the effects were not significant (see Table 3, Column 1 "Adequate housing"). We only found one significant effect for Wave 4 ($\beta = -1.18$; $z = -2.02$; $p = 0.044$), indicating that the primacy effect in the fourth wave was significantly lower than in the first wave of the study. However, the results did not show an overall pattern of significant decrease across waves.

Regarding the second response order experiment on social trust, we did not find a statistically significant primacy effect in Wave 1 ($\chi^2(1) = 1.67$; $p = 0.179$) with 44% of respondents stating that most people can be trusted when this option appeared first and 36% indicating that most people can be trusted when it appeared as the last response option. Comparing this effect of 7.7 p.p. to the equivalent primacy effect found within the GESIS Panel, we see that the identical experiment yielded a smaller but highly significant response effect of 6.5 p.p. ($\chi^2(1) = 19.04$; $p < 0.001$) across respondents of the benchmark study (see Table 2). For the following five waves of our non-probability panel study, we found similar results as in Wave 1 with rather small and non-significant response effects ranging from 2.5 to 4.3 p.p., except for Wave 3, which shows a non-significant but contradictory negative response effect of about 4.0 p.p. ($\chi^2(1) = 0.46$; $p = 0.499$) and Wave 4 showing a significant primacy effect of 12.0 p.p. ($\chi^2(1) = 4.23$; $p = 0.040$; see Table 2; for more detailed results, see Appendix, Table B2).

Assessing the change in satisficing across waves, results of the logistic panel model do not show a pattern of constant increase or decrease in primacy effects across waves. Moreover, we did not find statistically significant differences in the extent of primacy effect compared to Wave 1, so that *Hypothesis 1* is not supported (see Table 3, Column 2 "Trust").

Acquiescence

For the first acquiescence experiment on women in politics in Wave 1, we found a statistically significant response effect ($\chi^2(1) = 11.70$; $p = 0.001$) with 23% indicating that men are emotionally better suited for politics than women in the agree/disagree question format compared to only 8% stating that men are emotionally better suited when construct-specific response options were given. This response effect of 14.9 p.p. was stronger than the response effect of 6.1 p.p. ($\chi^2(1) = 22.46$; $p < 0.001$) found within the GESIS Panel

Table 2

Response effects in each panel wave with comparison to GESIS Panel benchmark

	Response order						Acquiescence						Don't know			
	Adequate housing			Trust			Women in politics			Crime			Courts		Smart leaders	
Question design experiment	Diff.	χ^2	*p*-value	Diff.	χ^2	*p*-value	Diff.	χ^2	*p*-value	Diff.	χ^2	*p*-value	Diff.	Fisher's exact *p*-value	Diff.	Fisher's exact *p*-value
Wave 1	27.98	22.74	0.000	7.67	1.67	0.197	14.86	11.70	0.001	13.34	5.03	0.025	9.31	0.001	16.63	0.000
GP Benchmark	15.85	52.77	0.000	6.54	19.04	0.000	6.05	22.46	0.000	17.60	65.19	0.000	24.77	0.000	17.62	0.000
Wave 2	17.96	8.97	0.003	2.52	0.19	0.667	3.24	0.60	0.439	12.20	4.24	0.040	13.57	0.000	19.90	0.000
Wave 3	21.44	12.85	0.000	−3.98	0.46	0.499	9.03	4.52	0.034	11.65	4.14	0.042	16.77	0.000	22.44	0.000
Wave 4	15.54	6.50	0.011	12.04	4.23	0.040	13.51	9.55	0.002	0.41	0.01	0.943	17.18	0.000	14.43	0.000
Wave 5	17.28	8.11	0.004	4.33	0.54	0.463	8.29	3.68	0.055	6.62	1.39	0.238	17.93	0.000	17.78	0.000
Wave 6	14.06	5.42	0.020	4.13	0.47	0.491	16.83	13.07	0.000	13.90	6.08	0.014	13.23	0.000	10.67	0.000

For some waves, the responses to the version of the "don't know" experiments without "don't know" response option did not comprise any item nonresponses. To statistically test the differences between the two question versions with respect to non-substantial vs. substantial answers, we converted a single observation to item nonresponse for each of those experiments.

(see Table 2). Over the next five waves, we found similar results with significant acquiescence bias in most waves and effect sizes ranging from 3.2 to 16.8 p.p. (see Table 2; for more detailed results, see Appendix, Table B3).

However, again we did not find significant changes in response effects across waves (see Table 3, Column 3 "Women in politics"). In addition, the results show no consistent pattern of increase or decrease over time, so that *Hypothesis 1* is not supported.

Regarding the second acquiescence experiment on causes of crime, we again found statistically significant response effects in the first wave of the study ($\chi^2(1) = 5.03$; $p = 0.025$). 67% of respondents reported that individuals are responsible for crime and lawlessness in the country in the agree/disagree format of the question, whereas only 54% indicated that individuals are responsible when they had to choose from explicitly stated construct-specific response categories. The resulting response effect of 13.3 p.p. was similar to the one found within the GESIS Panel (17.6 p.p. ($\chi^2(1) = 65.19$; $p < 0.001$; see Table 2). Results of the subsequent panel waves also showed significant effects in the majority of waves with a range in effect sizes from 0.4 p.p. to 13.9 p.p. (see Table 2; for more detailed results, see Appendix, Table B4).

Again, when investigating whether satisficing changes across waves (H1), we do not find an overall pattern of increase or decrease in the extent of acquiescence across waves and further, no significant differences between the response effects occurring in each wave compared to Wave 1 (see Table 3, Column 4 "Crime").

Saying "don't know"

With respect to the first experiment which manipulated the existence of a "don't know" response option, we found a statistically significant response effect in the first wave ($p = 0.001$) with 10% saying "don't know" instead of providing a substantial answer when a "don't know" response option was given, whereas only 1% chose to opt for nonresponse rather than providing a substantial answer when there was no "don't know" option offered. This response effect of 9.3 p.p. was weak compared to the benchmark study, which showed a response effect of 24.8 p.p. ($p < 0.001$; see Table 2). A possible explanation for this might be that "don't know" response categories are rarely used in the GESIS Panel, which might increase respondents' attention to this category and encourages them to say "don't know" instead of providing a substantial answer. For the subsequent five waves of our study, we found similar results with significant response effects regarding saying "don't know", varying from 13.2 to 17.9 p.p. (see Table 2; for more detailed results, see Appendix, Table B5).

Results of the logistic panel regression did not show a clear pattern of increase or decrease across waves. Thus, similar to our previous findings, we did not find significant differences in the extent of saying "don't know" between the panel waves, so that *Hypothesis 1* is not supported (see Table 3, Column 5 "Courts").

For the second "don't know" experiment on smart leaders we found a statistically significant response effect in the first wave of the study ($p < 0.001$). 17% of the respondents said "don't know" when the respective response category was offered compared to only 1% of respondents who opted for item nonresponse instead of providing a substantial answer when the "don't know" category was not given, resulting in a 16.6 p.p. response effect. For this experiment, the equivalent response effect within the benchmark study was similarly strong (17.6 p.p. ($p < 0.001$); see Table 2). The remaining five waves of our study showed consistently significant response effects of similar magnitude, ranging from 10.7 to 22.4 p.p. (see Table 2: for more detailed results, see Appendix, Table B6).

Consistent with the results of the preceding five question design experiments, we did not find an overall pattern of increase or decrease in saying "don't know" across waves. Again, response effects in subsequent panel waves were not significantly different from Wave 1, indicating that satisficing did not substantially change over the course of our panel study, so that *Hypothesis 1* was not supported (see Table 3, Column 6 "Smart leaders").

4.2 Learning mechanism underlying satisficing response behavior over time

The cross-sectional multilevel model estimating an overall response effect per respondent (as the single question design experiments were nested within respondents) as well as its difference between conditioned and unconditioned respondents shows similar results compared with the analyses within panel waves. Whereas the cross-sectional multilevel model provides additional evidence on the occurrence of satisficing in our data ($\beta = 0.53$; $z = 5.09$; $p < 0.001$; see Model 2 "Response effect" in Table 4), we did not find a significant difference in the extent of satisficing between the differently conditioned respondents ($\beta = -0.02$; $z = -0.15$; $p = 0.879$; see Model 3 "Learning of contents" in Table 4). This leads to a lack of support for *Hypothesis 2*.

Panel interval

Comparing respondents of the short interval (i.e., about one month between the waves) to respondents of the long interval (i.e., about three months), we did not find differences in the extent of satisficing between the two respondent groups ($\beta = 0.08$; $z = 0.52$; $p = 0.604$). Whereas the results again provide clear evidence for the existence of response effects due to satisficing in both intervals ($\beta = 0.54$; $z = 5.09$; $p < 0.001$), we did not find a statistically significant effect of panel interval on the change in satisficing caused by learning of the questionnaire content ($\beta = -0.11$; $z = -0.51$;

Table 3

Logistic random-effects panel regression models of question design experiments with response effects in each wave compared to Wave 1

	Response order		Acquiescence		Don't know	
	Adequate housing	Trust	Women in politics	Crime	Courts	Smart leaders
Question design experiment	State is responsible	Most people can be trusted	Most men better suited	Individuals	Item nonresponse	Item nonresponse
Question design×Wave 2	−1.005	−0.845	−0.597	0.104	0.995	−0.038
	(0.590)	(0.743)	(0.847)	(0.470)	(1.737)	(1.587)
Question design×Wave 3	−0.700	0.114	−0.151	0.219	0.185	0.534
	(0.580)	(0.726)	(0.864)	(0.471)	(1.687)	(1.599)
Question design×Wave 4	−1.179*	0.935	−0.106	−0.469	−0.054	0.098
	(0.585)	(0.750)	(0.851)	(0.470)	(1.687)	(1.580)
Question design×Wave 5	−1.026	−0.596	−1.125	−0.203	0.928	0.590
	(0.583)	(0.748)	(0.831)	(0.474)	(1.695)	(1.578)
Question design×Wave 6	−0.727	0.346	0.456	0.086	1.248	−0.379
	(0.576)	(0.740)	(0.839)	(0.479)	(1.719)	(1.597)
$ln(\sigma_u^2)$[a]	2.188***	3.680***	2.775***	1.098***	2.382***	1.630***
	(0.164)	(0.226)	(0.194)	(0.167)	(0.231)	(0.252)
Constant	0.085	−2.515***	−5.817***	0.247	−8.827***	−7.015***
	(0.322)	(0.602)	(0.598)	(0.251)	(1.263)	(1.153)
Wald	105.25	28.53	62.20	42.52	67.14	78.97
Observations	1,615	1,626	1,618	1,619	1,626	1,626
Number of respondents	271	271	271	271	271	271

Standard errors in parentheses
[a] Logged variance of the random effect
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

$p = 0.610$), so that *Hypothesis 3* is not supported (see Model 4 "Panel interval" in Table 4).

Cognitive ability

We find no significant impact of cognitive ability on the change in satisficing caused by learning of the questionnaire content for either respondents' actual cognitive performance at a cognitive reflection task nor for respondents' level of formal education (cognitive performance: $\beta = -0.15$; $z = -1.04$; $p = 0.299$; education, middle compared to low: $\beta = 0.20$; $z = 0.51$; $p = 0.608$; high compared to low: $\beta = -0.23$; $z = -0.63$; $p = 0.530$). Noteworthy, in addition to finding no significant effect of cognitive performance and education on the change in satisficing caused by learning of the questionnaire content, we did not find a significant effect of respondents' cognitive ability on the extent of satisficing in general (cognitive performance: $\beta = 0.00$; $z = 0.04$; $p = 0.970$; education, middle compared to low: $\beta = -0.22$; $z = -0.84$; $p = 0.403$; high compared to low: $\beta = 0.04$; $z = 0.13$; $p = 0.894$). Accordingly, the findings did not support *Hypothesis 4* and contradicted the repeatedly assumed effect of cognitive ability on satisficing (see Models 5 "Cognitive performance" and Model 6 "Education" in Table 4).

Motivation

The cross-sectional multilevel model estimating a three-way interaction effect of question design, conditioning frequency, and respondent motivation to assess the impact of respondent motivation on the change in satisficing did not show a significant effect. In particular, we did not find evidence for an effect of respondent motivation on the existence of satisficing in general ($\beta = -0.17$; $z = -1.55$; $p = 0.122$) nor the effect of respondent motivation on the change in satisficing caused by repeatedly answering identical question contents ($\beta = -0.08$; $z = -0.53$; $p = 0.593$). Accordingly, our findings did not support *Hypothesis 5* (see Model 7 "Motivation"

Table 4

Cross-sectional multilevel logistic regression models of satisficing response behavior per respondent with overall response effects for different conditioning frequencies and the moderators

	M1	M2	M3	M4[a]	M5	M6	M7
	Null model	Response effect	Learning of contents	Panel interval	Cognitive performance	Education	Motivation
Question design (QD)		0.531***	0.543***	0.543***	0.531***	0.605**	1.298**
(Ref.: Control version)	-	(0.075)	(0.107)	(0.107)	(0.141)	(0.188)	(0.501)
Conditioning freq. (Ref.:	-	-	0.093	0.093	0.029	0.243	0.204
Unconditioned) Conditioned	-	-	(0.112)	(0.112)	(0.153)	(0.195)	(0.538)
QD×cond. freq.	-	-	−0.023	−0.023	0.126	−0.009	0.318
	-	-	(0.151)	(0.151)	(0.205)	(0.260)	(0.709)
Panel interval (Ref.: Short)	-	-	-	0.085	-	-	-
Long	-	-	-	(0.115)	-	-	-
QD×interval	-	-	-	0.081	-	-	-
	-	-	-	(0.157)	-	-	-
Cond. freq.×interval	-	-	-	−0.091	-	-	-
	-	-	-	(0.162)	-	-	-
QD×cond. freq.×interval	-	-	-	−0.112	-	-	-
	-	-	-	(0.220)	-	-	-
Cognitive ability	-	-	-	-	0.035	-	-
(Sum score reflection task)	-	-	-	-	(0.072)	-	-
QD×cog. ability	-	-	-	-	0.004	-	-
	-	-	-	-	(0.099)	-	-
Cond. freq.×cog. ability	-	-	-	-	0.058	-	-
	-	-	-	-	(0.103)	-	-
QD×cond. freq.×cog. ability	-	-	-	-	−0.146	-	-
	-	-	-	-	(0.140)	-	-
Education (Ref.: Low)	-	-	-	-	-	0.118	-
Medium	-	-	-	-	-	(0.199)	-
High	-	-	-	-	-	0.155	-
	-	-	-	-	-	(0.192)	-
QD×medium	-	-	-	-	-	−0.222	-
	-	-	-	-	-	(0.266)	-

Continues on next page

Continued from last page

	M1	M2	M3	M4	M5	M6	M7
	Null model	Response effect	Learning of contents	Panel interval	Cognitive performance	Education	Motivation
QD×high	-	-	-	-	-	0.035	-
	-	-	-	-	-	(0.259)	-
Cond. freq. × medium	-	-	-	-	-	−0.286	-
	-	-	-	-	-	(0.285)	-
Cond. freq. × high	-	-	-	-	-	−0.184	-
	-	-	-	-	-	(0.266)	-
QD×cond. freq. × medium	-	-	-	-	-	0.195	-
	-	-	-	-	-	(0.380)	-
QD×cond. freq. × high	-	-	-	-	-	−0.227	-
	-	-	-	-	-	(0.361)	-
NFC (mean score)[b]	-	-	-	-	-	-	0.085
	-	-	-	-	-	-	(0.081)
QD×NFC	-	-	-	-	-	-	−0.166
	-	-	-	-	-	-	(0.107)
Cond. freq.×NFC	-	-	-	-	-	-	−0.022
QD×cond. freq.×NFC	-	-	-	-	-	-	−0.082
	-	-	-	-	-	-	(0.153)
$\sigma^2_{\text{Group level}}$[c]	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Constant	−0.744***	−1.023***	−1.069***	−1.069***	−1.097***	−1.162***	−1.459***
	(0.037)	(0.056)	(0.079)	(0.079)	(0.105)	(0.141)	(0.379)
Observations	3,304	3,304	3,304	6,175	3,286	3,304	3,298
Number of respondents	552	552	552	1,031	549	552	551

Standard errors in parentheses;

[a] Based on different sample by including both respondents of the short and long panel interval

[b] Need for cognition

[c] The variance measured at the group level is zero likely due to the six experiments being randomly assigned to each person individually.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

in Table 4).

4.3 Validity analysis

For the validity analysis, we only included correlations with at least $r = 0.2$ to investigate the impact of satisficing on substantial results. Only for three out of the six question design experiments we find sufficient correlations larger than $r = 0.2$. Both response order experiments and the first acquiescence experiment substantially correlated with a content-related variable. For the first response order experiment on adequate housing, we found a correlation with attitudes towards provision of jobs by the government ($r = 0.23$), for the second response order experiment on social trust, we found a correlation with an item of the Big Five Inventory (BFI-10) (Rammstedt et al., 2014), indicating whether a respondent easily trusts others ($r = 0.45$). Lastly, for the first acquiescence experiment on women in politics, we found a correlation with attitudes towards women's full-time employment and family life of $r = 0.25$.

The validity analysis conducted for these three experiments show a statistically significant effect of satisficing on the bivariate association of the first response order experiment on public housing and the content-related variable attitudes towards job provision by the government ($\beta = 0.32$; $z = 2.95$; $p = 0.003$) (for detailed results, see Table 5, Model 1). However, we did not find significantly different effects of satisficing between conditioned and unconditioned respondents on the correlation of two content-related variables (adequate housing: $\beta = 0.10$; $z = 0.43$; $p = 0.665$; trust: $\beta = -0.54$; $z = -1.34$; $p = 0.217$; women in politics: $\beta = -0.11$; $z = -0.37$; $p = 0.709$; for detailed results, see Table 5, Models 4 to 6 for the respective experiment). This indicates that learning the question contents across waves does not cause changes in the degree of satisficing which in turn could differently affect substantial results.

4.4 Robustness check

The robustness checks did not show a statistically significant difference in the extent of satisficing between respondents who dropped out of the panel and those who stayed. Thus, our findings are likely not affected by panel attrition; respondents who participated in all six panel waves are probably not significantly different from respondents who dropped out with respect to their satisficing response strategies (for detailed results, see Table 6).

5 Discussion

Satisficing is a known threat to the quality of survey responses. In this paper, we investigated satisficing response behavior and its underlying mechanisms within panel studies. We used data from a survey experiment within a German non-probability online access panel that manipulated the frequency with which respondents received identical questions over the course of six panel waves to examine whether satisficing response behavior changes across waves and whether learning of the questionnaire content (i.e., answering the same questions over multiple waves) is responsible for these changes. To measure the extent of satisficing, we used question design experiments indicating the extent of primacy effects, acquiescence, and saying "don't know".

Overall, we show that the panel data is considerably affected by the three forms of satisficing that we studied—choosing first response options, acquiescence, and saying "don't know"—and that the magnitude of satisficing in our data was similar to the magnitude of satisficing in the probability-based panel that we used as a benchmark study. However, we did not find support for the existence of learning effects across panel waves that cause changes in satisficing response behavior: First, we did not find significant changes in satisficing across the six panel waves. Second, we did not find significant differences in the extent of satisficing between the two experimental groups which differed in the frequency of receiving identical question contents (six times vs. one time). This finding indicates that learning of the questionnaire content is not responsible for variations in satisficing across waves. Furthermore, we did not find a significant impact of panel interval or respondent characteristics such as cognitive ability and motivation on changes in satisficing.

Consistent with previous studies (e.g., Bach & Eckman, 2018, 2020; Billiet & Davidov, 2008; Sun et al., 2019), our research provides additional support for the absence of changes in satisficing response behavior over the course of panel studies. Contrary to our findings, two previous studies by Schonlau and Toepoel (2015) and Warren and Halpern-Manners and Warren (2012) found evidence for learning effects in satisficing across waves of panel studies. However, these two studies investigated satisficing within panel studies for different indicators, such as straightlining and motivated misreporting. In contrast, results of the only previous study that investigated one of our indicators, acquiescence, similarly showed no changes in the tendency to answer affirmatively over time (Billiet & Davidov, 2008).

One of the strengths of our study is that the question design experiments allowed us to directly assess the extent of satisficing for each experiment in each wave. Thus, our study is one of the first that can quantify the extent of satisficing in each wave and illustrates that it often reached 10 percent or more of the given responses. Due to these strong response effects, distributions of variables are strongly affected by satisficing. Yet, researchers are often interested in associations between variables rather than univariate distributions. To address this aspect, we implemented a validity analysis in which we tested whether satisficing affects bivariate as-

Table 5

Validity analyses: Logistic regression models of question design experiments in Wave 6 with effect of question format on substantive correlations

	Response order		Acquiescence	Response order		Acquiescence
	Adequate housing (1)	Trust (2)	Women in politics (3)	Adequate housing (4)	Trust (5)	Women in politics (6)
Content-related variable	0.211**	0.899***	0.295**	0.131	0.705***	0.250
	(0.066)	(0.143)	(0.114)	(0.091)	(0.180)	(0.154)
Question Design (QD)	−0.391	−0.586	0.133	−0.248	−1.306	−0.418
(Ref.: Control version)	(0.431)	(0.753)	(0.726)	(0.592)	(1.062)	(0.990)
QD×content-related variable	0.316**	0.225	0.164	0.278	0.445	0.205
	(0.107)	(0.213)	(0.146)	(0.143)	(0.295)	(0.204)
Conditioning frequency	-	-	-	−0.924	−1.414	−0.774
(Ref.: Unconditioned)	-	-	-	(0.590)	(1.078)	(1.150)
Cond. freq.×content-related variable	-	-	-	0.166	0.507	0.103
	-	-	-	(0.134)	(0.304)	(0.231)
QD×cond. freq.	-	-	-	−0.348	1.738	1.235
	-	-	-	(0.883)	(1.558)	(1.475)
QD×cond. freq.×content-related variable	-	-	-	0.095	−0.544	−0.110
	-	-	-	(0.220)	(0.441)	(0.296)
Constant	−0.682*	−3.406***	−3.471***	−0.229	−2.875***	−3.127***
	(0.291)	(0.507)	(0.562)	(0.406)	(0.652)	(0.734)
Observations	518	551	528	518	551	528

Standard errors in parentheses
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

sociations (form-resistant correlation hypothesis). We show that satisficing can affect associations between variables; this finding increases concern with respect to the high degree of satisficing that we found in our panel data.

Our study has several limitations that need to be considered when generalizing our findings to other panel studies. First, we rely on data from respondents of a non-probability panel, and these respondents might show specific response behaviors with limited generalizability to the general population (Chang & Krosnick, 2009). However, our analyses show a similar degree of satisficing within the benchmark probability-based panel. This finding is in line with previous research that did not find significant differences between respondents of probability and non-probability-based studies for most satisficing indicators (Cornesse & Blom, 2020). Nevertheless, the comparison of the extent of satisficing between the non-probability study and the probability-based benchmark study only focuses on one single panel wave due to the data available. Thus, there was no benchmark for a change in satisficing across waves in a probability-based panel study. Moreover, for a few of the question design experiments, the differences in the magnitude of satisficing between the non-probability and probability-based study were large (e.g., experiment on "Courts" with the response effect being more than twice as large in the probability-based panel). We further believe that the high experience level of members of a non-probability online panel could have affected the generalizability of the study's conclusion: The changes across panel waves could be underestimated since respondents have likely already participated in many surveys and their level of satisficing could have become more stable compared to less experienced respondents. Future experimental research should investigate satisficing over time in probability-based panels to further examine found differences and the overall generalizability of our results. As an additional limitation, one could argue, that the underlying data does not meet the requirements to be used in random-effects models as it stems from a non-probabilistic online-access panel (see e.g., Gelman & Hill, 2006). We therefore tested whether fixed-effects models would produce compara-

Table 6

Robustness check: Comparison of extent of satisficing in each wave between respondents who participated in all waves of the study and the full sample including respondents who dropped out of the panel

		Response order						Acquiescence						Don't know	
		Adequate housing			Trust			Women in politics			Crime			Courts[a]	Smart leaders[a]
Experiment		Diff.	*z*	*p*	Diff.	*z*	*p*	Diff.	*z*	*p*	Diff.	*z*	*p*	Diff.	Diff.
Wave 1	Non-Att.[b]	27.98	−1.45	0.147	7.67	−1.42	0.156	14.86	−0.33	0.740	13.34	−1.28	0.199	9.31	16.63
	Full[c]	22.41			3.02			13.29			9.55			12.61	17.82
Wave 2	Non-Att.[b]	17.96	−0.13	0.900	2.52	−0.31	0.757	3.24	0.61	0.541	12.2	0.33	0.742	13.57	19.90
	Full[c]	17.31			1.92			5.94			12.94			15.37	20.83
Wave 3	Non-Att.[b]	21.44	1.10	0.272	−3.98	1.17	0.243	9.03	0.83	0.408	11.65	−0.43	0.667	16.77	22.44
	Full[c]	22.19			−1.44			11.07			10.19			18.52	23.60
Wave 4	Non-Att.[b]	15.54	0.81	0.418	12.04	1.16	0.245	13.51	−1.34	0.181	0.41	0.81	0.420	17.18	14.43
	Full[c]	16.55			13.73			13.07			1.72			16.76	15.04
Wave 5	Non-Att.[b]	17.28	−1.18	0.238	4.33	−0.32	0.751	8.29	0.66	0.507	6.62	−1.02	0.306	17.93	17.78
	Full[c]	15.40			3.77			9.45			5.00			17.25	17.25

The *z*-statistic and *p*-values are results of logistic regression models estimating an interaction effect of question design × attrition (0 /1).

[a] Interaction effects for both don't know experiments could not be estimated because there were no observations of item nonresponse in the question version without "don't know" response option among respondents who dropped out. [b] Non-attriters [c] Full sample including respondents who dropped out of the panel

ble results. In fact, our central findings still hold when using such models. However, due to our research question, it was necessary to use random-effects models to attribute any change in our latent variable to non-time-constant effects, knowing that such an approach might introduce bias into our results.

Another limitation is that respondents saying "don't know" might not necessarily satisfice but have no opinion on a topic and select this response category after they invested sufficient effort in the cognitive process of response generation (see, for example, Converse, 1970, on non-attitudes). Additionally, our study focused on only three out of a range of satisficing indicators. Future research should experimentally investigate whether the findings of the present study can be replicated with other indicators of satisficing response behavior, such as straightlining, choosing the middle response option, or speeding (Borgers et al., 2004; Kim et al., 2019; Zhang & Conrad, 2014). Moreover, our study uses a rather crude measure of respondent motivation by drawing on the concept of need for cognition. Need for cognition only serves as a proxy and does not directly capture a respondent's motivation to answer the survey questionnaire. Consequently, future studies should investigate whether our findings regarding the influence of respondent motivation on satisficing across waves are reproducible with more task-specific measures of respondents' motivation.

Future studies with online panels could further make use of the multitude of possibilities to employ innovative designs to test and measure satisficing, for example, by using paradata, such as timestamps and mouse movements as objective indirect measures of satisficing or by exploring possibilities of responsive designs such as tailored messages to respondents to reduce satisficing in longitudinal studies. As the between-subject design of our study only allows for comparisons between experimental groups and not for an investigation of situational factors causing changes in satisficing within respondents, future work could consider employing within-subject designs, for example, based on MTMM-experiments to disentangle the effects of respondent traits and situational factors on satisficing across panel waves.

Nevertheless, the present research contributes to the literature by providing experimental evidence on an underexplored phenomenon: changes in satisficing response behavior in longitudinal studies. Our findings of relatively large satisficing effects that do not diminish over the panel waves suggest that satisficing poses a continuous problem for panel studies. For survey practice, the considerable magnitude of response effects in all waves of the study implies that specific question designs foster the occurrence of satisficing response behavior and should be avoided wherever possible. For example, survey practitioners should try to replace questions using agree/disagree response options with questions using construct-specific response options, omit "don't know" response options if not absolutely necessary, and display grid questions in an item-by-item format (Krosnick et al., 2002; Roßmann et al., 2018; Saris et al., 2010). Furthermore, even experienced respondents, who participate in all waves of a longitudinal study still seem to have a relatively high likelihood of using satisficing as a response strategy. Our findings should encourage survey practitioners to systematically monitor the extent of satisficing in their panel studies by continuously checking response patterns in problematic question designs along with other indicators such as response time measurements. Additionally, question design experiments, similar to ours, could be part of a data quality reporting, which would allow to quantify the response effects in each wave. In general terms, incorporating a standardized response quality reporting that is comparable across panel waves and ideally across studies could be helpful for both data users and study designers.

References

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods & Research*, *20*(1), 139–181.

Bach, R. L., & Eckman, S. (2018). Motivated misreporting in web panels. *Journal of Survey Statistics and Methodology*, *6*(3), 418–430.

Bach, R. L., & Eckman, S. (2020). Rotation group bias in reporting of household purchases in the US Consumer Expenditure Survey. *Economics Letters*, *187*, 108889. https://doi.org/10.1016/j.econlet.2019.108889

Bailar, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, *70*(349), 23–30.

Beißert, H., Köhler, M., Rempel, M., & Beierlein, C. (2014). Eine deutschsprachige Kurzskala zur Messung des Konstrukts Need for Cognition [Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).]. https://doi.org/10.6102/zis230

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, *36*(4), 542–562.

Borgers, N., Sikkel, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and Quantity*, *38*(1), 17–33.

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*(5), 1008–1032.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, *73*(4), 641–678.

Converse, P. E. (1970). Attitudes and Non-Attitudes: Continuation of a Dialogue. In E. R. Tufte (Ed.), *The Quantitative Analysis of Social Problems* (pp. 168–189). Addison-Wesley.

Cornesse, C., & Blom, A. G. (2020). Response quality in nonprobability and probability-based online panels. *Sociological Methods & Research*. https://doi.org/10.1177/0049124120914940

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review/Revue Internationale de Statistique*, 67–82.

Eckman, S., & Kreuter, F. (2018). Misreporting to looping questions in surveys: Recall, motivation and burden. *Survey Research Methods*, *12*(1), 59–74.

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., & Presser, S. (2014). Assessing the mechanisms of misreporting to filter questions in surveys. *Public Opinion Quarterly*, *78*(3), 721–733.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25–42.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

GESIS. (2021). GESIS Panel—Standard Edition [GESIS Data Archive]. https://doi.org/10.4232/1.13785

Halpern-Manners, A., & Warren, J. R. (2012). Panel conditioning in longitudinal studies: Evidence from labor force items in the current population survey. *Demography*, *49*(4), 1499–1519.

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly*, *71*(3), 325–348.

Kalton, G., Kasprzyk, D., & McMillen, D. (1989). *Panel surveys*. Wiley.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74*(5), 956–984.

Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review*, *37*(2), 214–233.

Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The effects of asking filter questions in interleafed versus grouped format. *Sociological Methods & Research*, *40*(1), 88–104.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236.

Krosnick, J. A. (1999). Maximizing questionnaire quality. *Measures of Political Attitudes*, *2*, 37–58.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51*(2), 201–219.

Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*, *52*(4), 526–538.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Michael Hanemann, W., Kopp, R. J., Cameron Mitchell, R., Presser, S., Ruud, P. A., & Kerry Smith, V. (2002). The impact of" no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, *66*(3), 371–403.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In *Handbook of survey research*. Education Emerald.

Lazarsfeld, P. F. (1940). "panel" studies. *The Public Opinion Quarterly*, *4*(1), 122–128.

Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field Methods*, *24*(4), 409–428.

Lundmark, S., & Gilljam, M. (2013). Participation effects in panel surveys: Evidence from a seven-wave randomized experiment. *Annual conference of American Association for Public Opinion Research*, 16–19.

Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, *60*(1), 58–88.

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2014). Big Five Inventory (BFI-10) [Zusammenstellung sozialwissenschaftlicher Items und Skalen]. https://doi.org/10.6102/zis76

Roberts, C., Gilbert, E., Allum, N., & Eisner, L. (2019). Research synthesis: Satisficing in surveys: A systematic review of the literature. *Public Opinion Quarterly*, *83*(3), 598–626.

Roßmann, J., Gummer, T., & Silber, H. (2018). Mitigating satisficing in cognitively demanding grid questions: Evidence from two web-based experiments. *Journal of Survey Statistics and Methodology*, *6*(3), 376–400.

Saris, W. E., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree

response options to questions with construct-specific response options. *Survey Research Methods*, *4*(1), 61–79. https://doi.org/10.18148/srm/2010.v4i1.2682

Schaurer, I., Tanner, A., Cornesse, C., & Enderle, T. (2014). GESIS panel wave report, wave ba [GESIS—Leibniz Institute for the Social Sciences].

Scherpenzeel, A., & Saris, W. E. (2017). Multitrait-multimethod models for longitudinal research. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 381–401). Erlbaum.

Schonlau, M., & Toepoel, V. (2015). Straightlining in Web survey panels over time. *Survey Research Methods*, *9*(2), 125–137. https://doi.org/10.18148/srm/2015.v9i2.6128

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Silber, H., Schröder, J., Struminskaya, B., Stocké, V., & Bosnjak, M. (2019). Does panel conditioning affect data quality in ego-centered social network questions? *Social Networks*, *56*, 45–54. https://doi.org/10.1016/j.socnet.2018.08.003

Silber, H., Stark, T. H., Blom, A. G., & Krosnick, J. A. (2018). Implementing a multinational study of questionnaire design. *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)*, 161–179.

Struminskaya, B. (2016). Respondent conditioning in online panel surveys: Results of two field experiments. *Social Science Computer Review*, *34*(1), 95–115. https://doi.org/10.1177/0894439315574022

Struminskaya, B., & Bosnjak, M. (2021). Panel conditioning: Types, causes and empirical evidence of what we know so far. In P. Lynn (Ed.), *Advances in longitudinal survey methodology* (pp. 272–301). Willey. https://doi.org/10.1002/9781119376965.ch12

Sun, H., Tourangeau, R., & Presser, S. (2019). Panel efects: Do the eeports of panel respondents get better or worse over time? *Journal of Survey Statistics and Methodology*, *7*(4), 572–588. https://doi.org/10.1093/jssam/smy021

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

Warren, J. R., & Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research*, *41*(4), 491–534. https://doi.org/10.1177/0049124112460374

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, *8*(2), 127–135.

Appendix A
Experimental design

Table A1

Experimental setup with gross and net size of the experimental group across waves

	Panel interval short				Panel intervall long			
	Conditioned		Unconditioned		Conditioned		Unconditioned	
	Gross size	Net size	Gross size	Net size	Gross size	Net size	Gross size	Net size
Wave								
Wave 1	650	441	644	441	647	648	442	439
Wave 2	441	441	383	375	442	439	346	337
Wave 3	383	375	322	318	346	337	320	300
Wave 4	322	318	306	307	320	300	293	272
Wave 5	306	307	291	297	293	272	272	245
Wave 6	291	297	271	281	272	245	247	232

In every wave, respondents were randomly and independently assigned to one of two versions of six question design experiments to measure satisficing. For reasons of conciseness, this treatment level with the gross and net size of the experimental groups is not displayed by question design experiment and for each wave separately. However, the gross and net size of the treatment groups are comparable across the different question design experiments with net sample sizes ranging from $n = 203$ to about $n = 239$ in Wave 1 to $n = 99$ to $n = 147$ in Wave 6. For most analyses, we only used those respondents who completed each of the six panel waves to enable a direct comparison of response effect sizes; robustness checks using the full sample of the respective wave are provided in Table 6.

Appendix B
Question design experiments

Satisficing indicator: Primacy effect

Adequate housing

State first "Some people think that the state should provide adequate housing for everyone, while others think that everyone should take care of their own housing. Which of these views comes closest to your opinion?"

- The state should provide adequate housing
- Everyone should take care of their own housing

Individual first "Some people think that everyone should take care of their own housing, while others think that the state should provide adequate housing for everyone. Which of these views comes closest to your opinion?"

- Everyone should take care of their own housing.
- The state should provide adequate housing.

Trust

Trust most people first "Generally speaking: Do you think that most people can be trusted, or that you can't be careful enough when dealing with other people?"

- Most people can be trusted
- You can't be too careful.

Be careful first "Generally speaking: Do you think that you can't be careful enough when dealing with other people, or that most people can be trusted?"

- You can't be too careful.
- Most people can be trusted.

Satisficing Indicator: Aquiescence

Crime

Agree/Disagree "Do you agree or disagree with this statement? Individuals are more to blame than social conditions for crime and lawlessness in this country."

- Agree
- Disagree

Construct-specific "Which in your opinion is more to blame for crime and lawlessness in this country - individuals or social conditions?"

- Individuals
- Social conditions

Women in politics

Agree/Disagree "Do you agree or disagree with this statement? Most men are better suited emotionally for politics than are most women."

- Agree
- Disagree

Construct-specific "Would you say that most men are better suited emotionally for politics than are most women, that men and women are equally suited, or that women are better suited than men in this area?"

- Men are better suited
- Men and women are equally suited
- Women are better suited

Satisficing Indicator: Saying "don't know"

Courts

"Don't know" category existent "In general, do you think courts are too tough or not tough enough on criminals, or do you not know?"

- Too tough
- Not tough enough
- Don't know

Without "Don't know" category "In general, do you think courts are too tough or not tough enough on criminals?"

- Too tough
- Not tough enough

Smart leaders

"Don't know" category existent "Do you think that the government is run almost entirely by smart people, or do you think that quite a few in government don't seem to know what they're doing, or do you not know?"

- Run by smart people
- Don't know what they are doing
- Don't know

Without "Don't know" category "Do you think that the government is run almost entirely by smart people, or do you think that quite a few in government don't seem to know what they're doing?"

- Run by smart people
- Don't know what they are doing

Table B1

Response order experiment: "Adequate housing"

		State first	State second	Difference (in percentage points)	χ^2	p-value
Wave 1	State	77.61	49.63	27.98	22.74	0.000
	Individual	22.39	50.37	−27.98		
	Total	100%	100%	-		
	N	134	135	-		
Wave 2	State	68.35	50.39	17.96	8.97	0.003
	Individual	31.65	49.61	−17.96		
	Total	100%	100%	-		
	N	139	129	-		
Wave 3	State	70.29	48.85	21.44	12.85	0.000
	Individual	29.71	51.15	−21.44		
	Total	100%	100%	-		
	N	138	131	-		
Wave 4	State	60.58	45.04	15.54	6.50	0.011
	Individual	39.42	54.96	−15.54		
	Total	100%	100%	-		
	N	137	131	-		
Wave 5	State	60.14	42.86	17.28	8.11	0.004
	Individual	39.86	57.14	−17.28		
	Total	100%	100%	-		
	N	138	133	-		
Wave 6	State	64.06	50.00	14.06	5.42	0.020
	Individual	35.94	50.00	−14.06		
	Total	100%	100%	-		
	N	128	142	-		

Table B2

Response order experiment: "Trust"

		Trust first	Trust second	Difference (in percentage points)	χ^2	p-value
Wave 1	Trust	43.7	36.03	7.67	1.66	0.197
	Be careful	56.3	63.97	−7.67		
	Total	100%	100%	-		
	N	135	136	-		
Wave 2	Trust	38.13	35.61	2.52	0.19	0.667
	Be careful	61.87	64.39	−2.52		
	Total	100%	100%	-		
	N	139	132	-		
Wave 2	Trust	35.71	39.69	−3.98	0.46	0.499
	Be careful	64.29	60.31	3.98		
	Total	100%	100%	-		
	N	140	131	-		
Wave 4	Trust	42.75	30.71	12.04	4.23	0.040
	Be careful	57,25	69.29	−12.04		
	Total	100%	100%	-		
	N	131	140	-		
Wave 5	Trust	40.15	35.82	4.33	0.54	0.463
	Be careful	59.85	64.18	−4.33		
	Total	100%	100%	-		
	N	137	134	-		
Wave 6	Trust	43.66	39.53	4.13	0.47	0.491
	Be careful	56.34	60.47	−4.13		
	Total	100%	100%	-		
	N	142	129	-		

Table B3

Acquiescence experiment: "Women in politics"

		Agree/Disagree form	Construct-specific form	Difference (in percentage points)	χ^2	p-value
Wave 1	Men better	22.66	7.8	14.86	11.70	0.001
	Men not better	77.34	92.2	−14.86		
	Total	100%	100%	-		
	N	128	141	-		
Wave 2	Men better	15.38	12.14	3.24	0.60	0.439
	Men not better	84.62	87.86	−3.24		
	Total	100%	100%	-		
	N	130	140	-		
Wave 3	Men better	18.66	9.63	9.03	4.52	0.034
	Men not better	81.34	90.37	−9.03		
	Total	100%	100%	-		
	N	134	135	-		
Wave 4	Men better	22.14	8.63	13.51	9.55	0.002
	Men not better	77.86	91.37	−13.51		
	Total	100%	100%	-		
	N	131	139	-		
Wave 5	Men better	19.08	10.79	8.29	3.67	0.055
	Men not better	80.92	89.21	−8.29		
	Total	100%	100%	-		
	N	131	139	-		
Wave 6	Men better	26.32	9.49	16.83	13.07	0.000
	Men not better	73.68	90.51	−16.83		
	Total	100%	100%	-		
	N	133	137	-		

Table B4

Acquiescence experiment: "Crime"

		Agree/Disagree form	Construct-specific form	Difference (in percentage points)	Fisher's exact *p*-value
Wave 1	Individual	67.41	54.07	9.31	0.001
	Circumstances	32.59	45.93	−9.31	
	Total	100%	100%	-	
	N	135	135	-	
Wave 2	Individual	67.67	55.47	13.57	0.000
	Circumstances	32.33	44.53	−13.57	
	Total	100%	100%	-	
	N	133	137	-	
Wave 3	Individual	73.08	61.43	16.77	0.000
	Circumstances	26.92	38.57	−16.77	
	Total	100%	100%	-	
	N	130	140	-	
Wave 4	Individual	68.35	67.94	17.18	0.000
	Circumstances	31.65	32.06	−17.18	
	Total	100%	100%	-	
	N	139	131	-	
Wave 5	Individual	72.79	66.17	17.93	0.000
	Circumstances	27.21	33.83	−17.93	
	Total	100%	100%	-	
	N	136	133	-	
Wave 6	Individual	75.94	62.04	13.23	0.000
	Circumstances	24.06	37.96	−13.23	
	Total	100%	100%	-	
	N	133	137	-	

Table B5

Don't know experiment: "Courts"

		Agree/Disagree form	Construct-specific form	Difference (in percentage points)	Fisher's exact *p*-value
Wave 1	Nonresponse	10.07	0.76	9.31	0.001
	Substantial Answer	89.93	99.24	−9.31	
	Total	100%	100%	-	
	N	139	132	-	
Wave 2	Nonresponse	14.29	0.72	13.57	0.000
	Substantial Answer	85.71	99.28	−13.57	
	Total	100%	100%	-	
	N	133	138	-	
Wave 3	Nonresponse	17.52	0.75	16.77	0.000
	Substantial Answer	82.48	99.25	−16.77	
	Total	100%	100%	-	
	N	137	134	-	
Wave 4	Nonresponse	17.91	0.73	17.18	0.000
	Substantial Answer	82.09	99.27	−17.18	
	Total	100%	100%	-	
	N	134	137	-	
Wave 5	Nonresponse	18.66	0.73	17.93	0.000
	Substantial Answer	81.34	99.27	−17.93	
	Total	100%	100%	-	
	N	134	137	-	
Wave 6	Nonresponse	13.97	0.74	13.23	0.000
	Substantial Answer	86.03	99.26	−13.23	
	Total	100%	100%	-	
	N	136	135	-	

Table B6

Don't know experiment: "Smart leaders"

		Agree/Disagree form	Construct-specific form	Difference (in percentage points)	Fisher's exact *p*-value
Wave 1	Nonresponse	17.32	0.69	16.63	0.000
	Substantial Answer	82.68	99.31	−16.63	
	Total	100%	100%	-	
	N	127	144	-	
Wave 2	Nonresponse	20.61	0.71	19.90	0.000
	Substantial Answer	79.39	99.29	−19.90	
	Total	100%	100%	-	
	N	131	140	-	
Wave 3	Nonresponse	23.19	0.75	22.44	0.000
	Substantial Answer	76.81	99.25	−22.44	
	Total	100%	100%	-	
	N	138	133	-	
Wave 4	Nonresponse	15.15	0.72	14.43	0.000
	Substantial Answer	84.85	99.28	−14.43	
	Total	100%	100%	-	
	N	135	136	-	
Wave 5	Nonresponse	18.52	0.74	17.78	0.000
	Substantial Answer	81.48	99.26	−17.78	
	Total	100%	100%	-	
	N	135	136	-	
Wave 6	Nonresponse	11.43	0.76	10.67	0.000
	Substantial Answer	88.57	99.24	−10.67	
	Total	100%	100%	-	
	N	140	131	-	

Appendix C
Question wordings

Cognitive reflection task

Now we would like to ask you to give your assessment on the following questions.

- A racket and a ball cost a total of 1,10 euros. The racket costs 1,00 Euro more than the ball. How much does the ball cost?

  _____ cents

- If 5 machines take 5 minutes to produce 5 products, how long would it take 100 machines to produce 100 products?

  _____ minutes

- In a lake there is a spot with lily pads. Every day, the size of the spot doubles. If it takes 48 days for the water lilies to cover the entire lake, how long would it take for the water lilies to cover half of the lake?

  _____ days

Education

What is your highest educational degree?

- Still a student
- Finished school without degree
- Lower secondary school or polytechnic secondary school, degree 8th or 9th grade
- Secondary school or polytechnic secondary school, degree 10th grade
- Advanced technical college certificate
- General qualification for university entrance
- Other degree, namely:

Need for Cognition (NfC)

Below you will again find statements that relate to personal characteristics and may apply to you to a greater or lesser extent. For each statement, please indicate the extent to which it applies to you in general.

- It is enough for me simply to know the answer without understanding the reasons for the answer of a problem.
- I like my life to be full of tricky tasks to solve.
- I would prefer more complicated problems to simple problems.
- First and foremost, I think because I have to.

  1 = Does not apply at all; . . . ; 4 = Neither; . . . ; 7 = Applies completely

Content-related variables for validity analyses

Job provision

To what extent do you agree or disagree with the following statement? It is the responsibility of a government to provide a job for everyone who wants one.

1 = Strongly disagree; . . . ; 7 = Strongly agree; 99 = Don't know

Item 2 of Big Five Inventory (BFI-10)

I trust others easily, believe that there is good in people.

1 = Does not apply at all; 2 = Rather not true; 3 = Neither; 4 = Rather true; 5 = Applies completely

Women's full-time employment and family life

To what extent do you agree or disagree with the following statement? All in all, family life suffers when the woman works full time.

1 = Strongly disagree; . . . ; 7 = Strongly agree; 99 = Don't know

Appendix D
Outcome Rates

Table D1

Completion rates (COMR) per wave and respondent group (in percent)

	Conditioned Group (short interval)	Unconditioned Group (short interval)	Conditioned Group (long interval)	Unconditioned Group (long interval)
Wave 1	95.66	92.65	95.88	93.01
Wave 2	86.85	85.03	78.28	76.77
Wave 3	84.29	85.03	92.49	89.02
Wave 4	95.63	96.85	91.85	90.97
Wave 5	95.10	97.06	92.83	90.07
Wave 6	93.45	94.61	90.81	94.69

Table D2

Attrition rates (ATTR) per wave and respondent group (in percent)

	Conditioned Group (short interval)	Unconditioned Group (short interval)	Conditioned Group (long interval)	Unconditioned Group (long interval)
Wave 1	–	–	–	–
Wave 2	13.15	14.97	21.72	23.24
Wave 3	26.98	27.89	27.60	31.66
Wave 4	30.61	30.39	33.71	38.04
Wave 5	34.01	32.65	38.46	44.19
Wave 6	38.55	36.28	44.12	47.15

All attrition rates relate to the respondents who completed the first wave.

Appendix E
Analytic strategy

Terms and Definitions

A: Cognitive abilities

D: Question design

F: Conditioning frequency

I: Panel interval

M: Motivation

W: Panel wave

S: Satisficing response behavior

Y: Response distribution (of question design experiment)

Equations of estimated models

Hypothesis 1 $\log\left(\Pr\left(Y_{it}=1\right)\right)=\gamma_{00}+\gamma_{01}W_t+\gamma_{10}D_{it}+\gamma_{11}W_tD_{it}+u_{0t}+e_{it}$

Hypothesis 2 $\log\left(\Pr\left(S_{it}=1\right)\right)=\gamma_{00}+\gamma_{01}W_t+\gamma_{10}D_{it}F_{it}+u_{0t}+e_{it}$

Hypothesis 3: $\log\left(\Pr\left(S_{it}=1\right)\right)=\gamma_{00}+\gamma_{01}W_t+\gamma_{10}D_{it}F_{it}I_{it}+u_{0t}+e_{it}$

Hypothesis 4: $\log\left(\Pr\left(S_{it}=1\right)\right)=\gamma_{00}+\gamma_{01}W_t+\gamma_{10}D_{it}F_{it}A_{it}+u_{0t}+e_{it}$

Hypothesis 5: $\log\left(\Pr\left(S_{it}=1\right)\right)=\gamma_{00}+\gamma_{01}W_t+\gamma_{10}D_{it}F_{it}M_{it}+u_{0t}+e_{it}$