

Response Quality in Nonprobability and Probability-based Online Panels

Cornesse, Carina; Blom, Annelies G.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Cornesse, C., & Blom, A. G. (2020). Response Quality in Nonprobability and Probability-based Online Panels. *Sociological Methods & Research*, 1-30. <https://doi.org/10.1177/0049124120914940>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Response Quality in Nonprobability and Probability-based Online Panels

Sociological Methods & Research

1-30

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0049124120914940

journals.sagepub.com/home/smr

Carina Cornesse^{1,2}  and Annelies G. Blom^{1,3} 

Abstract

Recent years have seen a growing number of studies investigating the accuracy of nonprobability online panels; however, response quality in nonprobability online panels has not yet received much attention. To fill this gap, we investigate response quality in a comprehensive study of seven nonprobability online panels and three probability-based online panels with identical fieldwork periods and questionnaires in Germany. Three response quality indicators typically associated with survey satisficing are assessed: straight-lining in grid questions, item nonresponse, and midpoint selection in visual design experiments. Our results show that there is significantly more straight-lining in the nonprobability online panels than in the probability-based online panels. However, contrary to our expectations, there is no generalizable difference between nonprobability online panels and probability-based online panels with respect to item nonresponse. Finally, neither respondents in nonprobability online panels nor respondents in probability-based online panels are significantly affected by the visual design of the midpoint of the answer scale.

¹ Collaborative Research Center, SFB 884 “Political Economy of Reforms”, University of Mannheim, Germany

² GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany

³ Department of Political Science, School of Social Sciences, University of Mannheim, Germany

Corresponding Author:

Carina Cornesse, Collaborative Research Center, SFB 884 “Political Economy of Reforms”, University of Mannheim, Mannheim 68159, Germany.

Email: carina.cornesse@uni-mannheim.de

Keywords

nonprobability sample, probability-based sample, online panel, satisficing, response quality

The past decade has seen increasing debate about the quality of nonprobability online panels. This debate has primarily circled around whether or not these panels provide representative sets of respondents (for an overview of the debate see Cornesse et al. 2020). The apprehension of some researchers is that a biased subgroup of the population self-selects into nonprobability online panels (see Chang and Krosnick 2009; Legleye et al. 2015; Loosveldt and Sonck 2008; MacInnis et al. 2018; Malhotra and Krosnick 2007; Pasek 2016; Pennay et al. 2016; Yeager et al. 2011). Others argue that nonprobability online panels can accurately reflect the target population, especially after weighting (see Goel, Obeng, and Rothschild 2015; Kennedy, Keeter, and Weisel 2016; Wang et al. 2015). However, while the number of publications on nonprobability panel accuracy is increasing, less attention has been paid to response quality in nonprobability panels (for notable exceptions, see Chang and Krosnick 2009; Greszki, Meyer, and Schoen 2014; Hillygus, Jackson, and Young 2014).

This is surprising, because one might argue that, since their respondents participate in the panel mainly for monetary reasons (see e.g., GreenBook 2017), nonprobability online panel respondents may be less committed to the panel in terms of response quality. The focus on monetary rewards among the respondents is encouraged by the nonprobability online panel advertising industry that recruits the panel members. With advertising slogans like “Earn Cash With Quick Paid Surveys!” (www.quickpaysurvey.com), “Make Money Online With Paid Surveys” (www.cashcrate.com), or “Take Surveys for Cash” (www.takesurveysforcash.com) advertisers try to attract as many Internet users as possible. Being attracted to the online panel by the promise of easy money for little effort, nonprobability online panel respondents may show different care in answering survey questions than respondents that were recruited into an online panel by probability-based offline recruitment methods, independent of their sociodemographic characteristics. Consequently, response quality might differ between nonprobability and probability-based online panel respondents.

In this article, we investigate whether there is a difference in response quality between survey participants in nonprobability online panels and probability-based online panels. For this purpose, we look into three indicators that are often associated with response quality in the context of survey satisficing: straight-lining, item nonresponse, and midpoint selection.

Respondent Motivation and Survey Satisficing

Most social research based on survey data relies on the assumption that respondents answer the survey questions to the best of their ability. This requires the respondents to carefully carry out all the cognitive steps involved in answering a survey question. According to Tourangeau, Rips, and Rasinski (2000), cognitive response processing consists of four steps: question comprehension, information retrieval, judgment and estimation, and reporting an answer. Data analysts typically assume that all four response steps were carefully carried out by all respondents, that is, respondents optimize survey responses. Respondents, however, sometimes take shortcuts through the optimal cognitive response process. This behavior is called satisficing (Krosnick and Alwin 1987).

Krosnick (1991) defines two types of satisficing: weak satisficing and strong satisficing. Weak satisficing occurs when respondents execute all cognitive steps that are necessary to arrive at a response, but they do so only superficially. They might, for instance, read a question text carefully but skip the accompanying instruction text. Alternatively, when presented with a list of answer options, respondents might choose the first option that approximately fits their opinion without considering further answer options that might fit their opinion even better. Because weak satisficers carry out the response steps superficially, unmeaningful decision-making cues, like the visual design of an answer scale, can influence their responses.

While weak satisficers carry out all four cognitive response steps, even if they do so superficially, strong satisficers do not or only partially carry out the response steps. Respondents might, for instance, process only just enough information to arrive at a response that they consider generally reasonable without reading and considering the question carefully or without searching their memories and retrieving the relevant information. This strategy results in no responses at all, unmeaningful answers, nonsubstantive answers, or undifferentiated answers.

According to satisficing theory, there is a continuum of cognitive thoroughness of responses with perfectly optimized responses at one end of the continuum and strongly satisficed responses at the other end (see Krosnick

1991). Respondents can be in different positions on the continuum, with some being generally thorough in all of their responses and others being generally less careful in their responses (for empirical proof, see Kaminska, McCutcheon, and Billiet 2010; Knäuper 1999; Krosnick 1992; Malhotra 2008; Narayan and Krosnick 1996). In addition to this interpersonal variation, there can be intrapersonal variation in the level of satisficing observed during an interview, for example, when respondents fatigue during a long interview (for empirical proof, see Roberts et al. 2010).

According to Krosnick (1991), there are three factors that largely influence the occurrence and degree of satisficing: task difficulty, respondent ability, and respondent motivation. Task difficulty refers to the cognitive effort needed to answer a question optimally. The task difficulty depends on the complexity of a question and of the information asked for. When a question contains many words, long words, and/or uncommon or ambiguous terms, satisficing is more likely to occur than when questions are short, precise, and easy to comprehend (for empirical proof, see, e.g., Alwin and Krosnick 1991). Similarly, when a question asks respondents to evaluate multiple items or answer on a long, unlabeled scale, satisficing is more likely than when a question asks for the evaluation of only one item and provides few fully labeled answer options (see also Krosnick 1999; Krosnick and Berent 1993).

Respondent ability refers to the competences and skills involved in answering survey questions. It includes cognitive abilities such as cognitive sophistication, the amount of practice in thinking about a topic, and attitude strength, as well as practical abilities such as the ability to process and communicate answers. In self-administered questionnaires, respondent ability also includes reading and writing skills. Furthermore, in web surveys, computer literacy and technological skills necessary to start and navigate through the survey are important aspects of respondent ability (for empirical proof on satisficing in web surveys, see, e.g., Toepoel et al. 2009). Satisficing is more likely to occur when a respondent has low cognitive abilities and may therefore have problems comprehending the question (see also Kaminska et al. 2010; Krosnick 1992; Krosnick and Alwin 1987; Narayan and Krosnick 1996).

The respondents' motivation determines how much effort they are willing to invest in answering a question. To some extent, respondent motivation is a personality characteristic that is related to a person's need for cognition (for information on the concept of need for cognition, see Cacioppo and Petty 1982). Satisficing is more likely to occur if respondents have low need for cognition (see Kaminska et al. 2010). Respondents with high need for

cognition have an intrinsic motivation to fill out the questionnaire carefully because they generally enjoy thinking about questions and have fun expressing their opinions. High interest in a survey topic and the perceived importance of a survey can have a substantial impact on respondent motivation at the start of the survey (see also Groves, Presser, and Dipko 2004; Stoop 2005). However, respondent motivation may decrease over the course of the interview. Therefore, satisficing is more likely toward the end of a survey, especially in long surveys, than toward the start and in short surveys (see also Krosnick et al. 2002).

Measuring Satisficing

Thus, according to Krosnick (1991), the combination of task difficulty, respondent ability, and respondent motivation explains the occurrence and amount of satisficing. In our analyses, we keep task difficulty constant across panels by including exactly the same questions with exactly the same question format and response categories across all of the panels. In addition, we aim to keep respondent ability constant across panels by applying the same raking weighting procedure to all samples and thus controlling for sample composition differences in sociodemographic characteristics (i.e., marital status, household size, age, and education).

Regarding respondent motivation, we assume that the nonprobability panel participants are mainly motivated by the monetary incentives that they receive. Thus, we expect them to minimize effort in order to maximize their incentive-by-effort ratio. This expectation is supported by studies that demonstrate the importance of monetary incentives for nonprobability online panel members. For instance, in a study on the motives for joining nonprobability online panels, Keusch, Batinic, and Mayerhofer (2014:179) find that when asked to select all of their motives for participation, 40 percent of panelists indicated they “wanted to earn some extra money.” In addition, “monetary motives had the strongest correlation with survey participation” for nonprobability online panel participants (Keusch et al. 2014:185). Similarly, Sparrow (2006:5) finds that 52 percent of the new members of a nonprobability online panel participate because it is “an enjoyable way to earn money” as opposed to 20 percent who join because they “thought they would be interested in the topics covered,” and 19 percent who “enjoy answering questions.”

The probability-based online panelists are less driven by monetary incentives. For instance, analyzing survey data on the respondents’ most important reason to participate in the Dutch LISS Panel (www.lissdata.nl), a

probability-based online panel of the general population in the Netherlands, we find that 16.4 percent of the respondents participate in the panel because they “find it important to contribute to science.” The financial reward is only stated as the most important reason for participating by 15.2 percent of respondents. Furthermore, for the German Internet Panel (GIP; <https://www.uni-mannheim.de/gip>), a probability-based online panel of the German general population, we observe that about 13 percent of panelists even waive their incentive and choose to donate it to charity instead.

We, therefore, expect the share of people who care about giving optimized answers to be lower, and thus the amount of satisficing behavior to be higher, in the nonprobability online panels than in the probability-based online panels. To investigate differences in response quality across nonprobability and probability-based online panels, we examine three indicators of satisficing: straight-lining in grid questions, item nonresponse, and midpoint selection in a visual design experiment. In the following, we discuss the theoretical and empirical background in the literature of each of these indicators.

Straight-lining

The term straight-lining refers to the tendency of respondents to choose the same or a very similar answer option for each item in a grid (see Schonlau and Toepoel 2015). This phenomenon is sometimes also referred to as “non-differentiation” (see Malhotra, Miller, and Wedeking 2014). Straight-lining is a strong form of satisficing. It occurs in self-administered questionnaires because the grid format provides a visual cue that triggers a specific type of cognitive shortcut: While for the first item of the grid, respondents might still carry out all the cognitive steps necessary to arrive at an optimized response, the grid format suggests that the same answer will also be acceptable for the following items. Therefore, some respondents might abandon the full cognitive response processing in favor of a shortcut and give the same (or a very similar) answer to all other grid items. Research suggests that avoiding grid questions and asking each question separately instead, preferably with only one question per screen, can prevent straight-lining (see Couper 2008; Couper, Traugott, and Lamias 2001).

Item Nonresponse

Like straight-lining, item nonresponse is a form of strong satisficing. Respondents who choose this satisficing strategy skip one or all of the

cognitive response steps. In a web survey, respondents might not read the question text carefully and click on the “next” button instead in order to get to the end of the questionnaire more quickly. If they read the question carefully, they might still not be willing or able to go through the necessary information retrieval or the judgment and estimation processes. There are two types of item nonresponse: question skipping (QS) and giving a nonsubstantive answer, that is, answering “don’t know” (DK) or “don’t want to say” (DWS).

Nonsubstantive responses require some cognitive effort because respondents have to at least browse through the answer options or look for visual design cues that lead them to the DK or DWS answer category. Relative to nonsubstantive responses, QS is a much stronger type of satisficing because when choosing to skip a question altogether, respondents do not have to engage in any kind of cognitive response process.

A potential drawback to using nonsubstantive answers as a satisficing indicator is that respondents may carry out all the necessary response steps and in the end still decide to choose a nonsubstantive response, for example, because they honestly do not know the answer to a question (Converse 1974; Schuman and Presser 1981; Sturgis, Roberts, and Smith 2014). With respect to “no opinion” answers, however, Krosnick et al. (2002) show that respondents with low levels of education are more likely to choose this answer option, suggesting that people are more likely to choose “no opinion” responses when they perceive the processes of producing an optimal response as cumbersome. Furthermore, the authors find that the amount of “no opinion” answers increases with interview duration, suggesting that respondent motivation decreases toward the end of the interview resulting in less willingness to engage in the cognitive effort necessary to produce an optimal response. These findings are supported by further research (see, e.g., Bradburn and Sudman 1988; Feick 1989; Fowler and Cannell 1996; Gilljam and Granberg 1993; Holbrook, Green, and Krosnick 2003).

Midpoint Selection

The midpoint of an answer scale provides a superficial visual cue in self-administered questionnaires. Following Krosnick’s (1991) reasoning, some respondents might satisfice by selecting the middle category while optimizing respondents are not influenced by this visual cue and instead choose the answer option that best represents their “true” answer after carrying out all the necessary cognitive steps. Specifically, Krosnick and Fabrigar (1997:147) argue that “many people [...] might select an offered midpoint because it provides an easy choice that requires little effort and is easy to

justify.” On the continuum of cognitive thoroughness, midpoint selection can be interpreted as a weaker form of satisficing than straight-lining and item nonresponse when we assume that respondents go through the necessary response steps but are influenced in their decision by the visual design cue that the scale midpoint provides. Tourangeau, Couper, and Conrad (2004) show that, generally, the visual midpoint rather than the conceptual midpoint shapes the response distribution. Their explanation is that “the visual midpoint is seen as providing a benchmark, representing either the conceptual midpoint of the scale or the most typical response” (p. 390). Following satisficing theory, it is likely that satisficing respondents choose the visual midpoint to reduce the cognitive effort needed to evaluate the other response options against this midpoint. This behavior would constitute a shortcut to the information retrieval and/or the judgment and estimation processes. Therefore, the selection of the visual midpoint may be used as an indicator of satisficing behavior (see also Kaminska et al. 2010; Malhotra et al. 2014).

Online Panel Data

In this article, we assess response quality among respondents in 10 online panels (see Table 1). Three of them are probability-based online panels: the GIP, the GESIS Panel (GP), and one commercial probability-based online panel. The remaining seven are commercial nonprobability online panels. In each of the online panels, we fielded the same short multi-topic online survey in German language.

Our study has been conducted as part of a larger project on data quality in nonprobability and probability-based online surveys. Some of the survey questions in our questionnaire have specifically been designed to investigate survey satisficing while other questions have originally been designed to allow analyses on sample accuracy that are out of the scope of this article. In the following, we describe the various online panels and the study that we implemented in more detail.

The GIP

The GIP is based on a three-stage stratified probability area sample with subsequent face-to-face recruitment interviews. At the first sampling stage, a random sample of areas is drawn from a database that covers all areas in Germany. Within each primary sampling unit (PSU), lists record every household along a predefined random route until they have listed 200

Table 1. Online Panel Characteristics.

Panel ^a	Type	Sampling	Recruitment	Offline Households	Fieldwork Period	N
GIP	Academic	Probability	Face-to-face	Yes	March 1 to 31, 2015	3,075
GP	Academic	Probability	Face-to-face	Yes	February 18 to April 14, 2015	2,533
1	Commercial	Probability	Telephone	No	March 1 to 31, 2015	1,012
2	Commercial	Nonprobability	Online	No	March 5 to 18, 2015	1,038
3	Commercial	Nonprobability	Online	No	March 2 to 11, 2015	999
4	Commercial	Nonprobability	Online	No	March 1 to 18, 2015	1,002
5	Commercial	Nonprobability	Online	No	March 2 to 16, 2015	1,000
6	Commercial	Nonprobability	Online	No	March 25 to April 1, 2015	1,000
7	Commercial	Nonprobability	Online	No	March 3 to 9, 2015	994
8	Commercial	Nonprobability	Online	No	March 5 to 11, 2015	1,000

Note: GIP = German Internet Panel; GP = GESIS Panel.

^aRefer Note 5.

households. Subsequently, a random sample of households is drawn to be interviewed in face-to-face recruitment interviews. All age-eligible members of sampled households are invited to become online panelists (see Blom, Gathmann, and Krieger 2015). Furthermore, the GIP covers individuals without computer and/or Internet access by equipping them with the necessary devices (see Blom et al. 2017). All panel members are invited bi-monthly to participate in an online interview of about 20–25 minutes on a diversity of social, economic, and political topics. For the analyses in this article, we used wave 16 of the GIP. It had a completion rate¹ of 69.8 percent and a cumulative response rate² of 14.3 percent. Although the GIP covers the German population aged 16–75, we only use data from individuals that were aged 18–70 to make the data comparable across all panels in this study.

The GP

The GP is based on a two-stage stratified probability sample from population registers and subsequent face-to-face recruitment interviews. At the first sampling stage, the GP draws a random sample of areas from a database of municipalities in Germany. Then, the GP samples individuals from the local population registers within each of the sampled PSUs. Because 10 sampled municipalities refused to cooperate with the GP sampling request, these PSUs had to be substituted (see Bosnjak et al. 2017). Subsequently, the sampled individuals were contacted for face-to-face recruitment interviews. All interviewed individuals were invited to become panelists. Furthermore, the GP includes the offline population via paper-and-pencil mail surveys. Internet users that prefer to participate offline instead of online are also provided with paper-and-pencil mail surveys. All panel members of the GP are invited to participate in bi-monthly interviews of about 20 minutes on a wide variety of topics. For the analyses in this article, we used the so-called wave *ca* of the GP (for reference, see the report of wave *ca*, <https://www.gesis.org/en/gesis-panel/documentation/>). It had a completion rate³ of 77.72 percent and a cumulative response rate⁴ of 19.5 percent. The age range covered in the GP is 18–70 years. In our study, we exclude the GP mail respondents because the potential mode effect might bias our results (see, e.g., Green, Krosnick, and Holbrook 2001; Holbrook et al. 2003, for evidence on differences in satisficing by mode).

Panel 1

Panel 1 is a commercial probability-based online panel. To recruit panel members, the panel draws its sample from random digit dialing (RDD) telephone surveys conducted in-house by the same company. Individuals interviewed in an RDD telephone interview were subsequently invited to join the panel for regular online interviews if they had access to the Internet. Recruitment interview respondents that did not have access to the Internet were not invited for the subsequent online panel waves. For our study, panel 1 drew a quota subsample from its probability-based respondent pool. Panel participation rates could not be calculated due to the unavailability of the necessary information on how many panel participants were invited to our survey wave.

The Nonprobability Online Panels: Panels 2–8

For the recruitment of the nonprobability online panels, we published a call for applications in November 2014. The call explained that we sought to

implement a 10-minute questionnaire about traffic, politics, and health among 1,000 respondents that should be representative of the German population aged 18–70 years of age. The call further announced that the data were to be collected in March 2015. Regarding further design decisions (application of quotas, provision of weights, etc.), the panel providers were free to choose whichever approach they thought would provide the most representative data.⁵

In response to our call, we received 17 applications for conducting the specified survey. Of these, 16 survey providers explicitly offered a sample representative of the general population in Germany aged 18–70. Seven providers were considered fit for our purpose and within reasonable budgetary limits. We therefore commissioned them with collecting our data. The costs quoted for conducting the wave analyzed in this article plus two additional waves conducted in half-year intervals varied widely across the panels (see Online Appendix A, which can be found at <http://smr.sagepub.com/supplemental/>, for details). In the following, the nonprobability online panels are numbered from two to eight in sequence of ascending costs. Panel participation rates could not be calculated due to the unavailability of the necessary information on how many panel participants were invited to our survey wave.

Method

To assess differences in the amount and type of satisficing between the nonprobability online panels and the probability-based online panels, we consider three satisficing indicators: straight-lining in grids, item nonresponse, and midpoint selection in a visual design experiment. Their operationalization is described in the following. Information on the question texts and answer scales of the questions we used in our analyses can be found in Online Appendix B (which can be found at <http://smr.sagepub.com/supplemental/>). Furthermore, external validations of selected survey variables and sociodemographic characteristics can be found in Online Appendix C (which can be found at <http://smr.sagepub.com/supplemental/>).

Straight-lining

To assess straight-lining in our questionnaire, we implemented two psychological short scales with four items each in grid format. We do not use any grid questions in our questionnaire module except for these two psychological scales that we specifically use for the purpose of assessing straight-lining

in grids across the nonprobability and probability-based online panels. We define respondents as straight-liners if they choose the same answer category for every item on at least one of the two grid questions. Based on the literature described above, we test the following hypothesis:

Hypothesis 1: In nonprobability online panels, a higher proportion of respondents chooses to straight-line on grid questions than in probability-based online panels.

To test this hypothesis, we compare the proportions of straight-liners and the respective confidence intervals around these estimates across the respondents of each of the nonprobability and the probability-based online panels. Furthermore, we apply χ^2 -tests to examine averages across all nonprobability online panel respondents versus all probability-based online panel respondents.

Item Nonresponse

Our questionnaire contained several possibilities for generating item nonresponse, all of which were implemented in the same way across all panels. In our analyses, we can differentiate between different types of item nonresponse. At several questions, respondents were able to give nonsubstantive answers. At five questions, we provided the nonsubstantive answer option “DK,” at two questions we provided a “DWS” option, and at one question, we provided both DK and DWS answer options. In addition, at each question, respondents were able to skip the question by clicking on the “next” button, generating QS. Based on the literature described above, we test the following hypotheses:

Hypothesis 2: In nonprobability online panels, a higher proportion of respondents chooses to not provide any (substantive) response to a question than in probability-based online panels.

Hypothesis 2a: In nonprobability online panels, a higher proportion of respondents chooses to answer DK than in probability-based online panels.

Hypothesis 2b: In nonprobability online panels, a higher proportion of respondents chooses to answer DWS than in probability-based online panels.

Hypothesis 2c: In nonprobability online panels, a higher proportion of respondents chooses to skip a question than in probability-based online panels.

For each type of item nonresponse, we generated a continuous variable that operationalizes the proportion of each type of item nonresponse (DK, DWS, and QS) per respondent among all questions where these types of nonresponses were possible. In addition, we generated a variable that operationalizes the proportion of overall item nonresponse (INR) per respondent among all questions.

To test our hypotheses, we compare the average proportion of each type of item nonresponse and the respective confidence intervals around these estimates across the respondents of each of the nonprobability and probability-based online panels. Furthermore, we apply χ^2 -tests to examine average proportions of each type of item nonresponse among all nonprobability online panel respondents versus all probability-based online panel respondents.

Midpoint Selection

We implemented a visual design experiment in our questionnaire to investigate whether respondents answer consistently across different answer scales. Four experimental conditions were randomly assigned to respondents.

Condition 1: The conceptual midpoint was located at the visual midpoint.

Condition 2: The answer scale contained a conceptual midpoint but no visual midpoint.

Condition 3: The answer scale contained a visual midpoint but no conceptual midpoint.

Condition 4: The answer scale contained neither a conceptual nor a visual midpoint.

We conducted this experiment on two questions of our questionnaire and randomly assigned respondents independently at each question. The first question covered respondents' perceived health. The second question concerns respondents' opinion on environmental zones in cities.⁶

We varied the presence of a *conceptual* midpoint by including a "neither nor" or "average" answer option in the scale versus excluding this conceptual midpoint. We vary the presence of a *visual* midpoint by leaving a gap between the substantive answer options and the "I don't know" option versus not leaving a gap. The respective answer scales are depicted in Table 2 and Table 3.

Table 2. Experimental Conditions in the Midpoint Experiment on the Five-point Scale.

Conceptual & visual midpoint	Conceptual midpoint only	Visual midpoint only	No midpoint
Very good	Very good	Very good	Very good
Good	Good	Good	Good
Average	Average	Bad	Bad
Bad	Bad	Very bad	Very bad
Very bad	Very bad	DK	
	DK		DK
DK			

Note:  Conceptual midpoint;  visual midpoint. DK = don't know.

Table 3. Experimental Conditions in the Midpoint Experiment on the Seven-point Scale.

Conceptual & visual midpoint	Conceptual midpoint only	Visual midpoint only	No midpoint
Very good	Very good	Very good	Very good
Good	Good	Good	Good
Rather good	Rather good	Rather good	Rather good
Neither nor	Neither nor	Rather bad	Rather bad
Rather bad	Rather bad	Bad	Bad
Bad	Bad	Very bad	Very bad
Very bad	Very bad	DK	
	DK		DK
DK			

Note:  Conceptual midpoint;  visual midpoint. DK = don't know.

Following the literature described above, there are two reasons why respondents may choose an answer option from a response scale. The first reason for choosing an answer option is that, after carefully going through all necessary cognitive response processing steps, respondents decide that this answer option fits them best. If respondents choose an answer option for the first reason, the response distribution should not be affected by whether or not this answer option is located at the visual midpoint of the answer scale. The second reason for choosing an answer option is that respondents go through the cognitive response processing steps only superficially and therefore look for unmeaningful cues to decide which answer option to choose (i.e., weak satisficing as described in the literature above). If respondents satisfice and choose an answer option for the second reason, the proportion of respondents who choose an answer option should be higher when this answer option is located at the visual midpoint of the answer scale than when it is not located at the visual midpoint. Following our expectation that there is more satisficing in nonprobability online panels than in probability-based online panels, we test the following hypothesis in our midpoint selection experiment:

Hypothesis 3: In nonprobability online panels, a higher proportion of respondents chooses an answer option when it is located at the visual midpoint than in probability-based online panels.

As with choosing an answer option in general, there are two reasons for choosing the conceptual midpoint of an answer scale: The first reason for choosing the conceptual midpoint is that, after going through all cognitive response processing steps, respondents conclude that the conceptual midpoint fits them best. The second reason for choosing the conceptual midpoint is that the conceptual midpoint is located at the visual midpoint of the answer scale and therefore picking the conceptual midpoint serves as a satisficing strategy. Because the conceptual midpoint is usually located at the visual midpoint of the answer scale, we specify two additional hypotheses to Hypothesis 3. These additional hypotheses disentangle the effect of the visual design of the answer scale on the conceptual midpoint (Hypothesis 3a) and the effect of the visual design of the answer scale on a regular answer option (Hypothesis 3b) in nonprobability online panels compared to probability-based online panels.

Hypothesis 3a: The difference between the proportion of respondents choosing the conceptual midpoint when it matches the visual midpoint

(condition 1) and the proportion of respondents choosing the conceptual midpoint when there is no visual midpoint (condition 2) will be higher in nonprobability online panels than in probability-based online panels.

Hypothesis 3b: The difference between the proportion of respondents choosing an answer option when it is located at the visual midpoint (condition 3) and the proportion of respondents choosing the same answer option when it is not located at the visual midpoint (condition 4) will be higher in nonprobability online panels than in probability-based online panels.

We apply χ^2 -tests to examine averages across the nonprobability and probability-based online panel respondents. For each of the estimates in our analyses described above, we obtained bootstrapped standard errors by pooling results across 100 variance–covariance matrices.

Results

We examine whether there are significant differences in response quality between nonprobability online panels and probability-based online panels based on our hypotheses on the satisficing indicators described above (straight-lining, item nonresponse, and midpoint selection in a visual design experiment).

Straight-Lining

With regard to our hypothesis that a higher proportion of respondents chooses to straight-line on grid questions in nonprobability online panels than in probability-based online panels (Hypothesis 1), we indeed find significantly more straight-lining in the nonprobability online panels than in the probability-based online panels (on average 13.4 percent and 6.5 percent, respectively; $\chi^2(1) = 152.7, p = 0.00$).

When investigating the online panels in detail (see Figure 1), we find that in each of the nonprobability online panels, straight-lining is considerably more prevalent (between 10.2 percent in panel 2 and 16.2 percent in panel 5) than in the probability-based online panels (between 3.1 percent in GP and 9.2 percent in GIP). In fact, two of the probability-based online panels (GP and panel 1) show significantly less straight-lining than the best nonprobability online panel (panel 2).

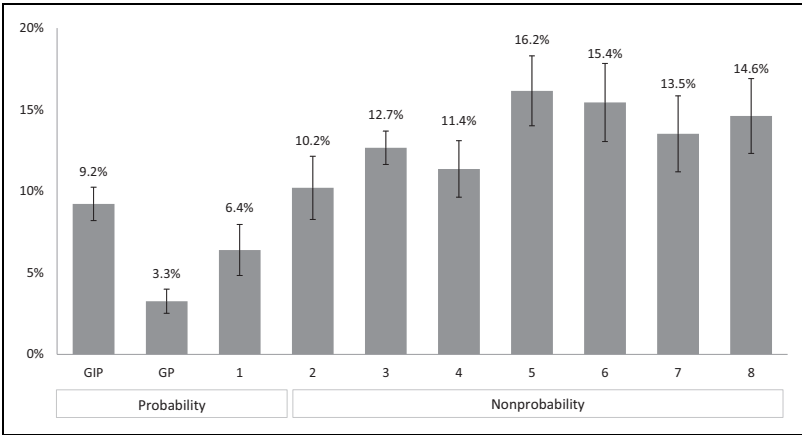


Figure 1. Proportion of straight-liners across panels (bars), bootstrapped 95 percent confidence intervals (spikes).

Item Nonresponse

Regarding our hypothesis that a higher proportion of respondents chooses to not provide any (substantive) response to a question in non-probability online panels than in probability-based online panels (Hypothesis 2), we find no generalizable evidence in support of our item nonresponse hypotheses across the three types of item nonresponse (DK, DWS, and QS).

When examining INR as an overall measure of item nonresponse that operationalizes the broader Hypothesis 2 (see INR bars in Figure 2), we find that, contrary to our expectations, the average proportion of nonresponses is significantly lower among the nonprobability online panel respondents compared to the probability-based online panel respondents (1.2 percent versus 1.6 percent, respectively; $\chi^2(1) = 50.3, p = 0.00$). In the following, we explore the different types of item nonresponse (DK, DWS, and QS) in more detail based on our Hypotheses 2a–2c.

With regard to our hypothesis that the average proportion of DK answers is higher among the nonprobability online panel respondents than among the probability-based online panel respondents (Hypothesis 2a), we find that, contrary to our expectations, the nonprobability online panel respondents had a significantly lower average percentage of DK answers than the probability-based online panel respondents (2.9 percent versus 3.4 percent, $\chi^2(1) = 64.9, p = 0.00$, see DK bars in Figure 2).

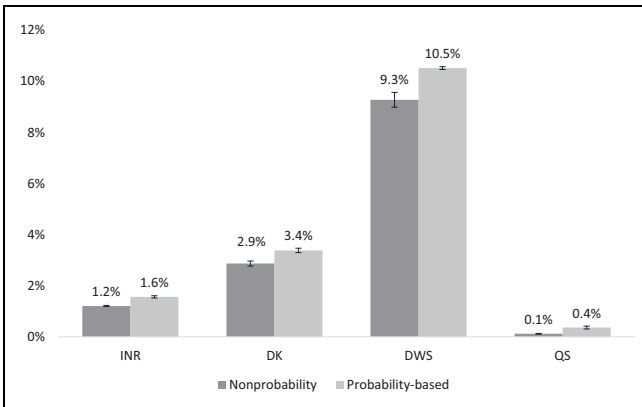


Figure 2. Proportion of item nonresponse in probability-based and nonprobability panels (bars), bootstrapped 95 percent confidence intervals (spikes).

However, looking at the proportions of DK across the online panels in detail (see left pane of Figure 3), we find some variability in the proportion of people who chose DK in the nonprobability online panels (between 1.8 percent in panel 4 and 3.7 percent in panel 8) and the probability-based online panels (between 2.8 percent in panel 1 and 4.0 percent in GP). While the average proportion of DK answers is significantly higher in the GP and significantly lower in panel 4 than in most other panels, the overall variability across the online panels is such that any differences between nonprobability and probability-based online panels seem coincidental.

Regarding our hypothesis that a higher proportion of respondents chooses to answer DWS in nonprobability online panels than in probability-based online panels (Hypothesis 2b), we find that there is a significant difference in the average proportion DWS answers between the nonprobability online panels and the probability-based online panels (9.3 percent versus 10.5 percent, $\chi^2(1) = 4.1, p < 0.1$).

Examining the proportions of DWS across the online panels in detail (see middle pane of Figure 3), we again do not find any generalizable evidence that may distinguish nonprobability online panels (between 7.3 percent in panel 4 and 11.7 percent in panel 8) from probability-based online panels (between 9.0 percent in GP and 12.0 percent in GIP).

With regard to our hypothesis that a higher proportion of respondents chooses to skip a question in nonprobability online panels than in probability-based online panels (Hypothesis 2c), we find a small but

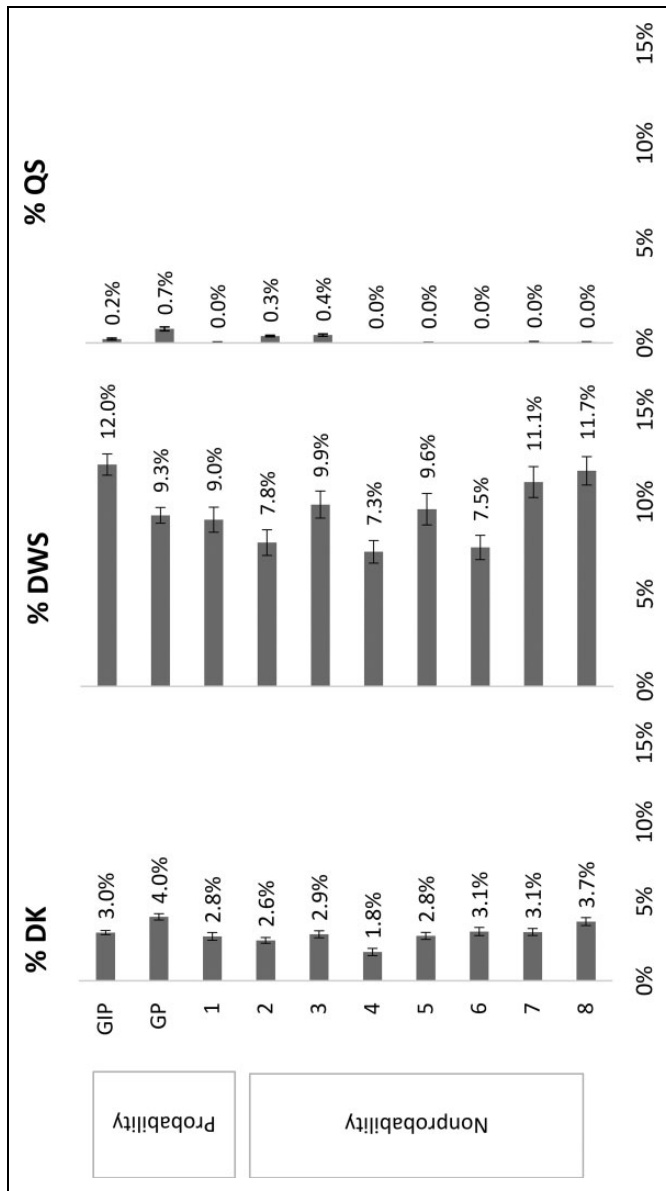


Figure 3. Average proportion of don't know (DK), don't want to say (DWS), and question skipping (QS) answers per person among all questions where the respective answer option was possible across panels (bars), bootstrapped 95 percent confidence intervals (spikes).

statistically significant difference between nonprobability online panels and probability-based online panels (0.1 percent versus 0.4 percent, $\chi^2(1) = 104.8$, $p < 0.1$).

When we examine QS across panels in detail (see right pane of Figure 3), we find no generalizable pattern of differences between the nonprobability online panels (between 0.0 percent in panels 4–8 and 0.4 percent in panel 3) and the probability-based online panels (between 0.0 percent in panel 1 and 0.7 percent in GP). Overall, we find that item nonresponse is generally very low in all examined online panels, especially with regard to DK answers and QS. In addition, item nonresponse seems unrelated to the sampling design of the panels.

Midpoint Selection

In Table 4, we present the results of our experiment on the influence of the answer scale design on the selection of the visual midpoint. Contrary to our hypothesis on midpoint selection (Hypothesis 3), we find no significant difference between nonprobability online panels and probability-based online panels in the effect of the visual design of the scale on midpoint selection. In the following, we investigate the differences in midpoint selection across nonprobability online panels and probability-based online panels in more detail based on our Hypotheses 3a and 3b.

With regard to our hypothesis that the difference between the proportion of respondents choosing the conceptual midpoint when it matches the visual midpoint (condition 1) and the proportion of respondents choosing the conceptual midpoint when there is no visual midpoint (condition 2) will be higher in nonprobability online panels than in probability-based online panels (Hypothesis 3a), we find that the difference in proportions is not significantly higher in the nonprobability online panels than in the probability-based online panels (0.8 percentage points and 0.4 percentage points, respectively, on the five-point scale, $\chi^2(1) = 0.01$, $p > 0.10$; 0.7 percentage points and -3.2 percentage points, respectively, $\chi^2(1) = -0.05$, $p > 0.1$, on the seven-point scale). This indicates that respondents in nonprobability online panels are not more influenced by the visual design of the midpoint than respondents in probability-based online panels.

Regarding our hypothesis that the difference between the proportion of respondents choosing an answer option when it is located at the visual midpoint (condition 3) and the proportion of respondents choosing the same answer option when it is not located at the visual midpoint (condition 4) will be higher in nonprobability online panels than in probability-based online

Table 4. Proportions and Absolute Numbers of Respondents in the Middle Category by Experimental Group in Probability-based Online Panels and Nonprobability Online Panels.

Experimental Condition	Five-point Scale				Seven-point Scale			
	Nonprobability		Probability Based		Nonprobability		Probability Based	
	Percent	N	Percent	N	Percent	N	Percent	N
Hypothesis 3a								
Condition 1: Conceptual midpoint matches visual midpoint	28.5	500	26.7	444	17.1	299	15.2	254
Condition 2: Conceptual midpoint but no visual midpoint	27.6	493	26.3	432	16.3	288	18.4	298
Difference between conditions 1 and 2	0.8	7	0.4	12	0.7	11	-3.2	-44
Hypothesis 3b								
Condition 3: Visual midpoint	14.7	254	9.2	154	10.4	181	11.7	196
Condition 4: No midpoint	14.0	246	11.9	194	11.6	205	10.6	175
Difference between conditions 3 and 4	0.7	8	-2.7	-40	-1.2	-24	1.1	21

panels (Hypothesis 3b), we find that the difference in proportions is not significantly higher in the nonprobability online panels than in the probability-based online panels (0.7 percentage points and -2.7 percentage points, respectively, on the five-point scale, $\chi^2 = 0.24, p > 0.1$; -1.2 percentage points and 1.1 percentage points, respectively, $\chi^2 = 0.05, p > 0.1$, on the seven-point scale). In accordance with our findings from Hypothesis 3a, the findings on Hypothesis 3b also indicate that respondents in

nonprobability online panels are not more influenced by the visual design of the midpoint than respondents in probability-based online panels.

Furthermore, when examining the results of the experiment in each of the online panels (see Online Appendix D, which can be found at <http://smr.sagepub.com/supplemental/>), we find that the visual design of the midpoint has no effect. Exceptions are three of the nonprobability online panels (panels 5–7), where we find a weakly significant effect in one comparison of experimental conditions each, but given the number of effects tested, this may well be just by chance.

Response Quality and Costs

Table A1 in Online Appendix A (which can be found at <http://smr.sagepub.com/supplemental/>) shows the total costs for data collection in the commercial online panels. Panel 1 participated without billing any costs. All other commercial online panels costed different amounts of money ranging from €5,392.97 in panel 2 to €10,676.44 in panel 8. There is no indication that the more costly panels perform better than the less costly panels in terms of data quality. For example, panel 2 as the least costly commercial online panel has a slightly lower proportion of straight-lining, DK answers, and DWS answers than panel 8 as the most costly commercial online panel. Panel 8, however, has a marginally lower percentage of QS than panel 2. Regarding the midpoint design experiment, both panels 2 and 8 do not show any significant differences across experimental subgroups. We therefore conclude that there is no association between costs and response quality.

Discussion

In this article, we investigate the effect of respondent motivation on response quality in nonprobability and probability-based online panels. We assume that respondents in nonprobability online panels are more focused on the monetary incentives provided by the panel providers than respondents in probability-based online panels, resulting in higher satisficing among nonprobability online panel respondents than among probability-based online panel respondents.

In our study, we implemented the same survey with the exact same questionnaire across 10 online panels (seven nonprobability online panels and three probability-based online panels) during the same fieldwork period. In our analysis, we used three satisficing indicators: straight-lining in grid questions, item nonresponse, and midpoint selection in a visual design

experiment. To be able to focus on whether differences in respondent motivation in nonprobability and probability-based online panels lead to differences in the occurrence of satisficing, we kept the task difficulty constant across the online panels and controlled for respondent ability in our weighting schemes.

In line with our expectations, we found significantly more straight-lining in the nonprobability online panels than in the probability-based online panels. This indicates that nonprobability online panel respondents are indeed more prone to satisficing than probability-based online panel respondents due to the nonprobability online panel respondents' focus on maximizing their incentive-by-effort ratio.

Contrary to our expectation, however, we found little to no DK answers (Hypothesis 2a) and QS (Hypothesis 2c) across the online panels examined. In addition, while we found substantial amounts of DWS (Hypothesis 2b) across the examined online panels, this cannot be explained by the online panel sampling designs. This might mean that the respondents in the examined online panels do not choose item nonresponse as a satisficing strategy. One reason for this might be that the online panel respondents perceive the effort of having to click their way through the probes that typically pop up when they skip a question as more cumbersome than just providing an answer to the question. In addition, respondents might choose the DK answer option to express their genuine failure to recall the information asked for in the survey question (for more information on recall error, see Eisenhower, Mathiowetz, and Morganstein 2004). Furthermore, respondents might choose the DWS answer option to express their genuine wish to keep some information secret, potentially due to online data protection concerns (for empirical proof on data protection concerns in online surveys, see Joinson et al. 2008). Future research should examine the mechanisms leading respondents to choose the different types of item nonresponse in online panels more closely.

Contrary to our expectations, we also did not find any effect of the visual design of the answer scale on midpoint selection (Hypothesis 3) in any of the online panels. Our findings therefore suggest, contrary to the findings from Tourangeau et al. (2004), that the visual midpoint of an answer scale does not serve as a superficial decision-making cue in online surveys. Furthermore, the results from our midpoint selection experiment are contrary to Krosnick's (1991) reasoning that respondents choose the midpoint of an answer scale as a satisficing strategy. However, our results are in accordance with the empirical finding by Krosnick and Fabrigar (1997) who do not find any evidence for midpoint selection as a satisficing strategy. Future research should

examine the exact circumstances under which the effect of the visual design of the answer scale on midpoint selection found by Tourangeau et al. (2004) replicates.

Finally, we find that there is no association between response quality and survey costs across the nonprobability online panels. This means that investing in a more expensive panel does not automatically lead to better data.

Our study is one of very few that explores differences between nonprobability and probability-based online panels regarding response quality. While we only detect differences in terms of straight-lining, future studies will need to investigate whether our findings and null-findings replicate. In addition, our study focused on three satisficing measures as indicators of response quality. Future research should explore other response quality indicators, such as response order effects (see Krosnick and Alwin 1987) and acquiescence (see McClendon 1991).

We would also like to point out that other factors than the sampling design might be responsible for the significantly higher amount of straight-lining in the nonprobability online panels compared to the probability-based online panels. One alternative explanation for the variability in response quality across the online panels might be that academic online panels generally invest more in strategies that lead to high response quality, while non-academic panels do not. However, our finding that the non-academic probability-based online panel in our study (panel 1) has similarly low straight-lining rates as the academic probability-based online panels (GIP and GP) speaks against this alternative explanation.

Last, with regard to the generalizability of our findings, our study is limited to online panel response quality. Surveys in online panels are usually shorter than offline surveys typically are (for empirical proof on the potential effect of questionnaire length on satisficing, see Roberts et al. 2010). Furthermore, unlike many offline surveys, surveys in online panels are self-administered by the respondents (for empirical proof on the effect of the survey mode on satisficing, see Fricker et al. 2005; Heerwegh and Loosveldt 2008). Whether and how survey length and survey mode interact with the sampling design of a survey remains another question for future research.


Declaration of Conflicting Interests


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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge the support from the Collaborative Research Center (SFB) 884 “Political Economy of Reforms” (projects A8 and Z1, Project-ID 139943784), funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and from the GESIS Panel, and funded by the German Federal Ministry of Education and Research.

ORCID iDs

Carina Corneise  <https://orcid.org/0000-0002-5437-2999>

Annelies G. Blom  <https://orcid.org/0000-0003-0377-301X>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Based on American Association for Public Opinion Research (AAPOR) Standard Definitions (AAPOR 2016), completion rate (COMR) = number of wave participants divided by the number of recruited panel members.
2. Based on AAPOR Standard Definitions (AAPOR 2016), cumulative response rate (CUMRR) = number of wave participants divided by the number of eligible persons
3. Completion rate = number of wave participants divided by the number of recruited panel members.
4. CUMRR = number of wave participants divided by the number of gross sample members.
5. All online panels we examine in our article claim to be able to produce representative data.
6. Environmental zones are areas in German cities, from which cars that emit high levels of respirable dust are banned.

References

- Alwin, Duane F. and Jon A. Krosnick. 1991. “Aging, Cohorts, and the Stability of Sociopolitical Orientations over the Life Span.” *The American Journal of Sociology* 97(1):169-95.
- American Association for Public Opinion. 2016. *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. Retrieved March 19, 2020 (https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf).

- Blom, Annelies G., Christina Gathmann, and Ulrich Krieger. 2015. "Setting up an Online Panel Representative of the General Population: The German Internet Panel." *Field Methods* 27(4):391-408.
- Blom, Annelies G., Jessica M. Herzing, Carina Cornesse, Joseph W. Sakshaug, Ulrich Krieger, and Dayana Bossert. 2017. "Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-based Online Panels? Evidence from the German Internet Panel." *Social Science Computer Review* 35(4): 498-520.
- Bosnjak, Michael, Tanja Dannwolf, Tobias Enderle, Ines Schaurer, Bella Struminskaya, Angela Tanner, and Kai W. Weyandt. 2017. "Establishing an Open Probability-based Mixed-mode Panel of the General Population in Germany." *Social Science Computer Review* 21(1):1-13.
- Bradburn, Norman M. and Seymour Sudman. 1988. *Polls & Surveys: Understanding What They Tell Us*. San Francisco, CA: Jossey-Bass.
- Cacioppo, John T. and Richard E. Petty. 1982. "The Need for Cognition." *Journal of Personality and Social Psychology* 42(1):116-31.
- Chang, Linchiat and Jon A. Krosnick. 2009. "National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 7(3):641-78.
- Converse, Philip E. 1974. "Nonattitudes and American Public Opinion: Comment: The Status of Nonattitudes." *American Political Science Review* 68(2):650-60.
- Cornesse, Carina, Annelies G. Blom, David Dutwin, Jon A. Krosnick, Edith D. De Leeuw, Stéphane Legleye, Josh Pasek, Darren Pennay, Benjamin Phillips, Joseph W. Sakshaug, Bella Struminskaya, and Alexander Wenz. 2020. "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research." *Journal of Survey Statistics and Methodology* 8(1): 4-36.
- Couper, Mick P. 2008. *Designing Effective Web Surveys*. New York: Cambridge University Press.
- Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65(2):230-53.
- Eisenhower, Donna, Nancy A. Mathiowetz, and David Morganstein. 2004. "Recall Error: Sources and Bias Reduction Techniques." Pp. 125-44 in *Measurement Errors in Surveys*, edited by Paul Biemer, Robert M. Groves, Lars E. Lyberg, and Nancy A. Mathiowetz. Wiley: Hoboken, New Jersey.
- Feick, Lawrence. 1989. "Latent Class Analysis of Survey Questions That Include Don't Know Responses." *Public Opinion Quarterly* 53(1):525-47.
- Fowler, Floyd J. and Charles F. Cannell. 1996. Using Behavioral Coding to Identify Cognitive Problems with Survey Questions. Pp. 15-36 in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey*

- Research*, edited by Norbert Schwarz and Seymour Sudman. San Francisco, CA: Jossey-Bass.
- Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan. 2005. "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69(3):370-92.
- Gilljam, Mikael and Donald Granberg. 1993. "Should We Take Don't Know for an Answer?" *Public Opinion Quarterly* 57(3):348-57.
- Goel, Sjarad, Adam Obeng, and David Rothschild. 2015. "Non-representative Surveys: Fast, Cheap, and Mostly Accurate." Working Paper No. 27. Retrieved March 19, 2020 (<http://researchdmr.com/FastCheapAccurate>).
- Green, Melanie C., Jon A. Krosnick, and Allyson L. Holbrook. 2001. *The Survey Response Process in Telephone and Face-to-face Surveys: Differences in Respondent Satisficing and Social Desirability Response Bias*. Manuscript, Ohio State University. Retrieved March 19, 2020 (<http://users.clas.ufl.edu/kenwald/pos6757/spring02/tch62.pdf>).
- GreenBook. 2017. *GRIT CPR Report—2017 Global Respondent Engagement Study*. Retrieved March 19, 2020 (https://www.greenbook.org/pdfs/GRIT_CPR_Final_42017.pdf).
- Greszki, Robert, Marco Meyer, and Harald Schoen. 2014. "The Impact of Speeding on Data Quality in Nonprobability and Freshly Recruited Probability-based Online Panels." Pp. 238-62 in *Online Panel Research: A Data Quality Perspective*, edited by Anja S. Göritz, Jelke Geert Bethlehem, Paul J Lavrakas, Reginald P. Baker, and Jon Krosnick. Chichester, England: John Wiley & Sons.
- Groves, Robert M., Stanley Presser, and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68(1):2-31.
- Heerwegh, Dirk and Geert Loosveldt. 2008. "Face-to-face versus Web Surveying in a High-internet-coverage Population: Differences in Response Quality." *Public Opinion Quarterly* 72(5):836-46.
- Hillygus, D. Sunshine, Natalie Jackson, and McKenzie Young. 2014. "Professional Respondents in Non-probability Online Panels." Pp. 219-37 in *Online Panel Research: A Data Quality Perspective*, edited by Anja S. Göritz, Jelke Geert Bethlehem, Paul J. Lavrakas, Reginald P. Baker, and Jon Krosnick. Chichester, England: John Wiley & Sons.
- Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone versus Face-to-face Interviewing of National Probability Samples with Long Questionnaires." *Public Opinion Quarterly* 67(1):79-125.
- Joinson, Adam N., Paine, Carina, Buchanan, Tom, and Reips, Ulf-Dietrich. 2008. "Measuring Self-disclosure Online: Blurring and Non-response to Sensitive Items in Web-based Surveys." *Computers in Human Behavior* 24(5):2158-171.

- Kaminska, Olena, Allan L. McCutcheon, and Jaak Billiet. 2010. "Satisficing among Reluctant Respondents in a Cross-national Context." *Public Opinion Quarterly* 74(5):956-84.
- Kennedy, Courtney, Andrew Mercer, Scott Keeter, Nick Hatley, Kyle McGeeney, and Alejandra Gimenez. 2016. "Evaluating Online Nonprobability Surveys. Vendor Choice Matters; Widespread Errors Found for Estimates Based on Blacks and Hispanics." *Pew Research Center*. Retrieved (<https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys/>).
- Kusch, Florian, Bernad Batinic, and Wolfgang Mayerhofer. 2014. "Motives for Joining Nonprobability Online Panels and Their Association with Survey Participation Behavior." Pp. 171-91 in *Online Panel Research: A Data Quality Perspective*, edited by Anja S. Göritz, Jelke Geert Bethlehem, Paul J Lavrakas, Reginald P. Baker, and Jon Krosnick. Chichester, England: John Wiley & Sons.
- Knäuper, Bärbel. 1999. "The Impact of Age and Education on Response Order Effects in Attitude Measurement." *Public Opinion Quarterly* 63(3):347-70.
- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5(3):213-36.
- Krosnick, Jon A. 1992. "The Impact of Cognitive Sophistication and Attitude Importance on Response-order and Question-order Effects." Pp. 203-18 in *Context Effects in Social and Psychological Research*, edited by Norbert Schwarz and Seymour Sudman. New York: Springer-Verlag.
- Krosnick, Jon A. 1999. "Maximizing Questionnaire Quality." Pp. 37-57 in *Measures of Political Attitudes*, edited by John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman. Oxford, England: Elsevier.
- Krosnick, Jon A. and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-order Effects in Survey Measurement." *Public Opinion Quarterly* 51(2):201-19.
- Krosnick, Jon A. and Matthew K. Berent. 1993. "Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format." *American Journal of Political Science* 73(3):941-64.
- Krosnick, Jon A. and Leandre R. Fabrigar. 1997. Designing Rating Scales for Effective Measurement in Surveys. Pp. 141-64 in *Survey Measurement and Process Quality*, edited by Lars Lyberg, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. Hoboken, NJ: John Wiley & Sons.
- Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert Cameron Mitchell, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green, and Michael Conaway. 2002. "The Impact of "No Opinion" Response Options on Data Quality." *Public Opinion Quarterly* 66(1):371-403.

- Legleye, Stéphane, Géraldine Charrance, Nicolas Razafindratsima, Nathalie Bajos, Aline Bohet, Caroline Moreau, and the FECOND Research Team. 2015. "The Use of a Nonprobability Internet Panel to Monitor Sexual Reproductive Health in the General Population." *Sociological Methods & Research* 47(2):1-35.
- Loosveldt, Geert and Nathalie Sonck. 2008. "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2(2): 93-105.
- Malhotra, Neil. 2008. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion Quarterly* 72(5):914-34.
- Malhotra, Neil and Jon A. Krosnick. . 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 15(3):286-323.
- Malhotra, Neil, Joanne M. Miller, and Justin Wedeking. 2014. "The Relationship between Nonresponse Strategies and Measurement Error: Comparing Online Panel Surveys to Traditional Surveys." Pp. 313-36 in *Online Panel Research: A Data Quality Perspective*, edited by Anja S. Göritz, Jelke Geert Bethlehem, Paul J Lavrakas, Reginald P. Baker, and Jon Krosnick. Chichester, England: John Wiley & Sons.
- McClendon, McKee J. 1991. "Acquiescence and Recency Response-order Effects in Interview Surveys." *Sociological Methods & Research* 20(1):60-103.
- Narayan, Sowmya and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60(1):58.
- Pasek, Josh. 2016. "When Will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence." *International Journal of Public Opinion Research* 28(2):269-91.
- Pennay, Darren Walter, Dina Neiger, Paul J. Lavrakas, Kim A. Borg, Sebastian Misson, and Nikki Honey. 2016. "2015–2016 Australian Online Panels Benchmarking Study: A Comparison of Surveys Using Probability and Nonprobability Samples in an Australian Research Context." Paper presented at the 69th World Association for Public Opinion Research (WAPOR) Conference, May, Austin, Texas.
- Roberts, Caroline, Gillian Eva, Nick Allum, and Peter Lynn. 2010. "Data Quality in Telephone Surveys and the Effect of Questionnaire Length: A Cross-national Experiment." *ISER Working Paper Series* No. 2010-36. Retrieved (<https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2010-36>).
- Schonlau, Matthias and Vera Toepoel. 2015. "Straightlining in Web Survey Panels over Time." *Survey Research Methods* 9(2):125-37.
- Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys*. New York: Academic Press.

- Smith, Tom W. 2003. "An Experimental Comparison of Knowledge Networks and the GSS." *International Journal of Public Opinion Research* 15(2):167-79.
- Sparrow, Nick. 2006. "Developing Reliable Online Polls." *International Journal of Market Research* 48(6):659.
- Stoop, Ineke A. 2005. *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague, the Netherlands: Social and Cultural Planning Office.
- Sturgis, Patrick, Caroline Roberts, and Patten Smith. 2014. "Middle Alternatives Revisited." *Sociological Methods & Research* 43(1):15-38.
- Toepoel, Vera, Corrie Vis, Marcel Das, and Arthur van Soest. 2009. "Design of Web Questionnaires. An Information-processing Perspective for the Effect of response Categories." *Sociological Methods & Research* 37(3):371-92.
- Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. 2004. "Spacing, Position, and Order Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68(3):368-93.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, England: Cambridge University Press.
- Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-representative Polls." *International Journal of Forecasting* 31(1):980-91.
- Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Lavitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-probability Samples." *Public Opinion Quarterly* 75(4):709-47.

Author Biographies

Carina Cornesse is a doctoral researcher with the SFB 884 "Political Economy of Reforms" at the University of Mannheim and GESIS—Leibniz Institute for the Social Sciences. Her research interests include survey methods, in particular nonresponse, representativeness, and response quality. Her first papers have been published in *Social Science Computer Review* and accepted for publication in *Survey Research Methods*.

Annelies G. Blom is a professor at the Department of Political Science, School of Social Sciences, University of Mannheim, and the SFB 884 "Political Economy of Reforms," University of Mannheim. Her research interests include representativeness and measurement quality of online surveys, methods of longitudinal research, non-response and attrition, and interviewer effects. Her research has, among other journals, been published in *Sociological Methods & Research*, *Public Opinion Quarterly*, *Journal of the Royal Statistical Society: Series A*, and *Social Science Computer Review*.