

Journal : PUBOPQ
Article Doi : 10.1093/poq/nfz035
Article Title : SATISFICING IN SURVEYS
First Author : CAROLINE ROBERTS
Corr. Author : CAROLINE ROBERTS

INSTRUCTIONS

We encourage you to use Adobe's editing tools (please see the next page for instructions). If this is not possible, fax it to: +44 (0)1865 355728 or e-mail a list of corrections to pubopq@oup.com. Please do not send corrections as track changed Word documents.

Changes should be corrections of typographical errors only. Changes that contradict journal style will not be made.

These proofs are for checking purposes only. They should not be considered as final publication format. The proof must not be used for any other purpose. In particular we request that you: do not post them on your personal/institutional web site, and do not print and distribute multiple copies (please use the attached offprint order form). Neither excerpts nor all of the article should be included in other publications written or edited by yourself until the final version has been published and the full citation details are available. You will be sent these when the article is published.

- 1. Licence to Publish:** If you have not already done so, please also complete the Licence to Publish form online by following the link sent to you in the 'Welcome' email.
- 2. Permissions:** Permission to reproduce any third party material in your paper should have been obtained prior to acceptance. If your paper contains figures or text that require permission to reproduce, please confirm that you have obtained all relevant permissions and that the correct permission text has been used as required by the copyright holders. Please contact jnls.author.support@oup.com if you have any questions regarding permissions.
- 3. Author groups:** Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online.
- 4.** Please check carefully the entire author note, including the spelling of names, institutional information, any acknowledgements, and correspondence details.
- 5. Figures:** Figures have been placed as close as possible to their first citation. Please check that they are complete and that the correct figure legend is present. Figures in the proof are low resolution versions that will be replaced with high resolution versions when the journal is printed.
- 6. Funding:** Please provide a Funding statement, detailing any funding received. Remember that any funding used while completing this work should be highlighted in a separate Funding section. Please ensure that you use the full official name of the funding body, and if your paper has received funding from any institution, such as NIH, please inform us of the grant number to go into the funding section. We use the institution names to tag NIH-funded articles so they are deposited at PMC. If we already have this information, we will have tagged it and it will appear as coloured text in the funding paragraph. Please check the information is correct.
- 7. Tables:** Please be sure to check all elements of tables very carefully.

MAKING CORRECTIONS TO YOUR PROOF

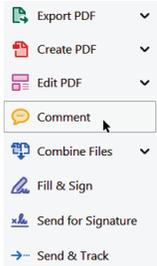
These instructions show you how to mark changes or add notes to your proofs using Adobe Acrobat Professional versions 7 and onwards, or Adobe Reader DC. To check what version you are using go to **Help** then **About**. The latest version of Adobe Reader is available for free from get.adobe.com/reader.

DO NOT OVERWRITE TEXT, USE COMMENTING TOOLS ONLY.

DISPLAYING THE TOOLBARS

Adobe Reader DC

In Adobe Reader DC, the Comment toolbar can be found by clicking 'Comment' in the menu on the right-hand side of the page (shown below).

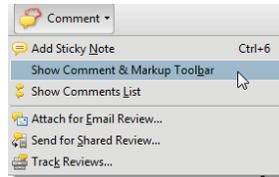


The toolbar shown below will then display along the top.



Acrobat Professional 7, 8, and 9

In Adobe Professional, the Comment toolbar can be found by clicking 'Comment(s)' in the top toolbar, and then clicking 'Show Comment & Markup Toolbar' (shown below).

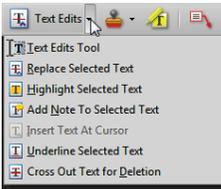


The toolbar shown below will then be displayed along the top.



USING TEXT EDITS AND COMMENTS IN ADOBE ACROBAT

This is the quickest, simplest and easiest method both to make corrections, and for your corrections to be transferred and checked.



1. Click **Text Edits**
2. Select the text to be annotated or place your cursor at the insertion point and start typing.
3. Click the **Text Edits** drop down arrow and select the required action.

You can also right click on selected text for a range of commenting options, or add sticky notes.

SAVING COMMENTS

In order to save your comments and notes, you need to save the file (**File, Save**) when you close the document.

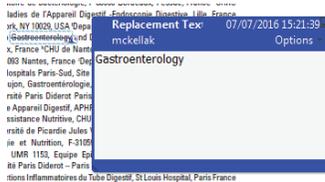
USING COMMENTING TOOLS IN ADOBE READER

All commenting tools are displayed in the toolbar. To edit your document, use the highlighter, sticky notes, and the variety of insert/replace text options.



POP-UP NOTES

In both Reader and Acrobat, when you insert or edit text a pop-up box will appear. In Acrobat it looks like this:



In Reader it looks like this, and will appear in the right-hand pane:



AUTHOR QUERY FORM

Journal : PUBOPQ

Article Doi : 10.1093/poq/nfz035

Article Title : SATISFICING IN SURVEYS

First Author : CAROLINE ROBERTS

Corr. Author : CAROLINE ROBERTS

AUTHOR QUERIES - TO BE ANSWERED BY THE CORRESPONDING AUTHOR

Please ensure that all queries are answered, as otherwise publication will be delayed and we will be unable to complete the next stage of the production process.

Please note that proofs will not be sent back for further editing.

The following queries have arisen during the typesetting of your manuscript. Please click on each query number and respond by indicating the change required within the text of the article. If no change is needed please add a note saying “No change.”

<p>AQ1</p>	<p>Please note that the cross-reference “Cannell, Miller, and Okun  1981” has not been provided in the reference list. Please provide the same.</p>
<p>AQ2</p>	<p>in paragraph beginning “Building on this foundational work” please clarify the source referred to as “ibid.” (Tourangeau, Ripps, and Rasiński 2000)</p>
<p>AQ3</p>	<p>please confirm citation for Gilbert 2015: Gilbert, Emily E. 2015. “A Comparison of Branched Versus Unbranched Rating Scales for the Measurement of Attitudes in Surveys.” <i>Public Opinion Quarterly</i> 79:443–70.] Correct </p>
<p>AQ4</p>	<p>Text for footnote b is missing in table 2. Please check and amend. </p>
<p>AQ5</p>	<p>Two footnotes are labeled as “a”  table 3. Please check and amend. There is no text for footnote b.</p>
<p>AQ6</p>	<p>under subheading “Predictors of  ficing,” please clarify that you mean “eleven of the shortlisted article”</p>
<p>AQ7</p>	<p>Please note that the cross-reference  ller et al. 2013” has not been provided in the reference list. Please provide the same.</p>
<p>AQ8</p>	<p>please confirm that Oppenheimer, Meyvis, and Davidenko 2009 is Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko. 2009. “Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power.” <i>Journal of Experimental Social Psychology</i> 45:867–72.] Correct.</p>

AQ9	please confirm that Gao, House, and Xie is Gao, Zhifeng, Lisa A. House, and Jing Xie. 2015. "Online Survey Data Quality and Its Implication for Willingness-to-Pay: A Cross-Country Comparison." Canadian Journal of Agricultural Economics 64:199–221.] Correct.
AQ10	please confirm that Hauser and Schwarz 2015 is Hauser, David J., and Norbert Schwarz. 2015. "It's a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on 'Tricky' Tasks." Sage Open 5:1–6.] Correct.
AQ11	please confirm that Revilla and Ochoa 2015 is Revilla, Melanie, and Carlos Ochoa. 2015. "What Are the Links in a Web Survey Among Response Time, Quality, and Auto-Evaluation of the Efforts Done?" Social Science Computer Review 33:97–114.] Correct - text reference corrected.
AQ12	in citation for Higgins and Green, please check url ">www.cochrane-handbook.org" Updated URL added.

SATISFICING IN SURVEYS

A SYSTEMATIC REVIEW OF THE LITERATURE

CAROLINE ROBERTS*

EMILY GILBERT

NICK ALLUM

LÉILA EISNER^o



1.5

1.10

Abstract Herbert Simon's (1956) concept of satisficing provides an intuitive explanation for the reasons why respondents to surveys sometimes adopt response strategies that can lead to a reduction in data quality. As such, the concept rapidly gained popularity among researchers after it was first introduced to the field of survey methodology by Krosnick and Alwin (1987), and it has become a widely cited buzzword linked to different forms of response error. In this article, we present the findings of a systematic review involving a content analysis of journal articles published in English-language journals between 1987 and 2015 that have drawn on the satisficing concept to evaluate survey data quality. Based on extensive searches of online databases, and an initial screening exercise to apply the study's inclusion criteria, 141 relevant articles were identified. Guided by the theory of survey satisficing described by Krosnick (1991), the methodological features of the shortlisted articles were coded, including the indicators of satisficing analyzed, the main predictors of satisficing, and the presence of main or interaction effects on the prevalence of satisficing involving indicators of task difficulty, respondent ability, and respondent motivation. Our analysis sheds light on potential differences in the extent to which satisficing theory holds for different types of response error, and highlights a number of avenues for future research.

1.15

1.20

1.25

1.30

CAROLINE ROBERTS is an assistant professor in the Institute of Social Sciences at the University of Lausanne, Lausanne, Switzerland. EMILY GILBERT is a survey manager at the Centre for Longitudinal Studies, UCL Institute of Education, London, UK. NICK ALLUM is professor of research methodology in the Department of Sociology at the University of Essex, Wivenhoe Park, Colchester, UK. LÉILA EISNER is a researcher in social psychology at the University of Lausanne, Lausanne, Switzerland. The authors thank the anonymous reviewers and the editors for their constructive feedback on earlier versions of this manuscript. *Address correspondence to Caroline Roberts, Institute for Social Sciences, University of Lausanne, Bâtiment Géopolis, Quartier Mouline, 1015 Lausanne, Switzerland; e-mail: caroline.roberts@unil.ch.

1.35

1.40

Background

Measurement error —defined as the “observational gap between the ideal measurement and the response obtained” (Groves et al. 2009, p. 51)— is often considered to be the most problematic source of survey error (Biemer and Lyberg 2003; Biemer 2010, p. 823). This is partly because measurement quality is so fundamental to the validity of the conclusions drawn from a survey (Alwin 1991, p. 5), and partly because the extent of its damaging effects cannot easily be ascertained or corrected. Error can reduce the overall efficiency of the data by introducing “noise” into measures of single variables and their relations with other variables, and it can lead to a substantial over- or underestimation of the prevalence of phenomena of interest (Alwin 2007).

Despite the threat it poses to the overall quality of data, measurement error can be reduced at relatively low cost compared to other survey errors (Fowler 1995, p. 150) by understanding its causes and taking remedial action (e.g., Foddy 1993; Payne 1951; Fowler 1995). Contributions from cognitive psychology have played a prominent role in this endeavor (e.g., Sudman, Bradburn, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000; Schwarz 2007), emphasizing the mental processes by which respondents complete the task of answering survey questions, the factors that can inhibit optimal processing, and, hence, offering clues as to how to improve question design (Willis 2004). The theory of satisficing (Krosnick and Alwin 1987; Krosnick 1991), which emphasizes the role of motivational and ability factors in determining how thoroughly cognitive processes are executed, provides a compelling explanation for how certain types of response errors may sometimes arise. Over the past three decades, researchers have been drawing on the framework provided by satisficing theory, as a means of indirectly assessing the relative extent of measurement error under divergent conditions and investigating its correlates. This work has contributed a substantial body of empirical findings to knowledge about what affects response quality and possible remedies. The aim of this paper is to systematically review this literature, in order to summarize what has been learned, and to identify fruitful avenues for future research. Before describing the aims in greater detail and the methods used, we briefly discuss different sources of measurement error, and present the key tenets of satisficing theory.

SOURCES OF MEASUREMENT ERROR

Measurement error arises from different sources: characteristics of the mode of data collection and the survey setting, the respondent, and the design of the questionnaire (Biemer and Lyberg 2003; Groves et al. 2009). Understanding of how the questionnaire influences response quality has largely been shaped by the findings of early contributors to the field of public opinion research, who pioneered the use of split-ballot experiments to test the effects of formulating

questions in different ways (e.g., [Cantril 1940](#); [Payne 1951](#)). Later researchers (notably, [Bishop, Oldendick, and Tuchfarber 1978](#); [Kalton, Collins, and Brook 1978](#); [Schuman and Presser 1981](#)) started to amass and replicate experimental evidence demonstrating how seemingly innocuous variations in wording or response alternatives could affect response distributions, highlighting the role of task characteristics in respondents’ answers to survey questions. 3.5

Building on this foundational work, cognitive psychologists focused on the respondent’s contribution to response quality, devising models of the mental processes involved in answering survey questions (originally, [Cannell, Miller, and Oksenburg 1981](#); later elaborated by, e.g., [Tourangeau, Ripps, and Rasinski 2000](#)) that typically identify four¹ main stages: (1) comprehending the survey question; (2) searching memory to retrieve considerations relevant to the answer; (3) integrating the retrieved information into a judgment; and (4) selecting from the available response categories. Errors can arise at each stage (ibid.), based on a variety of influences, including natural limits on respondents’ working memory, processing biases, as well as motivational factors, such as deliberately editing responses that may seem threatening to reveal (as in social desirability bias), or using heuristics or other shortcuts to arrive at a satisfactory, but possibly invalid, answer ([Tourangeau, Ripps, and Rasinski 2000](#)). The design of the questionnaire exerts an influence on the response process where questions are difficult to understand, present recall challenges, require complex mental calculations, or offer ambiguous, or potentially sensitive, response alternatives. Idiosyncratic interviewing styles and interviewer characteristics can similarly affect how respondents answer, either directly or indirectly ([Biemer and Lyberg 2003](#); [Holbrook, Green, and Krosnick 2003](#)). Meanwhile, other mode characteristics—such as computer-assisted or web-based administration, or being restricted to audio-only communication channels ([Couper 2011](#))—can influence how response tasks are executed and, hence, the quality of respondents’ answers. 3.10

Understanding the relative influence of these different sources of measurement error, and the psychological mechanisms by which they affect response, is key to finding optimal strategies for mitigating their effects. However, part of the challenge in predicting when measurement error will occur, and the particular form it will take, stems from the complex interaction between each of the components involved. The theory of satisficing provides a framework for understanding this interaction. 3.15

THE THEORY OF SATISFICING 3.20

[Krosnick and Alwin \(1987, 1988, 1989\)](#) and [Krosnick \(1991\)](#) first developed the application of Herbert [Simon’s \(1956\)](#) concepts of “satisficing” and 3.25

1. Some authors refer to a fifth stage, prior to comprehension, involving initial perception of the survey question (e.g., [Dillman, Smyth, and Christian 2014](#)). 3.30

3.35

THE THEORY OF SATISFICING

[Krosnick and Alwin \(1987, 1988, 1989\)](#) and [Krosnick \(1991\)](#) first developed the application of Herbert [Simon’s \(1956\)](#) concepts of “satisficing” and 3.40

1. Some authors refer to a fifth stage, prior to comprehension, involving initial perception of the survey question (e.g., [Dillman, Smyth, and Christian 2014](#)). 3.44

“optimizing” to the field of survey methodology (although see [Tourangeau 1984](#)). Satisficing refers to the expenditure of minimum effort to generate a satisfactory response, compared with expending a great deal of effort to generate a “maximally valid response” ([Alwin 1991](#), pp. 17–18). [Krosnick \(1991\)](#) distinguished between stronger and weaker forms of satisficing. In “weak satisficing,” respondents execute all the different stages of processing, but do so less thoroughly, resulting in response behaviors such as *selecting the first acceptable response alternative* (manifesting as response order [primacy and recency] effects) and *acquiescence* (the tendency to agree with assertions). By contrast, in “strong satisficing,” one or more stages of processing is skipped altogether, producing response errors such as *endorsing the status quo* (a preference for the middle “keep things the same” alternative in questions asking about support for policy change); *non-differentiation* (the tendency to select the same point on a rating scale to rate multiple items presented with the same response alternatives); saying “don’t know” instead of expressing an opinion; and “mental coin-flipping” (selecting response alternatives at random).

Each type of response error is said to be “more likely to occur under the conditions that foster satisficing” ([Krosnick 1991](#), p. 220), which are a function of three mediating factors: (1) high task difficulty; (2) low respondent ability; and (3) low motivation to perform the response task. To the extent that questionnaire design features, mode (including interviewers), and respondent characteristics contribute to these conditions, satisficing may result, with deleterious effects on response quality. Hypothetically, any variable indicating these conditions can exert a main effect on the prevalence of different types of response effect, and according to [Krosnick](#), these main effects may be additive, but “their relations are more likely to be multiplicative” (1991, p. 225). Formally, performance of the response task depends on the ratio of task difficulty and the product of respondent ability and motivation ([Krosnick 1991](#)):

$$p(\text{Satisficing}) = \frac{a_1(\text{Task difficulty})}{a_2(\text{Ability}) \times a_3(\text{Motivation})}$$

Testing the theory, therefore, implies measuring these different elements in a given survey context and assessing the extent to which the hypothesized model holds.

The Present Study

At first sight, the accumulated literature on satisficing in surveys can appear unwieldy. The findings are mixed, and the possibility to generalize from them is hindered by the variety of methods used for constructing indicators of response quality (which tend to be used as proxies for satisficing). It is also unclear whether the evidence, taken as a whole, is consistent with the theory. The present study was designed to address these concerns through a systematic

review. The key objective of the review was to *describe and summarize* the existing research findings relating to satisficing theory by achieving two goals. The first is to identify published research that has drawn on the theory of satisficing as a framework for investigating response quality in surveys. The second is to systematically record key features of the research, including the types of response effects hypothesized to result from satisficing, variables hypothesized to predict satisficing, and the presence of main and interaction effects of predictors on different types of response effect that either support or contradict satisficing theory.

5.5

In so doing, we aim to draw conclusions about the compatibility of satisficing theory with the empirical evidence, assist others working in this area to develop clearer recommendations about the optimal ways to identify response errors in survey data, and identify their underlying causes.

5.10

Methods

5.15

We designed our systematic review based on best practice in the field of health sciences and evidence-based medicine (see [Torgerson 2003](#); [Higgins and Green 2011](#)), as well as on guidelines for using the method in the social sciences (see [Petticrew and Roberts 2005](#)). The design entailed two main stages: (1) study identification and selection; and (2) data extraction and synthesis. Both involved content analytic procedures in which features of the texts were systematically coded using a purpose-designed coding frame (see [table A1](#) in the Online Appendix).

5.20

5.25

STAGE 1: STUDY IDENTIFICATION AND SELECTION

Initially, general and specialized bibliographic databases were searched using a combination of search terms (see below) to identify records referring to either “satisficing” in relation to survey measurement, or one of the original publications in which the theory was first developed (specifically, [Krosnick and Alwin 1987](#); [Krosnick 1991](#)).² The aim at this stage was to identify as many citations as possible to assess the scale of the review, including work subjected to editorial control or peer review, published in academic journals and books, and gray literature, such as institutional or technical reports, working papers, conference proceedings, dissertations, and theses. We placed no restrictions on language (except for using English search terms).

5.30

5.35

The following search engines were used: the Thomson Reuters ISI Web of Science Databases, a collection of seven databases, including the Social Sciences Citation Index and the Conference Proceedings Citation

5.40

2. Because the study was carried out over the course of several years, the searches were updated intermittently, using the same search strategies and search terms.

5.44

Index—Social Sciences and Humanities. The Web of Science enables cited reference searching, which was used to initiate the search, to identify records citing [Krosnick and Alwin \(1987\)](#) and [Krosnick \(1991\)](#). The same databases were then searched for literature containing combinations of the terms “satisficing,” “satisfice,” “survey,” and “questionnaire” (details of the precise search terms used in each database are available in [table A2](#) in the Online Appendix). The same search strings were used to search other academic databases, including Scirus, PsychInfo, and Academic Search Premier (occasional modifications to the search strings used were necessary depending on the contents of the database and how each search engine worked). The search was expanded using the same search terms in Google Scholar.

In addition to these large-scale database and web-based searches, online search engines for leading academic journals publishing articles in the field of survey methodology were targeted, as well as working paper series, and conference proceedings of the American Statistical Association Survey Methods Section.

Inclusion criteria: The searches produced a total of 3,581 records (before removing duplicates). After duplicates and obviously irrelevant reports were removed, a total of 1,526 unique records remained, including journal articles, books and book chapters, dissertations, working papers, and reports. These records (abstracts, and where available and necessary, full texts) were further screened to identify studies that met the inclusion criteria and were eligible for in-depth review, and to verify that they were fully in scope (i.e., included the relevant search terms).

Four inclusion criteria were applied sequentially, as follows. The first identified English-language research articles published in academic journals between the years of 1987 and 2015.³ The second identified articles with a methodological focus (as opposed to being focused solely on substantive research questions)—for example, studies comparing response quality in different modes of data collection (e.g., [Holbrook, Green, and Krosnick 2003](#); [Chang and Krosnick 2009](#)); and studies involving comparisons across variant question formulations (e.g., [Bishop and Smith 2001](#); Gilbert 2015). The third identified articles presenting new findings based on empirical data and analysis (as opposed to non-empirical papers with a theoretical focus—for example, [Tourangeau 2003](#); [Couper 2011](#)). The final inclusion criterion identified articles presenting comparisons of response quality across groups

AQ3

3. The search results included records of journal articles published in 2015 (or earlier) through Advance Access, but which were finally published in 2016. Because the data extraction procedure was still ongoing during 2016, three Advanced Access articles first published in 2015 were later excluded from the shortlist to ensure consistent application of the inclusion criterion relating to publication date.

assumed to differ in terms of their exposure to conditions hypothesized to foster satisficing. This criterion applied to studies that were guided explicitly by the framework provided by the theory of satisficing, as well as to studies that only referred briefly to the theory but were still concerned with variations in response quality. This distinction was not always clear-cut, so both types of study were retained in the sample for the second, more detailed stage of coding. Any article making only a passing reference to one of the search terms, or briefly describing the theory without presenting relevant new empirical evidence, was excluded.

7.5

7.10

STAGE 2: DATA EXTRACTION AND SYNTHESIS

Having completed this preliminary screening, the shortlisted studies were subjected to the second “data extraction” stage of coding, which focused on documenting the methodological features of the research. The main aims were to: (1) document the main features of the research designs; (2) identify which indicators of satisficing were analyzed and how they were constructed, as well as which independent variables (i.e., correlates or predictors of satisficing) were analyzed; and (3) record the main findings of the research relating to survey satisficing. For this, a purpose-designed coding frame was developed, consisting mainly of closed, pre-coded items, based on the theoretical model described by [Krosnick \(1991\)](#).

7.15

7.20

The following variables were coded: type of research design (experiment vs. non-experimental survey data); mode(s) of data collection; indicators of weak satisficing (primacy, recency, acquiescence); indicators of strong satisficing (endorsing the status quo, nondifferentiation, don’t know/no opinion, random reporting); other indicators of data quality (item nonresponse, middle alternatives, others); predictors of satisficing (indicators of task difficulty, respondent ability, and respondent motivation mentioned by [Krosnick 1991](#), and others not mentioned by him); and main effects and interaction effects (whether effects were consistent with or contradicted satisficing theory).

7.25

7.30

The coding frame was initially set up in an Excel spreadsheet, but was later programmed as an online questionnaire in Qualtrics to improve the usability and reliability of the instrument. The data were combined and analyzed in SPSS.

7.35

INTERCODER RELIABILITY

In content analysis, assessments of intercoder reliability (or, more specifically, rates of intercoder agreement—[Tinsley and Weiss 2000](#)) are key for testing and validating the coding scheme ([Kolbe and Burnett 1991](#); [Neuendorf 2002](#)), as well as for providing reassurance as to the validity of the results. All four authors were involved in both stages of coding. All coders independently coded a sample of articles assigned only to them, plus a randomly selected

7.40

7.44

subset of articles assigned to one of the other coders to permit an analysis of intercoder reliability. At the study selection stage, 32 percent of the articles retrieved from the searches were reviewed by two out of the four coders. At the data extraction stage, 23 percent of the shortlisted articles were double-coded. We report two indices of intercoder reliability: (1) the rate of agreement between coders, which has the advantage of being intuitive, but on its own is not considered adequate as it may give a misleading estimate of reliability between coders (Lombard, Snyder-Duch, and Campanella Bracken 2010); and (2) Cohen's kappa (as recommended, for example, by Landis and Koch 1977; Dewey 1983), which offers a number of advantages as an agreement index, including the fact that it accounts for levels of agreement that would be expected by chance.

The upper half of table 1 shows rates of agreement between coders for our three main inclusion criteria. For the "nature of reference" code (which refers to the nature of the reference to satisficing theory and the relevance of the empirical evidence presented), we permitted similar values to count as agreement. The main distinction of interest was between articles in which the search terms were only mentioned briefly, and articles presenting relevant empirical evidence relevant to the theory. The rates of agreement between coders for these codes were 93.7 percent for methodological versus substantive ($k = 0.84$), 96.4 percent ($k = 0.85$) for empirical versus theoretical, and 97.0 percent ($k = 0.91$) for nature of reference. These values were deemed to be more than acceptable. In any case, coders discussed and resolved all disagreements over the application of the inclusion criteria to ensure that no article was incorrectly excluded from the shortlist.

The lower half of table 1 shows the intercoder reliability indices for a selection of codes from the coding scheme used at the data extraction stage. Here, the rates of agreement varied more, ranging from 66.7 percent ($k = 0.35$) for whether the research analyzed independent variables that measured respondent motivation, to 96.8 percent ($k = 0.92$) for whether the research design was single or mixed mode. Overall, the mean percentage agreement across all variables for which intercoder reliability was assessed was 83.3 percent. The first author adjudicated on all discrepancies between coders, by referring to the article and deciding on a revised set of final codes. Modifications were made in the online version of the coding frame to improve reliability.

Results

After establishing the eligibility of the citations generated by our searches, we identified a total of 951 unique English-language journal articles referencing the search terms. After an initial "rush" of articles citing the search terms in the year following the publication of Krosnick's (1991) article, the number of articles citing the search terms remained reasonably constant until the mid-2000s (on average, 27 articles per year). At this time, there was a sharp increase in

Table 1. Intercoder reliability indices before adjudication

Codes	Percent agreement (%)	Cohen's kappa, <i>k</i>	
Stage 1 codes: Study selection			9.5
Inclusion criteria			
Methodological, substantive, or both	93.7	0.84	
Empirical, theoretical, or both	96.4	0.85	
Nature of reference (relevant empirical evidence vs. brief mention of search terms)	97.0	0.91	9.10
Stage 2 codes: Data abstraction and synthesis			
Study research design			
Survey experiment/Non-experimental	86.7	0.60	
Single/Mixed mode	96.7	0.92	9.15
Indicators of weak satisficing analyzed			
Primacy/Recency/Acquiescence	86.7	0.78	
Indicators of strong satisficing analyzed ^a			
Non-differentiation/Don't knows/Random reporting	80.0	0.68	
Independent variables analyzed			
Task difficulty	73.3	0.46	9.20
Respondent ability	83.3	0.70	
Respondent motivation	66.7	0.35	
Significant main effects			
Task difficulty	90.0	0.76	
Respondent ability	93.3	0.86	9.25
Respondent motivation	83.3	0.52	
Interactions			
Significant interaction effects reported	76.7	0.43	

^aEndorsing the status quo not included, as only a few studies used this indicator.

the number of publications (on average, 56 articles per year during the past decade) and, apart from a dip between 2011 and 2013, the number of publications citing the search terms has continued to rise. In 2015, there were 96 citations (the number of journal articles citing the search terms by year of publication for each subsample retained are shown in [figure 1](#)).

Of the 951 journal articles, 544 were excluded for having a substantive, non-methodological focus; and a further 55 were excluded for having purely theoretical, non-empirical content (see [figure 2](#) for a summary). This left 352 articles with a methodological, empirical focus (one of which we were unable to access, so it was not coded further). Of these, 207 articles were excluded because they contained only a passing reference to one of the search terms. A total of 144 were articles that presented new empirical data from studies comparing data quality between groups, either guided explicitly ($n = 87$) or

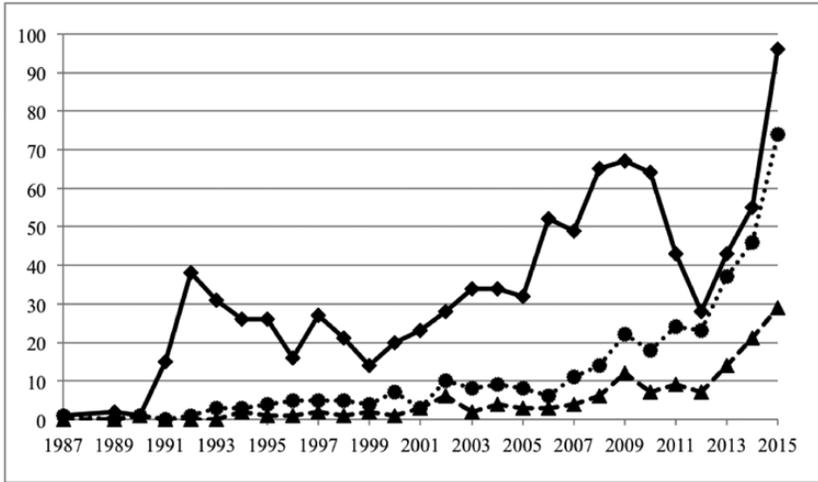


Figure 1. Number of studies published in academic journals, 1987–2015. The upper line shows the total number of articles ($n = 951$) by year of publication, the middle line shows the number of articles coded as methodological and empirical ($n = 352$), and the lower line shows the number of shortlisted articles ($n = 141$).

not explicitly ($n = 57$) by the theory of satisficing (three of which were excluded because their final publication date was 2016). Thus, the remaining analysis is based on this combined set of 141 articles.

The 141 shortlisted articles came from a wide range of publications (a total of 59 different journals spanning different academic disciplines), demonstrating the widespread interest in assessing survey response quality and the reach of satisficing theory (see table 2). A majority of the studies (65.2 percent) presented analyses of data from purpose-designed split-ballot survey experiments. The remaining studies were based on regular survey data (31.2 percent) or other data sources, including cognitive interviews (e.g., Darker and French 2009; Wagner and Zeglovitz 2014), eye-tracking studies (e.g., Galesic et al. 2008), register data (e.g., Brockington 2003), and paper-and-pencil questionnaires administered in schools (Wicker, Park, McCann, and Hamman 1995) (3.5 percent). Data mostly came from single-mode surveys (in 65.2 percent of the studies), while 22.7 percent of the studies involved experimental comparisons between two or more modes, and a further 12.1 percent involved replications in other modes. In terms of the modes analyzed, more than half of the studies (51.7 percent) analyzed data collected through web surveys. The next-most-analyzed mode was face-to-face interview data (31.2 percent), followed by mail (21.7 percent) and telephone interview (20.6 percent) data. Other modes (including mobile/smart phones, CASI, ACASI, and others) have received less attention.

65.3

3.6

65.3

52.5

22.0

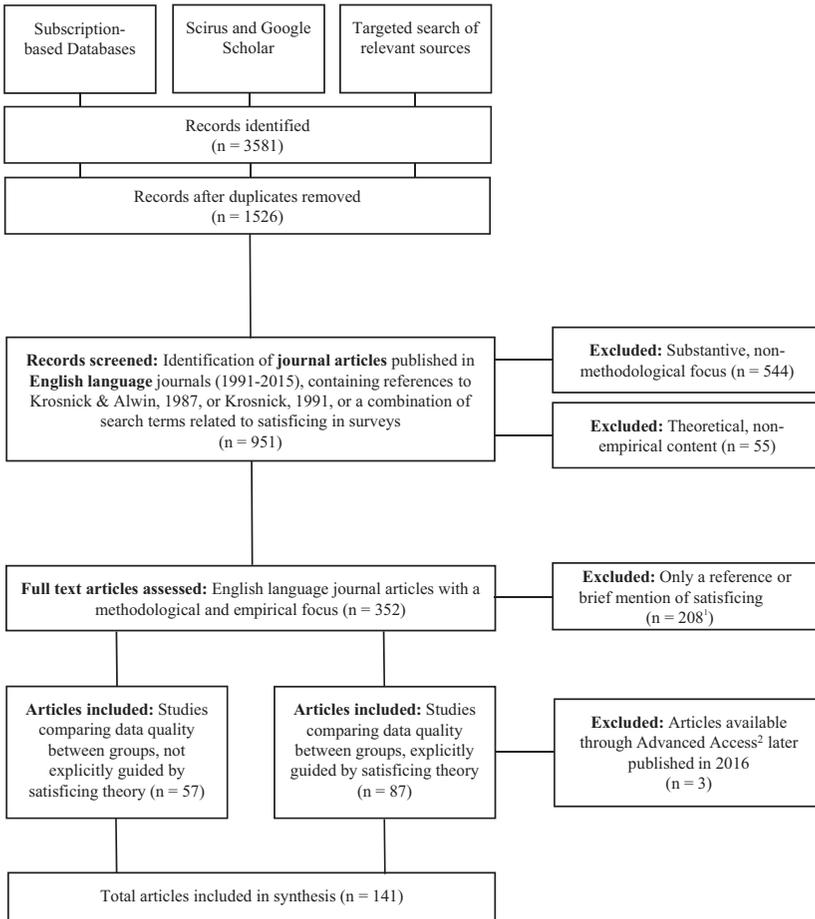


Figure 2. Flow diagram of study selection process. ¹Full text for one of the records was not freely available and neither were contact details for the authors, so it was also excluded at this stage. ²This decision was taken to ensure a clear definition of the inclusion criteria and improve reliability.

Indicators of satisficing: It was commonplace in the shortlisted studies to use multiple indicators of response quality. A total of 51 of the 141 shortlisted studies (36.2 percent) looked at weak forms of satisficing, while 64 studies (45.4 percent) looked at strong forms of satisficing (see table 3). Among the former, response-order effects (primacy and recency) received the most attention, while among the latter, nondifferentiation and selecting the “don’t know” alternative were considered most frequently. Random reporting and the tendency to endorse the status quo have been used as indi-

Table 2. Characteristics of reviewed shortlisted studies

Characteristic	N = 141	% Total	
No. of journals represented	59	–	
Reference to satisficing in title	13	9.2	
Reference to satisficing in abstract	45	31.9	
Research design ^a			
Survey experiment/ split-ballot	92	65.3	65.2
Non-experimental survey data	44		31.2
Other	5	3.6	3.5
Data collection modes ^a			
Single-mode studies	92	65.3	65.2
Mode comparison studies	32		22.7
Replication in a different mode	17		12.1
Mode(s) of data collection ^b Remove footnote b			
Web self-completion	74	52.5	51.7
Face-to-face interview	44		31.2
Paper self-completion	31	22.0	21.7
Telephone interview	29		20.6
Mobile phone interview	6	4.3	4.2
CASI	4		2.8
ACASI	1		0.7
Other	8	5.7	5.6

^aBase = all 141 articles. Values may deviate from 100 percent due to rounding.

AQ4

Change to: Sum of values may deviate from 100% due to rounding.

icators of satisficing in only a negligible number of the shortlisted studies. By contrast, a wide variety of alternative indicators of response quality have been used, and 44 (31.2 percent) of the shortlisted studies *only* used other indicators of response quality not mentioned in Krosnick’s original list (shown in [table 3](#)). These included, notably, item nonresponse, selecting middle alternatives in rating scales, interview pace (including overall interview duration and response latencies), selecting extreme responses, reliability or consistency of responses, and the length of answers to open-ended questions. Other less frequently used indicators of response quality were social desirability bias, response to trap questions, response accuracy, rounding and heaping, break-offs, and tests of validity (e.g., correlations with other variables).

All the shortlisted articles

AQ6 *Predictors of satisficing:* Eleven of the shortlisted articles presented new empirical findings relating to differences in response quality between subgroups of respondents. The observed, or experimentally manipulated, explanatory

12.30

12.35

12.40

12.44

Table 3. Indicators of response quality analyzed

	N = 141	% Total ^b		
Indicators of weak satisficing analyzed				
None	90		63.8	13.5
Primacy	32	22.7	22.4	
Acquiescence	24	17.0	16.8	
Recency	13	9.2	9.1	
Indicators of strong satisficing analyzed				
None	77		54.6	13.10
Don't know/No-opinion filters	38	27.0	26.6	
Non-differentiation	37		26.2	
Random reporting (mental coin-flipping)	2		1.4	
Endorsing the status quo	1		0.7	
Other indicators of response quality				
Item nonresponse	34	24.1	25.2	13.15
Middle alternatives in rating scales	25	17.7	17.5	
Interview pace (interview duration/response latencies)	20		14.2	
Extreme responses ^a	12		8.5	
Reliability/consistency	11		7.8	13.20
Length of responses to open-ended questions	8		5.7	
Social desirability bias	6		4.3	
Trap questions	6		4.3	
Accuracy	5		3.6	
Rounding and heaping	5		3.6	
Break-offs	4		2.8	13.25
Validity/correlations	4		2.8	
Other	20		14.2	

^aIncludes extreme plus middle responses. ^bStudies may include multiple indicators in each category, so the percentage shown is the percentage of all shortlisted studies using each type of indicator.

AQ5

13.30

variables for these differences were coded according to whether they related to task difficulty, respondent ability, or respondent motivation. Variables relating to task difficulty received the most attention in the shortlisted studies, appearing as predictors in over half of them (see table 4). Indicators of respondent characteristics appeared in fewer articles: variables relating to respondent ability were analyzed in 64 (44.8 percent) of the articles, while 45.4 percent variables capturing respondent motivation were analyzed in 57 (39.9 percent) 40.4 percent of the articles. Details of the number of articles using different indicators within these broader categories are presented in table 4 and summarized by article in table A3 in the Online Appendix. Note that some articles used multiple indicators within a broader category, so the totals within each do not match the total for the category as a whole. Less than half of the articles (64, or

13.35

13.40

13.44

Table 4. Predictors of satisficing—indicators of task difficulty, respondent ability, and respondent motivation

		N = 141	% ^a
14.5	Indicators of task difficulty	78	55.3
	Hypothesized differences between modes	23	16.3
	Features of the question format	21	14.8
	Response selection challenges	20	14.2
	Simple vs. complex judgments	10	7.1
14.10	Interpretability	8	5.7
	Recall task	6	4.3
	Other	10	7.1
	Indicators of respondent ability	64	45.4
	Cognitive sophistication	58	41.1
14.15	Highest level of education	36	25.5
	Number of years education	6	4.3
	Cognitive skills test	5	3.6
	Age	11	7.8
	Domain-relevant thinking/knowledge	4	2.8
	Pre-consolidated attitudes	5	3.6
14.20	Experience of taking surveys	2	1.4
	Other ^b	5	3.6
	Indicators of respondent motivation	57	40.4
	Personally important topic	15	10.6
14.25	Believing survey is worthwhile	10	7.1
	Interview duration	9	6.4
	Accountability	8	5.7
	Interviewer behavior	7	5.0
	Need for cognition/Need for evaluation	6	4.3
	Incentives	4	2.8
14.30	Self-reported effort	3	2.1
	Reluctant vs. cooperative respondents	2	1.4
	Other	6	4.3
	More than one type of predictor analyzed	64	45.4

14.35 ^aStudies may include multiple indicators in each category, so the percentage shown is the percentage of all shortlisted studies using each type of indicator. ^bIncludes other indicators of cognitive sophistication.

14.40 45.4 percent) addressed more than one category of explanatory variable simultaneously, enabling the analysis of their additive or multiplicative effects.

14.44 *Summary of results obtained from research based on satisficing theory:* Turning to the results of the studies, [table 5](#) shows the number reporting statistically significant main effects on response behavior associated with satisficing for task

Table 5. Summary of results relating to predictors satisficing

	<i>n</i>	%	% Total
Studies reporting significant main effects of total analyzing each predictor ^a			
Task difficulty	58/78	74.4	41.1
Respondent motivation	39/57	68.4	27.7
Respondent ability	39/64	60.9	27.7
No main effects observed	32/141	–	22.7
Studies reporting significant interaction effects involving predictors of satisficing			
Significant interaction effects observed	32/64	50.0	22.7

^aStudies may report more than one main or interaction effects.

difficulty, respondent ability, and respondent motivation. Of the total number of articles that looked at task difficulty, 74.4 percent reported significant main effects. Somewhat fewer articles found main effects of respondent ability and respondent motivation. Of the articles looking at variation in response quality by respondent ability, 60.9 percent reported significant effects, while of those looking across levels of motivation, 68.4 percent reported significant effects. Just over one-fifth (22.7 percent) of all the shortlisted articles found no significant main effects at all. Meanwhile, of the 64 studies that investigated more than one of the main predictors of satisficing and, hence, could have tested for their combined effects, only half (50.0 percent) reported significant interaction effects. [Table A4](#) in the Online Appendix individually summarizes the results for each of the shortlisted articles.

Another way to look at the results is to consider how many found evidence consistent with the theory of satisficing—that is, how many found evidence for an increased prevalence of satisficing under the conditions hypothesized to foster satisficing (i.e., increased task difficulty, and/or decreased respondent ability or motivation). In [table 6](#), for each of the main indicators of response quality considered, we present the number of statistically significant effects that are in the expected direction and the number that run in the opposite direction to that predicted by satisficing theory, alongside the total number of articles considering each indicator.⁴ From these results, it is evident that some indicators of satisficing have produced more mixed results with respect to the theory than others (in terms of the number of consistent and contradictory main effects reported). This is true for acquiescence and non-differentiation compared with response-order effects and “don’t know” responding, where

4. Note that in some articles, multiple main effects were reported for a given satisficing indicator, but only one is counted per article for the purposes of this table. Likewise, null findings were not coded on an indicator-by-indicator basis.

Table 6. Reported main effects consistent with or contradicting satisficing theory by satisficing indicator

	Total number of articles	Main effects consistent with theory ^b	Main effects contradicting theory
16.5	Weak satisficing indicators		
	Response-order effects	22	6
	Acquiescence	8	6
16.10	Strong satisficing indicators ^a		
	Don't know/No-opinion filters	24	8
	Non-differentiation	15	12
	Other indicators of response quality		
16.15	Item nonresponse	18	12
	Middle alternatives in rating scales	9	2
	Interview pace	10	8
	Extreme responses	7	3
	Reliability/consistency	5	2

^aEndorsing the status quo and random reporting (mental coin-flipping) are not considered here. ^bArticle reported at least one main effect for that indicator consistent with or contradicting satisficing theory.

the evidence for satisficing is more consistent. Among the other indicators of satisficing used, the consistency of the reported findings also varies as a function of the indicator used. These findings are illustrated in [figure 3](#), which shows the relative proportion of articles reporting consistent and contradictory findings for each indicator.

Discussion

In an effort to mitigate measurement error in surveys, methodologists have paid considerable attention to the cognitive processes involved in responding to questionnaires, the different factors that influence how these processes occur, and their effects on response quality. In this context, the theory of satisficing ([Krosnick and Alwin 1987](#); [Krosnick 1991](#)) has proved remarkably popular. This article has presented the results of a systematic review of empirical, methodological research that has explicitly invoked the concept of satisficing as an explanation for expected or observed differences in the prevalence of different types of response error between subgroups of respondents. Our coding frame was based on hypotheses outlined by [Krosnick \(1991\)](#), making it possible to assess the extent to which researchers drawing on the theory have pursued the agenda set forth in that article. A total of 141 studies were identified and retained in the full review.

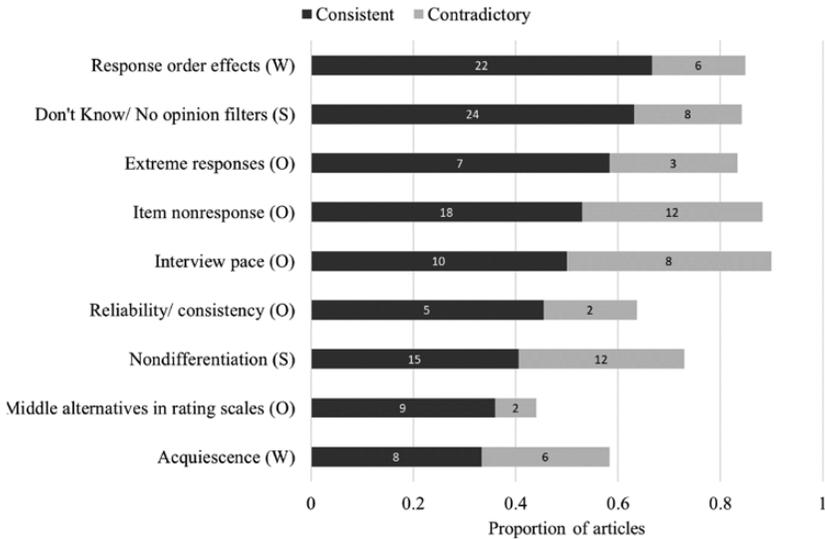


Figure 3. Proportion and number of articles reporting consistent and contradictory results by response quality indicator.

Looking at the number of publications per year confirmed our impression that satisficing theory has continued to gain in popularity. The last years examined saw an especially marked acceleration in the number of articles published. Part of this can be attributed to the increase in online survey data collection, and the concomitant need for researchers to examine mode differences in data quality. These studies tend to use indirect indicators of measurement error, of which many, but not all, match those mentioned in Krosnick’s original article. While main effects on satisficing were observed for all three factors (task difficulty, respondent ability, and respondent motivation) in the majority of studies that measured them, fewer studies investigated, and fewer still observed, significant interaction effects (though the possibility that nonsignificant interactions were found but not reported cannot be ruled out). Given Krosnick’s claim that the relations between task difficulty, respondent ability, and respondent motivation are likely to be “multiplicative” (1991, p. 225), this provides scope for future studies interested in testing the theory further. If satisficing is to be regarded as an adequate explanation for the variety of measurement effects for which it has been invoked as a cause, the relative paucity of studies that actually test for and find the interaction effects implied by the theory poses a potential challenge, and one that invites further research.

The review also suggests that certain response effects are more likely to be observed under conditions that foster satisficing than others. Specifically, there appears to be greater empirical support for the hypothesis that primacy, recency, and no-opinion reporting are more common in situations of high task

18.5 difficulty and low respondent ability and motivation than there is for acquiescence and nondifferentiation, and other frequently used alternative indicators of response quality. This is perhaps unsurprising given the variety of explanations that have been developed in the literature on response styles, which would suggest that multiple additional factors may simultaneously play a role besides those mentioned by Krosnick (see Roberts 2016 for a recent review). In the case of acquiescence, for example, there is likely considerable variation in individual and cultural propensity to acquiesce for reasons other than lack of motivation or ability, which could account for the inconsistency we find with this indicator. An additional explanation (raised earlier) for these inconsistencies comes from the variation in the methods used to construct indicators of satisficing (e.g., the types of constructs measured, item response format, number of items in a battery, methods to compile indices, and so on). The mixed pattern of findings may stem in part from this heterogeneity. More detailed analysis of the methods used and their implications for conclusions about response quality could reveal systematic elements to this variation.

18.10 A number of caveats to our conclusions, which relate to general challenges involved in systematic review, are worth mentioning. The first concerns the reliability of the coding procedure when using content analytic methods of the type used here. Even with a well-designed coding scheme, coders face numerous challenges when trying to decide how to apply it to specific units of analysis. In the case of the present study, the coding frame design was guided by the way in which Krosnick (1991) described the theory of satisficing. However, the shortlisting procedure used to select studies led us to err on the side of inclusivity when deciding whether articles met the eligibility criteria or not. As a result, the shortlist included a mix of articles that had worked squarely within the framework of satisficing theory, as well as articles investigating differences in response quality across groups, which were less closely guided by the theory. This led to difficulties in deciding how to code indicators and predictors of satisficing, where they were not explicitly labeled as such. For example, some studies compared variation in response quality across modes of data collection, or across different question formats, and task difficulty was either explicitly cited as the causal mechanism for expected or observed differences in response, or was only implied in the theoretical part of the article. Discrepancies in the codes assigned for the articles that were double-coded for our intercoder reliability analysis reflect the difficulties coders had deciding how to code these implicit explanatory variables. Nevertheless, the level of reliability between coders in our study was generally good (according to Landis and Koch's [1977] recommended interpretation of the kappa statistic), and we were able to effectively adjudicate on discrepancies, as well as learn from them in order to improve the design of the coding scheme used to code articles published later in the study period.

18.15
18.20
18.25
18.30
18.35
18.40
18.44 A second challenge relates to the selectivity of the sample of shortlisted articles, which resulted not only from the abovementioned subjectivity of the

coding procedure, but also from methodological decisions made relating to study selection. In relation to the former, the example just cited underlines the fact that part of the evidence base from which we draw conclusions relating to satisficing theory was never intended for this purpose, and the reader should be conscious of this when assessing the relative weight of contradictory or consistent evidence. Furthermore, the large number of citations retrieved in the literature searches led to the pragmatic decision to focus only on studies published in (English-language) academic journals between 1987 and 2015 (articles published under Advance Access in 2015 were excluded because it created ambiguity around the cutoff date and the eligibility of other articles published during the data extraction stage of the review). As well as limiting the scale of the data extraction task, the decision to only review journal articles was also taken partly out of concerns that the search results for other types of publication, and especially for gray literature, might be less reliable and less complete (there was also some duplication between unpublished material and material that was later published in journals).

19.5

19.10

19.15

While focusing on published (mostly, peer-reviewed) work guarantees a certain level of quality of the research reviewed, all of the well-rehearsed caveats about the risk of publication bias against null findings apply. Equally, our conclusions may not hold for articles drawing on satisficing theory published since 2015 (Beller et al. 2013).

19.20

AQ7

These limitations aside, we hope that the present review provides a useful starting point for future discussions about the relevance of the satisficing concept in survey research. It is especially noteworthy that although satisficing is a putative psychological mechanism thought to lie behind particular types of response behavior, very few of the articles we reviewed elaborate on the possible processes that instigate satisficing and the extent to which they are under the conscious control of the respondent (although see [Vannette and Krosnick 2014](#)). [Krosnick \(1991\)](#) states that respondents “perform” certain response behaviors because they satisfice, implying that it is the result of rational decision-making: “Rather than continuing to expend the mental effort necessary to generate optimal answers to question after question, respondents are likely to compromise their standards and expend less energy instead” (pp. 214–15). Yet the literature on judgment and decision-making would imply that the use of heuristics and shortcuts (and the biases they produce) emerges from the interplay between effortful and attentive mental activity (“system 2” thinking) and the relatively automatic, *involuntary*, low effort (“system 1”) thinking that both fuels and disrupts its counterpart ([Kahneman 2011](#)). A consideration of what is currently known about these respective modes of thinking may offer new insights into voluntary and involuntary triggers of satisficing in surveys.

19.25

19.30

19.35

19.40

Very few studies have attempted to measure satisficing directly by self-reports or other explicit means, preferring instead indirect indicators of response scale effects. In other words, satisficing is assumed to be the cause of

19.44

variation in such indicators. Exceptions include studies that include so-called “instructional manipulation checks” or “trap questions” to identify inattentive respondents (e.g., Oppenheimer, Meyvis, and Davidenko 2009; Berinsky, Margolis, and Sances 2014; Gao, House, and Xie 2015; Hauser and Schwarz 2015; Revilla and Ochoa 2015). These studies find such methods useful for identifying “bad” respondents, the removal of which from the analytic sample can improve the reliability of estimates. Instructional Manipulation Checks may also improve respondent attention to later questions, so this line of research offers promising guidance for how to measure satisficing and how to motivate respondents to optimize when responding to questionnaires.

AQ8

AQ9

AQ11

AQ10

In much of the literature reviewed, when negative or inconclusive results are found, the interpretation is that satisficing is not taking place in the context under investigation, not that the theory is incorrect or incomplete. Thus, one of the conclusions we draw from the review is that satisficing theory is widely *assumed* by survey researchers to be (1) useful; and (2) an appropriate description of the survey response process. There appears to be little appetite for formal attempts to test or falsify the theory. Nor is there much work that evaluates alternative theories that could potentially generate more accurate and consistent predictions about response quality. We hope that the present paper might stimulate thinking in this direction.

As well as providing insights into how the theory of satisficing has been used in survey methodological research, and into the empirical findings generated, we also hope that our study will serve as a useful resource for future researchers and practitioners. The systematic review approach not only helps impose some structure on the otherwise unwieldy literature relating to response quality in surveys, but the list of studies identified may also provide a useful sampling base for researchers seeking to undertake more focused analysis relating to particular aspects of satisficing theory. Such endeavors are of empirical interest, but more importantly, can play a role in validating and improving the theories that guide survey practice and methodology.

Appendices

List of Shortlisted Articles

Aichholzer, Julian. 2013. “Intra-Individual Variation of Extreme Response Style in Mixed-Mode Panel Studies.” *Social Science Research* 42:957–70.

Atkeson, Lonna Rae, Alex N. Adams, and R. Michael Alvarez. 2014. “Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys.” *Political Analysis* 22:304–20.

Auspurg, Katrin, and Annette Jäckle. 2015. “First Equals Most Important? Order Effects in Vignette-Based Measurement.” *Sociological Methods and Research* 46:490–539.

- Barber, Larissa K., Christopher M. Barnes, and Kevin D. Carlson. 2013. "Random and Systematic Error Effects of Insomnia on Survey Behavior." *Organizational Research Methods* 16:616–49.
- Barge, Scott, and Hunter Gehlbach. 2012. "Using the Theory of Satisficing to Evaluate the Quality of Survey Data." *Research in Higher Education* 53:182–200. 21.5
- Bassili, John N., and Jon A. Krosnick. 2000. "Do Strength-Related Attitude Properties Determine Susceptibility to Response Effects? New Evidence from Response Latency, Attitude Extremity, and Aggregate Indices." *Political Psychology* 21:107–32. 21.10
- Berinsky, Adam J., and Michele Margolis. 2011. "Missing Voices: Polling and Healthcare." *Journal of Health Politics, Policy and Law* 36:975–87.
- Bishop, George, and Andrew Smith. 2001. "Response-Order Effects and the Early Gallup Split-Ballots." *Public Opinion Quarterly* 65:479–505. 21.15
- Borgers, Natacha, and Joop Hox. 2001. "Item Nonresponse in Questionnaire Research with Children." *Journal of Official Statistics* 17:321–35. 21.15
- Borgers, Natacha, Dirk Sikkels, and Joop Hox. 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* 38:17–33. 21.20
- Brewer, Noel T., William K. Hallman, Nancy Fiedler, and Howard M. Kipen. 2004. "Why Do People Report Better Health by Phone Than by Mail?" *Medical Care* 42:875–83.
- Brockington, David. 2003. "A Low Information Theory of Ballot Position Effect." *Political Behavior* 25:1–27. 21.25
- Callegaro, Mario, Yongwei Yang, Dennison S. Bhola, Don A. Dillman, and Tzu-Yun Chin. 2009. "Response Latency as an Indicator of Optimizing in Online Questionnaires." *Bulletin of Sociological Methodology* 103:5–25.
- Castro, Rubén. 2013. "Inconsistent Respondents and Sensitive Questions. Field Methods." *Field Methods* 25:283–98. 21.30
- 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* 73:641–78.
- . "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment." *Public Opinion Quarterly* 74:154–67. 21.35
- Chen, Pu-Shih Daniel. 2011. "Finding Quality Responses: The Problem of Low-Quality Survey Responses and Its Impact on Accountability Measures." *Research in Higher Education* 52:659–74. 21.40
- Clifford, Scott, and Jennifer Jerit. 2015. "Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?" *Public Opinion Quarterly* 79:790–802. 21.44

Couper, Mick P. 1997. "Survey Introductions and Data Quality." *Public Opinion Quarterly* 61:317–38.

Couper, Mick P., Roger Tourangeau, Frederick G. Conrad, and Scott D. Crawford. 2004. "What They See Is What We Get—Response Options for Web Surveys." *Social Science Computer Review* 22:111–27.

Couper, Mick P., Roger Tourangeau, Frederick G. Conrad, and Chan Zhang. 2013. "The Design of Grids in Web Surveys." *Social Science Computer Review* 31:322–45.

Craig, Benjamin M., Shannon K. Runge, Kim Rand-Hendriksen, Juan Manuel Ramos-Goni, and Mark Oppe. 2015. "Learning and Satisficing: An Analysis of Sequence Effects in Health Valuation." *Value in Health* 18:217–23.

Darker, Catherine D., and David P. French. 2009. "What Sense Do People Make of a Theory of Planned Behaviour Questionnaire? A Think-Aloud Study." *Journal of Health Psychology* 14:861–71.

de Bruijne, Marika, and Arnaud Wijnant. 2014. "Improving Response Rates and Questionnaire Design for Mobile Web Surveys." *Public Opinion Quarterly* 78:951–62.

Díaz de Rada, Vidal. 2010. "Face-to-Face versus Telephone Surveys on Political Attitudes: A Comparative Analysis." *Quality and Quantity* 45:817–27.

Díaz de Rada, Vidal, and Juan Antonio Domínguez-Alvaraz. 2014. "Response Quality of Self-Administered Questionnaires." *Social Science Computer Review* 32:256–69.

———. 2015. "The Quality of Responses to Grid Questions as Used in Web Questionnaires (Compared with Paper Questionnaires)." *International Journal of Social Research Methodology* 18:337–48.

Dolnicar, Sara, and Bettina Grün. 2014. "Including Don't Know Answer Options in Brand Image Surveys Improves Data Quality." *International Journal of Market Research* 56:33–50.

Dumitrescu, Delia, and Johan Martinsson. 2015. "Surveys as a Social Experience: The Lingering Effects of Survey Design Choices on Respondents' Survey Experience and Subsequent Optimizing Behavior." *International Journal of Public Opinion Research* 28:534–61.

Eggs, Johannes, and Annette Jäckle. 2015. "Dependent Interviewing and Sub-Optimal Responding." *Survey Research Methods* 9:15–29.

Enns, Peter K., and Brian Richman. 2013. "Presidential Campaigns and the Fundamentals Reconsidered." *Journal of Politics* 75:803–20.

Fang, Jiaming, Chao Wen, and Victor Prybutok. 2013. "The Equivalence of Internet versus Paper-Based Surveys in IT/IS Adoption Research in Collectivistic Cultures: The Impact of Satisficing." *Behaviour & Information Technology* 32:480–90.

———. 2014. "An Assessment of Equivalence between Paper and Social Media Surveys: The Role of Social Desirability and Satisficing." *Computers in Human Behavior* 30:335–43.

Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan. 2005. “An Experimental Comparison of Web and Telephone Surveys.” *Public Opinion Quarterly* 69:370–92.

Galesic, Mirta, Roger Tourangeau, Mick P. Couper, and Frederick G. Conrad. 2008. “Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding.” *Public Opinion Quarterly* 72:892–913. 23.5

Gao, Zhifeng, Lisa A. House, and Jing Xie. 2015. “Online Survey Data Quality and Its Implication for Willingness-to-Pay: A Cross-Country Comparison.” *Canadian Journal of Agricultural Economics* 64:199–221. 23.10

Gehlbach, Hunter, and Scott Barge. 2012. “Anchoring and Adjusting in Questionnaire Responses.” *Basic and Applied Social Psychology* 34:417–33.

Gilbert, Emily E. 2015. “A Comparison of Branched Versus Unbranched Rating Scales for the Measurement of Attitudes in Surveys.” *Public Opinion Quarterly* 79:443–70. 23.15

Goldenbeld, Charles, and Saskia de Craen. 2013. “The Comparison of Road Safety Survey Answers Between Web-Panel and Face-to-Face; Dutch Results of SARTRE-4 Survey.” *Journal of Safety Research* 46:13–20.

Gooch, Andrew. 2015. “Measurements of Cognitive Skill by Survey Mode: Marginal Differences and Scaling Similarities.” *Research & Politics* 2:1–11. 23.20

Göritz, Anja S., and Susanne C. Luthé. 2013. “Effects of Lotteries on Response Behavior in Online Panels.” *Field Methods* 25:219–37.

Grauenhorst, Thomas, Michael Blohm, and Achim Koch. 2015. “Respondent Incentives in a National Face-to-Face Survey: Do They Affect Response Quality?” *Field Methods* 28:266–83. 23.25

Gray, Michelle, Margaret Blake, and Pamela Campanelli. 2014. “The Use of Cognitive Interviewing Methods to Evaluate Mode Effects in Survey Questions.” *Field Methods* 26:156–71.

Greszki, Robert, Marco Meyer, and Harald Schoen. 2015. “Exploring the Effects of Removing ‘Too Fast’ Responses and Respondents from Web Surveys.” *Public Opinion Quarterly* 79:471–503. 23.30

Guess, Andrew M. 2015. “Measure for Measure: An Experimental Test of Online Political Media Exposure.” *Political Analysis* 23:59–75.

Gummer, Tobias, and Joss Roßmann. 2015. “Explaining Interview Duration in Web Surveys: A Multilevel Approach.” *Social Science Computer Review* 33:217–34. 23.35

Guzy, Nathalie, and Heinz Leitgöb. 2015. “Assessing Mode Effects in Online and Telephone Victimization Surveys.” *International Review of Criminology* 21:101–31.

Hauser, David J., and Norbert Schwarz. 2015. “It’s a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on ‘Tricky’ Tasks.” *Sage Open* 5:1–6. 23.40

Heath, Oliver, Patten Smith, Nicholas Gilby, and Finn Hoolahan. 2015. “Partisan Priming and Subjective Evaluations of the Economy: Evidence from 23.44

a Survey Experiment.” *Journal of Elections, Public Opinion and Parties* 25:530–42.

Heerwegh, Dirk. 2009. “Mode Differences between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects.” *International Journal of Public Opinion Research* 21:111–21.

Heerwegh, Dirk, and Geert Loosveldt. 2002. “An Evaluation of the Effect of Response Formats on Data Quality in Web Surveys.” *Social Sciences Computer Review* 20:471–84.

———. 2002. “Web Surveys: The Effect of Controlling Survey Access using PIN Numbers.” *Social Science Computer Review* 20:10–21.

———. 2008. “Face-to-Face Versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality.” *Public Opinion Quarterly* 72:836–46.

Heerwegh, Dirk, Tim Vanhove, Koen Matthijs, and Geert Loosveldt. 2005. “The Effect of Personalization on Response Rates and Data Quality in Web Surveys.” *International Journal of Social Research Methodology* 8:85–99.

Hoehne, Karem Jan, and Timo Lenzner. 2015. “Investigating Response Order Effects in Web Surveys Using Eye Tracking.” *Psihologija* 48:361–77.

Holbrook, Allyson L., Sowmya Anand, Timothy P. Johnson, Young Ik Cho, Sharon Shavitt, Noel Chávez, and Saul Weiner. 2014. “Response Heaping in Interviewer-Administered Surveys: Is It Really a Form of Satisficing?” *Public Opinion Quarterly* 78:591–633.

Holbrook, Allyson L., Isabel C. Farrar, and Susan J. Popkin. 2006. “Surveying a Chicago Public Housing Development: Methodological Challenges and Lessons Learned.” *Evaluation Review* 30:779–802.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. “Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias.” *Public Opinion Quarterly* 67:79–125.

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

Hoogendoorn, Adriaan W. 2004. “A Questionnaire Design for Dependent Interviewing That Addresses the Problem of Cognitive Satisficing.” *Journal of Official Statistics* 20:219–32.

Hox, Joop J., Edith D. de Leeuw, and Hsuan-Tzu Chang. 2012. “Nonresponse Versus Measurement Error: Are Reluctant Respondents Worth Pursuing?” *Bulletin of Sociological Methodology* 113:5–19.

Hsieh, Yuli Patrick. 2015. “Check the Phone Book: Testing Information and Communication Technology (ICT) Recall Aids for Personal Network Surveys.” *Social Networks* 41:101–02.

Israel, Glenn D., and C. L. Taylor. 1990. “Can Response Order Bias Evaluations?” *Evaluation and Program Planning* 13:365–71.

Javeline, Debra. 1999. “Response Effects in Polite Cultures: A Test of Acquiescence in Kazakhstan.” *Public Opinion Quarterly* 63:1–28.

Johns, Robert. 2005. “One Size Doesn’t Fit All: Selecting Response Scales for Attitude Items.” *Journal of Elections, Public Opinions & Parties* 15:237–64.

Kaminska, Olena, Allan L. McCutcheon, and Jaak Billiet. 2010. “Satisficing Among Reluctant Respondents in a Cross-National Context.” *Public Opinion Quarterly* 74:956–84. 25.5

Kampen, Jarl K. 2007. “The Impact of Survey Methodology and Context on Central Tendency, Nonresponse and Associations of Subjective Indicator of Government Performance.” *Quality & Quantity* 41:793–813. 25.10

Kaplowitz, Michael D., Frank Lupi, Felix K. Yeboah, and Laurie G. Thorp. 2013. “Exploring the Middle Ground between Environmental Protection and Economic Growth.” *Public Understanding of Science* 22:413–26.

Kelly, Diane, David J. Harper, and Brian Landau. 2008. “Questionnaire Mode Effects in Interactive Information Retrieval Experiments.” *Information Processing & Management* 44:122–41. 25.15

Kieruj, Natalia D., and Guy Moors. 2010. “Variations in Response Style Behavior by Response Scale Format in Attitude Research.” *International Journal of Public Opinion Research* 22:320–42.

———. 2013. “Response Style Behavior: Question Format Dependent or Personal Style?” *Quality & Quantity* 47:193–211. 25.20

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

Knäuper, Barbel. 1999. “The Impact of Age and Education on Response Order Effects in Attitude Measurement.” *Public Opinion Quarterly* 63:347–70. 25.25

Knäuper, Barbel, Robert F. Belli, Daniel H. Hill, and A. Regula Herzog. 1997. “Question Difficulty and Respondents’ Cognitive Ability: The Effect on Data Quality.” *Journal of Official Statistics* 13:181–99.

Koch, Achim, and Michael Blohm. 2009. “Item Nonresponse in the European Social Survey.” *Research & Methods* 18:45–65. 25.30

Krebs, Dagmar, and Juergen H. Hoffmeyer-Zlotnik. 2010. “Positive First or Negative First? Effects of the Order of Answering Categories on Response Behavior.” *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6:118–27. 25.35

Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert Cameron Mitchel, et al. 2002. “The Impact of ‘No Opinion’ Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?” *Public Opinion Quarterly* 66:371–403. 25.40

Laurison, Daniel. 2015. “The Willingness to State an Opinion: Inequality, Don’t Know Responses, and Political Participation.” *Sociological Forum* 30:925–48. 25.44

Leeper, Thomas J. 2014. "Cognitive Style and the Survey Response." *Public Opinion Quarterly* 78:974–83.

26.5 Lelkes, Yphtach, Jon A. Krosnick, David M. Marx, Charles M. Judd, and Bernadette Park. 2012. "Complete Anonymity Compromises the Accuracy of Self-Reports." *Journal of Experimental Social Psychology* 48:1291–99.

Lelkes, Yphtach, and Rebecca Weiss. 2015. "Much Ado about Acquiescence: The Relative Validity and Reliability of Construct-Specific and Agree–Disagree Question." *Research & Politics* 2:1–8.

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

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

26.15 Lindhjem, Henrik, and Ståle Navrud. 2011. "Are Internet Surveys an Alternative to Face-to-Face Interviews in Contingent Valuation?" *Ecological Economics* 70:1628–37.

Lipps, Oliver. 2007. "Interviewer and Respondent Survey Quality Effects in a CATI Panel." *Bulletin of Sociological Methodology* 95:5–25.

26.20 Lugtig, Peter, and Gerty J. Lensveldt-Mulders. 2014. "Evaluating the Effect of Dependent Interviewing on the Quality of Measures of Change." *Field Methods* 26:172–90.

Mahon-Haft, Taj Alexander, and Don A. Dillman. 2010. "Does Visual Appeal Matter? Effects of Web Survey Aesthetics on Survey Quality." *Survey Research Methods* 4:43–59.

26.25 Malhotra, Neil. 2008. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion Quarterly* 72:914–34.

———. 2009. "Order Effects in Complex and Simple Tasks." *Public Opinion Quarterly* 73:180–98.

26.30 Matthijsse, Suzette M., Edith D. de Leeuw, and Joop J. Hox. 2015. "Internet Panels, Professional Respondents, and Data Quality." *Methodology* 11:81–88.

Mavletova, Aigul, and Mick P. Couper. 2013. "Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Difference?" *Survey Research Methods* 7:191–205.

26.35 Mayerl, Jochen. 2013. "Response Latency Measurement in Surveys. Detecting Strong Attitudes and Response Effects." *Survey Methods: Insights from the Field*. 1–27. doi:10.13094/SMIF-2013-00005

McCabe, Sean Esteban, Carol J. Boyd, Mick P. Couper, Scott Crawford, and Hannah d'Arcy. 2002. "Mode Effects for Collecting Alcohol and Other Drug Use Data: Web and US Mail." *Journal of Studies on Alcohol* 63:755–61.

26.40 McClamroch, Kristi J. 2011. "Evaluating the Usability of Personal Digital Assistants to Collect Behavioral Data on Adolescents with Paradata." *Field Methods* 23:219–42.

26.44

Mcdonald, Michael P., and Matthew P. Thornburg. 2012. "Interview Mode Effects: The Case of Exit Polls and Early Voting." *Public Opinion Quarterly* 76:326–49.

Medway, Rebecca L., and Roger Tourangeau. 2015. "Response Quality in Telephone Surveys: Do Prepaid Cash Incentives Make a Difference?" *Public Opinion Quarterly* 79:524–43. 27.5

Menold, Natalja, and Christoph J. Kemper. 2014. "How Do Real and Falsified Data Differ? Psychology of Survey Response as a Source of Falsification Indicators in Face-to-Face Surveys." *International Journal of Public Opinion Research* 26:41–65. 27.10

Menold, Natalja, Lars Kaczmirek, Timo Lenzner, and Aleš Neusar. 2014. "How Do Respondents Attend to Verbal Labels in Rating Scales?" *Field Methods* 26:21–39.

Murdoch, Maureen, John B. Pryor, Joan M. Griffin, Dian Cowper Ripley, Gary D. Gackstetter, Melissa A. Polusny, and James S. Hodges. 2011. "Unreliability and Error in the Military's 'Gold Standard' Measure of Sexual Harassment by Education and Gender." *Journal of Trauma & Dissociation* 12:216–31. 27.15

Muste, Christopher P. 2014. "The Implications of Survey Modes and Methods in Measuring Social-Group Closeness." *Public Opinion Quarterly* 78:859–88. 27.20

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

Nicolaas, Gerry, Pamela Campanelli, Steven Hope, Annette Jäckle, and Peter Lynn. 2015. "Revisiting 'Yes/No' versus 'Check All That Apply': Results from a Mixed Modes Experiment." *Survey Research Methods* 9: 189–204. 27.25

O'Halloran, Alissa, S. Sean Hu, Ann Malarcher, Robert McMillen, Nell Valentine, Mary A. Moore, Jennifer J. Reid, Natalie Darling, and Robert B. Gerzoff. 2014. "Response Order Effects in the Youth Tobacco Survey: Results of a Split-Ballot Experiment." *Survey Practice* 7:5. 27.30

Olson, Kristen, and Ipek Bilgen. 2011. "The Role of Interviewer Experience on Acquiescence." *Public Opinion Quarterly* 75:99–114.

Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45:867–72. 27.35

Pickery, Jan, and Geert Loosveldt. 1998. "The Impact of Respondent and Interviewer Characteristics on the Number of 'No Opinion' Answers." *Quality & Quantity* 32:31–45.

Pickett, Justin T., and Thomas Baker. 2014. "The Pragmatic American: Empirical Reality or Methodological Artefact?" *Criminology* 52:195–222. 27.40

Pickett, Justin T., Christi Falco Metcalfe, Thomas Baker, Marc Gertz, and Laura Bedard. 2014. "Superficial Survey Choice: An Experimental Test of 27.44

a Potential Method for Increasing Response Rates and Response Quality in Correctional Surveys.” *Journal of Quantitative Criminology* 30:265–84.

Prior, Markus. 2009. “Improving Media Effects Research through Better Measurement of News Exposure.” *Journal of Politics* 71:893–908.

28.5 Pustejovsky, James E., and James P. Spillane. 2009. “Question-Order Effects in Social Network Name Generators.” *Social Networks* 31:221–29.

Rasinski, Kenneth A., David Mingay, and Norman M. Bradburn. 1994. “Do Respondents Really ‘Mark All That Apply’ on Self-Administered Questions?” *Public Opinion Quarterly* 58:400–408.

28.10 Revilla, Melanie. 2012. “Impact of the Mode of Data Collection on the Quality of Answers to Survey Questions Depending on Respondent Characteristics.” *Bulletin of Sociological Methodology* 116:44–60.

Revilla, Melanie, and Carlos Ochoa. 2015. “What Are the Links in a Web Survey Among Response Time, Quality, and Auto-Evaluation of the Efforts Done?” *Social Science Computer Review* 33:97–114.

28.15 Robison, Joshua. 2014. “Who Knows? Question Format and Political Knowledge.” *International Journal of Public Opinion Research* 27:1–21.

Rogelberg, Steven G., Gwenith G. Fisher, Douglas C. Maynard, Milton D. Hakel, and Michael Horvath. 2001. “Attitudes Toward Surveys: Development of a Measure and Its Relationship to Respondent Behavior.” *Organizational Research Methods* 4:3–25.

28.20 Schonla, Matthias, and Vera Toepoel. 2015. “Straightlining in Web Survey Panels over Time.” *Survey Research Methods* 9:125–37.

Serenko, Alexander, and Nick Bontis. 2013. “First In, Best Dressed: The Presence of Order-Effect Bias in Journal Ranking Surveys.” *Journal of Infometrics* 7:138–44.

28.25 Shoemaker, Pamela J., Martin Eichholz, and Elizabeth A. Skewes. 2002. “Item Nonresponse: Distinguishing between Don’t Know and Refuse.” *International Journal of Public Opinion Research* 14:193–201.

28.30 Siminski, Peter. 2008. “Order Effects in Batteries of Questions.” *Quality & Quantity* 42:477–90.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Michael J. Stern. 2006. “Comparing Check-All and Forced-Choice Question Formats in Web Surveys.” *Public Opinion Quarterly* 70:66–77.

28.35 Smyth, Jolene, Kristen Olson, and Alian S. Kasabian. 2014. “The Effect of Answering in a Preferred Versus a Non-Preferred Survey Mode on Measurement.” *Survey Research Methods* 8:137–52.

Staszyńska, Katarzyna M. 2011. “Cognitive Determinants of Data Quality in Public Opinion Polls: Respondents Definition of the Survey.” *Polish Sociological Review* 176:493–514.

28.40 Steinbrecher, Markus, Joss Roßmann, and Jan Eric Blumenstiel. 2014. “Why Do Respondents Break Off Web Surveys and Does It Matter? Results from Four Follow-Up Surveys.” *International Journal of Public Opinion Research* 27:289–302.

28.44

Stern, Michael, Don A. Dillman, and Jolene D. Smyth. 2007. “Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey.” *Survey Research Methods* 1:121–38.

Stock, Volker. 2006. “Attitudes Toward Surveys, Attitude Accessibility and the Effect on Respondents’ Susceptibility to Nonresponse.” *Quality and Quantity* 40:259–88. 29.5

Stolte, John F. 1994. “The Context of Satisficing in Vignette Research.” *Journal of Social Psychology* 134:727–33.

Stratton, Terry D., Donald B. Witzke, Robert J. Jacob, Marlene J. Sauer, and Amy Murphy-Spencer. 2002. “Medical Students’ Ratings of Faculty Teaching in a Multi-Instructor Setting: An Examination of Monotonic Response Patterns.” *Advances in Health Sciences Education* 7:99–116. 29.10

Struminskaya, Bella, Kai Weyandt, and Michael Bosnjak. 2015. “The Effects of Questionnaire Completion Using Mobile Devices on Data Quality: Evidence from a Probability-Based General Population Panel.” *Methods, Data, Analyses* 9:261–92. 29.15

Sturgis, Patrick, Caroline Roberts, and Patten Smith. 2014. “Middle Alternatives Revisited: How the Neither/Nor Response Acts as a Way of Saying ‘I Don’t Know.’” *Sociological Methods and Research* 43:15–38.

Toepoel, Vera, Marcel Das, and Arthur Van Soest. 2008. “Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents.” *Public Opinion Quarterly* 72:985–1007. 29.20

Tourangeau, Roger, Robert M. Groves, Courtney Kennedy, and Ting Yan. 2009. “The Presentation of a Web Survey, Nonresponse and Measurement Error Among Members of Web Panel.” *Journal of Official Statistics* 25:299–321. 29.25

Turgeon, Mathieu. 2009. “‘Just Thinking’: Attitude Development, Public Opinion, and Political Representation.” *Political Behavior* 31:353–78.

Turner, Gosia, Patrick Sturgis, and David Martin. 2014. “Can Response Latencies Be Used to Detect Survey Satisficing on Cognitively Demanding Questions?” *Journal of Survey Statistics and Methodology* 3:89–108. 29.30

Vicente, Paula, Elizabeth Reis, and Maria Santos. 2009. “Using Mobile Phones for Survey Research.” *International Journal of Market Research* 51:613–33.

Vogl, Susanne. 2013. “Telephone Versus Face-to-Face Interviews: Mode Effect on Semistructured Interviews with Children.” *Sociological Methodology* 43:133–77. 29.35

Wagner, Markus, and Eva Zeglovits. 2014. “Survey Questions about Party Competence: Insights from Cognitive Interviews.” *Electoral Studies* 34:280–90.

Weijters, Bert, Hans Baumgartner, and Niels Schillewaert. 2013. “Reversed Item Bias: An Integrative Model.” *Psychological Methods* 18:320–34. 29.40

Wicker, Frank W., Yehsoon Park, Erin McCann, and Douglas Hamman. 1995. “Task Involvement and Satisficing with Motivational Scales.” *Psychological Reports* 77:211–23. 29.44

Yan, Ting, and Florian Keusch. 2015. "The Effects of the Direction of Rating Scales on Survey Responses in a Telephone Survey." *Public Opinion Quarterly* 79:145–65.

30.5 Yang, Yongwei, Mario Callegaro, Dennison S. Bholra, and Don A. Dillman. 2011. "IVR and Web Administration in Structured Interviews Utilizing Rating Scales: Exploring the Role of Motivation as a Moderator to Mode Effects." *International Journal of Social Research Methodology* 14:1–15.

30.10 Zhang, Chan, and Frederick Conrad. 2014. "Speeding in Web Surveys: The Tendency to Answer Very Fast and Its Association with Straightlining." *Survey Research Methods* 8:127–35.

Supplementary Data

30.15 Supplementary data are freely available at *Public Opinion Quarterly* online.

References

30.20 Alwin, Duane F. 1991. "Research on Survey Quality." *Sociological Methods and Research* 20:3–29.

———. 2007. *Margins of Error: A Study of Reliability in Survey Measurement*. Hoboken, NJ: Wiley-Interscience.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58:739–53.

30.25 Biemer, Paul P. 2010. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74:817–48.

Biemer, Paul P., and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons.

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber. 1978. "Effects of Question Wording and Format on Political Attitude Consistency." *Public Opinion Quarterly* 42:81–92.

30.30 Bishop, George, and Andrew Smith. 2001. "Response-Order Effects and the Early Gallup Split-Ballots." *Public Opinion Quarterly* 65:479–505.

Cantril, Hadley. 1940. "Experiments on the Wording of Questions." *Public Opinion Quarterly* 4:330–32.

30.35 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* 73:641–78.

Couper, Mick P. 2011. "The Future of Modes of Data Collection." *Public Opinion Quarterly* 5:889–908.

Dewey, M. E. 1983. "Coefficients of Agreement." *British Journal of Psychiatry* 143:487–89.

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

30.40 Foddy, William. 1993. *Constructing Questions for Interviews and Questionnaires*. Cambridge: Cambridge University Press.

Fowler, Floyd J. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.

AQ12

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology (Wiley Series in Survey Methods)*, 2nd ed. Hoboken, NJ: John Wiley & Sons.

Higgins, Julian P.T., and Sally Green. 2011. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available at www.cochrane-handbook.org.  31.5

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67:79–125.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kalton, Graham, Martin Collins, and Lindsay Brook. 1978. "Experiments in Wording Opinion Questions." *Journal of the Royal Statistical Society Series* 27:149–61. 31.10

Kolbe, Richard H., and Melissa S. Burnett. 1991. "Content-Analysis Research: An Examination of Applications with Directives for Improving Research Reliability and Objectivity." *Journal of Consumer Research* 18:243–50.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–36. 31.15

Krosnick, Jon A., and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:201–19.

———. 1988. "A Test of the Form-Resistant Correlation Hypothesis: Ratings, Rankings and the Measurement of Values." *Public Opinion Quarterly* 52:526–38.

———. 1989. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." Unpublished manuscript. Ann Arbor: Institute for Social Research, University of Michigan. 31.20

Landis, J. Richard, and Garry G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33:159–74.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2010. "Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects." Available at <http://matthewlombard.com/reliability/>, accessed June 2019. 31.25

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage.

Payne, Stanley Le Baron. 1951. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.

Petticrew, Mark, and Helen Roberts. 2005. *Systematic Reviews in the Social Sciences: A Practical Guide*. Malden, MA: Blackwell Publishing.

Roberts, Caroline. 2016. "Response Styles in Surveys: Understanding their Causes and Mitigating Their Impact on Data Quality." In *The Sage Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-Chih Fu, 579–96. London: Sage Publications. 31.30

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.

Schwarz, Norbert. 2007. "Cognitive Aspects of Survey Methodology." *Applied Cognitive Psychology* 21:277–87. 31.35

Simon, Herbert A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63:129–38.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers. 31.40

Tinsley, Howard E. A., and David J. Weiss. 2000. "Interrater Reliability and Agreement." In *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, edited by Howard E. A. Tinsley and Steven D. Brown, 95–124. San Diego, CA: Academic Press. 31.44

Torgerson, Carole. 2003. *Systematic Reviews and Meta-Analysis*. London: Continuum International Publishing.

Tourangeau, Roger. 1984. "Cognitive Science and Survey Methods." In *Cognitive Aspects of Survey Design: Building a Bridge Between Disciplines*, edited by Thomas B. Jabine, Miron L. Straf, Judith M. Tanur, and Roger Tourangeau, 73–100. Washington, DC: National Academies Press.

———. 2003. "Cognitive Aspects of Survey Measurement and Mismeasurement." *International Journal of Public Opinion Research* 15:3–7.

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

Vannette, David L., and Jon A. Krosnick. 2014. "A Comparison of Mindless and Survey Satisficing." In *The Wiley-Blackwell Handbook of Mindfulness*, edited by Amanda Ie, Christelle T. Ngnoumen, and Ellen J. Langer, 312–27. West Sussex, UK: John Wiley & Sons.

Willis, Gordon B. 2004. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications.

32.5

32.10

32.15

32.20

32.25

32.30

32.35

32.40

32.44