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).

**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.

**POPU-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.”

| AQ1 | Please note that the cross-reference “Cannell, Miller, and Oksenburg 1981” has not been provided in the reference list. Please provide the same. |
| AQ2 | in paragraph beginning “Building on this foundational work” please clarify the source referred to as “ibid.” |
| AQ4 | Text for footnote b is missing in table 2. Please check and amend. |
| AQ5 | Two footnotes are labeled as “a” in table 3. Please check and amend. There is no text for footnote b. |
| AQ6 | under subheading “Predictors of satisficing,” please clarify that you mean “eleven of the shortlisted article” |
| AQ7 | Please note that the cross-reference Miller et al. 2013” has not been provided in the reference list. Please provide the same. |
| 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
LÉÏLA EISNER*

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.
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.

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 four 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.

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.

**THE THEORY OF SATISIFICING**

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

---

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).
“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.

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.

Methods

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).

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). 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).

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 Index.

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.
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.

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.
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.

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).

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).

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.

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
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
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

<table>
<thead>
<tr>
<th>Codes</th>
<th>Percent agreement (%)</th>
<th>Cohen’s kappa, $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1 codes: Study selection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td>Methodological, substantive, or both</td>
<td>93.7</td>
</tr>
<tr>
<td></td>
<td>Empirical, theoretical, or both</td>
<td>96.4</td>
</tr>
<tr>
<td></td>
<td>Nature of reference (relevant empirical evidence vs. brief mention of search terms)</td>
<td>97.0</td>
</tr>
<tr>
<td>Stage 2 codes: Data abstraction and synthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study research design</td>
<td>Survey experiment/Non-experimental</td>
<td>86.7</td>
</tr>
<tr>
<td></td>
<td>Single/Mixed mode</td>
<td>96.7</td>
</tr>
<tr>
<td>Indicators of weak satisficing analyzed</td>
<td>Primacy/Recency/Acquiescence</td>
<td>86.7</td>
</tr>
<tr>
<td>Indicators of strong satisficing analyzed</td>
<td>Non-differentiation/Don’t knows/Random reporting</td>
<td>80.0</td>
</tr>
<tr>
<td>Independent variables analyzed</td>
<td>Task difficulty</td>
<td>73.3</td>
</tr>
<tr>
<td></td>
<td>Respondent ability</td>
<td>83.3</td>
</tr>
<tr>
<td></td>
<td>Respondent motivation</td>
<td>66.7</td>
</tr>
<tr>
<td>Significant main effects</td>
<td>Task difficulty</td>
<td>90.0</td>
</tr>
<tr>
<td></td>
<td>Respondent ability</td>
<td>93.3</td>
</tr>
<tr>
<td></td>
<td>Respondent motivation</td>
<td>83.3</td>
</tr>
<tr>
<td>Interactions</td>
<td>Significant interaction effects reported</td>
<td>76.7</td>
</tr>
</tbody>
</table>

*a*Endorsing the status quo not included, as only a few studies used this indicator.
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.
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-

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.
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).

Table 2. Characteristics of reviewed shortlisted studies

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N = 141</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of journals represented</td>
<td>59</td>
<td>–</td>
</tr>
<tr>
<td>Reference to satisficing in title</td>
<td>13</td>
<td>9.2</td>
</tr>
<tr>
<td>Reference to satisficing in abstract</td>
<td>45</td>
<td>31.9</td>
</tr>
</tbody>
</table>

Research design

<table>
<thead>
<tr>
<th>Research design</th>
<th>N = 141</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey experiment/ split-ballot</td>
<td>92</td>
<td>65.3</td>
</tr>
<tr>
<td>Non-experimental survey data</td>
<td>44</td>
<td>31.2</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Data collection modes

<table>
<thead>
<tr>
<th>Data collection modes</th>
<th>N = 141</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-mode studies</td>
<td>92</td>
<td>65.3</td>
</tr>
<tr>
<td>Mode comparison studies</td>
<td>32</td>
<td>22.7</td>
</tr>
<tr>
<td>Replication in a different mode</td>
<td>17</td>
<td>12.1</td>
</tr>
</tbody>
</table>

Mode(s) of data collection

<table>
<thead>
<tr>
<th>Mode(s) of data collection</th>
<th>N = 141</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web self-completion</td>
<td>74</td>
<td>52.5</td>
</tr>
<tr>
<td>Face-to-face interview</td>
<td>44</td>
<td>31.2</td>
</tr>
<tr>
<td>Paper self-completion</td>
<td>31</td>
<td>22.0</td>
</tr>
<tr>
<td>Telephone interview</td>
<td>29</td>
<td>20.6</td>
</tr>
<tr>
<td>Mobile phone interview</td>
<td>6</td>
<td>4.3</td>
</tr>
<tr>
<td>CASI</td>
<td>4</td>
<td>2.8</td>
</tr>
<tr>
<td>ACASI</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>5.7</td>
</tr>
</tbody>
</table>

AQ4 Base = 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.

AQ5 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...
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 variables capturing respondent motivation were analyzed in 57 (39.9 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
Table 4. Predictors of satisficing—indicators of task difficulty, respondent ability, and respondent motivation

<table>
<thead>
<tr>
<th>Indicators of task difficulty</th>
<th>N = 141</th>
<th>%*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesized differences between modes</td>
<td>23</td>
<td>16.3</td>
</tr>
<tr>
<td>Features of the question format</td>
<td>21</td>
<td>14.8</td>
</tr>
<tr>
<td>Response selection challenges</td>
<td>20</td>
<td>14.2</td>
</tr>
<tr>
<td>Simple vs. complex judgments</td>
<td>10</td>
<td>7.1</td>
</tr>
<tr>
<td>Interpretability</td>
<td>8</td>
<td>5.7</td>
</tr>
<tr>
<td>Recall task</td>
<td>6</td>
<td>4.3</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td>7.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicators of respondent ability</th>
<th>N = 141</th>
<th>%*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive sophistication</td>
<td>58</td>
<td>41.1</td>
</tr>
<tr>
<td>Highest level of education</td>
<td>36</td>
<td>25.5</td>
</tr>
<tr>
<td>Number of years education</td>
<td>6</td>
<td>4.3</td>
</tr>
<tr>
<td>Cognitive skills test</td>
<td>5</td>
<td>3.6</td>
</tr>
<tr>
<td>Age</td>
<td>11</td>
<td>7.8</td>
</tr>
<tr>
<td>Domain-relevant thinking/knowledge</td>
<td>4</td>
<td>2.8</td>
</tr>
<tr>
<td>Pre-consolidated attitudes</td>
<td>5</td>
<td>3.6</td>
</tr>
<tr>
<td>Experience of taking surveys</td>
<td>2</td>
<td>1.4</td>
</tr>
<tr>
<td>Otherb</td>
<td>5</td>
<td>3.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicators of respondent motivation</th>
<th>N = 141</th>
<th>%*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personally important topic</td>
<td>15</td>
<td>10.6</td>
</tr>
<tr>
<td>Believing survey is worthwhile</td>
<td>10</td>
<td>7.1</td>
</tr>
<tr>
<td>Interview duration</td>
<td>9</td>
<td>6.4</td>
</tr>
<tr>
<td>Accountability</td>
<td>8</td>
<td>5.7</td>
</tr>
<tr>
<td>Interviewer behavior</td>
<td>7</td>
<td>5.0</td>
</tr>
<tr>
<td>Need for cognition/Need for evaluation</td>
<td>6</td>
<td>4.3</td>
</tr>
<tr>
<td>Incentives</td>
<td>4</td>
<td>2.8</td>
</tr>
<tr>
<td>Self-reported effort</td>
<td>3</td>
<td>2.1</td>
</tr>
<tr>
<td>Reluctant vs. cooperative respondents</td>
<td>2</td>
<td>1.4</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>4.3</td>
</tr>
</tbody>
</table>

More than one type of predictor analyzed | 64 | 45.4 |

*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.

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

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

<table>
<thead>
<tr>
<th>Studies reporting significant main effects of total analyzing each predictora</th>
<th>n</th>
<th>%</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task difficulty</td>
<td>58</td>
<td>74.4</td>
<td>41.1</td>
</tr>
<tr>
<td>Respondent motivation</td>
<td>39</td>
<td>68.4</td>
<td>27.7</td>
</tr>
<tr>
<td>Respondent ability</td>
<td>39</td>
<td>60.9</td>
<td>27.7</td>
</tr>
<tr>
<td>No main effects observed</td>
<td>32</td>
<td>–</td>
<td>22.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Studies reporting significant interaction effects involving predictors of satisficing</th>
<th>n</th>
<th>%</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant interaction effects observed</td>
<td>32</td>
<td>50.0</td>
<td>22.7</td>
</tr>
</tbody>
</table>

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.
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.
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

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

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.

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).

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).

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.

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
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 Ochua 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.

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


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


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.


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


Research Synthesis—Satisficing in Surveys


**Supplementary Data**

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

**References**


