

Re-Examining the Middle Means Typical and the Left and Top Means First Heuristics Using Eye-Tracking Methodology

Höhne, Jan Karem; Lenzner, Timo; Neuert, Cornelia; Yan, Ting

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Höhne, J. K., Lenzner, T., Neuert, C., & Yan, T. (2021). Re-Examining the Middle Means Typical and the Left and Top Means First Heuristics Using Eye-Tracking Methodology. *Journal of Survey Statistics and Methodology*, 9(1), 25-50.

<https://doi.org/10.1093/jssam/smz028>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft

RE-EXAMINING THE MIDDLE MEANS TYPICAL AND THE LEFT AND TOP MEANS FIRST HEURISTICS USING EYE-TRACKING METHODOLOGY

JAN KAREM HÖHNE*

TIMO LENZNER

CORNELIA E. NEUERT

TING YAN

Web surveys are a common self-administered mode of data collection using written language to convey information. This language is usually accompanied by visual design elements, such as numbers, symbols, and graphics. As shown by previous research, such elements of survey questions can affect response behavior because respondents sometimes use interpretive heuristics, such as the “*middle means typical*” and the “*left and top means first*” heuristics when answering survey questions. In this study, we adopted the designs and survey questions of two experiments reported in Tourangeau, Couper, and Conrad (2004). One experiment varied the position of nonsubstantive response options in relation to other substantive response options and the second experiment varied the order of the response options. We implemented both experiments in an eye-tracking study. By recording respondents’ eye movements, we are able to observe how they read question stems and response options and we are able to draw conclusions about the survey response process the questions initiate. This enables us to investigate the mechanisms underlying the two interpretive heuristics and to test the assumptions of Tourangeau et al. (2004) about the ways in which interpretive heuristics influence survey responding. The eye-tracking data reveal mixed results

JAN KAREM HÖHNE is with the Collaborative Research Center 884 “Political Economy of Reforms,” University of Mannheim, B6 30-32, 68131 Mannheim, Germany and the Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra, Ramón Triás Fargas 25-27, 08005 Barcelona, Spain. TIMO LENZNER and CORNELIA E. NEUERT are with the Department of Survey Design and Methodology, GESIS - Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072 Mannheim, Germany. TING YAN is with Westat, 1600 Research Boulevard, Rockville, MD 20850.

*Address correspondence to Jan Karem Höhne, Collaborative Research Center 884 “Political Economy of Reforms,” University of Mannheim, B6 30-32, 68131 Mannheim, Germany; E-mail: hoehne@uni-mannheim.de.

for the two interpretive heuristics. For the *middle means typical* heuristic, it remains somewhat unclear whether respondents seize on the conceptual or visual midpoint of a response scale when answering survey questions. For the *left and top means first* heuristic, we found that violations of the heuristic increase response effort in terms of eye fixations. These results are discussed in the context of the findings of the original studies.

KEYWORDS: Eye tracking; Interpretive heuristics; Lab experiment; Question processing; Response behavior; Web survey research.

1. INTRODUCTION

Survey respondents are “cooperative communicators” who use any information provided by the survey instrument to understand and answer a question when completing a self-administered survey (Schwarz 1996). According to the principles of Gestalt psychology, information is communicated through four different visual elements (Dillman, Smyth, and Christian 2014): words (i.e., conveying meaning that helps to understand what is being asked), numbers (i.e., conveying additional meaning that helps to understand sequence or order), symbols (i.e., figures conveying additional meaning based on what they represent), and graphics (i.e., shapes and visual images conveying additional meaning based on what they represent).¹ Following this notion, respondents draw not only on words but also other visual elements in the form of numbers, symbols, and graphics (Christian and Dillman 2004; Couper, Tourangeau, and Kenyon 2004; Dillman et al. 2014; Schwarz, Grayson, and Knäuper 1998; Schwarz, Knäuper, Hippler, Noelle-Neumann, and Clark 1991; Smith 1995; Toepoel and Dillman 2011a, 2011b; Tourangeau et al. 2004, 2007; Tourangeau, Rips, and Rasinski 2000).

Web surveys make it easy to employ different visual elements, such as symbols and graphics (Couper et al. 2004; Dillman et al. 2014; Toepoel and Dillman 2011a). These visual elements help respondents to correctly fill out a questionnaire and make the survey experience more enjoyable. However, they can also influence respondents’ response behavior in unintended ways and thus influence their answers to the survey questions. One explanation for the impact of visual elements on response behavior is that respondents sometimes make use of so-called interpretive heuristics when answering survey questions. Expanding beyond the principles of Gestalt psychology, Tourangeau et al. (2004) proposed five such heuristics that assign meaning to spatial and/or visual elements: (i) “*middle means typical*” (i.e., respondents see the middle

¹ The Gestalt psychology covers several principles, such as the “principle of proximity” stating that objects that are closer together will be seen as a group. We refer interested readers to Dillman et al. (2014) and Toepoel and Dillman (2011a) for a comprehensive discussion of the principles of Gestalt psychology and their implications for questionnaire design.

response option as the most typical one), (ii) “*left and top means first*” (i.e., respondents see the leftmost or top response option as the first one in a conceptual sense), (iii) “*near means related*” (i.e., respondents see response options or questions that are physically close to each other to be related), (iv) “*up means good*” (i.e., respondents see the top response option as the most desirable one),² and (v) “*like means close*” (i.e., respondents see visually similar response options as conceptually closer).

One limitation of the Tourangeau et al. (2004) study is that only so-called “indirect data” (see Galesic, Tourangeau, Couper, and Conrad 2008; Galesic and Yan 2011), such as survey answers and response times, were examined. There is no direct evidence on how respondents process the survey questions. To overcome this limitation and to extend the research on interpretive heuristics, in this study we adopted the same designs and survey questions of two experiments reported by Tourangeau et al. (2004). The first experiment varied the position of nonsubstantive response options in relation to other substantive response options and the second experiment varied the order of the response options. We address the following main research question: how do violations of the *middle means typical* and the *left and top means first* heuristics influence respondents’ processing of survey questions?

In this article, we discuss and outline the theoretical considerations with respect to these two interpretive heuristics. For a detailed discussion of the other three heuristics, we refer interested readers to Toepoel and Dillman (2011a, 2011b) and Tourangeau et al. (2004, 2007).

2. BACKGROUND

Respondents are found to use response scales to infer the distribution of an attitude, opinion, or behavior in the general population (Schwarz and Hippler 1987). For instance, they conclude that the middle option of a response scale expresses the most typical or average value in the population and thus it serves as a standard of comparison (Schwarz, Hippler, Deutsch, and Strack 1985). According to the *middle means typical* heuristic, the middle option of a response scale serves as an anchor or reference point (see Tourangeau et al. 2000) because it is seen as the typical one. Tourangeau et al. (2004) investigated the use of this heuristic in web surveys by manipulating the presentation of the conceptual midpoint of a response scale; they specifically examined response scales in which the visual midpoint did not coincide with the conceptual midpoint (i.e., the visual midpoint of the response scale fell to one side of the conceptual midpoint). For this purpose, the authors manipulated how nonsubstantive options such as “don’t know” and “no opinion” were included in

² This heuristic can be seen as a special form of the “*left and top means first*” heuristic (Tourangeau et al. 2004).

response scales. The nonsubstantive options were either presented as additional radio buttons (resulting in a visual midpoint that was different from the conceptual midpoint) or were separated from the substantive options by a space or a divider line (resulting in a visual midpoint that coincided with the conceptual one). As predicted by the heuristic, for most questions respondents' answers shifted toward the visual midpoint in the condition in which nonsubstantive options were presented as additional radio buttons causing the visual midpoint falling to one side of the conceptual midpoint. The authors interpreted these results as supporting evidence of respondents' application of the *middle means typical* heuristic and concluded "that the meaning of each response option is partly based on its relative position within the array of response options" (Tourangeau et al. 2004, p. 376). In addition, more respondents selected the nonsubstantive responses in the conditions in which nonsubstantive responses were separated from other scale points. This indicates that the visual separation by a space or divider line drew respondents' attention to the nonsubstantive options shown as independent from the substantive options (see Toepoel and Dillman 2011a).

The *left and top means first* heuristic suggests that the first option of a response scale—either the leftmost one in horizontally arranged scales or the top one in vertically arranged scales—is conceptually seen as the first one. In general, this interpretive heuristic corresponds to the reading direction in most Western languages, such as English, French, Spanish, and German (see Rayner 1998; Rayner and Pollastek 2006; Tourangeau et al. 2004). It assumes that respondents expect the first response option in a series of ordered options to be the starting point (e.g., "very good"). In addition, they expect the successive options to follow in a logical progression (e.g., "good," "neither good nor bad," etc.) and that the lowermost option represents the opposite endpoint (e.g., "very bad"). Tourangeau et al. (2004) examined this heuristic by varying the order of the response options. In the first condition, the response options were presented in a way consistent with the *left and top means first* heuristic (i.e., the top option was one of the endpoints and the successive options followed in a logical order). In the second condition, the presentation of the response options was mildly inconsistent with the heuristic (i.e., the conceptual midpoint appeared at the bottom of the scale). In the third condition, the order of the response options was strongly inconsistent with the heuristic (i.e., "it depends," "strongly agree," "strongly disagree," "agree," and "disagree"). In line with the heuristic, response times increased with the degree of inconsistency, suggesting that the order discrepancies slowed respondents down (see also Holbrook, Krosnick, Carson, and Mitchell 2000). Furthermore, when the response option "it depends" was presented in the middle of the scale, more respondents selected this option than when it was presented at the top or at the bottom of the scale (Tourangeau et al. 2004). Again, the authors interpreted these results as evidence for the application of the *left and top means first* heuristic.

3. EYE-TRACKING METHODOLOGY

In the present study, we rerun the experiments on *separating scale points from nonsubstantive responses* and the *order of the response options* by Tourangeau et al. (2004) and re-examine the authors' assumptions about the ways in which these two interpretive heuristics influence survey responding. We extend the research of Tourangeau et al. (2004) by using eye-tracking methodology, which allows us to explore the underlying mechanisms affecting question-processing and respondent behavior (see Galesic and Yan 2011). In eye-tracking studies, participants' eye movements are captured by infrared cameras while reading questionnaire instructions, question stems, and response options. These cameras record respondents' exact eye location and the number, duration, and order of their fixations.

The connection between eye movements and cognitive processing is based upon two assumptions (Just and Carpenter 1980, p. 330). The "immediacy assumption" states that objects that are fixated by the eyes are processed immediately so that their interpretation is not deferred. The "eye-mind assumption" postulates that the eyes remain fixated on an object as long as it is being processed. Taken together, these two assumptions suggest that the time spent fixating on an object is approximately equal to the time it is being processed. As a result, eye movements provide direct information about what respondents process and how intensely they process it (Neuert and Lenzner 2017).

More fixations and longer fixations signal a longer response process (see Galesic et al. 2008; Galesic and Yan 2011; Höhne 2019; Höhne and Lenzner 2015, 2018; Kamoen et al. 2011, 2017; Lenzner, Kaczmirek, and Galesic 2011; Neuert and Lenzner 2017). A long response process could result from careful deliberation and thorough recall. It could also reflect respondents having difficulty with comprehension due to the use of a difficult or ambiguous word in a survey question, trouble recalling a specific event, and struggle with arriving at an answer or choosing between response options. For instance, Lenzner et al. (2011) showed that ambiguous noun phrases in survey questions produced higher fixation count and time than unambiguous noun phrases, suggesting that ambiguity increases response effort. In addition, the same ambiguous noun phrases were found to increase the selection of nonsubstantive responses and decrease response consistency, resulting in poor data quality (see Lenzner 2012). This finding suggests that eye-tracking parameters are good indicators of response effort and data quality.

In previous research, fixation count and time were used to determine whether respondents actually read all parts of a question and whether some parts receive more attention than others (Graesser et al. 2006; Höhne and Lenzner 2015, 2018; Lenzner, Kaczmirek, and Galesic 2014; Neuert 2017). Both Galesic et al. (2008) and Höhne and Lenzner (2015) investigated the occurrence and causes of primacy effects and found that respondents fixated more frequently and longer on the beginning of the scales. Höhne (2019) and

Höhne and Lenzner (2018) examined the response effort involved in answering agree/disagree and item-specific questions and observed that respondents fixated more frequently and longer on the item-specific response options than on the agree/disagree response options. Kamoen et al. (2011, 2017) examined the response effort of answering contrastive—positively or negatively formulated—survey questions and found that negatively worded questions yielded more and longer fixations than positively worded questions. Finally, Neuert (2017) analyzed the processing of forced-choice and check-all-that-apply question formats and reported higher fixation counts and times for forced-choice than check-all-that-apply questions, indicating a more deliberate response process initiated by the former format.

We use eye tracking to directly investigate the implications of the *middle means typical* and the *left and top means first* heuristics proposed by Tourangeau et al. (2004). In the next section, we outline the research designs and hypotheses for the two experiments, respectively. Subsequently, we describe the sample, the eye-tracking equipment, the procedure of the study, and the analytical strategies. We then present the results of each experiment. Finally, we discuss the theoretical and practical implications of the findings and suggest perspectives for future research.

4. EXPERIMENTAL DESIGNS AND HYPOTHESES

4.1 Experiment 1: Separating Scale Points from Nonsubstantive Responses

We used the same two questions as Tourangeau et al. (2004) that dealt with the performance of the government (see Appendices A and B). Respondents were randomly assigned to one of three experimental conditions. For the first group ($n = 44$), nonsubstantive response options were presented as additional radio buttons (nonseparation condition; condition 1), causing a mismatch between the conceptual and visual midpoint. The second group ($n = 46$) received the questions separating the nonsubstantive response options by a space (space condition; condition 2). The third group ($n = 41$) received the questions separating the nonsubstantive response options by a divider line (line condition; condition 3). In conditions 2 and 3, the visual midpoint of the response scale coincides with the conceptual midpoint, whereas the visual midpoint in condition 1 falls to the lower side of the conceptual midpoint.

The *middle means typical* heuristic indicates that the visual midpoint of a scale should receive comparatively more attention and that respondents' focus should shift toward the visual midpoint when it is placed to one side of the conceptual midpoint. Hence, we postulate the following hypotheses:

Hypothesis 1.1: Respondents fixate equally long and equally often on the visual midpoint of a response scale, irrespective of whether it coincides

with the conceptual midpoint; i.e. respondents pay the same amount of attention to the visual midpoint in conditions 2 and 3 (i.e., “about the right amount”) as to the visual midpoint in condition 1 (i.e., “too little”).

Hypothesis 1.2: Respondents fixate longer and more often on the visual midpoint than on the conceptual midpoint of a scale when the visual mid-point does not match the conceptual one (i.e., respondents pay more attention to the “too little” option than to the “about the right amount” option in condition 1).

Hypothesis 1.3: When the visual midpoint of a response scale is placed to one side of the conceptual midpoint (condition 1), respondents’ attention shifts toward this side of the scale, and they fixate longer and more often on the substantive options of this side (i.e., respondents pay more attention to the “too little” and “far too little” options in condition 1 than in conditions 2 and 3).

In addition to these hypotheses that were directly derived from the *middle means typical* heuristic, Tourangeau et al. (2004) found that respondents were more likely to select a nonsubstantive response option for the conditions in which these were separated from the other substantive options by a space or divider line. The authors speculated that dividing the nonsubstantive options would draw attention to these options and thus increase the likelihood of respondents selecting them. Hence, we postulate the following additional hypothesis:

Hypothesis 1.4: Respondents fixate longer and more often on the nonsubstantive response options if these are separated from the other options by a space or divider line (i.e., respondents pay more attention to the nonsubstantive response options in the conditions 2 and 3 than in condition 1).

4.2 Experiment 2: Order of the Response Options

The two questions for the experiment on *order of the response options* were taken from Tourangeau et al. (2004) and dealt with physician-patient relations (see Appendices A and B). Again, respondents were randomly assigned to one of three experimental conditions. The first group (n = 43) received the response options presented in a consistent order (consistent condition; condition 1). The second group (n = 46) was shown response options in a mildly inconsistent order (mildly inconsistent condition; condition 2). The third group (n = 42) received two questions presenting the response options in a strongly inconsistent order (strongly inconsistent condition; condition 3).

The *left and top means first* heuristic indicates that respondents will be confused when response options do not follow a logical order and thus will need more time to process the scale in comparison with response scales following a conventional order. Hence, we postulate the following hypotheses:

Hypothesis 2.1: Respondents fixate longer and more often on the response options the more inconsistent they are with the *left and top means first* heuristic; i.e., respondents fixate longest and most often on the response options when the order is strongly inconsistent with the heuristic (condition 3), followed by the mildly inconsistent order (condition 2), and the consistent order (condition 1).

Hypothesis 2.2: Respondents read more response options and show more re-fixations between the response options; that is, they re-fixate an option they have read previously after reading at least one other option the more inconsistent they are with the *left and top means first* heuristic. Specifically, respondents read most options and show most re-fixations in the strongly inconsistent condition (condition 3), followed by the mildly inconsistent condition (condition 2), and the consistent condition (condition 1).

In addition, Tourangeau et al. (2004) found that respondents were more likely to select the conceptual middle option “it depends” if it was presented in the middle of the response scale than if it was presented at the top or the bottom. The authors argued that the middle option’s meaning is unambiguous when its position suggests that it represents the midpoint of the scale. If this is the case, then respondents should need less time to infer the meaning of the response option “it depends” if it is placed in the center of the scale. Hence, we postulate the following additional hypothesis:

Hypothesis 2.3: Respondents fixate shorter and less often on the conceptual middle option (i.e., “it depends”) when it is presented in the middle of the scale than when it is presented at the top or the bottom of the scale.

5. METHOD

5.1 Participants

We recruited respondents from the respondent pool maintained by the institute and by word of mouth. In total, 131 respondents participated in both experiments. 38 percent were between 18 and 24 years old, 38 percent were between 25 and 44 years old, 19 percent were between 45 and 64 years old, and 5 percent were 65 years or older. Fifty percent of the respondents were female. Seven percent had graduated from a lower secondary school, 14 percent from an intermediate secondary school, and 79 percent from a college preparatory secondary school or university. Thirty-seven percent of the participants had participated in at least one web survey during the last three months.

In order to evaluate the effectiveness of random assignment and the sample composition between groups in both experiments, we conducted several χ^2 tests for the reported sociodemographic characteristics mentioned previously. No statistically significant differences could be observed.

Due to technical difficulties, the recordings of the eye movements for some respondents were not satisfactory because there was a systematic shift to the line below or above the one that was fixated. These respondents were excluded from the data, leaving 111 to 114 respondents for statistical analyses.

5.2 Eye-Tracking Equipment

The Senso Motoric Instruments (SMI) RED250mobile eye-tracking system was used to record participants' eye movements. To identify saccades and fixations, we used the SMI BeGaze version 2.3 built-in event detector for high-speed eye-tracking data (e.g., for data recorded at a sampling rate of 200 Hz or higher) in the default setting—minimum saccade duration: auto; peak velocity threshold: 40°/second; minimum fixation duration: 50 ms; peak velocity: (a) start: 20 percent of saccade length, and (b) end: 80 percent of saccade length. The RED250mobile Eye Tracker is a mobile device that can be mounted on the bottom frame of a desktop monitor or laptop display. The system is typically accurate within 0.4° and has a resolution of 0.3°. It permits head movements within a range of 32 × 21 cm at 60 cm distance. In the setup of this experiment, the eye tracker was mounted on the bottom frame of a 22-inch TFT monitor (resolution 1680 × 1050). Eye movements were recorded at a sampling rate of 250 Hz. The online questionnaire was programmed with a font size of 16 pixels and double-spaced text with a line height of 40 and 32 pixels for the question text and response options, respectively.

5.3 Procedure

The study was conducted at GESIS–Leibniz Institute for the Social Sciences in Mannheim, Germany, in April and May of 2017. Participants were seated in front of the SMI RED250mobile eye-tracking system so that their eyes were approximately 60 cm from it and they were instructed to sit down in a comfortable but stable position. Before the web survey started, the eye tracker was calibrated to ensure each participant's eyes could be accurately tracked.

The web survey contained several unrelated experiments that were independently randomized to avoid systematic carryover effects. The entire eye-tracking study was supervised by an experimenter who stayed in an adjacent room to observe participants' eye movements on a separate computer screen and to assist in case of problems. The average completion time was about

30 minutes. For their participation in the whole study, respondents received a compensation of 20 euros.

5.4 Analytical Strategies

To explore the mechanisms underlying interpretative heuristics and examine how respondents processed the different question versions, we looked at the following eye-tracking parameters: fixation count, fixation time, number of response options read, and re-fixations (the latter two only for the experiment on *order of the response options*). These four eye-tracking parameters can be defined as follows: (i) Fixation count is the total number of fixations on a specific area of interest (e.g., the response options) including re-readings. (ii) Fixation time is the total duration of fixations on a specific area of interest (e.g., the response options) including re-readings. (iii) Number of response options read refers to the total number of response options that respondents fixate (including re-fixations). (iv) Re-fixations on a specific area of interest (e.g., the response options) are the total number of areas that respondents re-fixate (e.g., fixating a response option again after reading at least one other option).

Due to technical limitations, the number of response options read and the number of re-fixations could not be automatically detected by the RED250mobile eye-tracking system. The questions were then coded by two coders; each coded the eye movements of one half of the respondents ($n = 65$ and $n = 66$, respectively). In addition, the eye movements of a randomly selected subset of 10 percent of the respondents ($n = 13$) were coded by both coders for the purpose of estimating reliability. Interrater agreement was excellent (Fleiss et al. 2003), with an intraclass correlation coefficient (ICC) of 0.98. Discrepancies between the two ratings were examined and discussed with the second author until a consensus was reached.

6. RESULTS

In both experiments, we conducted one-way analyses of variance (ANOVAs) using the Bonferroni α -inflation correction procedure for equal variances to deal with the problem of multiple comparisons to statistically analyze fixation counts, fixation times, number of response options read, and re-fixations. Furthermore, we calculated Cohen's d (see Cohen 1969) to determine the effect sizes. Because there are no distinct differences between specific questions, we decided to conduct all statistical analyses at the aggregated level in an attempt to reduce the number of statistical tests and efficiently summarize the results.

In addition, we report the response distributions (see Appendix C) and response times (see Appendix D) for the two questions of the experiments on *separating scale points from nonsubstantive responses* and *order of the response options*.

6.1 Experiment 1: Separating Scale Points from Nonsubstantive Responses

6.1.1 Fixation Count and Time on the Visual Midpoint. Supporting hypothesis 1.1, table 1 shows no statistically significant differences between the three experimental conditions with respect to the fixation count and time on the visual midpoint of the scale. This result is additionally supported by Cohen’s *d*, which indicates relatively small effect sizes ($d < 0.20$), except for fixation time between conditions 1 and 3 as well as conditions 2 and 3. All in all, the visual midpoints received the same amount of attention, irrespective of whether they fell on the conceptual midpoint of the scale (conditions 2 and 3) or whether it fell on the lower side of the conceptual midpoint of the scale (condition 1).³

We additionally investigated whether the visual midpoint was also the response option that received most attention in comparison with the other options. To do so, we looked at the mean fixation count and time on each option across all conditions. Figure 1 shows that respondents fixated longest and most often on the fourth option (“too little”) in all conditions. Hence, the visual midpoint did not necessarily receive most attention. However, it is important to keep in mind that the fixation count and time on each of the options is determined not only by how intensively respondents process them but also by the

Table 1. Mean Differences and Effect Sizes (in Parentheses) of Fixation Count and Time Between the Visual Midpoint (“Too Little”) in Condition 1 and the Visual Midpoint (“About the Right Amount”) in Conditions 2 and 3

Fixation count					
	Condition 1	Condition 2	<i>F</i> value ($df_1 = 2$)	df_2	<i>p</i> value
Condition 2	−0.37 (0.12)		0.38	111	0.685
Condition 3	−0.61 (0.19)	−0.25 (0.08)			
Fixation time (seconds)					
Condition 2	−0.37 (0.19)		1.99	111	0.141
Condition 3	−0.77 (0.49)	−0.40 (0.26)			

NOTE. — Mean differences were calculated by subtracting column means from row means. Cohen’s *d* (in parentheses) indicates the effect size. Condition 1: nonseparation; condition 2: space; condition 3: line.

³ We also controlled for the responses given in the analyses, but the main results did not change.

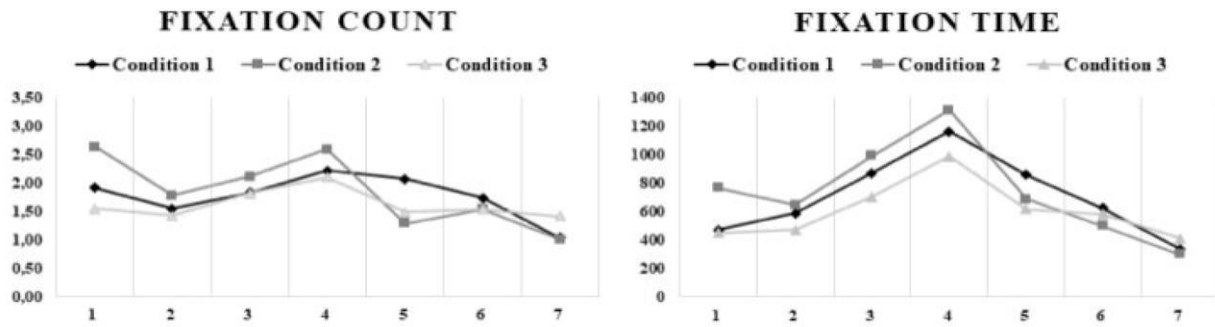


Figure 1. Mean Fixation Count and Time for the Seven Response Options Across the Three Experimental Conditions. NOTE. —Condition 1: nonseparation; condition 2: space; condition 3: line. Response options: (1) “far too much,” (2) “too much,” (3) “about the right amount,” (4) “too little,” (5) “far too little,” (6) “don’t know,” and (7) “no opinion.” Fixation time is stated in milliseconds (ms).

option they select. Options that end up being selected receive some attention simply by the fact that for selecting the option, respondents have to fixate them (see Galesic et al. 2008). An inspection of the response distributions (see Appendix C) reveals that the “too little” option was also selected most frequently, irrespective of the condition. This finding corresponds to the distributions reported by Tourangeau et al. (2004).

6.1.2 Fixation Count and Time on the Conceptual and Visual Midpoint in the First Condition.

In line with hypothesis 1.2, table 2 shows that respondents pay more attention to the visual midpoint (“too little”) than to the conceptual midpoint (“about the right amount”) of the response scale in condition 1 (see also figure 1). This was marginally significant for fixation count and significant for fixation time. Cohen’s *d* also indicates a small to medium effect size ($d > 0.20$).

6.1.3 Fixation Count and Time on the Fourth and Fifth Response Options.

According to hypothesis 1.3, we expected that placing the visual midpoint be low the conceptual midpoint (condition 1) would increase respondents’ fixation count and time on the two substantive options (“too little” and “far too little”) of this side of the response scale. As expected, table 3 shows that respondents consistently fixated longer and more often on these two substantive options in condition 1 than in the conditions 2 and 3. This finding is additionally supported by figure 1. However, the differences were not statistically significant. Accordingly, Cohen’s *d* indicates comparatively small effect sizes ($d < 0.30$).

6.1.4 Fixation Count and Time on the Nonsubstantive Response Options.

According to hypothesis 1.4, we examined whether the nonsubstantive

Table 2. Mean Differences and Effect Sizes (in Parentheses) of Fixation Count and Time Between the Conceptual Midpoint (“About the Right Amount”) and the Visual Midpoint (“Too Little”) in Condition 1

Condition 1	Eye-tracking parameters	
	Fixation count	Fixation time (seconds)
Conceptual–visual midpoint	–0.71 ⁺ <i>t</i> (40) = –1.62 (0.23)	–0.54* <i>t</i> (40) = –1.74 (0.31)

NOTE. — + $p < 0.10$; * $p < 0.05$. We conducted paired *t* tests to determine the differences in the allocation of attention between the conceptual and visual midpoint in condition 1. Cohen’s *d* (in parentheses) indicates the effect size. Condition 1: nonseparation.

Table 3. Mean Differences and Effect Sizes (in Parentheses) of Fixation Count and Time on the Fourth and Fifth Response Options (“Too Little” and “Far Too Little”) Between the Three Experimental Conditions

Fixation count			<i>F</i> value (<i>df</i> ₁ = 2)	<i>df</i> ₂	<i>p</i> value
	Condition 1	Condition 2			
Condition 2	–1.05 (0.18)		0.58	111	0.562
Condition 3	–1.12 (0.22)	–0.06 (0.01)			
Fixation time (seconds)					
Condition 2	–0.17 (0.06)		0.61	111	0.530
Condition 3	–0.68 (0.28)	–0.51 (0.20)			

NOTE. —Mean differences were calculated by subtracting column means from row means. Cohen’s *d* (in parentheses) indicates the effect size. Condition 1: nonseparation; condition 2: space; condition 3: line.

response options (“don’t know” and “no opinion”) received more attention if they were separated from the other options by a space (condition 2) or a divider line (condition 3) than if they were presented as additional radio buttons (condition 1). Contrary to our expectation, table 4 shows no significant differences in fixation count and time on the nonsubstantive options between the three

Table 4. Mean Differences and Effect Sizes (in Parentheses) of Fixation Count and Time on the Nonsubstantive Response Options (“Don’t Know” and “No Opinion”) Between the Three Experimental Conditions

Fixation count					
	Condition 1	Condition 2	<i>F</i> value (<i>df</i> ₁ = 2)	<i>df</i> ₂	<i>p</i> value
Condition 2	-0.68 (0.15)		0.53	111	0.589
Condition 3	0.46 (0.09)	1.14 (0.27)			
Fixation time (seconds)					
Condition 2	-0.36 (0.19)		0.69	111	0.503
Condition 3	0.12 (0.06)	-0.49 (0.27)			

NOTE. —Mean differences were calculated by subtracting column means from row means. Cohen’s *d* (in parentheses) indicates the effect size. Condition one: nonseparation; condition two: space; condition three: line.

conditions. This finding also corresponds to the average fixation count and time shown in figure 1. Cohen’s *d* indicates comparatively small effect sizes, except for fixation count and time between conditions 2 and 3. Hence, our findings are not in line with Tourangeau et al.’s (2004) speculation that nonsubstantive options receive more attention if they are separated from the other options.

6.2 Experiment 2: Order of the Response Options

6.2.1 Fixation Count and Time on the Full Scale. In line with hypothesis 2.1, table 5 shows that respondents fixated more often and longer on the response options the more inconsistent they were with the left and top means first heuristic. Significant differences in fixation count were found between conditions 1 and 3 and in fixation time between both conditions 1 and 2 as well as conditions 1 and 3. Cohen’s *d* additionally supports these results, indicating strong effect sizes ($d > 0.60$). All in all, it seems that the order inconsistencies indeed required respondents to engage in a more effortful processing. These findings are in line with Holbrook et al. (2000) and Tourangeau et al. (2004), who were able to show that order discrepancies slowed respondents down.

Table 5. Mean Differences and Effect Sizes (in Parentheses) of Fixation Count and Time Between the Three Experimental Conditions

Fixation count					
	Condition 1	Condition 2	<i>F</i> value (<i>df</i> ₁ = 2)	<i>df</i> ₂	<i>p</i> value
Condition 2	3.00 (0.31)		3.81	108	0.025
Condition 3	7.06* (0.61)	4.05 (0.35)			
Fixation time (seconds)					
Condition 2	1.79* (0.62)		4.24	108	0.017
Condition 3	-0.68 (0.28)	0.08 (0.02)			

NOTE. — * $p < 0.05$. Mean differences were calculated by subtracting column means from row means. Cohen’s *d* (in parentheses) indicates the effect size. Condition 1: consistent order; condition 2: mildly inconsistent order; condition 3: strongly inconsistent order.

6.2.2 Number of Response Options Read and Re-fixated. Consistent with hypothesis 2.2, table 6 shows that the number of response options read increases with the order discrepancies. Respondents read marginally significantly more options in condition 2 than in condition 1 and significantly more options in condition 3 than in condition 1. For these two comparisons, Cohen’s *d* shows medium strong effect sizes ($d > 0.50$).

Table 6 also shows that the number of re-fixations increases with the order inconsistencies. This effect was significant between condition 1 (consistent order) and condition 2 (mildly inconsistent order) and marginally significant between condition 1 and condition 3 (strongly inconsistent order). Cohen’s *d* reveals medium strong effect sizes ($d > 0.50$) for these comparisons.

6.2.3 Fixation Count and Time on the Conceptual Midpoint. With respect to hypothesis 2.3, we compared the fixation count and the time on the conceptual middle option (“it depends”). Contrary to our expectation, respondents in condition 1 (“it depends” presented in the middle) and condition 3 (“it depends” presented at the top) produced significantly more and longer fixations than respondents in condition 2 (“it depends” presented at the bottom), as shown in table 7. This result is again supported by Cohen’s *d*, which indicates strong effects sizes ($d > 0.70$).

Table 6. Mean Differences and Effect Sizes (in Parentheses) of Number of Response Options Read and Re-fixated Between the Three Experimental Conditions

Fixation count					
	Condition 1	Condition 2	<i>F</i> value (<i>df</i> ₁ = 2)	<i>df</i> ₂	<i>p</i> value
Condition 2	1.11 ⁺ (0.55)		4.50	108	0.013
Condition 3	1.44** (0.52)	0.34 (0.06)			
Fixation time (seconds)					
Condition 2	1.95* (0.56)		3.55	108	0.032
Condition 3	1.72 ⁺ (0.52)	-0.23 (0.07)			

NOTE. — ⁺*p*<0.10; **p*<0.05; ***p*<0.01. Mean differences were calculated by subtracting column means from row means. Cohen's *d* (in parentheses) indicates the effect size. Condition 1: consistent order; condition 2: mildly inconsistent order; condition 3: strongly inconsistent order.

7. DISCUSSION AND CONCLUSION

Tourangeau et al. (2004) provided intriguing examples and theoretical explanations of how the middle means typical and the left and top means first heuristics can influence response behavior. However, the conclusions drawn from their studies on the underlying mechanisms are based on so-called indirect data, such as response distributions and response times. The usefulness of such indirect data is somewhat limited because question-processing and respondent behavior are not directly observed and studied (Galesic et al. 2008; Galesic and Yan 2011). For this reason, we conducted the experiments on separating scale points from nonsubstantive responses and order of the response options by means of eye-tracking methodology to examine how violations of the middle means typical and the left and top means first heuristics influence respondents' processing of survey questions. Table 8 provides a summary of our findings in relation to the research hypotheses.

In line with our expectation (see H1.1), we found that the visual midpoint received the same amount of attention, irrespective of whether it fell on the conceptual midpoint of the scale (conditions 2 and 3) or whether it fell on the lower side of the conceptual midpoint of the scale (condition 1). In addition, we found that respondents paid more attention to the visual midpoint (“too

Table 7. Mean Differences and Effect Sizes (in Parentheses) of Fixation Count and Time on the Conceptual Middle Response Option Between the Three Experimental Conditions

Fixation count					
	Condition 1	Condition 2	<i>F</i> value (<i>df</i> ₁ = 2)	<i>df</i> ₂	<i>p</i> value
Condition 2	-3.50*** (0.99)		11.65	108	0.001
Condition 3	1.05 (0.19)	4.55*** (1.10)			
Fixation time (seconds)					
Condition 2	-0.98* (0.79)		7.53	108	0.001
Condition 3	0.13 (0.09)	1.11** (0.81)			

NOTE. —***p*<0.01; ****p*<0.001. Mean differences were calculated by subtracting column means from row means. Cohen’s *d* (in parentheses) indicates the effect size. Condition 1: consistent order; condition 2: mildly inconsistent order; condition 3: strongly inconsistent order.

Table 8. Summary of the Findings for the Hypotheses of the Two Experiments on Separating Scale Points From Nonsubstantive Responses and Order of the Response Options

Hypotheses	Findings
Experiment 1: Separating scale points from nonsubstantive responses	
H1.1	Supporting evidence
H1.2	Supporting evidence
H1.3	No supporting evidence
H1.4	No supporting evidence
Experiment 2: Order of the response options	
H2.1	Supporting evidence
H2.2	Supporting evidence
H2.3	No supporting evidence

little”) than to the conceptual midpoint (“about the right amount”) of the response scale in condition 1 (see H1.2). However, we did not find evidence for the attention shift toward the side of the scale where the visual midpoint is

located (see H1.3). There was also no evidence for more attention on nonsubstantive options when they are separated by a space or divider line from the substantive options (see H1.4). Our results indicate that respondents either do not utilize the middle means typical heuristic when answering the survey questions or only a comparatively small number of respondents is affected by the visual design changes. To investigate this issue further, future research could employ studies with more statistical power and investigate the effects among different respondent groups. It is also crucial to test questions that vary, for instance, with respect to the topic, verbalization, number of categories, scale labeling and polarity, and numerical labels. Altogether, more systematic research on the middle means typical heuristic and its implications is necessary to get a solid understanding of its importance.

For the experiment on order of the response options, we found that both fixation counts and fixation times increase with the order discrepancies (see H2.1), replicating findings reported by Holbrook et al. (2000) and Tourangeau et al. (2004). In addition, we observed that the number of options read and the number of re-fixations increase with the degree of inconsistency (see H2.2). However, there was no evidence indicating that respondents fixated on the conceptual midpoint (“it depends”) least intensively when it was placed in the middle of the scale (see H2.3). In general, the left and top means first heuristic seems to be at work for web survey responding. Response options that are not presented in a logical order affect question processing in terms of eye fixations, increase effort in responding, and affect responses obtained. Therefore, we highly recommend presenting response options in a way consistent with the *left and top means first* heuristic to decrease response effort and to enhance survey responses.

We note three limitations associated with this study. First, as suggested in the results section, the intensity with which a response option is fixated (e.g., in terms of fixation count and time) depends not only on how deeply respondents process it but also on whether they actually select it. For instance, in the case of the experiment on separating scale points from nonsubstantive responses, it is difficult to determine whether the option “too little” received more attention because of its function as visual midpoint or because it was the most frequently selected option. It would be beneficial if the visual midpoint was not the most popular option since this impedes the evaluation of its relevance in comparison with the other options by means of eye tracking. Second, more evidence on how respondents actually interpret response options, such as the middle option, is necessary to better understand the mechanisms of the middle means typical and left and top means first heuristics. Unfortunately, this cannot be solely achieved by eye-tracking methodology and goes beyond the scope of this study. For this reason, we recommend that future research address this point by combining eye-tracking methodology with cognitive interviewing (see Neuert and Lenzner 2017). Third, we only tested two questions on the experiments on separating scale points from nonsubstantive responses

and order of the response options, respectively. The main reason is that we adopted the experimental designs and survey questions used by Tourangeau et al. (2004). However, it would be useful if future research tests the implications of both heuristics by employing multiple questions.

Finally, our review of the existing survey literature on visual question design strategies and interpretive heuristics (see, for instance, Christian and Dillman 2004; Couper et al. 2004; Schwarz et al. 1998; Schwarz et al. 1991; Smith 1995; Toepoel and Dillman 2011a, 2011b; Tourangeau et al. 2004, 2007) indicates that most of these existing studies only address question processing and response behavior but do not address data quality, such as reliability and validity. Therefore, we recommend investigating the consequences for data quality in future studies. This would also facilitate evaluating the relevance of interpretive heuristics for quantitative social research.

References

- Christian, L. M., and D. A. Dillman (2004). "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions," *Public Opinion Quarterly*, 68, 57–80.
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Science*, New York, NY: Academic Press.
- Couper, M. P., R. Tourangeau, and K. Kenyon (2004). "Picture This! Exploring Visual Effects in Web Surveys," *Public Opinion Quarterly*, 68, 255–266.
- Dillman, D. A., J. D. Smyth, and L. M. Christian (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, Hoboken, NJ: Wiley.
- Fleiss, J. L., B. Levin, and M. C. Paik (2003). *Statistical Methods for Rates and Proportions*, Hoboken, NJ: Wiley.
- Galesic, M., R. Tourangeau, M. P. Couper, and F. 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.
- Galesic, M., and T. Yan (2011). "Use of Eye Tracking for Studying Survey Response Processes," in *Social and Behavioral Research and the Internet. Advances in Applied Methods and Research Strategies*, eds. M. Das, P. Ester, and L. Kaczmirek, pp. 349–370 (New York, NY: Routledge).
- Graesser, A. C., Z. Cai, M. M. Louwerse, and F. Daniel (2006). "Question Understanding Aid (QUAID). A Web Facility that Tests Question Comprehensibility," *Public Opinion Quarterly*, 70, 3–22.
- Holbrook, A., J. A. Krosnick, R. T. Carson, and R. C. Mitchell (2000). "Violating Conversational Conventions Disrupts Cognitive Processing of Attitude Questions," *Journal of Experimental Social Psychology*, 36, 465–494.
- Höhne, J. K. (2019). "Eye-Tracking Methodology: Exploring the Processing of Question Formats in Web Surveys," *International Journal of Social Research Methodology*, 22, 199–206.
- Höhne, J. K., and T. Lenzner (2015). "Investigating Response Order Effects in Web Surveys Using Eye Tracking," *Psihologija*, 48, 361–377.
- . (2018). "New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific Questions," *Journal of Survey Statistics and Methodology*, 6, 401–417.
- Just, M. A., and P. A. Carpenter (1980). "A Theory of Reading: From Eye Fixations to Comprehension," *Psychological Review*, 87, 329–354.

- Kamoen, N., B. Holleman, P. Mak, T. Sanders, and H. van den Bergh (2011). "Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions," *Discourse Processes*, 48, 355–385.
- . (2017). "Why Are Negative Questions Difficult to Answer? On the Processing of Linguistic Contrasts in Surveys," *Public Opinion Quarterly*, 81, 613–635.
- Lenzner, T. (2012). "Effects of Survey Question Comprehensibility on Response Quality," *Field Methods*, 24, 409–428.
- Lenzner, T., L. Kaczmirek, and M. Galesic (2011). "Seeing Through the Eyes of the Respondent: An Eye-Tracking Study on Survey Question Comprehension," *International Journal of Public Opinion Research*, 23, 361–373.
- . (2014). "Left feels right: A usability study on the position of answer boxes in web surveys," *Social Science Computer Review*, 32, 743–764.
- Neuert, C. E. (2017). "Processing Forced-Choice vs. Check-All-That-Apply Question Formats: Evidence from Eye Tracking," *Field Methods*, 29, 383–394.
- Neuert, C. E., and T. Lenzner (2017). "Incorporating Eye Tracking into Cognitive Interviewing to Pretest Survey Questions," *International Journal of Social Research Methodology*, 19, 501–519.
- Rayner, K. (1998). "Eye Movements in Reading and Information Processing: 20 Years of Research," *Psychological Bulletin*, 124, 372–422.
- Rayner, K., and A. Pollastek (2006). "Eye-Movement Control in Reading," in *Handbook of Psycholinguistic*, eds. M. Traxler and M. A. Gernsbacher, pp. 613–658 (Amsterdam: Elsevier).
- Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*, Mahwah, NJ: Lawrence Erlbaum Associates.
- Schwarz, N., C. E. Grayson, and B. Knäuper (1998). "Formal Features of Rating Scales and the Interpretation of Question Meaning," *International Journal of Public Opinion Research*, 10, 177–183.
- Schwarz, N., and H.-J. Hippler (1987). "What Response Scales May Tell Your Respondents: Information Functions of Response Alternatives," in *Social Information Processing and Survey Methodology*, eds. H.-J. Hippler, N. Schwarz, and S. Sudman, pp. 163–178 (New York, NY: Springer).
- Schwarz, N., H.-J. Hippler, B. Deutsch, and F. Strack (1985). "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments," *Public Opinion Quarterly*, 49, 388–395.
- Schwarz, N., B. Knäuper, H.-J. Hippler, E. Noelle-Neumann, and L. N. Clark (1991). "Rating Scales: Numeric Values May Change the Meaning of Scale Labels," *Public Opinion Quarterly*, 55, 570–582.
- Smith, T. W.; American Statistical Association (1995). "Little Things Matter: A Sampler of How Differences in Questionnaire Format Can Affect Survey Responses," in *Proceedings of the American Statistical Association: Survey Research Methods Section*, pp. 1046–1051 (New York, NY: Springer).
- Toepoel, V., and D. A. Dillman (2011a). "How Visual Design Affects the Interpretability of Survey Questions," in *Social and Behavioral Research and the Internet. Advances in Applied Methods and Research Strategies*, eds. M. Das, P. Ester, and L. Kaczmirek, pp. 165–190 (New York, NY: Routledge).
- . (2011b). "Words, Numbers, and Visual Heuristics in Web Surveys: Is There a Hierarchy of Importance?," *Social Science Computer Review*, 29, 193–207.
- Tourangeau, R., M. P. Couper, and F. G. Conrad (2004). "Spacing, Position, and Order. Interpretive Heuristics for Visual Features of Survey Questions," *Public Opinion Quarterly*, 68, 368–393.
- . (2007). "Colors, Labels, and Interpretive Heuristics for Response Scales," *Public Opinion Quarterly*, 71, 91–112.
- Tourangeau, R., L. L. Rips, and K. Rasinski (2000). *The Psychology of Survey Response*, New York, NY: Cambridge University Press.

Appendix A

Experimental questions (EQ) of the experiment on separating scale points from nonsubstantive responses (middle means typical):

EQ 1: Think of how much the federal government is doing to make sure women have the same job opportunities as men. Would you say the federal government is doing too much, about the right amount, or too little about this?

EQ 2: Think of how much the federal government is doing to provide day care centers for the children of working parents. Would you say the federal government is doing too much, about the right amount, or too little about this?

Response options to EQ 1 and EQ 2 are “far too much,” “too much,” “about the right amount,” “too little,” “far too little,” “don’t know,” and “no opinion.”

Experimental questions (EQ) of the experiment on order of the response options (left and top means first):

EQ 1: Do you agree or disagree with the following statement? It is sensible to do exactly what the doctors say.

EQ 2: Do you agree or disagree with the following statement? I have to be very ill before I go to the doctor.

Response options to EQ 1 and EQ 2 are “agree strongly,” “agree,” “it depends,” “disagree,” “disagree strongly” (consistent order).

Note. —The order of the questions corresponds to the presentation in Appendix A. The German translations of all questions including response options are available from the first author on request.

Appendix B

Think of how much the federal government is doing to make sure women have the same job opportunities as men. Would you say the federal government is doing too much, about the right amount, or too little about this?

- Far too much
- Too much
- About the right amount
- Too little
- Far too little
- Don't know
- No opinion

Continue

Figure 2a. First Question of the Experiment on Separating Scale Points from Nonsubstantive Responses (Nonseparation Condition).

Think of how much the federal government is doing to make sure women have the same job opportunities as men. Would you say the federal government is doing too much, about the right amount, or too little about this?

- Far too much
- Too much
- About the right amount
- Too little
- Far too little

- Don't know
- No opinion

Continue

Figure 2b. First Question of the Experiment on Separating Scale Points from Nonsubstantive Responses (Space Condition).

Think of how much the federal government is doing to make sure women have the same job opportunities as men. Would you say the federal government is doing too much, about the right amount, or too little about this?

- Far too much
- Too much
- About the right amount
- Too little
- Far too little

-
- Don't know
 - No opinion

Continue

Figure 2c. First Question of the Experiment on Separating Scale Points from Nonsubstantive Responses (Line Condition).

**Do you agree or disagree with the following statement?
It is sensible to do exactly what the doctors say.**

- Agree strongly
- Agree
- It depends
- Disagree
- Disagree strongly

Continue

Figure 2d. First Question of the Experiment on Order of the Response Options (Consistent Condition).

**Do you agree or disagree with the following statement?
It is sensible to do exactly what the doctors say.**

- Agree strongly
- Agree
- Disagree
- Disagree strongly
- It depends

Continue

Figure 2e. First Question of the Experiment on Order of the Response Options (Mildly Inconsistent Condition).

**Do you agree or disagree with the following statement?
It is sensible to do exactly what the doctors say.**

- It depends
- Agree strongly
- Disagree strongly
- Agree
- Disagree

Continue

Figure 2f. First Question of the Experiment on Order of the Response Options (Strongly Inconsistent Condition).

Appendix C

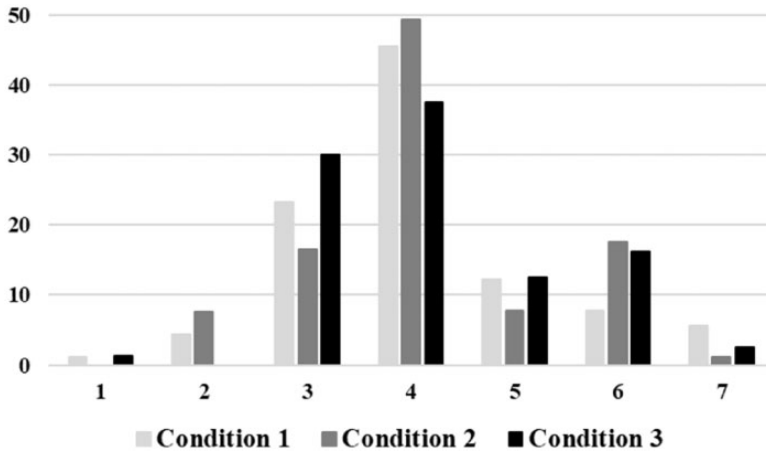


Figure 3a. Response Distributions for the Two Questions of the Experiment on Separating Scale Points from Nonsubstantive Responses.

NOTE.—N 5 131. Condition 1: nonseparation; condition 2: space; condition 3: line. Response options: (1) “far too much,” (2) “too much,” (3) “about the right amount,” (4) “too little,” (5) “far too little,” (6) “don’t know,” and (7) “no opinion.” Due to the small sample size, we do not report any test statistics.

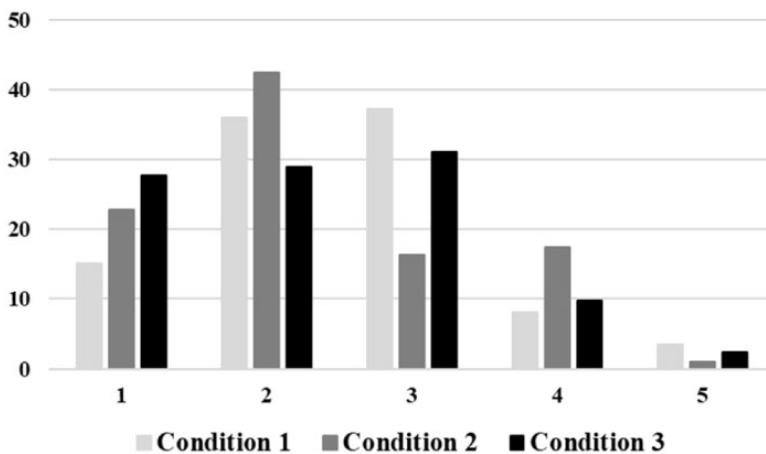


Figure 3b. Response Distributions for the Two Questions of the Experiment on Order of the Response Options.

NOTE.—N 5 131. Condition 1: consistent order; condition 2: mildly inconsistent order; condition 3: strongly inconsistent order. Response options: (1) “agree strongly,” (2) “agree,” (3) “it depends,” (4) “disagree,” (5) “disagree strongly.” Due to the small sample size, we do not report any test statistics.

Appendix D

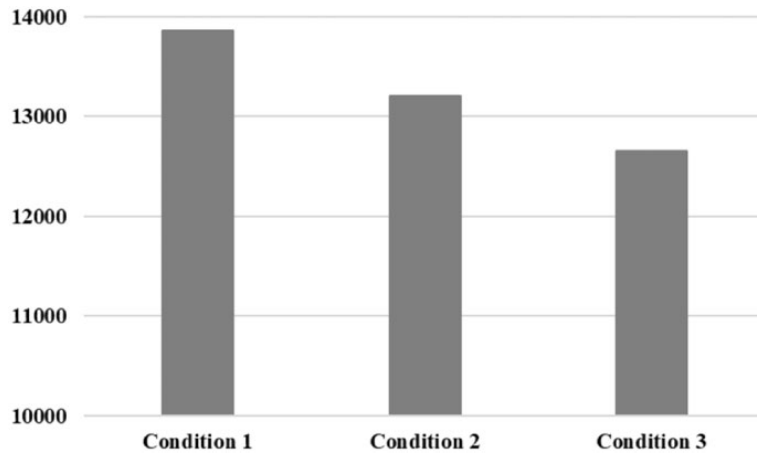


Figure 4a. Median Response Times (in Milliseconds) for the Two Questions of the Experiment on Separating Scale Points from Nonsubstantive Responses. NOTE. —N 5 131. Condition 1: nonseparation; condition 2: space; condition 3: line. The result of a Kruskal-Wallis test indicates no significant differences.

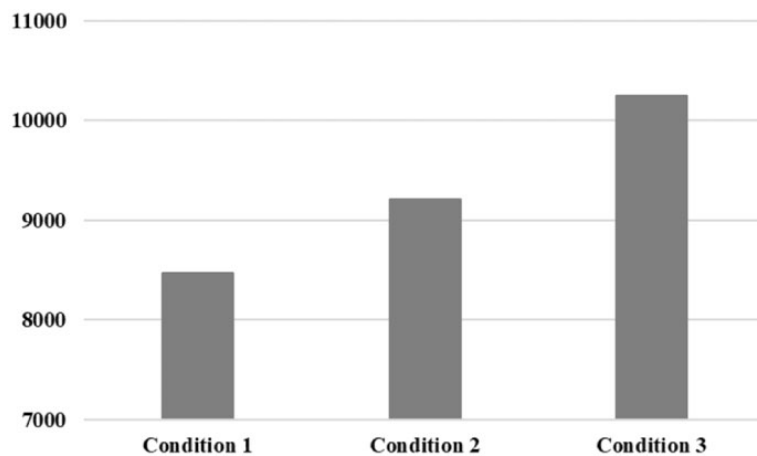


Figure 4b. Median Response Times (in Milliseconds) for the Two Questions of the Experiment on Order of the Response Options.

NOTE. —N 5 131. Condition 1: consistent order; condition 2: mildly inconsistent order; condition 3: strongly inconsistent order. The result of a Kruskal-Wallis test indicates no significant differences.