

New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific Questions

Höhne, Jan Karem; Lenzner, Timo

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. (2018). New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific Questions. *Journal of Survey Statistics and Methodology*, 6(3), 401-417. <https://doi.org/10.1093/jssam/smx028>

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.

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.

New Insights on the Cognitive Processing of agree/disagree and item-specific Questions

Jan Karem Höhne

Timo Lenzner

In empirical social research, using questions with an agreement scale, also known as agree/disagree (A/D) questions, is a popular technique for measuring attitudes and opinions. Methodological considerations, however, suggest that such questions require effortful cognitive processing and are prone to response bias, such as acquiescence. Therefore, many researchers recommend the use of item-specific (IS) questions, which are based on tailored response categories and seem to imply less response burden. In this study, we investigate the cognitive processing of A/D and IS questions in web surveys, using eye-tracking methodology. On the basis of recordings of respondents' eye movements, we are able to draw conclusions about how respondents process survey questions and to evaluate how they process information. Our results indicate that IS questions require deeper processing than A/D questions. Interestingly, the eye-tracking data reveals that this phenomenon is only observable for the response categories but not for the question stems; this indicates that the stems do not differ in terms of cognitive effort. We therefore argue that the observed differences are directly attributable to a more intensive processing of the IS response categories. Practically speaking, this

*Address correspondence to Jan Karem Höhne, Center of Methods in Social Sciences, Faculty of Social Sciences, University of Göttingen, Goßlerstraße 19, 37073 Göttingen, Germany; E-mail: jhoehne@uni-goettingen.de.

JAN KAREM HÖHNE is a doctoral candidate and research associate at the Center of Methods in Social Sciences, Faculty of Social Sciences, University of Göttingen, Goßlerstraße 19, 37073 Göttingen, Germany. TIMO LENZNER is a senior researcher at the Department of Survey Design and Methodology, GESIS—Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072 Mannheim, Germany.

The authors are grateful to Jon A. Krosnick and the Political Psychology Research Group (Stanford University), Willem E. Saris (RECSM-Universitat Pompeu Fabra), and Cornelia E. Neuert (GESIS—Leibniz Institute for the Social Sciences) for their advice and support in writing this article. They also would like to thank Clara Beitz and Stefanie Gebhardt (GESIS—Leibniz Institute for the Social Sciences) for coding the eye-tracking videos, as well as the anonymous reviewers and the editors for their constructive suggestions for improving this article.

additionally indicates a more thoughtful processing of the response categories and, thus, might lead to more well-considered and appropriate responses.

KEYWORDS: Asking manner; Cognitive effort; Eye-tracking experiment; Question format; Response behavior; Web survey.

1. INTRODUCTION AND BACKGROUND

The measurement of attitudes and opinions by means of agree/disagree (A/D) questions (i.e., response categories are based on an agreement/disagreement continuum) is a prevalent methodological measurement technique in behavioral and social science research. A/D questions ask respondents to report their attitudes and opinions by agreeing or disagreeing with a statement. For example, to measure political efficacy, respondents can be asked whether they agree or disagree with the following statement taken from the American National Election Study (2012): “Sometimes politics and government seem so complicated that a person like me can’t really understand what’s going on.” Since the publication of Likert’s (1932) well-known article on the measurement of attitudes, this technique has become increasingly popular over recent decades.¹ Major national and international surveys, such as the American National Election Study (ANES), the Eurobarometer, and the International Social Survey Program (ISSP), have made use of the A/D question format. As a consequence, many empirical findings in the social sciences are based on A/D questions. According to the catalog of rating scales from Robinson, Shaver, and Wrightsman (1999), approximately 81% of the empirical findings in social science research are based on A/D questions. The reasons for the popularity of this question format among survey researchers are twofold (Saris, Revilla, Krosnick, and Shaeffer 2010): First, A/D questions allow researchers to ask about several unrelated topics (e.g., social inequality and xenophobia) using the same response scale for all questions. Second, A/D questions save space in self-administered surveys and time in both self- and interviewer-administered surveys, particularly if grids are used.

Despite these methodological benefits, the survey literature discusses several problems associated with A/D questions (for a detailed literature review, see Höhne, Schlosser, and Krebs 2017 or Saris et al. 2010). Various researchers argue that item-specific (IS) questions (i.e., response categories address the underlying content dimension directly) are easier for respondents to answer and, thus, produce higher quality data (Fowler 1995; Fowler and Cosenza 2008; Krosnick and Presser 2010; Revilla, Saris, and Krosnick 2013; Saris et al. 2010). For example, the statement on political efficacy above could also be

¹ In this context, it must be mentioned that Likert (1932) tested five-point, fully labeled “approval/disapproval” response scales.

asked in the IS question format as follows: “How often do politics and government seem so complicated that you can’t really understand what’s going on? Always, most of the time, about half of the time, some of the time, or never?” According to Fowler (1995), as well as Krosnick and Presser (2010), the IS format represents a simpler, more direct, and more informative way of asking survey questions than the A/D format.

One major problem associated with the A/D format is that responding to such questions inherently involves more cognitive complexity than responding to IS questions. Answering A/D questions requires respondents to accomplish multiple specific mental tasks (Carpenter and Just 1975; Fowler 1995; Fowler and Cosenza 2008; Höhne et al. 2017; Revilla et al. 2013; Saris et al. 2010). For instance, when responding to the A/D statement “I am interested in politics,” respondents are required to perform several tasks. They must (1) comprehend the literal meaning of the question (i.e., what the individual words mean), (2) identify the underlying dimension of interest to the researcher (e.g., the in-tensity of interest in politics), (3) place themselves on that dimension (i.e., to what degree are they interested in politics?), (4) calculate where, on that dimension, the stem of the question lies (i.e., decide where “interested” lies on a continuum ranging from “very interested” to “not at all interested”), (5) evaluate the distance between their own position on the dimension and the position of the question stem, and finally, (6) translate this judgement into the A/D response categories. In the IS question format, mapping an answer to the response scale (6) is undoubtedly less difficult, because the response categories match the underlying dimension of interest. In addition, performing task 5 is usually not required when answering IS questions. Hence, it is indeed reasonable to assume that answering questions in the IS format is a less complex endeavor than answering questions in the A/D format.

A second serious problem associated with A/D questions is that many studies have shown that these questions tend to produce response bias², such as acquiescence (Baumgartner and Steenkamp 2001; Converse and Presser 1986; Holbrook 2008; Krosnick 1991; Krosnick, Narayan, and Smith 1996; Saris et al. 2010; Schuman and Presser 1981; van Vaerenbergh and Thomas 2013). Although there are several theoretical explanations for the occurrence of response bias in A/D questions, the reason why this question format causes such bias has not been conclusively identified. The satisficing theory, however, provides a convincing argument, because it posits that respondents are not always willing or motivated to expend the effort needed to answer a survey question optimally (Krosnick 1991; Krosnick and Alwin 1987). Instead, they try to shortcut the cognitive response process, for example, by agreeing with statements presented to them in the A/D question format (Krosnick 1991). A/D questions seem to promote a superficial cognitive response process, due to an

² The term response bias refers to a systematic distortion of the cognitive response process, due to the design of survey instruments or interview settings (Groves et al. 2004).

invariant form of the questions (i.e., employing the same response scale for all questions), which forces respondents to perform the same answering task repeatedly (Höhne and Krebs 2017; Höhne et al. 2017). Furthermore, answers to A/D statements do not refer directly to the underlying dimension of interest (e.g., “interest in politics”). This implies an indirect manner of asking, which additionally impedes responding to A/D questions. By contrast, IS questions ask directly about the content dimension and employ response scales tailored to the specific dimension. The use of different response scales for different content dimensions reduces the possible redundancies and boredom associated with the A/D question format.

Indeed, from a psychological perspective, A/D questions require intricate and sophisticated cognitive processing. However, respondents do not have to read the response categories repeatedly because the manner of asking does not change across questions, and they are able to mentally extrapolate the A/D response continuum (Höhne and Lenzner 2015), which encourages a superficial answering process. Although simpler to process, IS questions require permanent reconsideration of the underlying content dimension, inciting respondents to engage in a more active and more intensive response process for each question, in terms of its specific content. Therefore, the processing of IS questions entails a greater consideration of responses than the processing of A/D questions.

In the present study, we investigate whether the cognitive processing of A/D and IS questions differs and whether answering questions in one format is, in-deed, more effortful than in the other. In contrast to previous studies comparing A/D and IS questions, which reported somewhat mixed empirical findings (see Hanson 2015; Höhne and Krebs 2017; Höhne et al. 2017; Kuru and Pasek 2016; Lelkes and Weiss 2015; Liu, Lee, and Conrad 2015; Saris et al. 2010; Scherpenzeel and Saris 1997; Schuman and Presser 1981), we use eye-tracking methodology to investigate our research question. While reading questionnaire instructions, question stems, and response categories, respondents’ eye movements are captured by infrared cameras. These cameras record respondents’ exact eye location, as well as the number, duration, and order of their fixations. Eye tracking thus enables researchers to directly investigate hypotheses about response processes and respondent behavior (Galesic and Yan 2011). For instance, both Galesic, Tourangeau, Couper, and Conrad (2008) and Höhne and Lenzner (2015) investigated the occurrence and causes of response order effects in survey responding. Kamoen, Holleman, Mak, Sanders, and van den Bergh (2011) examined the cognitive burden of answering contrastive survey questions, whilst Menold, Kaczmirek, Lenzner, and Neusar (2014) analyzed the influence of scale length and scale labeling based on the attention that (verbal) labels received. These studies confirm that eye tracking is a useful methodological approach to investigate response behavior and information processing in surveys.

2. RESEARCH HYPOTHESES

If IS questions do, in fact, promote more conscientious responding than A/D questions, due to a more active and more intensive cognitive response process, then this should manifest itself in the eye-tracking data in the form of higher fixation counts, longer fixation times, and more re-fixations. Fixation count is defined as the total number of fixations on a specific area of interest (e.g., the question stem), including re-readings. Fixation time is defined as the total duration of fixations on a specific area of interest (e.g., the question stem), again including re-readings. Re-fixations of response categories are defined as the total number of response categories that respondents re-fixate (i.e., fixate again after reading at least one other response category). These eye-tracking parameters—fixation count, fixation time, and re-fixations—have been proven to be good indicators of cognitive effort in responding to survey questions (Galesic et al. 2008; Galesic and Yan 2011; Höhne and Lenzner 2015; Kamoen et al. 2011; Lenzner, Kaczmirek, and Galesic 2014; Menold et al. 2014). Furthermore, Lenzner (2012) was able to show that several linguistic text features impede question understanding; these were detected by means of two of the eye-tracking parameters mentioned above (fixation count and fixation time; Lenzner, Kaczmirek, and Galesic 2011), and they affect data quality (e.g., amount of non-substantive responses and response consistency). Hence, these parameters seem to be good predictors of survey data quality.

Our reasoning is based on two basic assumptions about the relationship between eye fixations and cognitive processing (Just and Carpenter 1980): First, the immediacy assumption posits that words or objects that are fixated by the eyes are processed directly; their interpretation is not deferred. Second, the eye-mind assumption posits that the eyes remain fixated on a word or an object as long as it is being processed. Taken together, these assumptions postulate that there is a close connection between fixation behavior and mental processing: the fixation count and time spent on a word or an object are, approximately, equal to the count and time required for processing it. Adopting these two assumptions, we investigate whether answering IS questions is indeed characterized by higher fixation counts, longer fixation times, and more re-fixations, indicating a more conscientious response process.

Under optimal conditions, we would expect that respondents actually carry out all six of the mental tasks described above when answering A/D questions. This, in turn, should show up in the eye-tracking data in the form of higher fixation counts, longer fixation times, and more re-fixations for the A/D format, because the question processing is cognitively more complex than for the IS format. However, assuming that this kind of optimal responding only occurs rarely and that the invariant and indirect manner of asking A/D questions promotes a superficial response process, we expect, instead, more fixations and re-fixations, as well as longer fixations in the IS question format.

Based on our reasoning described above, we postulate the following three hypotheses: First, respondents fixate more frequently and longer on the question stems and the response categories when answering IS questions than when answering A/D questions (hypothesis 1). Second, respondents read more response categories in the IS question format than in the A/D question format (hypothesis 2). Finally, respondents show more re-fixations between the response categories when answering IS questions compared to A/D questions; i.e., they re-read response categories they have read previously (hypothesis 3).

3. METHOD

3.1 Study Design

We conducted an eye-tracking experiment to investigate the cognitive processing of A/D and IS questions during completion of a web survey. Respondents were randomly assigned to one of two experimental groups. The first group ($n = 44$) received three individual A/D questions with a five-point, fully labeled response scale (agree/disagree condition). The second group ($n = 40$) received three individual IS questions with a five-point, fully labeled response scale (item-specific condition).

3.2 Survey Questions

The three questions used were adapted from the European Social Survey (2008), as well as the International Social Survey Program (2004), and dealt with political issues, such as political interest. For each IS question adapted from these surveys, we developed an A/D counterpart that preserved the question's content as much as possible.³ The questions were designed in German, which was the mother tongue of 93% of the participants. Respondents answered both A/D and IS questions on five-point, fully labeled response scales with a vertical arrangement of the response categories (see Appendix for details about the questions used).⁴

³ Strictly speaking, in the German questionnaire, both question formats are based on unipolar response scales (except the third IS question), which is the most common way to ask A/D questions in German (see, for instance, the German ISSP questionnaires). In his pioneering study, Rohrmann (1978) also demonstrated that both types of the German A/D scales do not differ regarding equidistance.

⁴ The order of the questions was not randomized and, thus, the occurrence of question order effects cannot be precluded with certainty. However, the order was the same in both experimental conditions.

3.3 Participants

In total, 84 participants took part in the experiment. Due to technical difficulties, the eye movements of 2 respondents could not be recorded accurately, and the recorded eye fixations of 7 were not satisfactory because there was a systematic shift to the line below or above the one that was fixated. These participants were excluded from the data, leaving 75 in the analyses. The respondents were between 17 and 76 years old, with a mean age of 35.7 (SD = 14.6), and 53% of them were female. 20% had graduated from a lower secondary school, 12% from an intermediate secondary school, and 68% from a college preparatory secondary school or university. The great majority used a computer and the internet every day or almost every day (89% and 88%, respectively), and 81% had participated in at least one web survey prior to this study. To evaluate possible differences in the sample composition between the two experimental groups, we additionally conducted χ^2 tests. The results showed no statistically significant differences regarding the following socio-demographic characteristics: age [$\chi^2(2) = 2.23$, $p = 0.33$], gender [$\chi^2(1) = 2.26$, $p = 0.13$], education [$\chi^2(2) = 0.76$, $p = 0.68$], computer usage [$\chi^2(1) = 0.26$, $p = 0.61$], internet use-age [$\chi^2(1) = 0.64$, $p = 0.42$], and survey experience [$\chi^2(1) = 0.59$, $p = 0.44$].

3.4 Eye-Tracking Equipment

Participants' eye movements were recorded by a Tobii T120 Eye Tracker, and the data was analyzed with the Tobii Studio 3.2.1 software. The Tobii T120 is a remote eye tracker embedded in a 17" TFT monitor (resolution 1280 × 1024), with two binocular infrared cameras located underneath the computer screen. The system is accurate within 0.5° with less than 0.3° drift over time and permits head movements within a range of 30 × 22 × 30 centimeters. Eye movements were recorded at a sampling rate of 120 hertz. The online questionnaire was programmed with a font size of 18 and 16 pixels and double-spaced text with a line height of 40 and 32 pixels for the question text and response categories, respectively. Before analyzing the eye-tracking data, we applied Tobii Studio's I-VT fixation filter in the default setting (gap fill-in: enabled, 75 milliseconds; eye selection: average; noise reduction: disabled; velocity calculator window length: 20 milliseconds; I-VT classifier: 30°/s; merge adjacent fixations: enabled, max time between fixations: 75 milliseconds, max. angle between fixations: 0.5°; discard short fixations: enabled, minimum fixation duration: 60 milliseconds) to identify "true" fixations in the raw data.⁵ In a sensitivity check, we repeated the analyses of the fixation counts and times on the

⁵ The Tobii I-VT filter is an update of older fixation filters and allows for more sophisticated data cleaning. The default values were selected to provide the best possible fixation classification across recordings with different levels of noise. Detailed descriptions of the general principles behind the I-VT fixation filter can be found in Tobii Technology (2012).

question stems and response categories using Tobii's ClearView fixation filter that was set to include only fixations that lasted at least 100 milliseconds and encompassed 20 pixels. The results were similar to those we obtained when applying the I-VT filter in the default setting, and all of our conclusions remained unchanged.

3.5 Procedures

The study was conducted at GESIS—Leibniz Institute for the Social Sciences in Mannheim, Germany, in October and November of 2012 and was part of a larger study with several unrelated experiments (Höhne and Lenzner 2015; Lenzner et al. 2014). All experiments were independently randomized to re-duce the possibility of any systematic carryover effects. One test session lasted approximately 90 minutes, 30 minutes of which were devoted to eye tracking and 60 minutes to cognitive interviewing. The present experiment was embedded in a web questionnaire that participants completed after taking part in a cognitive interview during the second half of the test session.

After participating in the cognitive interview, participants were seated in front of the eye tracker so that their eyes were approximately 60 centimeters from the computer screen. After completing a standardized calibration procedure (during which they were asked to follow a moving red dot on the screen with their eyes), they started the web survey. The calibration procedure was carried out by an experimenter who also oversaw the experiment from a separate observer room next to the laboratory. The experimenter monitored respondents' eye movements on a computer screen in real time. Respondents were instructed to read at a normal pace while trying to understand the questions as well as they could. Only one question at a time was displayed on the screen, and the questions were written in black text against a white background. At the beginning, all participants answered the same two questions, which were used to calculate their individual fixation rate, reading rate, and re-fixation rate⁶ (these parameters were used as covariates in the statistical analyses to control for inter-individual differences). The whole questionnaire took about 12 minutes to complete. For their participation in the whole study (including the cognitive interview), respondents received a compensation of €30.

3.6 Analytical Strategies

The results of this experimental study will be reported as follows: we first look separately at the fixation counts and times for the question stems and response

⁶ Fixation rate refers to the average number of fixations on these questions, reading rate refers to the average fixation time on these questions, and re-fixation rate refers to the average number of response categories that were re-fixated in these questions.

categories to investigate whether these two question parts affect the process of responding differently. Afterwards, we investigate the number of response categories that were read and re-fixated. Because there are no substantial differences at the question level, and to reduce the number of the subsequent statistical procedures as well as to efficiently summarize the results, we conducted our analyses on the means of the three A/D and IS questions over all respondents.

Due to technical limitations, the number of response categories read and the number of re-fixations could not be detected by the Tobii eye-tracking system, so that the questions had to be coded by two coders, each of whom coded the eye movements of one half of the respondents ($n = 1/42$). In addition, the eye movements of a randomly selected subset of 10% of the respondents ($n = 1/48$) were coded by both coders for the purpose of estimating reliability. Inter rater agreement was excellent (see Fleiss, Levin, and Paik 2003), with an Intraclass Correlation Coefficient (ICC) of 0.95. Discrepancies between the two ratings were examined and discussed with the second author until a consensus was reached.

The question stems and response categories of the A/D and IS questions differed in the number of words; this was necessary to avoid formulating survey questions that sounded artificial. In accordance with Ferreira and Clifton (1986), we corrected for length differences of question stems and response categories between the two question formats (A/D versus IS) by dividing fixation count, fixation time, and re-fixations by the number of characters. Hence, fixation count and time for the question stems and response categories, as well as re-fixations for the response categories per character, are reported in our results.

4. RESULTS

In order to determine the cognitive effort associated with A/D and IS questions, we employed general linear models for the question stems and response categories and used fixation rate, reading rate, and re-fixation rate as covariates to control for inter-individual differences in respondents' reading and response behavior.

4.1 Fixation Count and Fixation Time

In line with our first hypothesis, table 1 shows that IS questions indeed lead to (on average) a higher fixation number than their A/D counterparts, as well as longer fixation times. However, there are substantial differences between question stems and response categories: whereas for question stems, we cannot find any mean differences between the two question formats, we observe large mean differences regarding fixation count and time for response categories. With respect to the question stems, there are no significant differences between the A/D and IS questions for fixation count [$F(1,72) = 0.03, p = 0.87, \text{partial } \eta^2 = 0.00$],

Table 1. Means and Standard Errors (in Parentheses) of Fixation Count and Time per Character for Question Stems and Response Categories of A/D and IS Questions

Eye-tracking parameter	Question part	Agree/disagree (A/D)	Item-specific (IS)
Fixation count	Question stems	0.22 (0.01)	0.22 (0.02)
	Response categories	0.16 (0.02)	0.29 (0.02)
Fixation time (sec)	Question stems	0.04 (0.00)	0.04 (0.00)
	Response categories	0.05 (0.01)	0.10 (0.01)

NOTE. —The table displays estimated marginal means after controlling for the covariates fixation rate and reading rate. To control for length differences in question stems and response categories between the two question formats, we divided the two eye-tracking parameters by the number of characters (see Ferreira and Clifton 1986).

and fixation time [$F(1,72) = 0.08, p = 0.78, \text{partial } \eta^2 = 0.00$]. In contrast, for the response categories, there are highly significant differences between the two question formats for fixation count [$F(1,72) = 21.49, p < 0.001, \text{partial } \eta^2 = 0.23$], and fixation time [$F(1,72) = 13.97, p < 0.001, \text{partial } \eta^2 = 0.16$].

Furthermore, to identify differences in the allocation of attention to question stems and to response categories, we subtracted the fixation count and time on the response categories from the fixation count and time on the question stems within the A/D and IS question format groups, respectively. Then we conducted analyses of variance on these fixation count and time differences between the A/D and IS question format groups. The results show highly significant differences for fixation count [$F(1,73) = 15.93, p < 0.001, \text{partial } \eta^2 = 0.18$] and fixation time [$F(1,73) = 11.50, p < 0.01, \text{partial } \eta^2 = 0.14$] between the two experimental groups; respondents expend much more cognitive effort when processing the response categories of IS questions than A/D questions compared to the respective question stems. This is also suggested by the estimated marginal means in table 1.

4.2 Response Categories Read and Re-Fixated

With respect to our second hypothesis, we compared the average number of response categories read in both question formats. Table 2 shows that, contrary to our expectations, respondents do not read more response categories in the IS question format than in the A/D question format. An analysis of variance of

Table 2. Means and Standard Errors (in Parentheses) of Response Categories Read and Re-fixated (per Character) in the A/D and IS Question Format

Eye-tracking parameter	Agree/disagree (A/D)	Item-specific (IS)
No. of response categories read	3.10	3.06
	(0.12)	(0.13)
No. of response categories re-fixated	0.02	0.07
	(0.01)	(0.01)

NOTE. —The table displays the response categories read (on average), as well as the estimated marginal means for the number of response categories re-fixated after controlling for the covariate re-fixation rate. To control for length differences in response categories between the two question formats, we divided the number of response categories re-fixated by the number of characters (see Ferreira and Clifton 1986).

the means of the three questions revealed no significant differences in the number of response categories read between the two experimental groups [$F(1,73) = 0.04$, $p = 0.85$, partial $\eta^2 = 0.00$]. Figure 1 presents six gaze plots of different respondents for the three questions on political issues for both experimental groups (A/D versus IS). Gaze plots display the scan path (i.e., sequence of the eye movements) across visual stimuli, in this case across the question stems and response categories of A/D and IS questions. The circles denote fixations, and the lines between them denote saccades. The size of the circles is proportional to the duration of the fixation itself. The gaze plots reveal that, in both groups, the center of the response scales (i.e., the middle categories) was fixated most intensively. This finding is in line with the response distributions and demonstrates that respondents most frequently selected the middle response category. This, in turn, indicates a relationship between the intensity of looking at a specific area of the response scale and selecting a response category from this area (see Höhne and Lenzner 2015). Furthermore, the respondents did not fixate on the last response categories at the bottom of the response scales and thus did not read all response categories, irrespective of the question format. Respondents selected these response categories less frequently. Also, the IS response categories, in particular, were fixated more intensively than their A/D counterparts. More precisely, it is evident that respondents have more and longer fixations, as well as more re-fixations, when answering IS questions, compared to A/D questions. These findings correspond to the results presented in tables 1 and 2. Our third hypothesis postulates that respondents re-fixate the response categories more often when answering IS questions compared to A/D questions, indicating more conscientious processing of the underlying response categories. In line with our expectations, table 2 shows considerable differences in the mean number of re-fixations between the A/D and IS question formats

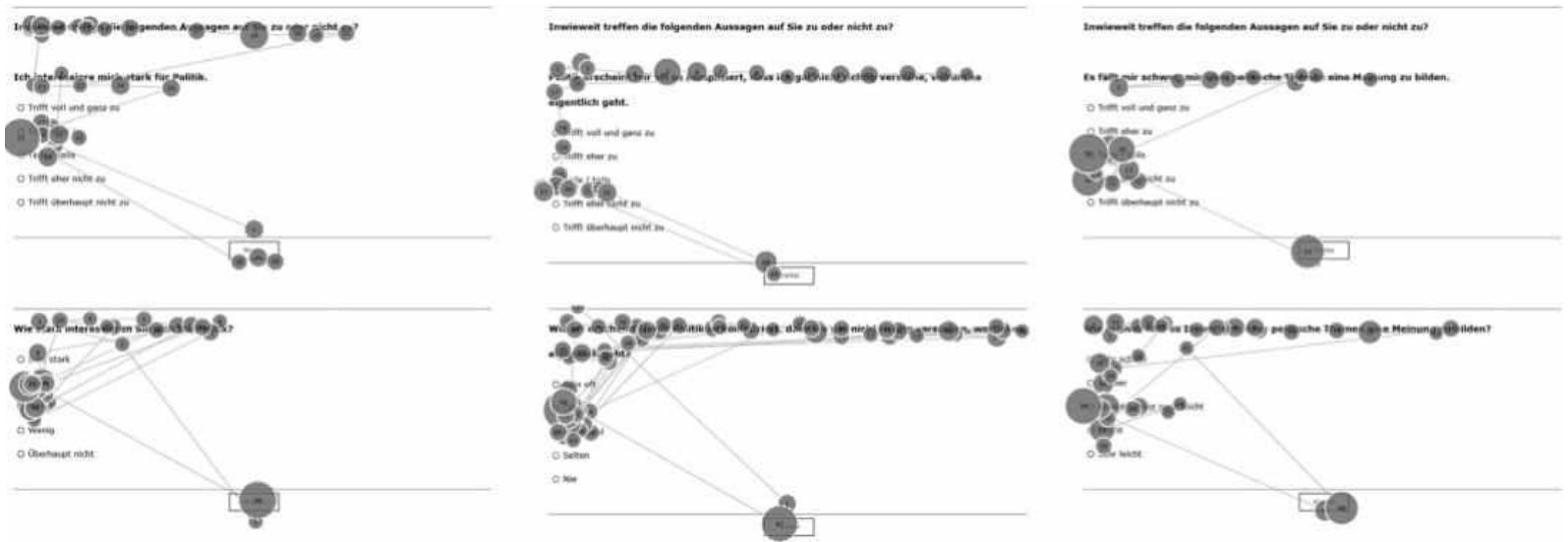


Figure 1. Gaze Plots of Different Respondents for the Three A/D and IS Questions

NOTE. — The three upper gaze plots correspond to the first experimental group (agree/disagree condition), and the three lower gaze plots correspond to the second experimental group (item-specific condition). Each gaze plot displays the eye movements of one respondent. The circles indicate fixations, and the lines between the circles indicate saccades. The numbers within the circles indicate the order of the fixations, and the size of the circles is proportional to the fixation time.

Our statistical analyses yield highly significant differences between the two experimental groups [$F(1,72) = 11.43, p < 0.001, \text{partial } \eta^2 = 0.14$]. Hence, it again appears that respondents process the IS response categories much more intensively than the A/D response categories.

5. DISCUSSION AND CONCLUSION

The aim of this eye-tracking study was to investigate the processing of A/D and IS questions, in order to evaluate the cognitive effort associated with these two question formats. Our results show no differences between the A/D and IS questions in respect to respondents' fixation on the question stems. Considering the characteristics of A/D and IS questions in general, and the questions tested in particular (see Appendix for details about the questions), it is obvious that there is no substantial semantic and/or syntactic difference between them (for a general overview of question comprehensibility, see Graesser et al. 2006). The main difference between these two question formats is that the stems are formulated either as statements (A/D question format) or as "real" questions (IS question format).

Regarding the processing of the response categories, however, our results show that the processing of IS questions is characterized by more conscientious responding than the processing of A/D questions. This is indicated by the fact that the response categories of the IS questions are fixated more frequently and longer than those of the A/D questions. In addition, the response categories of IS questions are more frequently re-fixated than their A/D counterparts. These findings support the notion of asking manner postulated by Höhne et al. (2017). According to this reasoning, A/D questions—though theoretically requiring intricate and sophisticated cognitive processing—promote a state of boredom and weariness and, therefore, superficial cognitive processing. IS questions, in contrast, usually change the manner of asking and, thus, presumably demand and encourage a more conscientious consideration of the response categories.

Contrary to our expectations, we did not find any significant differences between A/D and IS questions related to the number of response categories read. The gaze plots presented in figure 1 reveal that respondents in both question formats do not fixate (and therefore do not read) all response categories, but relatively frequently skip the last response categories. Similar to the findings of Höhne and Lenzner (2015) in their study of A/D questions, respondents also seem to be able to extrapolate the response continuum of IS questions after reading the initial response categories. This circumstance might be attributed to the fact that the response categories in both A/D and IS questions form rating scales, which follow an ordered and closed response continuum. This implies that respondents do not have to fixate all response categories. Overall, there are two limitations to this study. First, our experiment does not investigate the data quality obtained from A/D and IS questions. Although

we found evidence that IS questions trigger a more considerate response process than A/D questions, it remains unclear whether this has also a positive effect on the reliability and validity of respondents' answers. Saris and his colleagues (2010), for instance, found that responses to IS questions are of a higher quality than responses to comparable A/D questions. Nevertheless, further research is necessary in order to systematically evaluate the connection between the cognitive processing of both question formats and their data quality. Second, our experimental study only investigated IS questions that differed in the manner of asking (i.e., employing different response categories). However, IS questions do not necessarily have to change the manner of asking if all questions deal with the same dimension of interest (e.g., frequency and intensity). Hence, it is yet unclear whether the crucial difference between A/D and IS questions is the repetition of the response categories or the directness of the question format—or both. In an attempt to explore this issue further, and as suggested by one of the anonymous reviewers, we compared the fixations on A/D and IS questions for each of the three experimental questions. If it is correct that the crucial difference between the two question formats is the repetition of response categories, then, for the first question, the differences in fixation times and counts should be smaller than for the subsequent questions or even non-existent. Our analysis did not reveal substantial differences between the three experimental questions with respect to means and effect sizes. This might indicate that not only the continuity (i.e., repetitiveness of response categories) but also the directness (i.e., addressing the content dimension) of the manner of asking affects the processing of the questions. This, however, is only a hypothetical explanation and lacks empirical evidence. It is also possible—again, as suggested by the anonymous reviewer—that we did not find these differences between our experimental questions because our participants were skilled survey respondents (i.e., 81% had participated previously in at least one web survey) who might have been familiar with the A/D question format and recognized it from their past experience. Therefore, future research is needed to investigate whether the continuity and/or directness of the question format is responsible for the differences in processing A/D and IS questions.

The empirical findings of this study have theoretical and practical implications for social science research. From a theoretical point of view, we found evidence supporting the notion that the asking manner (Höhne et al. 2017) affects the process of responding to A/D and IS questions. More precisely, this implies that a continuous and indirect manner of asking survey questions, as is the case with the A/D question format, negatively affects diligence and thoughtfulness in responding. Furthermore, we have demonstrated that there are substantial differences between the presumed cognitive complexity of question formats and the cognitive effort expended in responding. Therefore, we argue that the notion of asking manner should be taken into consideration in future studies to obtain a better understanding of how respondents process

A/D and IS questions. From a practical perspective, our data indicates that IS questions are characterized by a more active and intensive response process than A/D questions. Given that this finding is compatible with earlier research attesting higher data quality to IS questions than to A/D questions (see Saris et al. 2010), we encourage survey researchers and practitioners to give preference to IS questions over A/D questions when developing survey instruments.

Appendix. Question Stems and Response Categories for Baseline Speed (BS), Agree/Disagree (A/D), and Item-Specific (IS) Questions (English translation of the original German questions)

Questions to compute baseline speed (covariate)

BS 1: How successful do you think the government is nowadays in dealing with threats to Germany's security?

Very successful, quite successful, neither successful nor unsuccessful, quite unsuccessful, very unsuccessful

BS 2: And how successful do you think the government is nowadays in fighting unemployment?

Very successful, quite successful, neither successful nor unsuccessful, quite unsuccessful, very unsuccessful

Agree/Disagree Questions

A/D 1: I am very interested in politics.

A/D 2: Politics very often seem so complicated that I can't really understand what is going on.

A/D 3: I find it very difficult to make my mind up about political issues.

Response categories to A/D 1 – A/D 3 are agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly

Item-Specific Questions

IS 1: How interested would you say you are in politics?

Very interested, fairly interested, somewhat interested, hardly interested, not at all interested

IS 2: How often does politics seem so complicated that you can't really understand what is going on?

Very often, often, sometimes, rarely, never

IS 3: How difficult or easy do you find it to make your mind up about political issues?

Very difficult, difficult, neither difficult nor easy, easy, very easy

The order of the questions as well as response categories correspond to the presentation in the Appendix. The response categories of all questions were presented vertically below the question stem. The original German wordings of the questions are available from the first author upon request.

References

- Baumgartner, H., and J. B. Steenkamp (2001), "Response Styles in Marketing Research: A Cross-National Investigation," *Journal of Marketing Research*, 38, 143–156.
- Carpenter, P. A., and M. A. Just (1975), "Sentence Comprehension: A Psycholinguistic Processing Model of Verification," *Psychological Review*, 82, 45–73.
- Converse, J. M., and S. Presser (1986), *Survey Questions: Handcrafting the Standardized Questionnaire*, Beverly Hills, CA: Sage.
- Ferreira, F., and C. Clifton (1986), "The Independence of Syntactic Processing," *Journal of Memory and Language*, 25, 348–368.
- Fleiss, J. L., B. Levin, and M. C. Paik (2003), *Statistical Methods for Rates and Proportions*, Hoboken, NJ: Wiley.
- Fowler, F. J. (1995), *Improving Survey Questions*, Thousand Oaks, CA: Sage.
- Fowler, F. J., and C. Cosenza (2008), "Writing Effective Questionnaires," in *International Handbook of Survey Methodology*, eds. E. de Leeuw, J. Hox, and D. A. Dillman, pp. 136–160, New York, NY: Erlbaum.
- 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*, eds. M. Das, P. Ester, and L. Kaczmarek, 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.
- Groves, R. M., F. L. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2004), *Survey Methodology*, Hoboken, NJ: John Wiley & Sons.
- Hanson, T. (2015), "Comparing Agreement and Item-Specific Response Scales: Results from an Experiment," *Social Research Practice*, 1, 17–26.
- Höhne, J. K., and D. Krebs (2017), "Scale Direction Effects in Agree/Disagree and Item-Specific Questions: A Comparison of Question Formats," *International Journal of Social Research Methodology*, DOI: 10.1080/13645579.2017.1325566.
- Höhne, J. K., and T. Lenzner (2015), "Investigating Response Order Effects in Web Surveys using Eye Tracking," *Psihologija*, 48, 361–377.
- Höhne, J. K., S. Schlosser, and D. Krebs (2017), "Investigating Cognitive Effort and Response Quality of Question Formats in Web Surveys Using Paradata," *Field Methods*, DOI: 10.1177/1525822X17710640.
- Holbrook, A. L. (2008), "Acquiescence Response Bias," in *Encyclopedia of Survey Research Methods*, ed. P. J. Lavrakas, pp. 34, London, UK: Sage.
- 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.
- Krosnick, J. A. (1991), "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology*, 5, 213–236.
- Krosnick, J. A., and D. F. Alwin (1987), "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement," *Public Opinion Quarterly*, 51, 201–219.
- Krosnick, J. A., S. Narayan, and W. R. Smith (1996), "Satisficing in Surveys: Initial Evidence," in *New Directions for Evaluation: Advances in Survey Research*, eds. M. T. Braverman, and J. K. Slater, pp. 29–44, San Francisco, CA: Jossey-Bass.
- Krosnick, J. A., and S. Presser (2010), "Question and Questionnaire Design," in *Handbook of Survey Research*, eds. P. V. Marsden, and J. D. Wright, pp. 263–313, Bingley, UK: Emerald.
- Kuru, O., and J. Pasek (2016), "Improving Social Media Measurement in Surveys: Avoiding Acquiescence Bias in Facebook Research," *Computers in Human Behavior*, 57, 82–92.

- Lelkes, Y., and R. Weiss (2015), "Much ado about Acquiescence: The relative Validity and Reliability of Construct-Specific and Agree-Disagree Questions," *Research and Politics*, 1–8. DOI: 10.1177/2053168015604173
- 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.
- Likert, R. (1932), "A Technique for the Measurement of Attitudes," *Archives of Psychology*, 140, 1–55.
- Liu, M., S. Lee, and F. G. Conrad (2015), "Comparing Extreme Response Styles between Agree-Disagree and Item-Specific Scales," *Public Opinion Quarterly*, 79, 952–975.
- Menold, N., L. Kaczmirek, T. Lenzner, and A. Neusar (2014), "How Do Respondents Attend to Verbal Labels in Rating Scales?" *Field Methods*, 26, 21–39.
- Revilla, M., W. E. Saris, and J. A. Krosnick (2013), "Choosing the Number of Categories in Agree-Disagree Scales," *Sociological Methods & Research*, 43, 73–97.
- Robinson, J. P., P. R. Shaver, and L. S. Wrightsman (1999), *Measures of Political Attitudes*, San Diego, CA: Academic Press.
- Rohrmann, B. (1978), "Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung," *Zeitschrift Für Sozialpsychologie*, 9, 222–245.
- Saris, W. E., M. Revilla, J. A. Krosnick, and E. M. Shaeffer (2010), "Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options," *Survey Research Methods*, 4, 61–79.
- Scherpenzeel, A., and W. E. Saris (1997), "The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies," *Sociological Method and Research*, 25, 341–383.
- Schuman, H., and S. Presser (1981), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, Thousand Oaks, CA: Sage.
- Tobii Technology (2012), "Determining the Tobii I-VT Fixation Filter's Default Values," Available at <http://www.tobii.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/determining-the-tobii-pro-i-vt-fixation-filters-default-values.pdf>.
- van Vaerenbergh, Y., and T. D. Thomas (2013), "Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies," *International Journal of Public Opinion Research*, 25, 195–217.