Eye-Tracking Methodology: Exploring the Processing of Question Formats in Web Surveys

Jan Karem Höhne
University of Mannheim (Germany)
Universitat Pompeu Fabra (Spain)

Abstract

In social research, the use of agree/disagree (A/D) questions is a popular method for measuring attitudes. Research has shown that A/D questions require complex cognitive processing and are susceptible to response bias. Thus, some researchers recommend the use of item-specific (IS) questions. This study examines the processing of A/D and IS questions, using eye-tracking methodology. By recording respondents' eye movements, how respondents process survey questions can be evaluated. The results reveal that IS questions cause more and longer fixations. However, this only applies to the response categories. There are no differences regarding the question stems. Altogether, it seems that IS response categories trigger deeper cognitive processing than A/D response categories.

Keywords: agree/disagree questions, item-specific questions, lab experiment, web surveys, rating scales, survey methodology

Introduction

In social research, the agree/disagree (A/D) question format is intensively employed and major social surveys, such as the International Social Survey Program (ISSP), make use of it. A/D questions commonly start with a pre-request (e.g., *Do you agree or disagree with the following statement?*), accompanied by an indirect statement (e.g., *I am interested in politics.*), followed by a response scale in which the categories are based on an agreement/disagreement continuum (e.g., *"agree strongly" to "disagree strongly"*). One reason for this popularity is that the format, in principal, allows researchers to measure constructs with identical response scales, which optimizes questionnaire design, especially if the questions are employed in the grid presentation mode (Saris, Revilla, Krosnick, & Shaeffer, 2010).

The literature, however, suggests that A/D questions are associated with serious methodological drawbacks, such as being susceptible to response bias (see Baumgartner & Steenkamp, 2001; Converse & Presser, 1986; Fowler, 1995; Fowler & Cosenza, 2008; Höhne & Lenzner, 2015; Krosnick, 1991; Krosnick & Presser, 2010; Revilla, Saris, & Krosnick, 2013; Saris et al., 2010; Schuman & Presser, 1981). Therefore, many survey researchers recommend the use of the item-specific (IS) question format instead. This format is based on a direct question (e.g., *How interested would you say you are in politics?*) and a tailored response scale that matches the content dimension (e.g., *"very interested" to "not at all interested"*).

This document is a preprint and thus it may differ from the final version: Höhne, Jan K. (2019). Eye-tracking methodology: Exploring the processing of question formats in web surveys. International Journal of Social Research Methodology, 22, 199–206. DOI: 10.1080/13645579.2018.1515533.

It is frequently argued that the A/D question format forces respondents to accomplish complex cognitive tasks (see Carpenter & Just, 1975; Fowler, 1995; Fowler & Cosenza, 2008; Höhne & Lenzner, 2017; Höhne, Schlosser, & Krebs, 2017; Krosnick & Presser, 2010; Saris et al., 2010): Respondents have to discover (1) the semantic and (2) the literal meaning of the A/D statement. Subsequently, they have to (3) locate themselves on the dimension of interest, (4) determine where, on the content dimension, the statement lies, and (5) evaluate the range of their placement and the placement of the statement on the content dimension. Finally, they have to (6) translate their judgment into the A/D response scale. All in all, responding to IS questions seems to be simpler and more direct because the points 2 and 5 are usually superfluous. Furthermore, point 6, in IS questions, seems to be simpler because the response scale is tailored for each individual question.

As already mentioned, many studies have shown that the A/D question format is susceptible to response bias (see Baumgartner & Steenkamp, 2001; Converse & Presser, 1986; Fowler, 1995; Fowler & Cosenza, 2008; Höhne & Lenzner, 2015; Krosnick, 1991; Krosnick & Presser, 2010; Revilla, Saris, & Krosnick, 2013; Saris et al., 2010; Schuman & Presser, 1981). One convincing explanation for its susceptibility to response bias was provided by the satisficing theory (Krosnick, 1991), which argues that respondents are not always motivated or willing to invest the response effort necessary to respond to a question appropriately. In fact, they frequently try to shortcut the response process, which, in turn, results in the occurrence of response bias (Krosnick, 1991). Höhne, Schlosser, and Krebs (2017) also posit that A/D questions support perfunctory processing due to an indirect and repetitive manner of asking. The questions employ indirect statements and non-tailored response scales, forcing respondents to carry out identical response tasks repeatedly. IS questions, by contrast, are based on a direct question and usually vary the manner of asking. Presumably, they do not reduce the attention and/or motivation of respondents and, thus, might produce higher response quality.

Although the A/D question format theoretically demands complex and elaborated processing, respondents have to read the response categories only once. In principle, they are able to mentally extrapolate the A/D response continuum (Höhne & Lenzner, 2015), which promotes perfunctory question processing. The IS question format is, theoretically, simpler to respond to but it usually requires a constant reconsideration of the dimension of interest, which encourages respondents to perform relatively active and intensive processing. Hence, it seems reasonable that responding to IS questions is more well-considered but also more effortful than responding to A/D questions.

Hypotheses

As argued above, the A/D question format seems to promote superficial rather than optimal responding, due to the indirect and unchanging manner of asking. This should reflect itself in respondents' gaze behavior and, thus, in the eye-tracking data; in particular, in a comparatively low fixation number (i.e., the total count of fixations on a region of interest) and a comparatively short fixation time (i.e., the total duration of fixations on a region of interest). The IS question format, by contrast, seems to promote active and intensive rather than superficial responding, due to the direct and changing manner of asking. This should also reflect itself in the eye-tracking data; in particular, in a comparatively high fixation number and a comparatively long

fixation time. Previous research has demonstrated that these two eye-tracking parameters are good indicators of effort in responding (Galesic, Tourangeau, Couper, & Conrad, 2008; Höhne & Lenzner, 2015; Höhne & Lenzner, 2017; Kamoen, Holleman, Mak, Sanders, & van den Bergh, 2017; Lenzner, Kaczmirek, Galesic, 2011).

This argumentation is based on two conjectures about the relationship between eye-tracking data and mental processes (Just & Carpenter, 1980, p. 330): The *immediacy assumption* states that the processing of objects that are fixated is not deferred because it occurs as soon as possible. The *eye-mind assumption* states that a considerable delay between the fixation of an object and its processing does not occur. Hence, it can be assumed that the fixation number and time for an object are similar to the fixation number and time required for processing it.

Two research hypotheses are postulated: First, it is hypothesized that respondents fixate more frequently and longer on the question stems when responding to IS questions than when responding to A/D questions (*Hypothesis 1*). Second, it is hypothesized that respondents fixate more frequently and longer on the response categories when responding to IS questions than when responding to A/D questions (*Hypothesis 2*).

Method

Design

An eye-tracking experiment was conducted to investigate the processing of A/D and IS questions. Participants were randomly assigned to one of two groups: The first group (n = 43) received three A/D questions in a grid (*agree/disagree condition*). The second group (n = 41) received three individual IS questions presented on the same page (*item-specific condition*).

Questions

The three survey questions were adapted from existing social surveys. In each case, both an A/D and IS counterpart that preserved question content as much as possible were developed (see Appendix for the questions used). The questions were designed in German, which was the mother tongue of 93% of the participants. All questions were presented with five-point, completely verbalized response scales. The A/D questions were presented in a grid, which is the predominant way of employing them (Couper, Tourangeau, Conrad, & Zhang, 2013; Saris et al., 2010). The IS questions, by contrast, were presented individually with horizontally arranged response categories below each question. All questions were displayed on the same page.

Respondents

84 participants took part in the experiment. Due to technical difficulties, the eye movements of two participants could not be recorded accurately. The recorded eye fixations of six other participants were not satisfactory because there was a systematic shift in the eye-tracking recordings. These participants were excluded from the data, leaving 76 in the analyses. Participants were between 17 and 76 years old with a mean age of 35.9 (SD = 14.5). 54% of the participants were female. 22% had graduated from a lower secondary school, 11% from an intermediate secondary school, and 67% from a college preparatory secondary school or

university. The majority used a computer and the Internet every day or almost every day (88% and 87%, respectively). 80% had participated in at least one web survey prior to this study.

Procedures

This experiment was conducted at GESIS – Leibniz Institute for the Social Sciences (Germany) in 2012 and was part of a larger study including cognitive interviewing and several independently randomized eye-tracking experiments (see, for instance, Höhne & Lenzner, 2015; Höhne & Lenzner, 2017). Participants were seated in front of the eye-tracking system[1] and completed a standardized calibration procedure (i.e., following a moving dot on the screen with the eyes). After a successful calibration, they began the web survey. The entire experiment was supervised by an experimenter who also observed respondents' eye movements on a computer screen. Participants were asked to read at a normal pace and to try to understand the questions as well as possible. At the beginning of the web survey, two questions were asked to determine the individual fixation and reading rate of respondents (see Appendix for the questions used). Both parameters served as covariates in the subsequent analyses. Completing the entire web survey lasted approximately 12 min and participants received a compensation of €30 for taking part in the entire study.

Results

Question stems and response categories of the A/D and IS questions differed in the number of words. This could only have been avoided by developing artificial questions. To take length differences into account, fixation count and time of all question stems and response categories, respectively, were corrected for length differences by dividing them by the number of characters (see Ferreira & Clifton, 1986).

To evaluate the information processing of the A/D and IS question format, general linear models for the question stems and response categories were calculated. In addition, fixation rate and reading rate were employed as covariates to control for inter-individual differences.[2]
The analyses were conducted for the three aggregated questions. This strategy was adopted because the A/D grid questions could not simply be separated. It also makes it possible to reduce the statistical tests and to efficiently summarize the results.

Fixation Count and Fixation Time

In accordance with the research hypotheses, the analysis focused on whether the three IS questions cause more and longer fixations than the three A/D questions. Table 1 reveals that this is only partially supported by the eye-tracking data, because there are differences with respect to the question stems and response categories. For the question stems, the fixation number [$F(1,73) = 0.81$, $p < 0.37$, partial $\eta^2 = 0.01$] and fixation time [$F(1,73) = 0.13$, $p < 0.72$, partial $\eta^2 = 0.00$] do not significantly differ between the A/D and IS question formats. This

[1] Participants' eye movements were recorded by a Tobii T120 Eye Tracker and the data were analyzed with the Tobii Studio 3.2.1 software (for a more detailed description, see Höhne & Lenzner, 2015; Höhne & Lenzner, 2017).
[2] Whereas fixation rate is defined as the mean number of fixations on these questions, reading rate is defined as the mean time of fixation on these questions.

indicates that both stems are processed similarly. For the response categories, by contrast, the fixation number [$F(1,73) = 16.66$, $p < 0.001$, partial $\eta^2 = 0.19$] and fixation time [$F(1,73) = 7.05$, $p < 0.01$, partial $\eta^2 = 0.09$] differ significantly between the two question formats. These results indicate that the IS response categories are processed more intensively than their A/D counterparts.

Table 1. Means and standard errors (in parentheses) of fixation number 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	0.18	0.20
		(0.01)	(0.01)
	Response Categories	0.10	0.15
		(0.01)	(0.01)
Fixation Time (sec)	Question Stems	0.03	0.03
		(0.00)	(0.00)
	Response Categories	0.03	0.04
		(0.00)	(0.00)

Note. The table reports estimated marginal means after controlling for the covariates fixation rate and reading rate. To control for length differences of question stems and response categories between the two question formats, the eye-tracking parameters were divided by the number of characters.

In addition, the differences in fixation number and time between the question stems and response categories within the A/D and IS group were calculated, respectively, to analyze differences in the attention allocated to both question parts. Based on these differences in means, analyses of variance for fixation number and time between the group with A/D questions and the group with IS questions were conducted. The statistical results reveal marginally significant differences for fixation number [$F(1,74) = 3.43$, $p < .10$, partial $\eta^2 = .04$] and significant differences for fixation time [$F(1,74) = 5.00$, $p < .05$, partial $\eta^2 = .06$]. Hence, respondents seem to invest more effort when processing the response categories of IS questions than of A/D questions, compared to the respective question stems.

To explore the processing of A/D and IS questions, the scan paths of respondents were additionally inspected by means of gaze plots. Figure 1 contains two exemplary gaze plots from two respondents, one of whom answered the three A/D questions presented in a grid, whilst the other respondent answered the three IS questions presented on the same page. Gaze plots depict the order of respondents' eye movements across objects, such as question stems and response categories. The circles indicate fixations and the size of the circles is proportional to the fixation time. The lines indicate saccades (i.e., quick eye movements that initiate fixations).

Closer inspection of Figure 1 reveals that especially the categories at the beginning and/or the middle of the A/D and IS response scales are fixated most intensively. It is also apparent that both respondents do not fixate all response categories, irrespective of the question format. In fact, the response categories at the end of the scales are mostly overlooked. Whereas the respondent in the IS group processed the questions relatively sequentially (i.e., from the question stem to the response categories), the respondent in the A/D group showed relatively many re-fixations between the question stems and response categories.

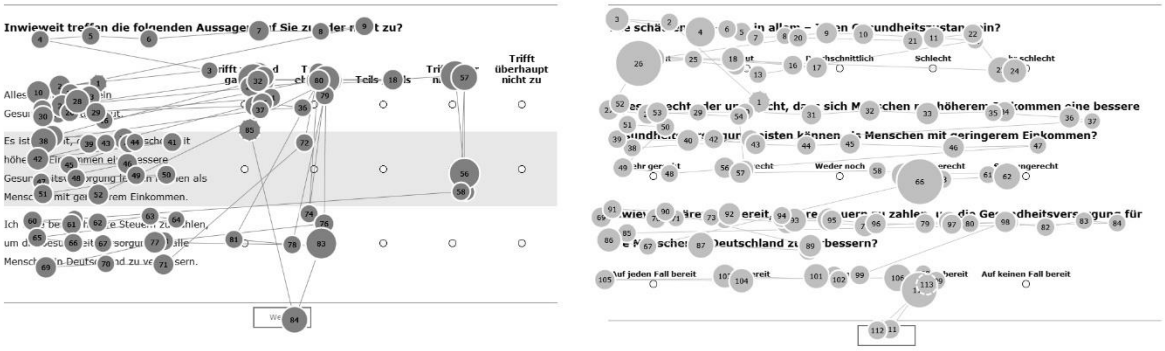

Figure 1. Gaze plots of two respondents for the three A/D and IS questions

Note. The gaze plot on the right side corresponds to the first group (agree/disagree condition) and the gaze plot on the left side corresponds to the second group (item-specific condition). The circles represent fixations and the lines between the circles represent saccades. The numbers in the center of the circles represent the sequence of the fixations and the size of the circles is proportional to the fixation time.

Discussion and Conclusion

The main goal of this eye-tracking experiment was to explore how respondents process A/D and IS questions and to draw conclusions about the respective response effort associated with both question formats. In line with Höhne and Lenzner (2017), the empirical findings indicate that there are no substantial differences with respect to the processing of the A/D and IS question stems. This can thus be considered as an indicator for equality in terms of question stem processing. A closer look at the questions used in this study reveals that they do not differ substantially in terms of semantic and/or syntactic issues (see Appendix for the questions used). In fact, they only differ with respect to the form of wording: Whilst the A/D questions are worded in a declarative form (i.e., as indirect statements), the IS questions are worded in an interrogative form (i.e., as direct questions).

In contrast to the question stems, the A/D and IS response categories differ significantly in terms of fixation number and time. In particular, this indicates that the response categories of the IS questions are more intensively processed than the response categories of the A/D questions. This finding supports the notion of the manner of asking (see Höhne & Lenzner, 2017; Höhne, Schlosser, & Krebs, 2017). Technically, the A/D question format demands complex and elaborated processing. However, its indirect and repetitive manner of asking seem to promote perfunctory responding, which manifests itself in a lower fixation number and time. The IS question format, by contrast, is characterized by a direct and varied manner of asking, which seems to promote intensive processing. This manifests itself in a higher fixation number and time.

Interestingly, the gaze plots show that respondents do not fixate all response categories and, thus, do not read all of them. In fact, they do not consider the response categories at the end of the scale, irrespective of the question format. It seems that respondents usually fixate the categories at the beginning and the center of horizontally arranged response scales (see Höhne & Lenzner, 2015). Consequently, this indicates that respondents seem to be able to mentally extrapolate the response continuum of A/D and IS questions. One explanation for this phenomenon might be that both question formats are based on rating scales. In other words,

they follow an ordered and closed response continuum that, in principle, allows an extrapolation of subsequent response categories.

A further interesting point is that the gaze plots reveal that the IS questions apparently cause more re-fixations between the response categories than the A/D questions. This would additionally indicate a more intensive processing of the response categories.

There are some limitations to this study. First of all, the eye-tracking experiment only explores the processing of A/D and IS questions without considering the quality of responses. Therefore, further research that investigates the respective response effort and the response quality associated with the A/D and IS question formats would be desirable. Second, this experimental study compared A/D questions employed in grid presentation mode with IS questions employed in single presentation mode. However, previous research has shown that questions employed in grids are frequently accompanied by several undesirable outcomes, such as superficial responding and low response quality (Couper et al., 2013). Furthermore, this study only used IS questions that change the manner of asking (i.e., addressing different content dimensions). Therefore, it would be interesting if future studies could also employ IS questions without changing the manner of asking (i.e., addressing the same content dimension, such as intensity or importance).

To conclude: This eye-tracking study provides empirical evidence for the notion of the manner of asking survey questions. It seems that an indirect and invariant manner of asking, as is the case with the A/D question format, promotes relatively superficial responding. In general, this suggests a difference between the theoretically presumed complexity of question formats and the actual effort expended in responding. The findings also indicate that the IS question format seems to encourage respondents to engage in more thoughtful and deliberate responding to each question in the light of its specific content. In line with previous research that points to the superiority of IS questions over A/D questions, in terms of reliability and validity (see Saris et al., 2010), together with the results of this study, it seems wiser to make use of the IS question format when designing questionnaires.

References

Baumgartner, H., & Steenkamp, J. B. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143-156.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82*, 45-73.

Converse, J. M., & Presser S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Beverly Hills, CA: Sage.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review, 31*, 322-345.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Fowler, F. (1995). *Improving Survey Questions: Design and Evaluation.* Thousand Oaks, CA: Sage.

Fowler, F., & Cosenza, C. (2008). Writing effective questions. In E. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 136-160). New York, NY: Taylor & Francis.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*, 892-913.

Höhne, J. K., & Lenzner, T. (2015). Investigating response order effects in web surveys using eye tracking. *Psihologija*, *48*, 361–377.

Höhne, J. K., & Lenzner, T. (2017). New insights on the cognitive processing of agree/disagree and item-specific questions. *Journal of Survey Statistics and Methodology*. DOI: 10.1093/jssam/smx028

Höhne, J. K., Schlosser, S., & Krebs, D. (2017). Investigating cognitive effort and response quality of questions formats in web surveys using paradata. *Field Methods, 29*, 365–382.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kamoen, N., Holleman, B., Mak, P., Sanders, T., & van den Bergh, H. (2017). Why are negative questions difficult to answer? On the processing of linguistic contrasts in surveys. *Public Opinion Quarterly, 81*, 613-635.

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., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (pp. 263-313). Bingley, UK: Emerald.

Lenzner, T., Kaczmirek, L., & Galesic, M. (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.

Revilla, M., Saris, W. E., & Krosnick, J. A. (2013). Choosing the number of categories in agree-disagree scales. *Sociological Methods & Research, 43*, 73-97.

Saris, W. E., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with item-specific response options. *Survey Research Method, 4*, 61-79.

Schuman, H., & Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage.

Appendix

Question stems and response categories for Baseline Speed (BS), Agree/Disagree (A/D), and Item-Specific (IS) questions (the original wordings of the questions are available from the author on request)

Baseline Speed Questions (Covariates)

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

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: All in all, my health is good.

A/D 2: It is fair that people with higher incomes can afford better health care than people with lower incomes.

A/D 3: I am willing to pay higher taxes in order to improve health care for all people in Germany.

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 would you rate your health overall?

Very good, good, neither good nor bad, bad, very bad

IS 2: Is it fair or unfair that people with higher incomes can afford better health care than people with lower incomes?

Very fair, fair, neither fair nor unfair, unfair, very unfair

IS 3: To what extent would you be willing to pay higher taxes to improve health care for all people in Germany?

In any case willing, fairly willing, somewhat willing, hardly willing, not at all willing