

ARTICLES

Which Scale Direction is More Difficult for Respondents to Use? An Eye-tracking Study

Ting Yan¹ 

¹ Westat

Keywords: eye-tracking, cognitive burden, scale direction, questionnaire design

<https://doi.org/10.29115/SP-2023-0015>

Survey Practice

Vol. 16, Issue 1, 2023

Scale direction is found to affect response distributions by yielding more selections of scale points closer to the beginning of the scale. Although this phenomenon has been empirically demonstrated in different modes of data collection for respondents with different demographic characteristics, it is not clear which scale direction is cognitively difficult for respondents to use. Eye-tracking is used to address this research question because it provides a direct window into how respondents process survey questions. I compared dilation and fixation measures across scale direction and found that the satisfaction scale and the frequency scale running from "Never" to "Very Often" had more and longer fixations. In addition, the dissatisfaction options incurred larger dilations, more peak dilations, and more and longer fixations than the other parts of the scale. Findings of this paper shed light on how respondents process response scales and have practical implications for questionnaire design. Furthermore, findings of this paper demonstrate the utility and potential of using eye-tracking in general and dilation measures in particular to understand cognitive burden of answering survey questions.

Introduction

The direction in which a response scale is presented to respondents is found to affect the resultant answers, leading to scale direction effects (Yan and Keusch 2015). Scale direction effects are observed in telephone surveys (e.g., Yan and Keusch 2015), face-to-face surveys (e.g., Carp 1974), web surveys (e.g., Garbarski, Schaeffer, and Dykema 2015; Keusch and Yan 2018), and paper questionnaires (e.g., Höhne and Krebs 2018). Research on scale direction effects demonstrates a stronger impact of scale direction for longer scales than shorter scales (Yan, Keusch, and He 2018), for items asked early in a survey than later in the survey (Carp 1974; Yan, Keusch, and He 2018), and for items that cannot be easily coded as either attitudinal or behavioral than for items that are clearly attitudinal (Yan, Keusch, and He 2018). In addition, respondents with less knowledge of the question topic and speeders who went through the survey too fast tend to show stronger scale direction effects than those with more knowledge (Keusch and Yan 2018; Yan and Keusch 2015).

This paper attempts to examine which scale direction is more difficult and cognitively more burdensome for respondents to process and to use, using eye-tracking technology, for two specific scales. The answer to this question has important implications for questionnaire design. If the findings indicate that a particular direction is more difficult for respondents, question writers should

avoid using that direction. Furthermore, this paper illustrates how eye-tracking and, in particular, pupil dilation measures can be used to investigate cognitive difficulty and burden associated with answering survey questions.

Eye-tracking Methodology

In an eye-tracking study, respondents' eye movements are tracked by infrared cameras while they are reading and answering survey questions. These cameras record where people look and how long they look. As a result, eye-tracking provides a direct window into respondents' survey response process (Galesic and Yan 2011). Fixation count and fixation duration are typical measures of attention and cognitive processing in eye-tracking studies (Galesic and Yan 2011). Survey questions incurring more fixations and longer fixations indicate that these questions have a higher cognitive demand on respondents and require a longer survey response process than questions incurring less and shorter fixations, signaling that these questions are burdensome and difficult to answer. For example, Lenzer and colleagues (2011) showed that text features that reduce readability of survey questions (such as complex syntax, ambiguous nouns) incurred longer fixation times and more fixations. Clifton, Staub, and Rayner (2017) reported longer fixations on difficult words than on easy ones. Kamoen et al. (2011, 2017) combined fixation count and duration with reread—also captured by eye-tracking—and found that negative questions are reread longer and more often, indicating that negative questions pose more burden for respondents.

Task-evoked pupillary responses (TEPRs) captured by eye-tracking can also be used to understand difficulty and cognitive burden of answering survey questions. Typically, pupils begin dilating immediately after a stimulus is presented and constricts after the task is completed either gradually or immediately (Beatty 1982; Eckstein et al. 2017). This process is automatic and involuntary. Larger pupil dilations are found for more difficult tasks across a wide range of tasks such as short-term memory (e.g., Kahneman and Beatty 1966), arithmetic operations (e.g., Hossain and Yeasin 2014), sentence comprehension and document editing (e.g., Just and Carpenter 1993), and visual search (Attar, Schneps, and Pomplun 2016). Despite its great potential as a tool to evaluate cognitive burden of survey items, TERPs are not widely used in the survey context. There are only two studies that empirically explored the use of pupil dilation in evaluating cognitive burden. Neuert (2020) compared pupil dilation, fixation time, and fixation count between problematic versions and improved versions of six survey items, hypothesizing larger dilations, longer fixations, and more fixations for the problematic survey items than for the improved versions. Although dilations were generally larger for the problematic items than for the improved versions, the difference between the two question versions achieved statistical significance at the .10 level for only three items. By contrast, significantly longer durations were found under the problematic condition for all six items at the .10 level. Furthermore, the problematic items induced significantly more fixations than

the improved versions for three items. She concluded that pupil dilation was less effective than the two fixation measures in identifying problematic survey items. Yan et al. (2016) compared average pupil dilations and peak dilations across 24 questions. These questions are classified into “hard” or “easy” questions based on breakoff rates observed in earlier web surveys. Yan et al. (2016) found that hard questions that elicited higher breakoff rates in earlier web surveys elicited larger dilations than easy questions. In addition, questions that respondents self-reported to be more difficult produced larger dilations than questions that had a lower rating on difficulty. Yan et al. (2016) also showed that peak dilations were more likely to occur at response options for hard questions and at question stems for easy questions.

This paper used two dilation measures (average dilation and peak dilation) together with two fixation measures (fixation count and fixation duration) to examine which scale direction is more difficult for survey respondents and to identify specific components of scales posing the greatest burden for respondents. Since survey literature has shown that scale direction plays a crucial role in the formation of answers to attitudinal items but not to behavioral items (Carp 1974; Keusch and Yan 2019), I examined two types of scales in this paper—a satisfaction scale commonly used for attitudinal items and a frequency scale for behavioral items. An experiment was implemented to randomly assign respondents to scales that run either in an ascending order (e.g., from “very dissatisfied” to “very satisfied” for the satisfaction scale and from “never” to “very often” for the frequency scale) or in a descending order (from “very satisfied” to “very dissatisfied” or from “very often” to “often”). Everything else about the scale (e.g., scale labels and number of scale points) as well as the interview setting were held constant across the experimental conditions. Given the experimental set-up, I hypothesize that the scale direction eliciting larger dilations, more fixations, and longer fixations are more difficult for respondents to process and to use than the scale direction incurring smaller dilations, fewer fixations, and shorter fixations.

Data and Methods

Data

Data used in this paper are from an eye-tracking study conducted in January 2016. A total of 20 respondents from the Washington DC metropolitan area were invited to come to the Westat lab in Rockville. Respondents had a mixed background with regard to age, gender, education, and race and ethnicity. Upon arriving in the lab, respondents were asked to complete a web survey while wearing the ASL mobile eye-XG eye-tracking glasses, which record eye-tracking videos. The web survey includes 34 target questions and displays one question per screen (see Figure 1 in the Supplementary Materials for an example of how survey questions are displayed on a computer screen).

A scale direction experiment was embedded in the web survey, varying the direction of two scales—a satisfaction scale and a frequency scale. Both scales are presented vertically on the computer screen. The satisfaction scale is a 5-point fully labeled scale, starting from “very satisfied” to “very dissatisfied” for a random half of the respondents (descending order) and from “very dissatisfied” to “very satisfied” for the other random half (ascending order). Five survey items used this satisfaction scale, asking respondents to rate their satisfactions with health, diet, the grocery store where they shop the most often, their neighborhood, and the city they live in.

The frequency scale is also a 5-point fully labeled scale, ranging from “Very Often” to “Never” for a random half (descending order) and “Never” to “Very Often” for the other half (ascending order). Eight items employed this scale and asked respondents how often in the past 12 months they drank beer, wine or wine coolers, liquor or mixed drinks, coffee (caffeinated or decaffeinated), iced tea (caffeinated or decaffeinated), milk, orange juice, or apple juice.

Eye-tracking Measures

As a standard practice for eye-tracking research, I first created four areas of interest (AOIs) on each computer screen: question stem, top two response options, middle response option, and bottom two response options (see Figure 2 of the Supplementary Materials for how AOIs were designated). For each survey item involved in the scale direction experiment, I computed two fixation measures. Fixation count is the number of fixations on a specific AOI and one fixation is approximately 100 or more milliseconds of attention. Fixation duration is the total duration of all fixations (in seconds) on a specific AOI. Fixation count and fixation duration are computed over the full question-answering process (fixations on all four AOIs), on the full scale (fixations on the three scale-related AOIs), and on specific parts of the scale—the top two scale points, the bottom two scale points, and the scale midpoint.

To compute dilation measures, I compared the pupil diameter recorded at each fixation to a base pupil diameter and then divided the difference by the base pupil size, resulting in a relative dilation expressed as the percentage of increase in pupil size over the base pupil size. For each respondent, the base pupil size is the minimum pupil diameter recorded throughout the web survey. Although this base pupil size could potentially inflate dilation measures, it makes it possible to compare relative dilation across all survey items and across all individuals. For each respondent, average dilation is the mean of relative dilations over all fixations on an AOI. I computed average dilations for the full question-answering process for each survey item, and for fixations on the full scale, and on specific parts of the scale. Peak dilation is the maximum dilation observed during the period of answering each survey item for each respondent.

Table 1. Average dilations, fixation counts, and fixation durations by scale direction for satisfaction scale

	Average Dilations		Fixation Counts		Fixation durations	
	Ascending	Descending	Ascending	Descending	Ascending	Descending
Overall	11% ^a	13% ^a	16.6	14.3	6.4	5.1
Full scale	27% ^a	38% ^a	7.8	7.2	3.8	3.0
Dissatisfied options	17% ^{b,c}	19% ^{b,c}	5.6 ^{a,b,c}	3.8 ^{a,b,c}	2.7 ^{a,b,c}	1.7 ^{a,b,c}
Neutral	6% ^c	8% ^c	1.0 ^c	1.0 ^c	0.6 ^c	0.5 ^c
Satisfied options	4% ^{a,b}	11% ^{a,b}	0.9 ^{a,b}	2.4 ^{a,b}	0.5 ^b	0.8 ^b

Note: ^a comparison by scale direction;

^b comparison between dissatisfied options and satisfied options within scale direction;

^c comparison between dissatisfied options and neutral within scale direction;

Italicized superscripts denote significance at $p=.10$ level;

Bolded superscripts denote significance at $p=.05$ level

Results on Satisfaction Scale

[Table 1](#) displays the means of average dilations, fixation counts, and fixation durations by scale direction. The descending satisfaction scale starting with “very satisfied” incurred marginally significantly larger dilations than the ascending scale beginning with “very dissatisfied” for the overall question-answering process ($p=.08$) and fixations on the full scale ($p=.07$). The average number and duration of fixations did not differ by scaled direction for the overall question-answering process and for the full scale.

In terms of differential attention to and processing of specific scale AOIs, the satisfied options incurred significantly larger dilation ($p=.003$) and more fixations ($p=.01$) when they were presented first than when they came last. Respondents had significantly more fixations ($p=.03$) and longer fixations ($p=.04$) on the dissatisfied options under the ascending condition than the descending condition.

Regardless of scale direction, dissatisfied options incurred significantly larger dilations and received significantly more and longer fixations than the other parts of the scale. Take the ascending scale as an example. Average dilation is 17% when respondents processed the dissatisfaction options, but only 6% for neutral and 4% for satisfied options. Respondents had 5.6 fixations on dissatisfied options, and 1.0 on neutral and 2.4 on satisfied options. Respondents also spent two more seconds on dissatisfied options than on neutral and satisfied options. It seems that dissatisfied options are cognitively more difficult for respondents than neutral and satisfied options. The same trend is observed for the descending scale condition.

Peak dilation pinpoints the location where respondents struggled the most. [Figure 1](#) plots the percentage of respondents who had peak dilation on different AOIs. About 60% of respondents had peak dilation when fixating on dissatisfied options regardless of scale direction. By contrast, around 10% of peak dilations occurred at the question stem. This again suggests that the

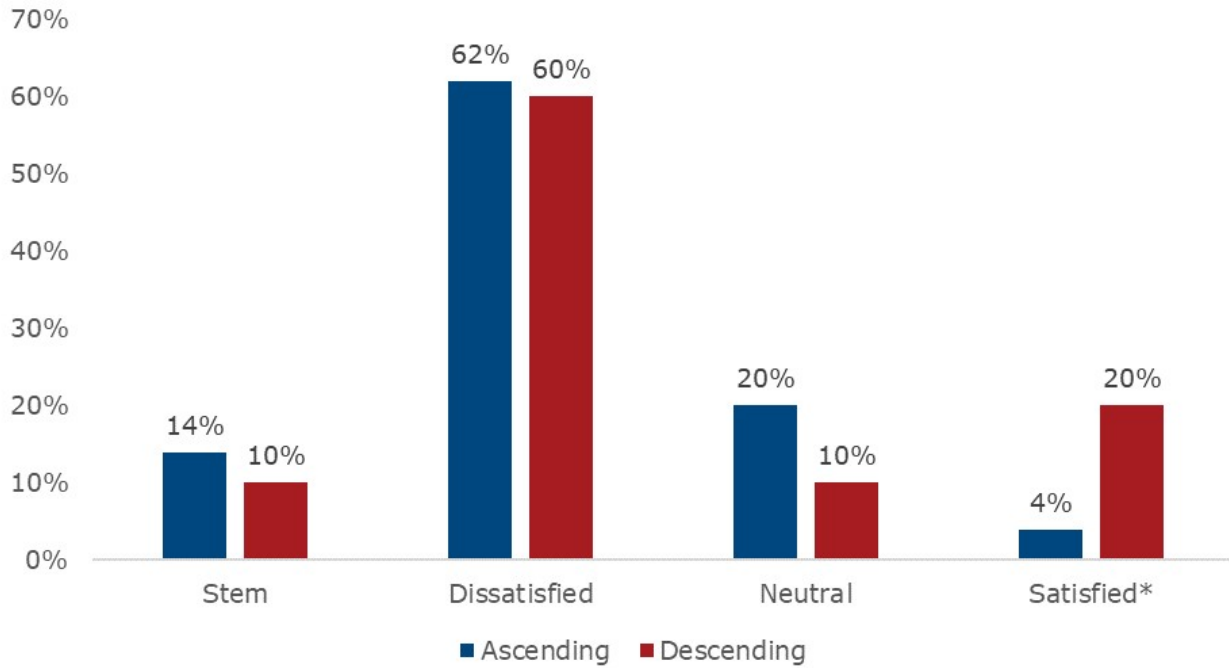


Figure 1. Peak dilation by scale direction for satisfaction scale

Table 2. Average dilations, fixation counts, and fixation durations by scale direction for frequency scale.

	Average dilations		Fixation counts		Fixation durations	
	Ascending	Descending	Ascending	Descending	Ascending	Descending
Overall	12%	14%	19.6 ^a	14.6 ^a	6.8 ^a	5.0 ^a
Full scale	31%	32%	6.6 ^a	4.3 ^a	3.1 ^a	2.4 ^a
Never/seldom	16% ^{a,b,c}	11% ^{a,b,c}	4.0 ^{a,b,c}	1.4 ^{a,b,c}	1.8 ^{a,b,c}	0.8 ^{a,b,c}
Not too often	6% ^c	8% ^c	1.2 ^c	1.0 ^c	0.5 ^c	0.5 ^c
Pretty often/ Very often	9% ^{a,b}	14% ^{a,b}	1.4 ^b	1.9 ^b	0.8 ^b	1.1 ^b

Note: ^acomparison by scale direction;

^bcomparison between “Never/seldom” and “Pretty often/Very often” within scale direction;

^ccomparison between “Never/seldom” and “Not too often” within scale direction;

Italicized superscripts denote significance at $p=.10$ level;

Bolded superscripts denote significance at $p=0.5$ level

dissatisfied options posed the greatest challenge for respondents. Furthermore, the satisfied options incurred more peak dilations (20%) when they were presented earlier in the descending order than in the ascending order (4%).

Results on Frequency Scale

Unlike the satisfaction scale, the amount of dilation does not differ by the direction of the frequency scale for the overall question-answer process and for the processing and use of the full scale, as shown in [Table 2](#). The ascending scale condition elicited significantly longer and more fixations than the descending scale condition for the overall question-answering process (fixation count: $p=.001$, fixation duration: $p=.0004$) and processing of the full scale (fixation count: $p=.003$, fixation duration: $p=.03$).

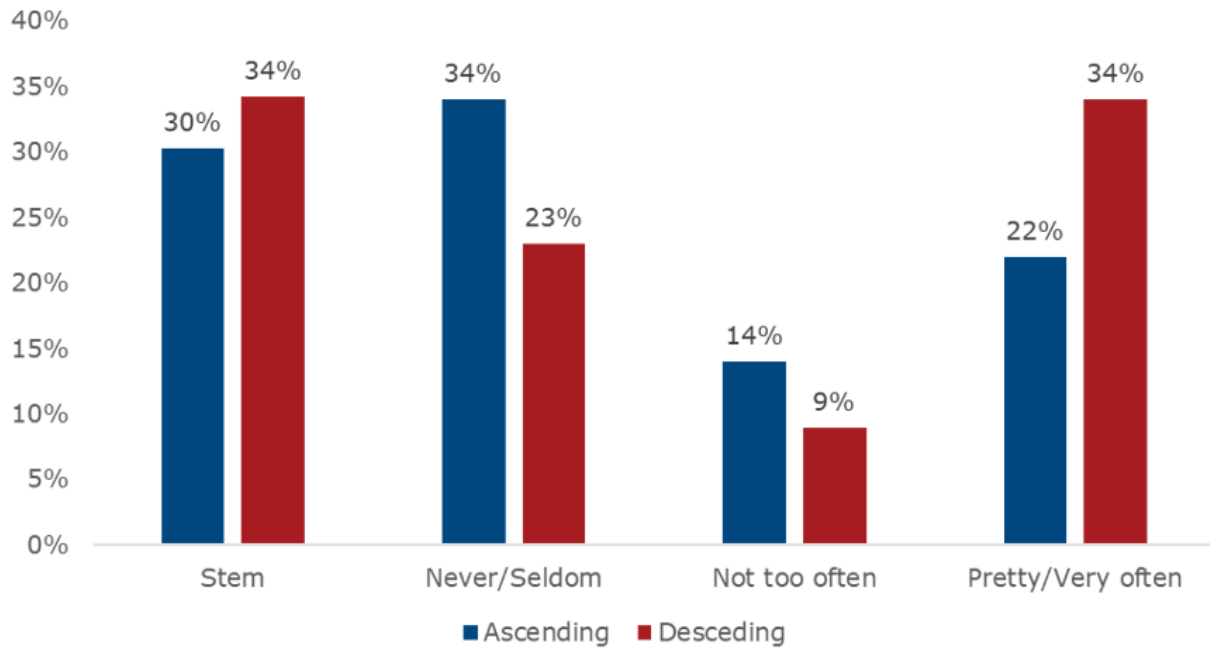


Figure 2. Peak dilation by scale direction for frequency scale

In addition, the “Never” and “Seldom” options incurred significantly larger dilations ($p=.02$), more fixations ($p<.0001$), and long fixations ($p<.0001$) when they are presented first in the ascending order than in the descending order. The two high frequency options (“Pretty Often” and “Very Often”) incurred significantly larger dilation ($p=.002$) under the descending condition than under the ascending condition.

Regardless of scale direction, scale options closer to the start of the frequency scale incurred significantly larger dilation, received significantly more and longer fixations than those in the latter part of the scale.

Furthermore, about one third of peak dilations occurred on a question stem, whereas another one third peaked at the first two scale options regardless of scale direction, as shown in [Figure 2](#). This is very different from the satisfaction scale. For questions using satisfaction scales, two thirds of peak dilations occurred at dissatisfaction options. In addition, the first two scale options had 10% more peak dilations than the last two scale options, regardless of scale direction.

Discussion

This paper illustrates how eye-tracking can be used to explore which scale direction is more difficult for respondents to process and to use, holding other scale features constant. Four eye-tracking measures are examined. Average pupil dilation and peak dilation are used to assess the amount of cognitive effort required of respondents; larger dilations and more peak dilations indicate larger cognitive demand and greater cognitive difficulty. Fixation count and fixation duration are used to quantify the amount of attention and cognitive

processing with longer fixations and more fixations speaking to longer processing. Two scales were examined in this eye-tracking study. Both scales are fully labeled 5-point scales that are presented vertically on the computer screen.

For the satisfaction scale, there is no difference in fixation count and fixation duration in the processing of the full scale by scale direction. However, pupils dilated more for the descending satisfaction scale than for the ascending order, suggesting that the descending scale required more effort and presented greater difficulty for respondents. In particular, the two satisfaction options incurred larger dilations and more fixations when presented first in the descending order than in the ascending order. As a result, an ascending order is recommended for the satisfaction scale.

By contrast, average pupil dilation when processing the full frequency scale did not differ by scale direction. However, the descending frequency scale led to fewer fixations and shorter fixations than the ascending scale. Furthermore, the differences in attention to and effort on both ends of the scale are smaller for the descending order than on the ascending order. These findings collectively indicate that the descending frequency scale is somewhat easier for people to use than the ascending scale; consequently, a descending order is recommended for the frequency scale.

More importantly, this paper found that scale points required different amounts of effort and attracted varying amounts of attention. For the satisfaction scale, regardless of scale direction, the two dissatisfaction options incurred larger dilations, more peak dilations, more fixations, and longer fixations than the two satisfaction options. These findings indicate that dissatisfaction options took more effort and longer time to process than the other parts of the scale. However, for the frequency scale, scale points incurred larger dilations, more fixations, and longer fixations when they were presented earlier than when they came later, consistent with predictions by the satisficing notion. It seems that respondents processed and used the two scales differently.

This paper demonstrates again the utility of eye-tracking to provide a direct window to the survey response process. Eye-tracking technology records eye movements as respondents undergo the question-answering process. Fixation measures are commonly used to understand response mechanisms, to explain observed responses, and to identify problematic questions. This paper illustrates the use of a less commonly used metric that is readily obtained from eye-tracking software. The results support the potential and benefits of using pupil dilation measures to learn what is difficult and cognitively burdensome for respondents and to pinpoint where exactly respondents struggle in the survey response process. Pupil dilation measures complement the usual fixation measures in understanding how respondents process and use response scales. Consequently, the dilation measures have a great potential for question testing and evaluation and are recommended to measure and monitor response burden (Yan and Williams 2022).

Pupil dilation measures can be used alone to identify questions or response options susceptible to problems and warranting further investigation. As shown earlier, the dissatisfaction options elicited much larger dilations and more peak dilations than the other options of the scale, which calls for a revisit to these scale labels. Pupil dilation measures can also be combined with other question testing and evaluation techniques. For instance, cognitive interviews could be conducted to understand the root cause of the difficulty respondents seemed to have with the dissatisfaction options—is it because respondents had trouble understanding the meaning of dissatisfaction or it is because respondents had difficulty mapping their experience to the dissatisfaction options?

This study only looked at two particular scales—a satisfaction scale and a frequency scale, and the findings do not extend to other scales such as the commonly used agreement scale. However, the eye-tracking technique and pupil dilations can be readily and easily applied to other types of response scales, other questions, other question formats, and even other design formats (e.g., a grid, a scroll design where multiple questions are displayed on one screen). Future research should continue exploring the effectiveness of dilation measures to test and evaluate survey questions. The paper has two limitations. First, the study was conducted on 20 respondents. Although the small sample size is not uncommon for qualitative research, it did limit the study's power to detect small differences by scale direction or by different parts of a scale. Second, this paper only examined which scale direction required longer processing. I did not try to link processing with the quality of answers provided. As longer processing does not automatically or necessarily lead to answers of better quality (e.g., Yan et al. 2015), I encourage future research to examine the relationship between fixations, dilations, and data quality.

Submitted: June 14, 2023 EST, Accepted: August 12, 2023 EST

REFERENCES

- Attar, Nada, Matthew H. Schneps, and Marc Pomplun. 2016. "Working Memory Load Predicts Visual Search Efficiency: Evidence from a Novel Pupillary Response Paradigm." *Memory & Cognition* 44 (7): 1038–49. <https://doi.org/10.3758/s13421-016-0617-8>.
- Beatty, Jackson. 1982. "Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources." *Psychological Bulletin* 91 (2): 276–92. <https://doi.org/10.1037/0033-2909.91.2.276>.
- Carp, Frances M. 1974. "Position Effects on Interview Responses." *Journal of Gerontology* 29 (5): 581–87. <https://doi.org/10.1093/geronj/29.5.581>.
- Clifton, Charles, Jr., Adrian Staub, and Keith Rayner. 2017. "Eye Movements in Reading Words and Sentences." In *Eye Movements: A Window on Mind and Brain*, edited by Roger P. G. Van Gompel, Martin H. Fischer, Wayne S. Murray, and Robin L. Hill, 341–71. Amsterdam, Netherlands: Elsevier. <https://doi.org/10.1016/b978-008044980-7/50017-3>.
- Eckstein, Maria K., Belén Guerra-Carrillo, Alison T. Miller Singley, and Silvia A. Bunge. 2017. "Beyond Eye Gaze: What Else Can Eyetracking Reveal About Cognition and Cognitive Development?" *Developmental Cognitive Neuroscience* 25 (June): 69–91. <https://doi.org/10.1016/j.dcn.2016.11.001>.
- Galesic, Mirta, and Ting 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*, edited by M. Das, P. Ester, and L. Kaczmarek, 349–70. New York: Routledge. <https://doi.org/10.4324/9780203844922-14>.
- Garbarski, Dana, Nora Cate Schaeffer, and Jennifer Dykema. 2015. "The Effects of Response Option Order and Question Order on Self-Rated Health." *Quality of Life Research* 24 (6): 1443–53. <https://doi.org/10.1007/s11136-014-0861-y>.
- Höhne, Jan Karem, and Dagmar Krebs. 2018. "Scale Direction Effects in Agree/Disagree and Item-Specific Questions: A Comparison of Question Formats." *International Journal of Social Research Methodology* 21 (1): 91–103. <https://doi.org/10.1080/13645579.2017.1325566>.
- Hossain, Gahangir, and Mohammed Yeasin. 2014. "Understanding Effects of Cognitive Load from Pupillary Responses Using Hilbert Analytic Phase." In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, Ohio*, 381–86. <https://doi.org/10.1109/cvprw.2014.62>.
- Just, Marcel A., and Patricia A. Carpenter. 1993. "The Intensity Dimension of Thought: Pupillometric Indices of Sentence Processing." *Canadian Journal of Experimental Psychology* 47 (2): 310–39. <https://doi.org/10.1037/h0078820>.
- Kahneman, Daniel, and Jackson Beatty. 1966. "Pupil Diameter and Load on Memory." *Science* 154 (3756): 1583–85. <https://doi.org/10.1126/science.154.3756.1583>.
- Kamoen, Naomi, Bregje Holleman, Pim Mak, Ted Sanders, and Huub van den Bergh. 2011. "Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions." *Discourse Processes* 48 (5): 355–85. <https://doi.org/10.1080/0163853x.2011.578910>.
- Kamoen, Naomi, Bregje Holleman, Pim Mak, Ted Sanders, and Huub Van Den Bergh. 2017. "Why Are Negative Questions Difficult to Answer? On the Processing of Linguistic Contrasts in Surveys." *Public Opinion Quarterly* 81 (3): 613–35. <https://doi.org/10.1093/poq/nfx010>.
- Keusch, Florian, and Ting Yan. 2018. "Is Satisficing Responsible for Response Order Effects in Rating Scale Question?" *Survey Research Methods* 12 (3): 299–270. <https://doi.org/10.18148/SRM/2018.V12I3.7263>.

- . 2019. “Impact of Response Scale Features on Survey Responses to Behavioral Questions.” In *Experimental Methods in Survey Research: Techniques That Combine Random Sampling with Random Assignment*, edited by Paul J. Lavrakas, Michael W. Traugott, Courtney Kennedy, Alyson Holbrook, Edith de Leeuw, and Brady T. West, 131–49. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781119083771.ch7>.
- Lenzner, Timo, Lars Kaczmirek, and Mirta Galesic. 2011. “Seeing Through the Eyes of the Respondent: An Eye-Tracking Study on Survey Question Comprehension.” *International Journal of Public Opinion Research* 23 (3): 361–73. <https://doi.org/10.1093/ijpor/edq053>.
- Neuert, Cornelia E. 2020. “How Effective Are Eye-Tracking Data in Identifying Problematic Questions?” *Social Science Computer Review* 38 (6): 793–802. <https://doi.org/10.1177/0894439319834289>.
- Yan, Ting, and Florian Keusch. 2015. “The Effects of the Direction of Rating Scales on Survey Responses in a Telephone Survey.” *Public Opinion Quarterly* 79 (1): 145–65. <https://doi.org/10.1093/poq/nfu062>.
- Yan, Ting, Florian Keusch, and Lirui He. 2018. “The Impact of Question and Scale Characteristics on Scale Direction Effects.” *Survey Practice* 11 (2): 1–10. <https://doi.org/10.29115/sp-2018-0008>.
- Yan, Ting, Lindsay Ryan, Sandra E. Becker, and Jacqui Smith. 2015. “Assessing Quality of Answers to a Global Subjective Well-Being Question Through Response Times.” *Survey Research Methods* 9 (2): 101–9. <https://doi.org/10.18148/srm/2014.v9i2.6107>.
- Yan, Ting, and Douglas Williams. 2022. “Response Burden – Review and Conceptual Framework.” *Journal of Official Statistics* 38 (4): 939–61. <https://doi.org/10.2478/jos-2022-0041>.
- Yan, Ting, Douglas Williams, Aaron Maitland, and Roger Tourangeau. 2016. “Use of Eye-Tracking to Measure Response Burden.” Paper presented at the 2016 AAPOR Annual Conference, Austin, Texas.

SUPPLEMENTARY MATERIALS

Supp. Figure 1

Download: <https://www.surveypractice.org/article/87944-which-scale-direction-is-more-difficult-for-respondents-to-use-an-eye-tracking-study/attachment/180743.png>

Supp. Figure 2

Download: <https://www.surveypractice.org/article/87944-which-scale-direction-is-more-difficult-for-respondents-to-use-an-eye-tracking-study/attachment/180739.png>
