Survey Research Methods (2018)
Vol. 12, No. 3, pp. 259-270
doi:10.18148/srm/2018.v12i3.7263

ISSN 1864-3361
http://www.surveymethods.org

Is Satisficing Responsible for Response Order Effects in Rating Scale Questions?

Florian Keusch
University of Mannheim
Germany

Ting Yan
Westat
Rockville, MD, USA

Rating scales are used extensively in surveys. A rating scale can descend from the highest to the lowest point or from the positive to the negative pole. A rating scale can also start with the lowest point (or the negative pole) and ascend to the highest point (or the positive pole). Previous research has shown that the direction of the scale, i.e., the order of the response options, has an impact on responses, and that respondents are more likely to select response options close to the starting point of the scale, regardless of whether the scale starts with the lowest or the highest point. This paper advances the literature by examining empirically whether or not the response order effect in rating scale questions is driven by satisficing. Drawing on data from an experiment on five multi-item grids, we found that scale direction had a significant and extreme impact on response distributions. Although the effect of scale direction was stronger among speeders than non-speeders, the effect was observed across the board among those who were at a high risk of satisficing and those who were not.

Keywords: scale direction, response order effect, rating scales, satisficing, Web survey

1 Introduction

In surveys, the measurement of many constructs, such as attitudes, opinions, behaviors, and personal states, relies on the use of rating scales. When developing such rating scales for self-administered surveys, questionnaire designers have to make design decisions concerning the number of scale points offered (Krosnick & Fabrigar, 1997), the use of verbal and/or numeric labels assigned to all or some of the scale points (O'Muircheartaigh, Gaskell, & Wright, 1995; Schwarz, Grayson, & Knäuper, 1998; Schwarz & Hippler, 1995; Schwarz, Knäuper, Hippler, Noelle-Neumann, & Clark, 1991), the spacing of response options (Daamen & de Bie, 1992; Tourangeau, Couper, & Conrad, 2004), the alignment of the scale, i.e., the decision to present the scales horizontally or vertically on a screen or paper (Christian, Parsons, & Dillman, 2009), and whether to present individual items one per screen or group them and place them in a matrix format (Tourangeau et al., 2004; van Schaik & Ling, 2007). Studies have demonstrated empirically that the design features of rating scales affect how survey respondents process the scale and use the scale to construct their responses.

Another design feature of a rating scale pertains to its direction, that is, the order of the scale points. A rating scale can descend from the positive to the negative pole (e.g., "strongly agree" to "strongly disagree") or from the highest to the lowest point (e.g., "all of the time" to "never"). The scale can also ascend from the negative to the positive pole (e.g., "strongly disagree" to "strongly agree") or the lowest to the highest point (e.g., "never" to "all of the time"). Hofmans et al. (2007) also refer to the two formats as "decremental scales" and "incremental scales," respectively. Compared to other rating scale features – such as the number of scale points and the use of verbal labels – for which questionnaire textbooks provide clear design guidelines (e.g., Bradburn, Sudman, & Wansink, 2004; Fowler, 1995; Krosnick & Presser, 2010), the choice of scale direction largely "seems to be a matter of taste" (Ramstedt & Krebs, 2007, p. 33).

Existing research demonstrated that scale direction can have an impact on responses in self-administered paper-pencil surveys (e.g., Belson, 1966; Chan, 1991; Friedman, Herskovitz, & Pollack, 1993; Friedman, Weiser Friedmann, & Gluck, 1988; Krebs & Hoffmeyer-Zlotnik, 2010; Sheluga, Jacoby, & Major, 1978), but some studies did not find an effect of scale direction (Dickson & Albaum, 1975; Israel & Taylor, 1990; Powers, Morrow, Goudy, & Keith, 1977; Ramstedt & Krebs, 2007; Weng & Cheng, 2000). For Web surveys, four studies showed a significant impact of scale direction on answers (Hofmans et al., 2007; Liu & Keusch, 2017; Stapleton, 2013; Toepoel, Das, & van Soest, 2009), and one study failed to demonstrate this effect (Keusch & Yan, forthcoming). Results are mixed in Christian et al. (2009), Höhne and Krebs (2018), Höhne and Lenzner (2015), Krebs (2012),

Corresponding author: Florian Keusch, Professorship for Statistics and Methodology, University of Mannheim, 68131 Mannheim, Germany (E-Mail: f.keusch@uni-mannheim.de).

Malhotra (2009), who found a significant impact of scale direction on some questions, but not others. In studies where a significant response order effect in rating scales was observed, responses were found to lean toward the start of the rating scale. That is, the high or positive end of the scales was endorsed more often in descending scales and the low or negative end of the scales attracted more endorsement in ascending scales.

As to the mechanisms responsible for the response order effect in rating scales, most existing research did not attempt to examine the underlying mechanism causing this shift of responses to the start of the scale (see Yan & Keusch, 2015, for an exception). Some researchers simply considered scale direction effect to be a special case of primacy effect reported for unordered response options in survey modes with visual presentation (Krosnick & Alwin, 1987) and resorted to satisficing when trying to account for it (e.g., Krebs & Hoffmeyer-Zlotnik, 2010). That is, when respondents are unable or unwilling to fulfill the potentially complex task of retrieving information from memory and editing it to fit the categories of survey response options, they take a cognitive shortcut by selecting the first acceptable or satisfactory response that is offered rather than selecting an optimal answer that matches their true attitude (Krosnick, 1991). For questions with unordered response categories, a number of studies have demonstrated that satisficing leads to response order effects (e.g. Holbrook, Krosnick, Moore, & Tourangeau, 2007; Meredith & Salant, 2013; Narayan & Krosnick, 1996).

However, rating scales are different from unordered response categories. Rating scales represent scale points on a continuum that follows an inherent order; therefore, rating scales could evoke different cognitive processes leading to different response strategies than when a list of unordered response options are used. We suspect that the satisficing notion does not entirely account for response order effect in rating scales for three reasons. First, a primacy effect in rating scales was also observed in surveys employing an aural administration (e.g. Kalton, Collins, & Brook, 1978; Mingay & Greenwell, 1989; Yan & Keusch, 2015). Second, scale direction effect was *not* found under conditions that are conducive to satisficing such as fast interview pace in Mingay and Greenwell (1989) and survey questions placed towards the end of the questionnaire as in Carp (1974) and Yan and Keusch (2015). One exception is Malhotra (2008), who reports that respondents with low education who speeded through the questionnaire showed the strongest scale direction effect on items employing unipolar rating scales. Third, Yan and Keusch (2015) demonstrates empirically that respondents' use of anchoring-and-adjustment heuristics leads to scale direction effects.

In this paper, we aim to directly examine whether or not the response order effect in rating scales can be attributed to satisficing. We first look at the impact of scale direction on endorsement of individual scale points. We then examine scale direction effect among respondents who are at different risk of satisficing. The satisficing perspective argues for a stronger scale direction effect among respondents with limited cognitive capacity and lower motivation (Krosnick, 1991, 1999). We use the speed at which respondents completed a survey as an indicator of motivation. Previous research demonstrates that speeders (i.e., respondents who sped through an interview and completed the survey at a much faster speed) are more likely to exhibit primacy effects in unordered response options and show more straightlining behavior in multi-item scales (Callegaro, Yang, Bhola, Dillman, & Chin, 2009; Kaminska, McCutcheon, & Billiet, 2010; Malhotra, 2008; Zhang & Conrad, 2014). Therefore, speeders are considered to have a lower motivation to complete the survey and a higher risk to satisfice. In addition, we use age and education as proxies for cognitive capacity. Previous research found that older respondents (Knauper, 1999; Krosnick, 1991, 1999) and respondents with lower formal education (Holbrook et al., 2007; Krosnick, 1991, 1999; Mingay & Greenwell, 1989; Narayan & Krosnick, 1996) are more likely to satisfice than younger respondents and respondents with higher level of education, respectively. The focus of the study is to investigate whether scale direction effect is stronger among respondents prone to satisficing (e.g., speeders, older respondents, and respondents with a lower level of education) than those less prone to satisficing.

2 Methods

2.1 Data

Data for the study come from an experiment embedded in a Web survey among members of an Austrian nonprobability online access panel.[1] In June 2010, 502 panel members participated in a 10-minute Web survey. The email invitation announced the topic of the survey as "Brands & advertising". The quota sample consisted of 50 percent males and 50 percent females. The age of the respondents ranged from 12 to 75 years with a mean of 40.6 years (median=39 years). 58 percent of the respondents reported to have at least a high school degree.

Among others, respondents were asked to rate their attitudes towards two brand logos and three brand advertisements on a six-point end-labeled unipolar rating scale. For

[1]The goal of this study is to estimate the causal effect of scale direction on responses to questions in an experiment embedded in a web survey. Mullinix, Leeper, Druckman, and Freese (2015) compare the results of experimental studies conducted on the probability-based online panel KnowledgePanel to those from experiments conducted on several convenience samples (e.g., from Amazon's Mechanical Turk). They find that results from the convenience samples generally provide estimates of causal effects comparable to those found on the population-based samples from KnowledgePanel.

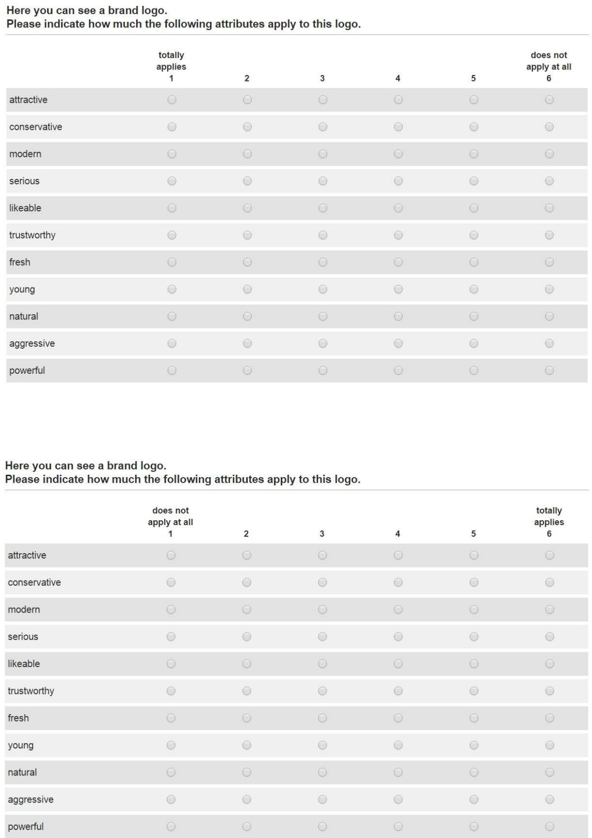

Figure 1. Example screenshots of scale direction experimental conditions (top panel: descending scale; bottom panel: ascending scale; translated from German to English)

each brand, attributes were presented in a grid on one screen. The number of items in a grid ranged between 10 and 12 per brand. The horizontal display of the rating scale in a grid format makes our study different from earlier research, which studies scale direction effects mostly on individual items with vertical scales.

We experimentally varied the direction of the rating scale used in the five grids in a between-subject design.[2] A random half of the respondents received a scale running from “totally applies” to “does not apply at all” (descending scale condition; top panel in Figure 1), and the other half received a scale running from “does not apply at all” to “totally applies” (ascending scale condition; bottom panel in Figure 1)[3]. The order of the grids presented to the respondents was randomized.

2.2 Analysis plan

To assess the influence of scale direction on survey responses, we first created a binary outcome variable, for each individual item, indicating whether a response fell on the low side of the scale (=1), i.e., the three scale points on the side of the scale anchored with “does not apply at all”, or the high side of the scale (=0), i.e., the three scale points on the side of the scale anchored with “totally applies”. Then we compared differences in the distribution of this binary outcome variable at the individual item level across scale direction conditions (0=“ascending scale”, 1=“descending scale”) with Chi-squared tests using the CrossTab function from the GMODEL package (Warnes, Bolker, Lumley, & Johnson, 2015) in R version 3.5.1 (R Core Team, 2018).

To analyze the influence of satisficing on the response order effect in rating scales, we specified multiple general linear mixed-effects models on the binary outcome variable treating each grid as an independent replication. Since the binary outcome variable was coded for each item and for each individual respondent, the outcome variable was cross-classified by respondents and item. As a result, we fit a cross-classified random effects model using the GLMER function from the LME4 package (Bates, Maechler, Bolker, & Walker, 2015) in R. In these models, we specified respondents and items as random effects and indicators for scale direction, age (0=”under 65 years”, 1=”65 years and older”), education (0=”at least high school degree”, 1=”less than high school degree”), and speeding (0=”Non-speeders”,1=”Speeders”) as well as interaction terms of scale direction with age, education, and speeding as fixed effects. We followed Zhang and Conrad (2014) in their approach of defining speeders as respondents who took less than 300ms per word to complete a survey question. Due to the experimental design of the questionnaire, response time was only available on the entire questionnaire level and not on the page-level. The questionnaire comprised 1,230 words, setting the initial threshold for speeding to 369 seconds (= $300\text{ms/word} \times 1,230$ words). We then added five seconds per grid (a total of 25 seconds) to the initial threshold because respondents were asked to thoroughly look at the logos and ads before answering the

[2] Our study has a between-subject design component (scale direction is varied at the respondent level) and also a within-subject design component (all respondents have to answer questions in all five grids). Because the scale direction manipulation is between-subject, our analyses mostly employ this between-subject design feature and treat the five grids as independent replications. We acknowledge that this analytic decision is suboptimal for a within-subject design.

[3] Independent of the scale direction manipulation, we also experimentally varied the use of numeric labels. One random half of the respondents received a scale with numeric labels (1 to 6) for all scale points, and the other half received a scale with only verbal labels for the two end points and no numeric labels. There were neither a significant main effect of numeric labels nor significant interaction effect between numeric labels and scale direction in the multivariate models presented below. As a result, we combined cases across the numeric labels conditions.

attitude items. We also added one second per screen of the questionnaire to account for transitioning between screens (43 seconds in total), yielding a final threshold of 437 second. About 8 percent of respondents answered the questionnaire faster than this threshold and were thus labeled as speeders.[4]

3 Results

Table 1 shows the proportion of respondents who selected a response option on the low side of the scale (i.e., the three scale points on the side anchored with "does not apply at all") by scale direction. For all eleven items in Grid 1, respondents were significantly more likely to select a response option from the low side when the scale started with the low side (i.e., the ascending scale condition) than when the scale started with the high side (i.e., the descending scale condition). For some items, the difference in endorsement of the low side between the ascending and the descending condition is more than 35 percentage points. The results for the other four grids are highly comparable with all individual Chi-squared test being statistically significant at $p < 0.001$.

To analyze the multivariate effect of scale direction on endorsement of the low side of the scale for respondents with different levels of satisficing, cross-classified multilevel models were fit with the binary outcome variable indicating whether or not a response from the low side of the scale (i.e., the three response options on the side anchored with "does not apply at all") was selected as the dependent variable. For each of the five grids, we first fit a base model to partition the total variance of the binary outcome variable into within-cell and between-cell components (not shown). About 43 percent of the total variation in the likelihood of selecting from the low side of the scale for Grid 1, for instance, was across respondents and about 6 percent was across items. For the other four grids, about 28 to 37 percent of the total variation was across respondents and 4 to 14 percent was across items. Overall there is a larger variation at the respondent level than at the item level.

Tables 2 shows the results of the cross-classified random effects models fitted on the binary outcome variables created for items for all five grids. In terms of the fixed effects, scale direction consistently showed a significant influence on the endorsement of the low side of the scale, which is in accordance with the results of the univariate analysis – a descending scale reduced the likelihood to select from the low side of the scale. With regard to the indicators of satisficing, age does not seem to affect the likelihood of selecting from the low side of the scale. Education had a significant main effect on the likelihood of selecting from the low side for two (out of five) models and the speeding indicator had a significant main effect for two models too. Even when changing the threshold for speeding from 300 to 250 and 350 ms per word, results of the models are highly comparable to our initial models (see Appendix), showing that the effect of scale

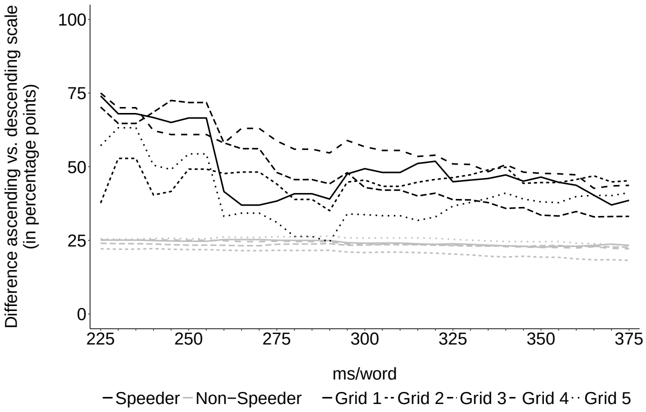

Figure 2. Difference in proportion of selecting low side of the scale between ascending and descending scale condition by speeding across different speeding thresholds

direction is independent of the definition of speeding used in our analysis.

To determine whether the observed scale direction effect is caused by satisficing, we turn to examining two-way interactions between scale direction and the satisficing indicators. For four (out of five) grids (Grid 1, Grid 2, Grid 3, and Grid 4), the interaction between scale direction and speeding was significant. Figure 2 plots the differences in endorsement of the low side of the scale between the ascending and the descending scale condition in a given grid for speeders (black lines) and non-speeders (grey lines) across 31 different definitions of speeding. The average difference in the proportion selecting from the low side of the scale was larger for speeders than for non-speeders in all five experiments across all speeding thresholds but one. In other words, speeders consistently show larger scale direction effect but scale direction effect is observed in both speeders and non-speeders, regardless of how we define speeding. There is no significant interaction between scale direction and the other two indicators of satisficing (age and education).

Finally, we specified models including three-way interactions between scale direction, speeding, and education.[5] The

[4]To test the robustness of our findings, we varied the threshold for speeding and reran the cross-classified models. We used a stricter definition with a threshold for speeding of 250ms per word and a less strict definition with a threshold of 350ms per word. In both cases, the final threshold values include time for looking at the brand and moving from page to page. The stricter definition (speeding ≤ 375sec) flagged three percent of respondents as speeders while the less strict definition of speeding (speeding ≤ 498sec) flagged 14 percent of respondents as speeders. The results from the models including the alternative definitions of speeding are highly consistent with the findings of the models including the original definition of speeding (see Appendix).

[5]We also specified and reran models with a three-way interaction term for scale direction × speeding × age, but the models did

Table 1
Proportion of responses from the low side of the scale by scale condition

Item	Ascending Scale (in %)	Descending Scale (in %)	Difference
Grid 1			
attractive	51.2	25.4	25.8
conservative	74.4	51.6	22.8
modern	58.8	30.2	17.6
serious	64.4	30.2	31.2
likeable	54.4	23.4	31.0
trustworthy	68.0	32.5	35.5
fresh	57.2	32.1	25.1
young	52.4	26.2	16.2
natural	69.6	44.0	25.6
aggressive	74.8	49.6	25.2
powerful	45.6	19.4	26.2
N	250	252	
Grid 2			
attractive	47.6	26.6	21.0
conservative	83.2	67.5	15.7
modern	39.2	15.1	24.1
serious	56.8	29.8	27.0
likeable	48.4	24.2	24.2
trustworthy	60.4	31.0	29.4
obtrusive	72.4	49.2	23.2
cheap	72.8	53.2	19.6
charming	64.4	36.5	27.9
romantic	73.2	59.1	14.1
feminine	75.2	49.6	25.2
N	250	252	
Grid 3			
attractive	49.6	18.3	21.3
conservative	36.0	19.4	16.6
modern	71.6	54.0	17.6
serious	28.8	5.5	23.3
likeable	52.8	17.9	34.9
trustworthy	40.0	11.9	28.1
fresh	75.2	46.4	28.8
young	81.2	56.3	24.9
natural	60.0	30.2	29.8
elegant	37.2	15.9	21.3
demur	51.2	33.3	17.9
bourgeois	63.2	45.2	18.0
N	250	252	

Continues on next page

Continued from last page

Item	Ascending Scale (in %)	Descending Scale (in %)	Difference
Grid 4			
attractive	56.8	32.5	23.3
conservative	58.8	36.5	22.3
modern	57.6	33.7	23.9
serious	44.8	19.0	25.8
likeable	65.6	39.3	26.3
trustworthy	57.6	27.4	30.2
aggressive	73.2	50.0	23.2
powerful	52.4	23.8	28.6
cold	55.6	29.4	26.2
masculine	70.4	40.9	29.5
N	250	252	
Grid 5			
attractive	56.8	29.0	27.8
conservative	61.6	42.5	19.1
modern	54.8	29.8	25.0
serious	46.6	17.5	28.9
likeable	58.0	26.6	31.4
trustworthy	58.0	25.4	22.6
elegant	54.0	31.0	23.0
stylish	52.8	28.6	24.2
extravagant	79.2	50.4	28.8
cold	70.8	38.9	21.9
masculine	95.6	80.6	15.0
N	250	252	

All differences between scales are statistically significant ($p < 0.001$).

three-way interaction term was not statistically significant in four of the five models using 300 ms per word as the threshold for speeding. The only significant three-way interaction found in the model for Grid 3 is displayed in Figure 3. Contrary to what would be expected from the satisficing notion, speeders with a high school degree are more prone to scale direction effects than speeders without a high school degree.

4 Conclusion

The study presented in this paper found a rather strong tendency of answers to be pushed to the start of rating scales that were presented in a grid format to respondents; respondents were more likely to select a response option from the low side when the scales were ascending from "does not apply at all" to" totally applies" than when the scales were descending from "totally applies" to "does not apply at all." This study used a sample from a nonprobability online access panel, but it replicates what Yan and Keusch (2015) have found on a probability sample of the general population. Our findings add one more piece of empirical evidence to the literature on the impact of scale direction on survey answers.

Despite the significant and large impact of scale direction on responses, we only found limited evidence that the scale direction effect was solely driven by satisficing. We employed three commonly used indicators of satisficing in

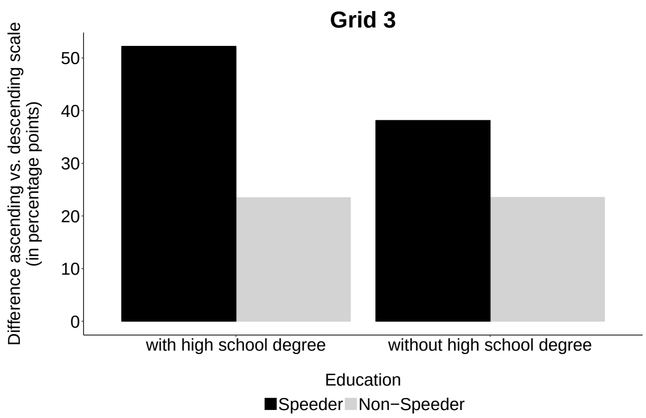

Figure 3. Difference in proportion of selecting low side of the scale between ascending and descending scale condition by speeding by education for Grid 3

not converge.

Table 2
Multivariate analyses of selecting the low side of the scale

	Grid 1		Grid 2		Grid 3		Grid 4		Grid 5	
Variables	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Fixed effects										
Intercept	0.58	0.49	0.88*	0.43	−0.29	0.43	0.76	0.39	1.17*	0.54
Scale direction (Asc. scale)	−2.18***	0.68	−1.10*	0.52	−1.79***	0.52	−0.96	0.51	−1.91**	0.64
Education (with HS degree)	−0.60**	0.21	−0.12	0.16	−0.07	0.16	0.32*	0.16	−0.32	0.20
Age (65+ ys)	0.21	0.46	−0.16	0.36	0.53	0.34	−0.49	0.36	−0.25	0.45
Speeding (Non-speeders)	0.71	0.52	0.64	0.41	1.10**	0.40	1.21**	0.41	0.38	0.48
Two-way interaction of scale direction and:										
Education	0.27	0.30	−0.23	0.23	0.13	0.22	−0.19	0.23	−0.22	0.28
Age	0.66	0.68	−0.07	0.52	0.42	0.52	−0.18	0.52	0.45	0.64
Speeding	−1.75**	0.62	−1.56**	0.49	−1.36**	0.48	−2.04***	0.49	−0.53	0.58
Random effects										
Respondent variance	2.03		1.07		1.03		1.02		1.77	
Item variance	0.38		0.61		0.82		0.18		0.98	
Residuals	3.29		3.29		3.29		3.29		3.29	
Model fit statistics										
Log Likelihood	-3,074.0		-3,234.8		-3,366.7		-3,035.1		-3,032.6	
AIC	6,168.0		6,489.7		6,753.4		6,090.2		6,085.3	
BIC	6,234.2		6,555.8		6,820.5		6,155.5		6,151.4	
Total observations	5,522		5,522		6,024		5,020		5,522	

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$
Reference categories in parentheses.

our study: the tendency to speed through the questionnaire as a measure of respondent motivation, and respondent age as well as educational attainment as measures of respondent ability. Respondents who sped through the questionnaire did – to some extent – show stronger response order effects for some of the items, indicating that motivation is associated with scale direction effects. However, we did not find older respondents and respondents with lower level of formal education to be more prone to response order effects in rating scales than younger respondents and those with higher level of formal education, respectively. Only in one of five models did we find a significant three-way interaction between scale direction, speeding, and education. However, that significant interaction effect seems to be counter to what would be expected if satisficing were responsible for scale direction effect.

In addition, we found that respondents who did not speed through the questionnaire also showed significant and substantial response order effects in rating scale questions. We therefore cannot conclude that the observed response order effect in rating scales was completely accounted for by the notion of satisficing. We acknowledge that satisficing might arise from a lack of personal interest in a survey topic, difficult tasks or other factors. Unfortunately, we do not have access to these measures in our study and we encourage other researchers to further investigate the role of these factors on scale direction effects.

The findings presented in this paper have important implications for survey methodologists, researchers, and practitioners who rely on the use of rating scales in their studies. First, our results demonstrated that the direction of a rating scale matters, especially when univariate analyses of the answers are the key analytic goal. Therefore, survey methodologists and researchers should be aware of the response order effect when designing rating scales for survey questionnaires. Unfortunately, the survey field does not have enough evidence concluding which scale direction yields data of better quality.

Second, the field needs to better understand what causes the responses to shift to the start of a scale. Response order effects in rating scales tend to be brushed off as a special case of primacy effects attributed to respondents satisficing during the question-answering process (Krosnick, 1991; Krosnick & Presser, 2010). However, our findings at best partially support the satisficing account. The effects of scale direction on responses to rating scale questions are sometimes stronger for speeders, but *not* necessarily stronger in other situations conducive to satisficing, consistent with Carp (1974), Mingay and Greenwell (1989), Yan and Keusch (2015). Yan and Keusch (2015) suggest that the effect of scale direction may be due to respondents' use of anchoring-and-adjustment when constructing and mapping their answers to one of the scale points. Using eye-tracking technology, (Höhne & Lenzner, 2015) found that respondents fixate the first half of a response scales more intensively than the second half and that the amount of time spent looking at the first half correlates with the probability of selecting a response category from that side. Salzberger and Koller (2013) attributed the influence of the direction of rating scales on responses to the 'near means related' heuristic respondents use when answering survey questions(Tourangeau et al., 2004, 2007). More research is needed to examine the applicability of these explanations.

One limitation of the scale direction experiment presented in this paper is that the scale features are fixed; the experiments used six-point unipolar end-labeled rating scales that were presented horizontally in a multi-item grid. As a result, we acknowledge the potential limitations with the generalizability of our findings to scales with different features and the inability to investigate the moderating properties of different other scale features on response order effects in rating scale questions. However, scale direction effects are observed on scales with different number of scale points (e.g., Yan & Keusch, 2015), fully-labeled scales (e.g., Garbarski, Schaeffer, & Dykema, 2015), and bipolar scales (e.g., Höhne & Krebs, 2018). In addition, certain question characteristics and scale features moderate the influence of scale direction on survey responses (Yan, Keusch, & He, 2018). For instance, scale direction effects are stronger for non-attitudinal items, earlier survey items, and items with longer scales. Furthermore, the moderating impact of question type, question location, and scale length on scale direction effects is more pronounced for items administered via Computer-Assisted Personal Interviewing than in self-administration. Others found that scale direction effects are more pronounced in vertical scales than horizontal scales (Höhne & Lenzner, 2015). More experimental research is needed that systematically varies these scale features so as to thoroughly understand scale direction effects.

Another limitation of the paper is that only univariate distributions of responses are examined. Future research should investigate the influence of scale direction on the correlation structures of survey responses (e.g., relationships among several items of a multi-item scale) and how scale direction interacts with the polarity of items (i.e., whether a survey item is worded positively or negatively) in order to advance the literature on scale direction effects.

Acknowledgements

The data and code used in this article are available for reuse as supplementary to this paper, and in addition also from http://data.aussda.at at AUSSDA – The Austrian Social Science Data Archive. The data are published under a Creative Commons Attribution 4.0 International License and can be cited as: Keusch, F. and Yan, T. (2018). Replication Data for: "Is satisficing responsible for response order effects

in rating scale questions?", doi:10.11587/Z1Z34H, *AUSSDA Dataverse*, V1.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67(1)*, 1–48.

Belson, W. A. (1966). The effect of reversing the presentation order of verbal rating scales. *Journal of Advertising Research*, *6*, 30–37.

Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions: the definitive guide to questionnaire design for market research, political polls, and social and health questionnaires*. San Francisco: Jossey-Bass.

Callegaro, M., Yang, Y., Bhola, D., Dillman, D., & Chin, T.-Y. (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin of Sociological Methodology*, *103(1)*, 5–25.

Carp, F. (1974). Position effects on interview response. *Journal of Gerontology*, *29(5)*, 581–587.

Chan, J. (1991). Response-order effects in likert-type scales. *Educational and Psychological Measurement*, *51(3)*, 530–541.

Christian, L., Parsons, N., & Dillman, D. (2009). Designing scalar questions for web surveys. *Sociological Methods & Research*, *37(3)*, 393–425.

Daamen, D. & de Bie, S. (1992). Serial context effects in survey items. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 91–114). New York: Springer.

Dickson, J. & Albaum, G. (1975). Effects of polarity on semantic differential scales in consumer research. *Advances in Consumer Research*, *2*, 507–514.

Fowler, J. (1995). *Improving survey questions. design and evaluation*. Thousand Oaks: Sage Publications.

Friedman, H., Herskovitz, P., & Pollack, S. (1993). The biasing effects of scale-checking styles on response to a likert scale. Proceedings of the American statistical association annual conference: survey research methods, pp. 792–795.

Friedman, H., Weiser Friedmann, L., & Gluck, B. (1988). The effects of scale-checking styles on responses on a semantic differential scale. *Journal of the Market Research Society*, *30(4)*, 477–481.

Garbarski, D., Schaeffer, N., & Dykema, J. (2015). The effects of question order and response option order on self-rated health. *Quality of Life Research*, *24(6)*, 1443–1453.

Hofmans, J., Theuns, P., Baekelandt, S., Mairesse, O., Schillewaert, N., & Cools, W. (2007). Bias and change in perceived intensity of verbal qualifiers effected by scale orientation. *Survey Research Methods*, *1(2)*, 97–108.

Höhne, J. & Krebs, D. (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. doi:10.1080/13645579.2017.1325566

Höhne, J. & Lenzner, T. (2015). Investigating response order effects in web surveys using eye tracking. *Psihologija*, *48(4)*, 361–377.

Holbrook, A., Krosnick, J., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally. the impact of question and respondent attributes. *Public Opinion Quarterly*, *71(3)*, 325–348.

Israel, G. & Taylor, C. (1990). Can response order bias evaluations? *Evaluation and Program Planning*, *13(4)*, 365–371.

Kalton, G., Collins, M., & Brook, L. (1978). Experiments in wording opinion questions. *Journal of the Royal Statistical Society (Series C)*, *27(2)*, 149–161.

Kaminska, O., McCutcheon, A., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, *74(5)*, 956–984.

Keusch, F. & Yan, T. (forthcoming). Impact of response scale features on survey responses to factual behavioral questions. In P. Lavrakas, E. de Leeuw, A. Holbrook, C. Kennedy, M. Traugott, & B. West (Eds.), *Experimental methods in survey research: techniques that combine random sampling with random assignment*. Hoboken: John Wiley & Sons.

Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Research*, *63(3)*, 347–370.

Krebs, D. (2012). The impact of response format on attitude measurement. In S. Salzborn, E. Davidov, & J. Reinecke (Eds.), *Methods, theories, and applications in the social sciences* (pp. 105–113). Wiesbaden: Springer VS.

Krebs, D. & Hoffmeyer-Zlotnik, J. (2010). Positive first or negative first? effects of the order of answering categories on response behavior methodology. *European Journal of Research Methods for the Behavioral and Social Sciences*, *6(3)*, 118–127.

Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5(3)*, 213–236.

Krosnick, J. (1999). Survey research. *Annual Review of Psychology*, *50*, 537–567.

Krosnick, J. & Alwin, D. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51(2)*, 201–219.

Krosnick, J. & Fabrigar, L. (1997). *Designing rating scales for measurement in surveys*. New York: Wiley-Interscience.

Krosnick, J. & Presser, S. (2010). Questionnaire design. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (pp. 263–313). England: Emerald Group.

Liu, M. & Keusch, F. (2017). Effects of scale direction on response style of ordinal rating scales. *Journal of Official Statistics*, *33(1)*, 137–154.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, *72(5)*, 914–934.

Malhotra, N. (2009). Order effects in complex and simple tasks. *Public Opinion Quarterly*, *73(1)*, 180–798.

Meredith, M. & Salant, Y. (2013). On the causes and consequences of ballot order effects. *Political Behavior*, *35(1)*, 175–197.

Mingay, D. & Greenwell, M. (1989). Memory bias and response-order effects. *Journal of Official Statistics*, *5(3)*, 253–263.

Mullinix, K., Leeper, T., Druckman, J., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2(2)*, 109–138.

Narayan, S. & Krosnick, J. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, *60(1)*, 58–88.

O'Muircheartaigh, C., Gaskell, G., & Wright, D. (1995). Weighing anchors: verbal and numeric labels for response scales. *Journal of Official Statistics*, *11(3)*, 295–307.

Powers, E., Morrow, P., Goudy, W., & Keith, P. (1977). Serial order preference in survey research. *Public Opinion Quarterly*, *41(1)*, 80–85.

R Core Team. (2018). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Ramstedt, B. & Krebs, D. (2007). Does response scale format affect the answering of personality scales? assessing the big five dimensions of personality with different response scales in a dependent sample. *European Journal of Psychological Assessment*, *23(1)*, 32–38.

Salzberger, T. & Koller, M. (2013). Towards a new paradigm of measurement in marketing. *Journal of Business Research*, *66(9)*, 1307–1317.

Schwarz, N., Grayson, C., & Knäuper, B. (1998). Formal features of rating scales and the interpretation of question meaning. *International Journal of Public Opinion Research*, *10(2)*, 177–184.

Schwarz, N. & Hippler, H. (1995). The numeric values of rating scales: a comparison of their impact in mail surveys and telephone interviews. *International Journal of Public Opinion Research*, *7(1)*, 72–74.

Schwarz, N., Knäuper, B., Hippler, H., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, *55(4)*, 570–582.

Sheluga, D., Jacoby, J., & Major, B. (1978). Whether to agree-disagree or disagree-agree: the effect of anchor order on item response. *Advances in Consumer Research*, *5*, 109–113.

Stapleton, C. (2013). The smartphone way to collect survey data. *Survey Practice*, *6(2)*, 1–7.

Toepoel, V., Das, M., & van Soest, A. (2009). Design of web questionnaires: the effect of layout in rating scales. *Journal of Official Statistics*, *25(4)*, 509–528.

Tourangeau, R., Couper, M., & Conrad, F. (2004). Spacing, positioning, and order. interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68(3)*, 368–393.

Tourangeau, R., Couper, M., & Conrad, F. (2007). Colors, labels, and interpretative heuristics for response scales. *Public Opinion Quarterly*, *71(1)*, 91–112.

van Schaik, P. & Ling, J. (2007). Design parameters of rating scales for web sites. *ACM Transactions on Computer-Human Interaction*, *14*, 4.

Warnes, G., Bolker, B., Lumley, T., & Johnson, R. (2015). Gmodels: various r programming tools for model fitting. R package version 2.16.2. Retrieved from https://CRAN.R-project.org/package=gmodels

Weng, L.-J. & Cheng, C.-P. (2000). Effects of response order on likert-type scales. *Educational and Psychological Measurement*, *60(6)*, 91–112.

Yan, T. & Keusch, F. (2015). The effects of the direction of rating scales on survey responses in a telephone survey. *Public Opinion Quarterly*, *79(1)*, 145–165.

Yan, T., Keusch, F., & He, L. (2018). The impact of question and scale characteristics on scale direction effects. *Survey Practice*, *11(2)*, 3126.

Zhang, C. & Conrad, F. (2014). Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Survey Research Methods*, *8(2)*, 127–135.

Appendix
Tables

(*Appendix tables follow on next page*)

Table A1
Multivariate analyses of selecting the low side of the scale with speeding = faster than 250ms/word

Variables	Grid 1 Coeff.	Grid 1 SE	Grid 2 Coeff.	Grid 2 SE	Grid 3 Coeff.	Grid 3 SE	Grid 4 Coeff.	Grid 4 SE	Grid 5 Coeff.	Grid 5 SE
Fixed effects										
Intercept	0.59	0.50	0.89*	0.43	0.28	0.43	0.78*	0.39	1.17*	0.54
Scale direction (Ascending scale)	−2.18***	0.68	−1.11*	0.52	−1.79***	0.52	−0.98	0.52	−1.91**	0.64
Education (with HS degree)	−0.62**	0.21	−0.14	0.16	−0.09	0.15	0.27	0.16	−0.33	0.20
Age (65+ ys)	0.22	0.46	−0.13	0.36	0.54	0.34	−0.46	0.36	−0.25	0.44
Speeding (Non-speeders)	1.00	0.83	0.33	0.65	3.13***	0.82	1.19	0.67	1.43	0.82
Two-way interaction of scale direction and:										
Education	0.28	0.29	−0.22	0.23	0.14	0.22	−0.16	0.23	−0.21	0.28
Age	0.61	0.68	0.01	0.52	0.41	0.52	−0.26	0.52	0.47	0.64
Speeding <250ms/word	−3.05**	1.02	−1.91*	0.79	−3.89***	0.92	−2.21**	0.79	−2.12*	0.96
Random effects										
Respondent variance	2.01		1.07		0.99		1.05		1.76	
Item variance	0.38		0.61		0.82		0.18		0.98	
Residuals	3.29		3.29		3.29		3.29		3.29	
Model fit statistics										
Log Likelihood	-3,072.3		-3,235.2		-3,360.5		-3,039.9		-3,030.6	
AIC	6,164.5		6,490.5		6,741.0		6,099.8		6,081.2	
BIC	6,230.7		6,556.7		6,808.0		6,165.0		6,147.4	
Total observations	5,522		5,522		6,024		5,020		5,522	

* $p < .05$ ** $p < .01$ *** $p < .001$
Reference categories in parentheses.

Table A2

Multivariate analyses of selecting the low side of the scale with speeding = faster than 350ms/word

	Grid 1		Grid 2		Grid 3		Grid 4		Grid 5	
Variables	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Fixed effects										
Intercept	0.57	0.50	0.86^{*}	0.43	−0.29	0.43	0.74	0.39	1.15^{*}	0.54
Scale direction (Ascending scale)	-2.17^{**}	0.68	-1.09^{*}	0.51	-1.79^{***}	0.52	−0.94	0.51	-1.89^{**}	0.64
Education (with HS degree)	-0.58^{**}	0.21	−0.09	0.16	−0.07	0.16	0.34^{*}	0.16	−0.29	0.20
Age (65+ ys)	0.18	0.46	−0.19	0.36	0.52	0.35	−0.52	0.36	−0.29	0.45
Speeding (Non-speeders)	0.65	0.35	0.67^{*}	0.27	0.51	0.26	0.85^{**}	0.27	0.62	0.32
Two-way interaction of scale direction and:										
Education	0.25	0.30	−0.26	0.23	0.13	0.23	−0.21	0.23	−0.25	0.28
Age	0.75	0.68	0.16	0.51	0.44	0.53	−0.11	0.52	0.53	0.64
Speeding <350ms/word	-1.68^{***}	0.45	-1.58^{***}	0.35	-0.73^{*}	0.34	-1.60^{***}	0.35	-0.94^{*}	0.72
Random effects										
Respondent variance	2.01		1.05		1.04		1.02		1.76	
Item variance	0.38		0.61		0.82		0.18		0.98	
Residuals	3.29		3.29		3.29		3.29		3.29	
Model fit statistics										
Log Likelihood	-3,071.2		-3,230.5		-3,368.6		-3,033.7		-3,030.5	
AIC	6,162.4		6,481.1		6,757.2		6,087.4		6,081.1	
BIC	6,228.6		6,547.2		6,824.3		6,152.6		6,147.2	
Total observations	5,522		5,522		6,024		5,020		5,522	

* $p < .05$ ** $p < .01$ *** $p < .001$

Reference categories in parentheses.