

Labelling and direction of slider questions

Results from web survey experiments

Mingnan Liu
Facebook

Using a web survey experiment, this study examines measurement comparability between two radio button questions (fully labelled and endpoint labelled) with slider questions. The slider question is unique to web surveys, displaying a horizontal or vertical line with a bar on the line. Respondents need to click and drag the bar to the desired position on the line in order to register their answers. The study described in this paper found that mean scores, break-off rates, time to complete, reliability and respondents' evaluations are similar across question types, but that the item non-response rate for slider questions is significantly higher than for the radio buttons. In a second experiment, the direction of slider (positive–negative vs negative–positive) is compared. With few exceptions, all measures, including the mean scores, break-off rates, item non-response rates, time to complete, reliability and respondents' evaluations are similar between the two directions. The implications and limitations of this study are also discussed.

Introduction

The slider question is a unique question type for web surveys. A slider question displays a horizontal or vertical line with a bar on the line. Respondents need to click and drag the bar to the desired position on the line in order to register their answers.¹ This study reports findings

¹ A very similar question type is the visual analogue scale (VAS). The only visual difference between a VAS and a slider question is that a VAS does not have a bar. Instead of dragging and dropping a bar along the line, the VAS requires pointing and clicking. In several major commercial web survey platforms, slider questions can also be answered in the same fashion as the VAS (by pointing and clicking), which creates a similar user experience. Despite this, these two terms have been used interchangeably to some extent in the industry.

Received (in revised form): 5 October 2016

from two web survey experiments on the comparison of measurements between slider questions and radio button multiple-choice questions with different labelling techniques, and a comparison of survey estimates between two directions of slider question (either from positive to negative or from negative to positive) to determine whether this has an impact on the survey estimates. When designing survey questions, particularly for discrete variables (as opposed to continuous variables), a question can be written in a few different formats, including radio button and slider questions. However, it is important to ensure that the survey estimates collected from these two question types are comparable and answers are not altered by question type. A second consideration in designing slider questions is the direction of the slider. Slider questions are often used for ordinal scales, and a positive adjective can be placed either at the left/top or the right/bottom. This is often referred to as scale direction. Although scale direction has been studied in other contexts, especially using radio button questions, it has not been examined using slider questions. Given that the respondent's interactions with the survey interface between the slider and radio button questions are different, it is critical to examine the impact of scale direction for slider questions.

Literature review

Slider vs radio button

Several studies examined slider questions and compared them to radio button questions. In a web survey experiment, Cook *et al.* (2001) examined three versions of slider questions, 5-point, 9-point and 100-point, with a radio button question with 9-point response options. In all conditions, the two endpoints of the scale were verbally labelled where the other response options were labelled with numbers. They found that there was a higher reliability from the 9-point radio button condition than all the slider conditions. Within slider questions, the 100-point sliders had the highest reliability. The authors concluded that a less coarse (more response options) slider had better measurement. Funke *et al.* (2011) compared endpoint labelled 7-point radio button questions with endpoint labelled slider questions. They also varied the display orientations, either vertically or horizontally. The break-off rate for the slider questions was significantly higher than that for the radio button questions, and the difference between the two question types was even larger for respondents with lower education. They also found a

significantly longer completion time for the slider than the radio button questions, and the effect was moderated by the display orientation (vertical display took longer than horizontal display). In addition, for the two survey questions they asked, one had a significant higher mean score for slider than the radio button condition for the vertically displayed scales. For the horizontally displayed scale, no mean difference was found. Later, Funke (2016) tested radio button, slider and VAS questions, all with endpoint labels. Also, within each question type, he varied the length of the scale (3-point, 5-point and 7-point). The study found a higher break-off rate, item non-response rate and longer completion time for slider questions than radio button and VAS questions. The differences on mobile devices were larger than on computers. The mean scores, however, were reasonably similar to each other across the conditions. More recently, Roster *et al.* (2015) compared a 5-point radio button question (with endpoint labels) with slider questions (also with endpoint labels) and found no significant mean difference. Also, respondents' willingness to participate in the survey did not differ by question type. Different designs of slider questions have also been studied. For example, Buskirk *et al.* (2015) compared the starting position of the slider handle and found that, depending on where the handle was positioned initially (left, middle, right or no handle), the survey estimates and item non-response rates differed as well.

Scale direction

For any scale, one needs to decide on the labels for the starting and ending points, and whether to put the positive/high adjective first/top or last/bottom. The direction of the scale can sometimes affect the responses and hence the survey estimates. A few studies have examined the scale direction on the univariate distribution; these date back to as early as the 1960s (Belson 1966). For example, Dillman *et al.* (1995) examined 82 scale direction experiments in 12 surveys conducted over the phone or by mail. They found that only three of the telephone experiments, one of the mail experiments and two of the telephone/mail mixed-mode experiments showed significant scale direction effect. Christian *et al.* (2008) tested the scale direction using 5-point ordinal scales in web and telephone surveys and found no univariate difference. Stapleton (2013) found a significant impact of the scale direction on univariate distribution in both smartphone and PC surveys, and the impact was larger for the smartphone surveys than for the PC surveys.

The scale direction could also have an impact on the latent structure of survey estimates. For example, Chan (1991) found the factor analysis model fit tended to be higher in the positive–negative scale than in the negative–positive scale, although the item discrimination was the opposite. In another study, Krebs and Hoffmeyer-Zlotnik (2010) showed that the factor loadings, latent means and dimensional structures did not differ substantially between the two scale directions. In a more recent study, Salzberger and Koller (2013) showed that, depending on the analytical techniques, confirmatory factor analysis and the Rasch model showed differences in the effect of scale direction, and the interaction between scale direction and completion time. Lastly, Saris and Gallhofer (2007) showed that, in a meta-analysis, negative–positive scales had lower reliability but higher validity, than positive–negative scales.

This study

This study explores additional design features of slider questions and expands the literature in three directions. First, this study examines the impact of labels of radio button and slider questions on the responses of respondents. The few studies that have compared radio button with slider questions have all used endpoint labels for both slider and radio button questions (Cook *et al.* 2001; Funke *et al.* 2011; Funke & Reips 2012; Funke 2016). However, fully labelled radio buttons are also frequently used in surveys and have been shown to produce different estimates to endpoint labelled scales (Krosnick & Presser 2010; Menold *et al.* 2014). The slider questions are almost always labelled at the endpoints only. Radio button questions, particularly fully labelled ones, generate comparable estimates to slider questions. This is important for practitioners to inform them in relation to decisions on design change (switch between the two question types) and comparison of direction results. Second, the direction of slider question is also examined in this study. As reviewed above, previous research has shown an impact of the direction of radio button questions on survey responses. The user web interface of slider questions is different from the radio button questions and therefore the findings from the previous studies may or may not hold for the slider questions. Third, most of the studies in the literature examined slider questions from PC respondents. In this study, mobile respondents and PC respondents will be analysed separately to examine whether the survey device moderates the experimental effects. As increasing number of respondents are taking web surveys on their mobile devices, this will provide useful information to survey designers.

The following sections report findings from two web survey experiments designed to answer the two research questions. First, whether slider questions (with endpoint labels) produce similar or different answers from radio button questions (with both full labels and endpoint labels). Second, whether the direction of the slider questions, either positive adjective first or negative adjective first, will change the survey response.

Experiment 1

Research population

This experiment was conducted using the SurveyMonkey platform between 26 and 29 July 2016. The sample came from the SurveyMonkey Audience, an online non-probability web panel. In total, some 2,016 respondents completed this survey. They were randomly assigned to one of the six experimental conditions, as described below. The survey included 18 questions and was split across five web pages. The mode gender of respondents was female (63.7%), the mode age was 30–44 (45.3%) and the mode household income was \$50,000–\$99,999 (36.2%); 63.8% responded via mobile device and 36.2% via PC.

Design of experiment

This study used a 2×3 design with six conditions:

- fully labelled radio button 5-point scale,
- endpoint labelled radio button 5-point scale,
- slider 5-point scale,
- fully labelled radio button 7-point scale,
- endpoint labelled radio button 7-point scale and
- slider 7-point scale.

Respondents were randomly assigned to one of these six conditions. Each condition had the same 13 questions (see Appendix 1). The first six questions were from the Multidimensional Personality Questionnaire and measured self-control/impulsiveness (Caspi *et al.* 1997) using an agree/disagree Likert scale. The next four questions were satisfaction scales. The last three questions were about politics and elections, and their response options were specific to the questions. See Appendix 2 for a sample screen shot of the survey questions from the survey platform.

Measures

In the analysis, the mean score of each question in each condition was calculated and compared first across all sample respondents, as well as by response device (mobile vs PC). Second, item non-response and break-off rates were calculated across conditions. Third, the time to complete for the entire survey was also compared across conditions and by response device. A last step was to compute the Cronbach's alpha for the first six questions on self-control/impulsiveness, and to compare them across conditions and devices.

Results

The mean score for each question under each condition is presented in Figures 1 and 2. In both figures, the solid line represents slider questions, the dotted line represents endpoint labelled rating scales and the dashed line represents the fully verbally labelled rating scales. As can be seen, the mean scores were, for the most part, very close to one another. The ANOVA test showed that, other than Q7 ($p < 0.01$) and Q8 ($p < 0.01$), the difference in the mean across the three question types was not statistically significant for the 5-point scale in Figure 1. For both Q7 and Q8, the means for the fully labelled scales were lower than for the other two conditions, although the absolute difference was very small.

Figure 1 also presents the mean scores for the same set of questions by response device. As can be seen, within each device the mean scores for each question across conditions were very similar to one another. For mobile respondents, Q7 ($p < 0.01$) was the only question that showed a significant difference and the fully labelled condition had a lower mean. For PC respondents, Q7 ($p < 0.01$) and Q8 ($p < 0.01$) showed significant differences and again the fully labelled scale means were significantly lower than the other two question types.

For 7-point scales (Figure 2), there were more questions that showed significant mean difference across conditions. Specifically, for Q7, Q8, Q9, Q11 and Q12 ($p < 0.001$) the mean scores in the slider condition were significantly higher than in the other two conditions in the whole sample analysis. When examining the data by device, the exact same pattern was found among mobile respondents. For PC respondents, Q9 ($p = 0.01$) and Q11 ($p = 0.03$) were the only two significant questions, with the slider condition having the highest mean scores.

The survey break-off rate and item non-response rate were also compared across conditions. For the 5-point scale, the break-off rates for the fully labelled, endpoint labelled and slider questions were 5.2%, 4.1%

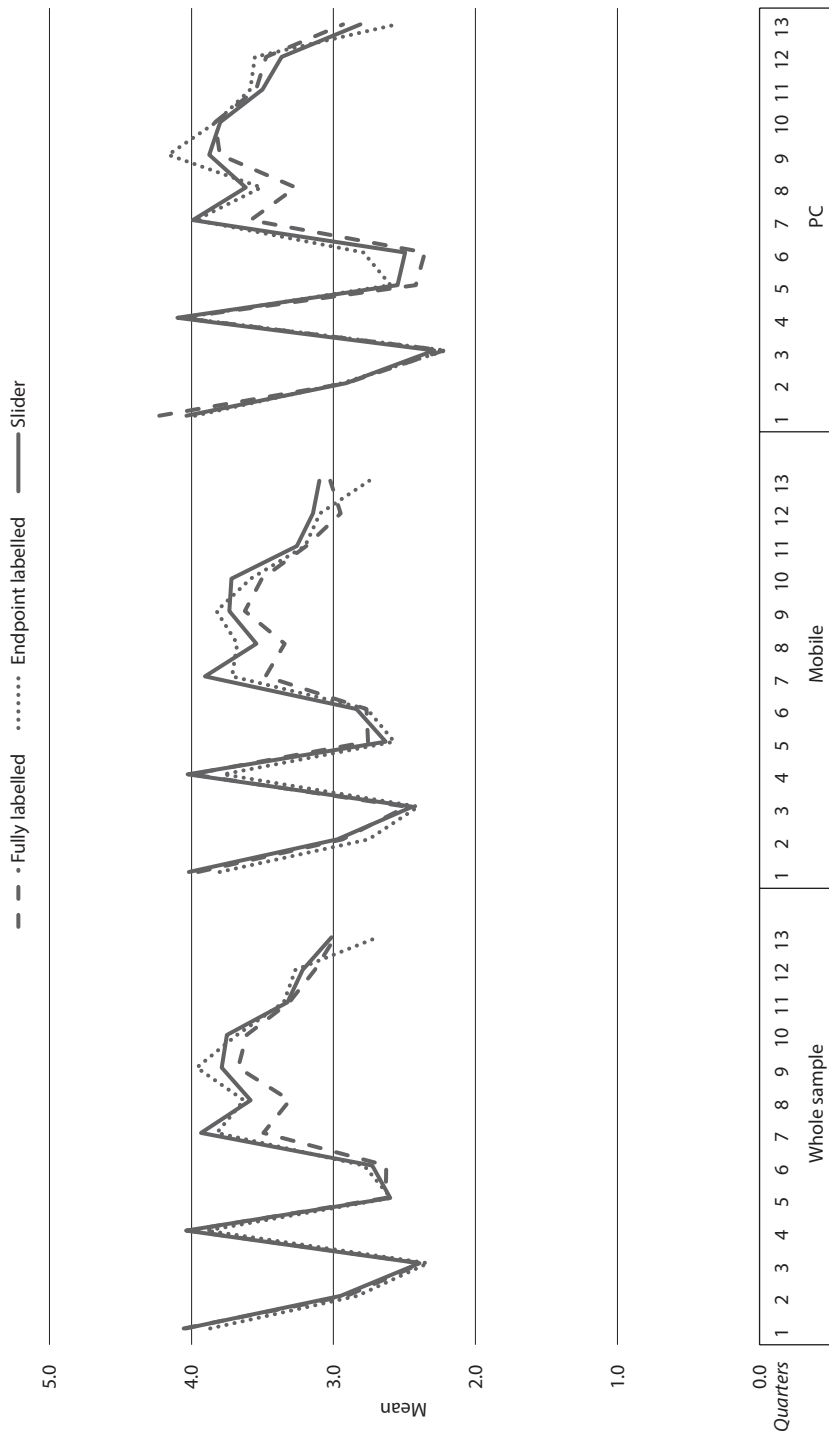


Figure 1 Response mean for 5-point scales (fully labelled, endpoint labelled, slider)

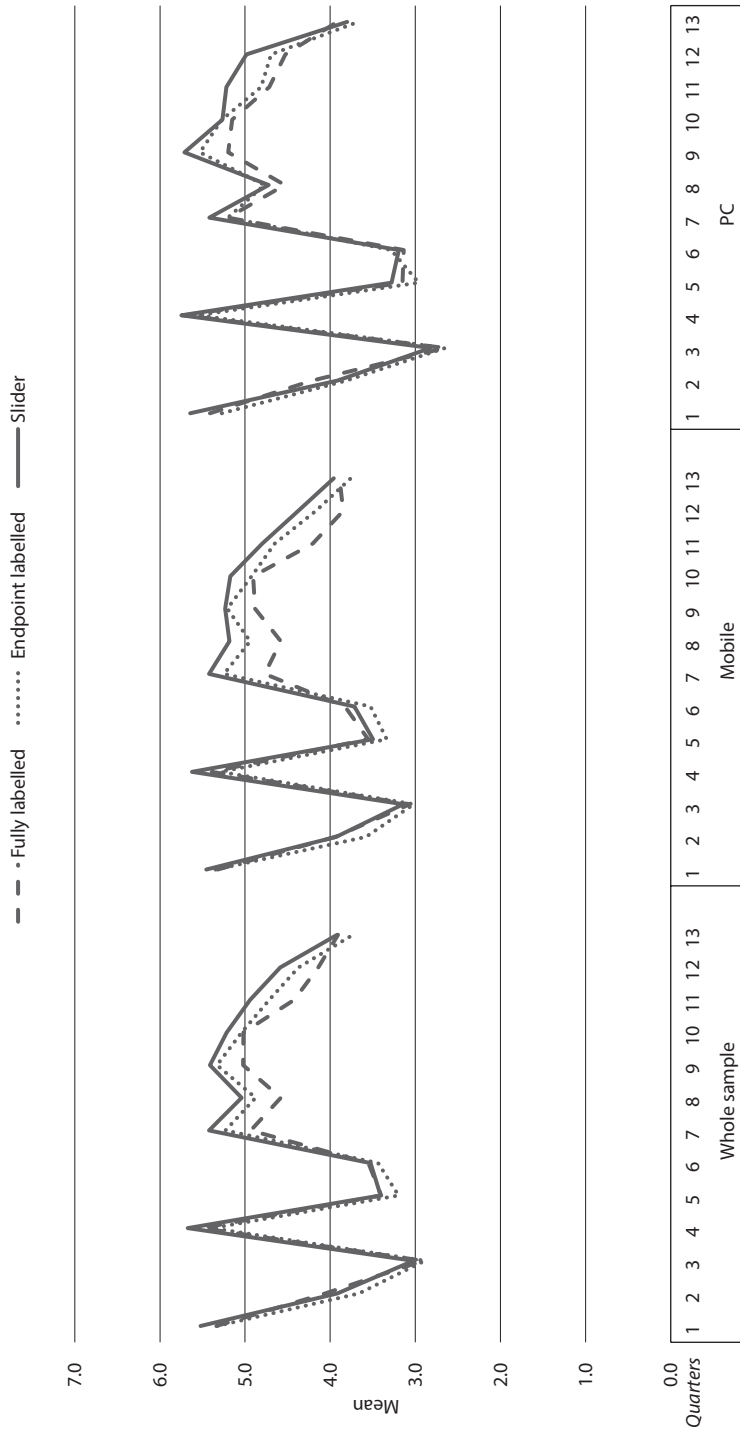


Figure 2 Response mean for 7-point scales (fully labelled, endpoint labelled, slider)

Table 1 Item non-response by experimental conditions and devices (Experiment 1)

	Whole sample				Mobile				PC			
	Fully labelled (%)	Endpoint labelled (%)	Slider (%)	F	Fully labelled (%)	Endpoint labelled (%)	Slider (%)	F	Fully labelled (%)	Endpoint labelled (%)	Slider (%)	F
<i>5-point scales</i>												
Q1	0.6	0.0	0.3	0.52	0.0	0.0	0.0	NA	1.7	0.0	0.9	0.47
Q2	0.6	0.3	2.8	7.35**	0.5	0.5	3.0	5.70*	0.8	0.0	2.6	1.66
Q3	0.3	0.3	1.7	4.63*	0.5	0.5	1.3	1.08	0.0	0.0	2.6	4.86*
Q4	0.3	0.3	1.1	2.20	0.5	0.5	0.9	0.32	0.0	0.0	1.7	3.22
Q5	0.9	0.6	1.7	1.18	0.5	0.5	1.7	2.08	1.7	0.8	1.7	0.00
Q6	0.0	0.3	3.7	18.36**	0.0	0.5	3.0	8.87**	0.0	0.0	5.2	9.94**
Q7	0.9	0.6	3.1	5.95*	0.9	0.9	3.5	4.20*	0.8	0.0	2.6	1.66
Q8	0.6	0.3	2.8	7.35**	0.5	0.5	3.0	5.70*	0.8	0.0	2.6	1.66
Q9	1.2	0.6	2.0	0.99	1.4	0.9	2.2	0.50	0.8	0.0	1.7	0.57
Q10	0.6	1.4	3.1	6.71*	0.5	1.4	3.0	4.61*	0.8	1.6	3.5	2.18
Q11	0.0	0.9	2.3	8.71**	0.0	1.4	2.2	4.42*	0.0	0.0	2.6	4.86*
Q12	0.9	0.6	4.0	9.57**	0.9	0.5	3.5	4.61*	0.8	0.0	5.2	5.95*
Q13	0.6	0.3	4.8	17.14***	0.0	0.5	3.9	11.83**	1.7	0.0	6.1	4.69*
<i>7-point scales</i>												
Q1	0.6	0.3	0.6	0.00	0.0	0.0	0.4	1.38	1.5	0.8	0.8	0.27
Q2	0.0	0.3	2.0	9.17**	0.0	0.0	3.0	9.90**	0.0	0.8	0.0	0.00
Q3	0.0	0.3	2.3	10.68**	0.0	0.0	2.6	8.46**	0.0	0.8	1.7	2.29
Q4	0.6	0.6	1.4	1.42	0.5	0.4	2.2	2.96	0.7	0.8	0.0	0.65
Q5	0.3	0.8	2.8	8.62**	0.5	0.0	2.6	4.81**	0.0	2.3	3.4	4.06*
Q6	0.6	0.3	3.4	9.97**	1.0	0.0	3.0	3.59	0.0	0.8	4.2	7.15**
Q7	0.0	0.8	3.1	13.02***	0.0	0.9	3.9	10.25**	0.0	0.8	1.7	2.29
Q8	0.0	0.3	3.4	16.81***	0.0	0.0	3.5	11.36**	0.0	0.8	3.4	5.48*
Q9	0.6	0.3	2.8	7.31**	0.5	0.0	2.6	4.81**	0.7	0.8	3.4	2.70
Q10	0.0	0.3	2.8	13.72***	0.0	0.0	3.5	11.36**	0.0	0.8	1.7	2.29
Q11	0.3	1.1	1.7	3.31	0.5	1.3	2.2	2.18	0.0	0.8	0.8	0.89
Q12	0.3	0.3	3.1	11.55**	0.5	0.0	3.9	8.85**	0.0	0.8	1.7	2.29
Q13	0.0	1.1	3.7	15.03***	0.0	1.3	3.9	9.29**	0.0	0.8	3.4	5.48*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and 3.5%, respectively, and they were not significantly different from one another ($\chi^2 = 1.26$, $p = 0.53$). For the 7-point scale, the break-off rates for the fully labelled, endpoint labelled and slider questions were 5.3%, 4.3% and 3.8%, respectively, and they were also not significantly different from one another ($\chi^2 = 0.98$, $p = 0.61$). The item non-response for each question under each condition and device is presented in Table 1. The item non-response rates for both the fully labelled and endpoint labelled rating scales were very low and similar – under 1% for the majority of the cases. The item non-response rate for the slider questions tended to be significantly higher, although for the most part it was around 3–4%, a reasonably low item non-response rate.

The median time to complete the entire survey is shown in Table 2. For the 5-point scale, the median time to finish the slider questions was 118.5 seconds, significantly longer than for the other two rating scale conditions. For the 7-point scale, the slider questions also took slightly longer to complete, but the difference was not significant. Similarly, for mobile respondents, the slider questions took significantly longer to complete for the 5-point scale but not for the 7-point scale. The median time to complete for the PC respondents was similar across conditions for both 5- and 7-point scales.

The Cronbach's alpha was calculated for the first six questions on self-control/impulsiveness from the Multidimensional Personality Questionnaire. The results for both 5- and 7-point scales under each condition for the whole sample, as well as by device, are presented in Table 3. Among all the pairwise comparisons, there were only two significant differences: for the 7-point scale, the Cronbach's alpha was higher for the slider than the endpoint labelled rating scales for the whole sample ($p = 0.03$) and for PC respondents ($p = 0.01$).

Table 2 Completion time by experimental conditions (Experiment 1)

	Fully labelled	Endpoint labelled	Slider	χ^2	p
<i>Whole sample</i>					
5-point	108.0	101.5	118.5	15.02	0.00
7-point	113.0	113.5	120.0	1.89	0.39
<i>Mobile</i>					
5-point	106.0	95.0	117.0	11.53	0.00
7-point	106.0	111.0	117.0	1.74	0.42
<i>PC</i>					
5-point	116.0	108.5	124.0	3.45	0.18
7-point	119.0	117.0	125.0	3.39	0.18

Table 3 Cronbach's alpha by experimental conditions (Experiment 1)

	Alpha			<i>p</i>		
	Fully labelled	Endpoint labelled	Slider	Fully vs endpoint	Fully vs slider	Endpoint vs slider
<i>Whole sample</i>						
5-point	0.46	0.37	0.34	0.16	0.06	0.65
7-point	0.39	0.30	0.44	0.18	0.39	0.03
<i>Mobile</i>						
5-point	0.49	0.41	0.34	0.28	0.06	0.43
7-point	0.40	0.33	0.40	0.41	0.96	0.43
<i>PC</i>						
5-point	0.43	0.25	0.27	0.13	0.20	0.86
7-point	0.36	0.24	0.51	0.28	0.16	0.01

At the end of the survey, the respondents were asked to evaluate the survey design, the survey-taking experience and the survey difficulty. For the 5-point scale, the respondents' evaluation of the survey design ($F = 0.90$, $p = 0.34$), the survey experience ($F = 0.09$, $p = 0.77$) and the easiness ($F = 2.44$, $p = 0.12$) were similar across the conditions. For the 7-point scale, the survey design ($F = 0.36$, $p = 0.55$) and experience ($F = 0.20$, $p = 0.65$) did not differ significantly across conditions, but respondents found the slider questions to be easier than the radio buttons ($F = 7.01$, $p < 0.01$).

Experiment 2

Research population

The experiment was conducted using the SurveyMonkey platform and the samples were recruited from the SurveyMonkey Audience panel between 20 and 21 September 2016. In total, some 1,211 respondents completed the survey and were randomly assigned to one of the four conditions, as described below. The mode gender was female (56.4%), the mode age was 30–44 (37.7%) and the mode household income was \$25,000–\$49,999 (17.69%); 38.1% responded via mobile device and 61.9% via PC.

Design of experiment

This study used a 2×2 experiment design. The first factor was the scale length, either a 5-point or a 7-point scale. The second factor was the scale

direction, either from positive to negative or from negative to positive. The question wording was identical to that in Experiment 1, as shown in Appendix 1. Appendix 3 shows screen shots for selected questions from the survey platform.

Measures

In this study, similar measures were used to those for Experiment 1. They included mean scores, item non-response and break-off, time to complete and Cronbach's alpha. Data were analysed aggregately as well as by survey-taking device.

Results

Figures 3 and 4 present the mean scores for both scale directions. The solid line represents the negative–positive scale and the dashed line represents the positive–negative scale. The response options for the positive first scales were re-coded so that, for both scale directions, a higher mean score suggests a more positive response. For the 5-point scale, the two lines for the two scale directions aligned very well, suggesting similar responses to both scale directions. Q4 ($p = 0.01$) and Q7 ($p = 0.03$) were the only two with significant differences in the whole sample analysis. For both questions, when the negative option was presented on the left side of the slider question, respondents were more likely to provide a positive response (right side of the slider). For mobile respondents, none of the differences was significant. For PC respondents, Q4 ($p < 0.001$) again showed a significant difference in the same trend as in the whole sample.

For the 7-point scale, in the whole sample analysis, the mean scores for Q1 ($p = 0.03$), Q4 ($p = 0.04$) and Q9 ($p < 0.001$) were significantly higher (more positive) in the negative–positive scales than the positive–negative scale. For mobile respondents, there were also higher mean scores in the negative first scale for Q9 ($p < 0.001$) and Q12 ($p = 0.04$). For PC respondents, the mean score for Q9 ($p = 0.03$) was significantly higher in the negative first condition than in the positive first condition.

For the 5-point scale, the break-off rates were 6.5% for the negative first scale and 5.5% for the positive scale, and the difference was not significant ($\chi^2 = 0.11$, $p = 0.74$). For the 7-point scale, the break-off rates for the negative first and positive first scales were 4.0% and 6.1%, respectively ($\chi^2 = 1.08$, $p = 0.30$). The item non-response rates are shown in Table 4. For

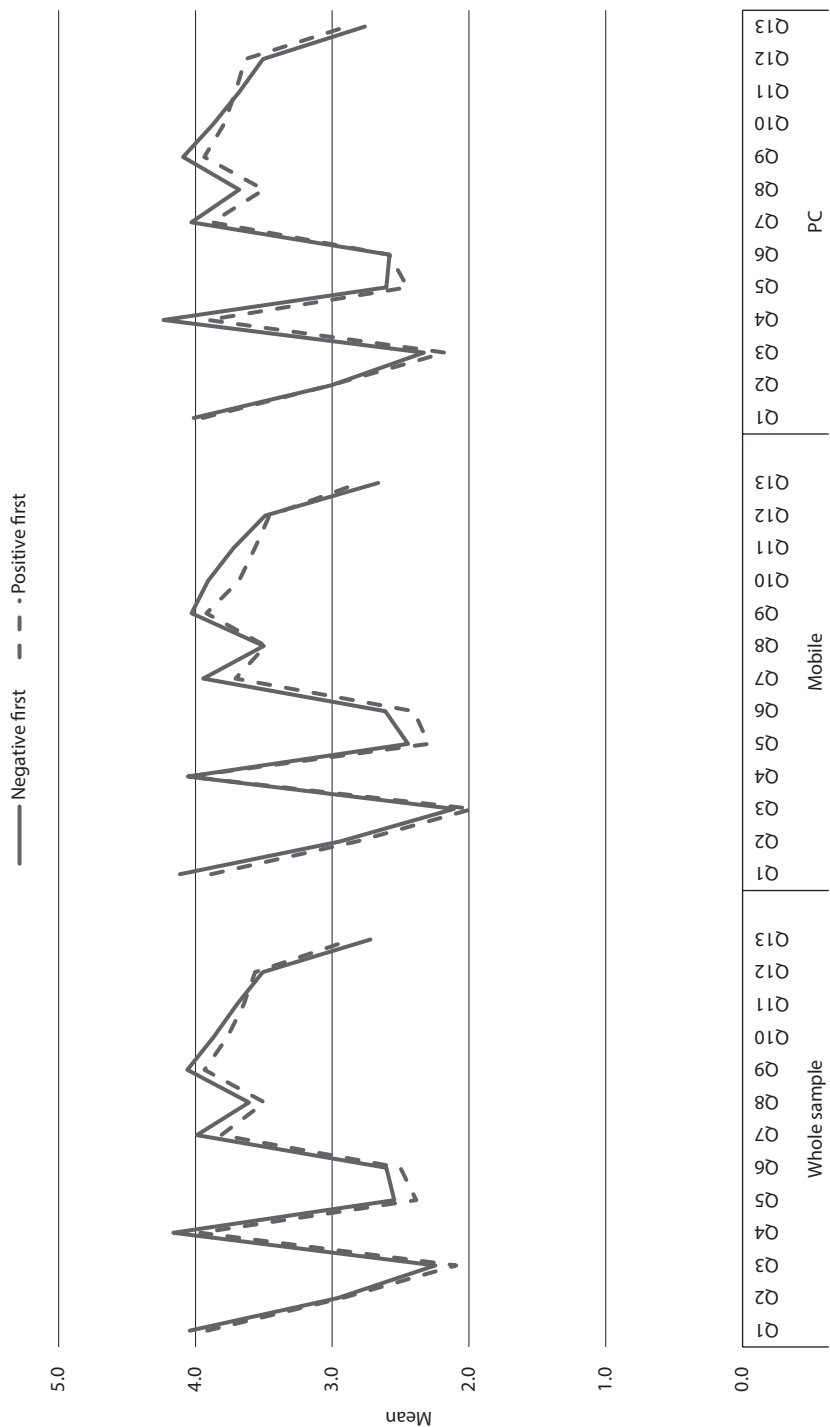


Figure 3 Response mean for 5-point scales (negative first vs positive first)

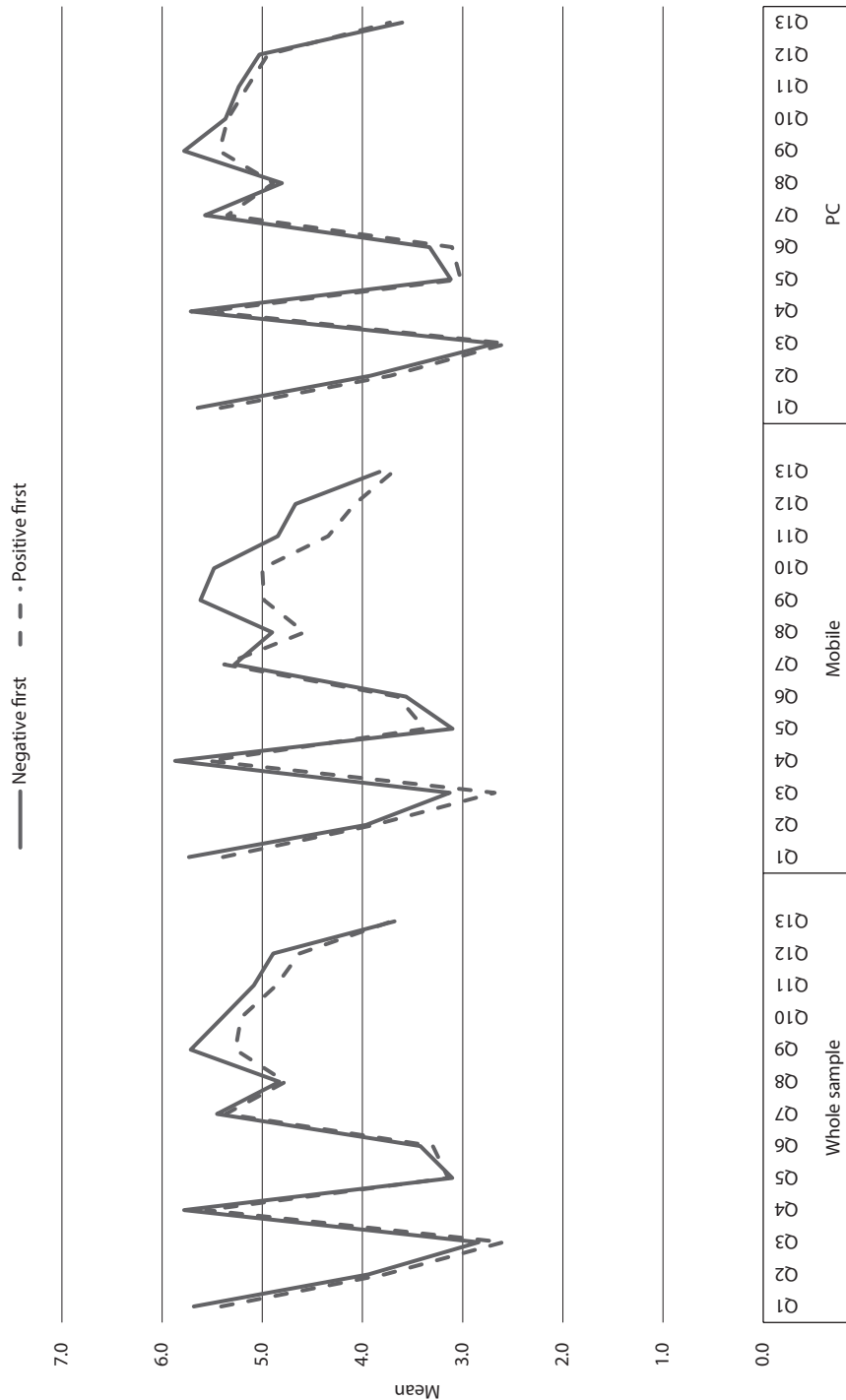


Figure 4 Response mean for 7-point scales (negative first vs positive first)

Table 4 Item non-response by experimental conditions and devices (Experiment 2)

	Whole sample			Mobile			PC		
	Negative first (%)	Positive first (%)	F	Negative first (%)	Positive first (%)	F	Negative first (%)	Positive first (%)	F
<i>5-point scales</i>									
Q1	2.4	1.3	1.04	3.6	0.0	4.26*	1.2	2.1	0.47
Q2	4.5	1.9	3.15	2.7	0.9	1.09	5.3	2.6	1.70
Q3	2.8	2.6	0.02	0.9	1.8	0.30	3.5	3.2	0.04
Q4	1.4	1.6	0.06	0.0	0.0	NA	1.8	2.6	0.31
Q5	2.4	3.9	1.05	0.0	1.8	1.95	3.5	5.3	0.63
Q6	6.2	4.9	0.53	6.4	4.4	0.43	5.9	5.3	0.07
Q7	4.5	5.5	0.33	3.6	0.9	1.95	4.7	8.4	1.99
Q8	4.5	3.6	0.33	3.6	0.9	1.95	4.7	5.3	0.06
Q9	1.7	1.6	0.01	0.0	0.9	0.97	2.4	2.1	0.03
Q10	3.5	3.9	0.08	1.8	3.5	0.61	4.1	4.2	0.00
Q11	2.8	4.2	0.93	1.8	0.9	0.37	2.9	6.3	2.27
Q12	4.5	3.2	0.63	3.6	0.9	1.95	4.7	4.7	0.00
Q13	7.3	5.8	0.49	8.2	3.5	2.24	6.5	7.4	0.11
<i>7-point scales</i>									
Q1	2.8	0.6	4.40*	2.7	0.8	1.14	2.8	0.5	3.28
Q2	2.4	1.9	0.23	1.8	0.8	0.40	2.8	2.5	0.04
Q3	3.1	3.4	0.04	2.7	4.2	0.40	3.4	3.0	0.05
Q4	2.8	1.9	0.56	2.7	0.8	1.14	2.8	2.5	0.04
Q5	4.1	3.1	0.49	2.7	1.7	0.27	5.1	4.0	0.27
Q6	4.5	4.6	0.01	2.7	3.4	0.09	5.6	5.5	0.01
Q7	4.8	4.6	0.01	2.7	4.2	0.40	6.2	5.0	0.27
Q8	3.4	3.7	0.03	0.9	5.1	3.40	5.1	3.0	1.09
Q9	3.1	2.2	0.54	0.9	1.7	0.28	4.5	2.5	1.17
Q10	3.1	3.7	0.17	0.9	1.7	0.28	4.5	5.0	0.04
Q11	2.4	2.2	0.04	0.9	0.8	0.00	3.4	2.5	0.27
Q12	3.1	2.2	0.54	1.8	1.7	0.00	4.0	2.0	1.28
Q13	4.5	3.7	0.24	4.5	2.5	0.65	4.5	4.0	0.07

* $p < 0.05$

Table 5 Completion time by experimental condition (Experiment 2)

	Negative first	Positive first	χ^2	<i>p</i>
<i>Whole sample</i>				
5-point	125.0	131.5	1.94	0.16
7-point	123.0	127.0	0.83	0.36
<i>Mobile</i>				
5-point	115.5	147.5	9.80	0.00
7-point	132.0	128.0	0.02	0.89
<i>PC</i>				
5-point	130.5	123.5	0.28	0.60
7-point	119.0	125.0	0.91	0.34

most questions, the item non-response rates were relatively low and the difference between the two scale directions was small and not significant. This was the case for the whole sample, and for mobile and PC separately too.

The median time to complete was compared between the two slider directions in Table 5. For the 5-point scale, the positive–negative scales took longer to complete than the opposite direction ($\chi^2 = 9.80, p < 0.001$). For the other comparisons, the differences were small and not significant.

Next, the Cronbach's alpha for the first six questions (self-control/impulsiveness) was calculated for the 5- and 7-point scales by experimental conditions and devices (Table 6). The differences were relatively small and not significant for all comparisons.

Lastly, as in Experiment 1, the three survey evaluation questions were analysed. For both 5-point scales, respondents rated the negative–positive

Table 6 Cronbach's alpha by experimental condition (Experiment 2)

	Alpha		<i>p</i>
	Negative first	Positive first	
<i>Whole sample</i>			
5-point	0.38	0.34	0.60
7-point	0.29	0.27	0.79
<i>Mobile</i>			
5-point	0.41	0.29	0.33
7-point	0.31	0.22	0.48
<i>PC</i>			
5-point	0.36	0.37	0.92
7-point	0.27	0.30	0.76

slider to be better designed ($F = 5.10, p = 0.02$) and easier ($F = 2.33, p = 0.13$), but survey-taking experience was similar between these two directions ($F = 2.33, p = 0.13$). For the 7-point scale, the survey design ($F = 0.65, p = 0.42$), the survey experience ($F = 0.87, p = 0.35$) and the survey difficulty ($F = 0.29, p = 0.59$) were similar between the scale directions.

Discussion

Using data collected from two web survey experiments, this study examined the impact of labelling and the scale direction of slider questions on survey responses. In the first experiment, three question types, including radio button with full labels, radio button with endpoint labels and slider questions, were compared. It reached the following conclusions.

First, for both 5-point and 7-point scales, the mean scores for fully labelled radio button, endpoint labelled radio button and slider questions were very similar to one another. With very few exceptions, the differences were not statistically significant. Even for the significant cases, the size of the difference was extremely small. Also, both PC and mobile respondents showed similar mean scores across the question types. Second, the break-off rate for the slider questions was slightly but not significantly lower than that for both the radio button questions. However, the item non-response rate was significantly higher for slider questions than for both the radio button (fully labelled and endpoint labelled) conditions. Third, the time to complete was slightly longer for the slider than for the radio button questions, although not substantially. Fourth, the scale reliability, as measured by the Cronbach's alpha, was similar across the three question types. Lastly, the respondents' subjective judgements of the surveys were largely similar across question types.

In the second experiment, the scale direction, either positive–negative or negative–positive, was crossed with the scale length (5-point or 7-point scales). Similar measures were used as in Experiment 1 and reached the following conclusions. First, with very few exceptions, the mean scores between the two scale directions were not significantly different for both the 5-point and 7-point scales. The patterns for mobile and PC respondents were very similar as well. Second, both the survey break-off rates and item non-response rates were similar between the two scale directions for 5-point and 7-point scales. Third, the time to complete was very close between the two scale directions. Fourth, the Cronbach's alpha showed similar scale reliability between the two scale directions. Lastly, respondents gave similar evaluations to the two scale directions.

For both experiments, the mean score differences were more prevalent among 7-point scales than 5-point scales, and the differences were driven primarily by mobile respondents. One possible explanation was the limited screen size and readability of the fully labelled scales on mobile devices. The end-labelled scales were more similar to the slider questions in that only endpoints were labelled verbally. Consequently, the mean scores for the end-labelled scales and sliders were more similar than the mean scores between fully labelled scales and sliders. For Experiment 2, the experimental effects of scale direction were larger for mobile than PC respondents, which was consistent with the findings reported by Stapleton (2013). Also, interestingly, some of the positive first scales had lower mean scores than negative first scales. Although it is not clear why this happened, future research can further explore this area by using techniques like eye tracking or mouse movement tracking. It will be valuable for researchers to dive deep into the mechanism of this phenomenon. It's also interesting to note that Q7 and several subsequent questions are those with significant mean differences. Q7 asked about respondents' experience with the SurveyMonkey Contribute, the online survey panel from which the data were collected for this study. Since the survey respondents were all part of this panel, and the survey was conducted using the SurveyMonkey platform, the responses could be affected by the question content. The difference for questions after Q7 could be due to a halo effect. To test this, future research could randomise the order of the questions to see if the result will still hold.

There are some important limitations to this study. First, it focused only on 5- and 7-point discrete variables, while slider questions can also be used for scales with more response options and continuous variables. Future research should examine whether the number of response options moderates the effects shown in this study. Second, the device was not randomly assigned but self-selected. Although this reflects what happens in the real world, it prevents us from making causal inference regarding the device effect. Third, the VAS was not tested in this study. Given the similar functionality and user experience between these two question types, the findings in this study should, it is hoped, still hold for VAS questions, but this needs to be tested empirically. Fourth, there are other design features, such as horizontal scales and different scale labels (numeric vs verbal, endpoint vs full labels), that were not tested in the experiments. This study chose to test the fully labelled scale as it has been shown to produce better data quality (Krosnick & Presser 2010). Also, a horizontal scale may not work well on mobile devices, especially with a 7-point scale, as some categories may not be visible for smaller screens. Last, but not least, the

initial position of the slider handle could also impact the survey response, as found by Buskirk *et al.* (2015). In this study, the handle was positioned in the middle for all questions so that it would not confound the results. However, future studies should consider varying the starting position of the slider and comparing this with the radio button questions.

Despite the limitations, this study showed that slider questions provided similar measurements to radio button questions, either with full labels or endpoint labels, with one caveat: the item non-response rate for the slider questions was somewhat higher than for the radio buttons. Also, the direction of slider questions, or the endpoint verbal labels, did not change the survey estimates of the slider questions. Given this, survey researchers and practitioners should feel safe to choose between slider and radio button questions, or placing the positive or negative adjective on either endpoint of the slider questions.

Appendix 1: Question wordings and response options

1. I keep close track of where my money goes.
2. I often stop one thing before completing it and start another.
3. I often act without thinking.
4. Before I get into a new situation, I like to find out what to expect from it.
5. I am often not as cautious as I should be.
6. I often prefer to 'play things by ear' rather than to plan ahead.

Response options for questions 1–6:

7-point scale: Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree

5-point scale: Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree

7. How satisfied are you in general with your experience using SurveyMonkey Contribute?
8. How satisfied are you in general with your current internet service provider?
9. How satisfied are you in general with your current primary mode of transportation (e.g. car, bike, walk, train)?

10. How satisfied are you in general with your current living arrangement?

Response options for questions 7–10:

7-point scale: Not at all satisfied, A tiny bit satisfied, A little satisfied, Somewhat satisfied, Satisfied, Very satisfied, Extremely satisfied

5-point scale: Not at all satisfied, A little satisfied, Somewhat satisfied, Very satisfied, Extremely satisfied

11. How closely have you been following the U.S. presidential election?

7-point scale: Not at all closely, A tiny bit closely, A little closely, Somewhat closely, Closely, Very closely, Extremely closely

5-point scale: Not at all closely, A little closely, Somewhat closely, Very closely, Extremely closely

12. How interested are you in national politics?

7-point scale: Not at all interested, A tiny bit interested, A little interested, Somewhat interested, Interested, Very interested, Extremely interested

5-point scale: Not at all interested, A little interested, Somewhat interested, Very interested, Extremely interested

13. In general, how would you describe your views on most political issues?

7-point scale: Very liberal, Liberal, Somewhat liberal, Moderate/Middle of the road, Somewhat conservative, Conservative, Very conservative

5-point scale: Very liberal, Somewhat liberal, Moderate/Middle of the road, Somewhat conservative, Very conservative

Survey evaluation

What did you think of the design of this survey? Poor, Fair, Good, Great, Excellent.

How positive was your overall experience taking this survey? Not at all positive, A little positive, Somewhat positive, Very positive, Extremely positive.

How easy was it to answer the questions in this survey? Very difficult, Somewhat difficult, Neither easy nor difficult, Somewhat easy, Very easy.

Appendix 2: Screen shots of sample survey questions (Experiment 1)

SurveyMonkey Research Survey

How much do you agree or disagree with the following statements?

I keep close track of where my money goes.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Somewhat disagree
- ☐ Neither agree nor disagree
- ☐ Somewhat agree
- ☐ Agree
- ☐ Strongly agree

I often stop one thing before completing it and start another.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Somewhat disagree
- ☐ Neither agree nor disagree
- ☐ Somewhat agree
- ☐ Agree
- ☐ Strongly agree

SurveyMonkey Research Survey

How much do you agree or disagree with the following statements?

I keep close track of where my money goes.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Somewhat disagree
- ☐ Neither agree nor disagree
- ☐ Somewhat agree
- ☐ Agree
- ☐ Strongly agree

I often stop one thing before completing it and start another.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Somewhat disagree
- ☐ Neither agree nor disagree
- ☐ Somewhat agree
- ☐ Agree
- ☐ Strongly agree

SurveyMonkey Research Survey

How much do you agree or disagree with the following statements?

I keep close track of where my money goes.

1 = Strongly disagree

7 = Strongly agree

☐

I often stop one thing before completing it and start another.

1 = Strongly disagree

7 = Strongly agree

☐

Appendix 3: Screen shots of sample survey questions (Experiment 2)

SurveyMonkey Research Survey

How much do you agree or disagree with the following statements?

I keep close track of where my money goes.

Strongly agree

Strongly disagree

☐

I often stop one thing before completing it and start another.

Strongly agree

Strongly disagree

☐

I often act without thinking.

Strongly agree

Strongly disagree

☐

SurveyMonkey Research Survey

How much do you agree or disagree with the following statements?

I keep close track of where my money goes.

Strongly disagree

Strongly agree

☐

I often stop one thing before completing it and start another.

Strongly disagree

Strongly agree

☐

I often act without thinking.

Strongly disagree

Strongly agree

☐

References

- Belson, W.A. (1966) Effects of reversing presentation order of verbal rating scales. *Journal of Advertising Research*, 6, 4, pp. 30–37.
- Buskirk, T.D., Saunders, T. & Michaud, J. (2015) Are sliders too slick for surveys? An experiment comparing slider and radio button scales for smartphone, tablet and computer based surveys. Available online at: www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.9_Heft_2/MDA_Vol9_2015-2_Buskirk.pdf (accessed 30 June 2017).
- Caspi, A., Begg, D., Dickson, N., Harrington, H., Langley, J., Moffitt, T.E. & Silva, P.A. (1997) Personality differences predict health-risk behaviors in young adulthood: evidence from a longitudinal study. *Journal of Personality and Social Psychology*, 73, 5, pp. 1052–1063.
- Chan, J.C. (1991) Response-order effect in Likert-type scales. *Educational and Psychological Measurement*, 51, pp. 531–540.
- Christian, L.M., Dillman, D.A. & Smyth, J.D. (2008) The effects of mode and format on answers to scalar questions in telephone and web surveys, in Lepkowski, J.M., Tucker, N.C., Brick, J.M., de Leeuw, E.D., Japec, L., Lavrakas, P.J., Link, M.W. & Sangster, R.L. (eds) *Advances in Telephone Survey Methodology*, New York: Wiley-Interscience, pp. 250–275.
- Cook, C., Heath, F., Thompson, R.L. & Thompson, B. (2001) Score reliability in web or internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, 61, 4, pp. 697–706.
- Dillman, D.A., Brown, T.L., Carlson, J.E., Carpenter, E.H., Lorenz, F.O., Mason, R., Saltiel, J. & Songster, R.L. (1995) Effects of category order on answers in mail and telephone surveys. *Rural Sociology*, 60, 4, pp. 674–687.
- Funke, F. (2016) A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*, 34, 2, pp. 244–254 (first published 9 March 2015).
- Funke, F. & Reips, U.-D. (2012) Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*, 24, 3, pp. 310–327.
- Funke, F., Reips, U.-D. & Thomas, R.K. (2011) Sliders for the smart: type of rating scale on the web interacts with educational level. *Social Science Computer Review*, 29, 2, pp. 221–231.
- Krebs, D. & Hoffmeyer-Zlotnik, J.H.P. (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, pp. 118–127.
- Krosnick, J.A. & Presser, S. (2010) Question and questionnaire design. *Handbook of Survey Research*, 2, pp. 263–314.
- Menold, N., Kaczmirek, L., Lenzner, T. & Neusar, A. (2014) How do respondents attend to verbal labels in rating scales? *Field Methods*, 26, 1, pp. 21–39.
- Roster, C.A., Lucianetti, L. & Albaum, G. (2015) Exploring slider vs categorical response formats in web-based surveys. *Journal of Research Practice*, 11, 1. Available online at: <http://jrp.icaap.org/index.php/jrp/article/view/509/413>.
- Salzberger, T. & Koller, M. (2013) Towards a new paradigm of measurement in marketing. *Journal of Business Research*, 66, 9, pp. 1307–1317.
- Saris, W.E. & Gallhofer, I.N. (2007) Estimation of the effects of measurement characteristics on the quality of survey questions, in Saris, W.E. & Gallhofer, I.N., *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, NJ: John Wiley & Sons, Inc., pp. 237–253.
- Stapleton, C. (2013) The smart (phone) way to collect survey data. *Survey Practice*, 6, 2. Available online at: <http://surveypractice.org/index.php/SurveyPractice/article/view/75> (accessed 30 June 2017).

About the authors

Mingnan Liu is a quantitative researcher at Facebook. Prior to that, he worked at SurveyMonkey as a survey scientist. He received his doctoral degree in survey methodology from the University of Michigan. His publication on survey research and methodology has appeared in journals including *Public Opinion Quarterly*, *Computers in Human Behavior*, *Social Science Computer Review*, *Journal of Cross-Cultural Psychology*, among others.

Email: mingnanliu@fb.com

Copyright of International Journal of Market Research is the property of Warc LTD and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.