Effects of Scale Direction on Response Style of Ordinal Rating Scales

Mingnan Liu[1] and Florian Keusch[2]

Although ordinal rating scales have received much research attention in survey methodology literature, the direction of the rating scales has not been as extensively studied as other design features. Research on scale direction effect has mainly focused on the influence on response distribution, while largely overlooking its impact on latent constructs. This study examines the scale direction effect on extreme and acquiescent response style latent class variables in an experiment embedded in a national probability sample. We found a higher level of acquiescent response style from scales starting with positive adjective words using a web survey. No significant effect of scale direction was detected on extreme response style or in a face-to-face survey (with show cards). This study also demonstrates that scale direction does not affect the substance latent class variables, once the response style latent class variables are included in the model. Implications of these findings and future research directions are discussed.

Key words: Rating scale; scale direction; response style; latent class analysis; data collection mode.

1. Background and Literature Review

Attitudes, as different from behaviors, are not directly observable. In surveys, attitudes are typically measured by a series of items with ordinal rating scales. Given the popularity of attitudinal questions, many research efforts have been devoted to examining measurement error associated with this type of question, which leads to a very rich body of literature on this topic (for a review, see Krosnick and Presser 2010). When designing a rating scale, many decisions have to be made by researchers, one of which is the scale direction. In rating scales, the two endpoints are usually defined by two adjective words. One way to design the rating scale is to put the positive/high adjective on the left side/top of the scale and the negative/low adjective on the right side/bottom of the scale. Another way is to reverse the order of the response options by putting the negative/low adjective first and the positive/high adjective last. This seemingly trivial design decision could affect response behavior and result in a shift in the response distribution – the 'scale direction effect' (Belson 1966).

   Research on the scale direction effect dates back more than half a century (Belson 1966) and several studies focused on the impact on the univariate distribution. For example,

---

[1] SurveyMonkey, 101 Lytton Avenue, Palo Alto, CA 94301, U.S.A. Email: mingnanliu@gmail.com
[2] University of Mannheim, 68131 Mannheim, Germany. Email: f.keusch@uni-mannheim.de

He et al. (2014) report in their summary of empirical studies involving an experiment on scale direction that twelve out of 27 experiments found a shift in responses based on the starting point. That is, the high/positive end of the scale was endorsed more often when the scale started with the high/positive end, and the low/negative end of the scale attracted more endorsement when the scale began with the low/negative end. Dillman et al. (1995) examined scale direction effects in 82 experiments in twelve surveys using ordinal scales over the telephone and in mail questionnaires. Out of 21 experiments involving mail surveys, only one showed a significant scale direction effect. Three out of 22 telephone experiments showed a significant effect. Furthermore, of the fifteen experiments using both mail and telephone, only two showed a significant effect across both modes. The authors concluded that the scale direction effect "has been overestimated by past research" (p. 674). Another set of experiments by Christian et al. (2008) tested presenting the positive option versus the negative option first using five-point ordinal scales in both telephone and web surveys. None of the comparisons in each mode yielded significant differences. A recent experiment by Stapleton (2013) on scale direction effects among respondents who could choose from one of two modes (PC and mobile device) showed that the univariate distribution shifted by the scale direction in both modes, although mobile respondents were more prone to the scale direction effect than PC respondents.

In addition to the distribution of response options, researchers have also examined the impact of scale direction on the latent structure of estimates. For example, Chan (1991) administered five items with five response options to a group of high school students and in a factor analysis, the results showed that the positive-negative scale had a better model fit than the negative-positive scale. However, item discrimination is higher in the latter than the former scale. These findings mean that the two forms of scale direction produce different estimations of the latent trait in the same group of individuals. A study by Krebs and Hoffmeyer-Zlotnik (2010) showed no substantial difference in terms of the dimensional structure, factor loadings, and the latent means between the two scale directions. A recent study adopted two analytical approaches – the Rasch model and confirmatory factor analysis – to examine the interaction effect between scale direction and response speed (Salzberger and Koller 2013). The two modeling approaches revealed different findings, including differential interaction effects with response speed. Saris and Gallhofer (2007) conducted a meta-analysis of multitrait–multimethod experiments on question reliability and validity. Scale direction is one of the factors they examined, and the findings showed that providing the negative option first reduced the reliability but improved the validity of the responses.

Taken together, empirical studies report mixed findings with regard to the influence of the direction of rating scales on response. Some studies found no significant difference in terms of univariate distribution (Dillman et al. 1995; Christian et al. 2008), while others did find a difference (Krebs and Hoffmeyer-Zlotnik 2010; Belson 1966; Stapleton 2013). Also, some studies found no latent structure/underlying structure difference between the two scale directions (Krebs and Hoffmeyer-Zlotnik 2010), while others concluded that the latent variables differed by the scale direction (Chan 1991; Saris and Gallhofer 2007) and different modeling approaches reached different conclusions (Salzberger and Koller 2013). There was no obvious difference between

interviewer-administered and self-administered modes (Dillman et al. 1995; Christian et al. 2008).

Response style is another form of measurement bias that is particularly associated with ordinal rating scales. Response style refers to the tendency of choosing certain response options from an ordinal rating scale based on some question content-irrelevant information rather than the question itself (Paulhus 1991). Two of the most frequently studied response styles are acquiescent response style (ARS) and extreme response style (ERS). ARS refers to the tendency to agree with a statement regardless of the question content (Baumgartner and Steenkamp 2001). ERS refers to the tendency to select the endpoints of a rating scale more frequently than other response options (Paulhus 1991). Previous research has shown that several design features of ordinal rating scales (e.g., the presence of a midpoint, scale length, labeling) can impact response styles (Moors 2008; Moors et al. 2014; Kieruj and Moors 2010). For example, Moors (2008) found that extreme response style existed regardless of the presence or absence of a middle response category. Kieruj and Moors (2013) examined the impact of scale length on both ARS and ERS. Specifically, they studied Likert scales that ranged from 5–11 points and found that both types of response styles existed regardless of scale length. More recently, Moors et al. (2014) examined the labeling of a Likert scale and its impact on response styles. The authors demonstrated that both ARS and ERS existed, regardless of whether numeric or verbal labels were used, and whether the scales were fully or endpoint labeled. As will be discussed in detail below, survey satisficing and anchoring-and-adjustment are potential causes of scale direction effects. Reinforced by acquiescence, these mechanisms can manifest themselves as response styles.

2. Conceptual Framework and Expectations

Different theories have been brought forward when trying to explain the influence of scale direction on response behavior to rating scales. One explanation sees the scale direction effect as a special case of the primacy effect observed in survey modes where lists of categorical response options are presented visually. Respondents are more likely to choose options offered first than later in such lists of unordered response options, due to response satisficing (Krosnick 1991). This is because, unlike optimizers, satisficers choose a response option that is good enough without going through the whole response option list and selecting the best answer. This approach is cognitively less burdensome than providing an optimized response option. As a result, satisficing has also been used as a potential explanation for scale direction effects in ordinal rating scales (e.g., Krebs and Hoffmeyer-Zlotnik 2010).

However, several studies also show that the scale direction effect is not stronger in situations conducive to satisficing (Carp 1974; Mingay and Greenwell 1989), and scale direction effects are also observed in surveys employing aural administration (Kalton et al. 1978; Mingay and Greenwell 1989; Yan and Keusch 2015). Yan and Keusch (2015) propose that the scale direction effect may also be caused by respondents' use of anchoring-and-adjustment (Tversky and Kahneman 1974) when constructing and mapping their answers to one of the scale points. The basic idea is that people make numerical estimates when under uncertainty, by anchoring on an initial value (e.g., the

start of a rating scale), and then adjusting to that anchor until a plausible estimate is reached. Because the adjustments made to the anchor are more often incomplete and insufficient, the final estimate is usually biased toward the anchor. Although anchoring-and-adjustment was initially proposed as a heuristic approach for numerical estimates, Yan and Keusch (2015) showed that it can also be used to explain scale direction effects in rating scales. Whether anchoring-and-adjustment also applies to Likert-type agree-disagree questions is yet unknown. Salzberger and Koller (2013) attributed the scale direction effect in multi-item grid questions to the 'near means related' heuristic that respondents use when answering rating scale questions (Tourangeau et al. 2004; Tourangeau et al. 2007). That is, the spatial proximity of the survey items and the starting point of a response scale lead to a higher endorsement of this side of the scale in self-administered modes.

The direction of a response order is an important design feature of rating scales. However, the relationship between scale direction and response styles, including ARS and ERS, is yet unknown. The goal of this study is to evaluate and compare response styles across scale directions through experimental data. Survey satisficing, anchoring-and-adjustment, and acquiescence are all potential causes of measurement bias for ordinal rating scales. Satisficing can lead to a primacy effect, that is, higher endorsement of the response option that is presented first, whether positive or negative. Similarly, the anchoring-and-adjustment process used for responding to ordinal rating scales also predicts a higher level of agreement in the positive-negative scale than the negative-positive scale, since respondents are more likely to use the first presented option as the anchoring point, and it is most likely to be selected if the adjustment process is insufficient. Acquiescence, different from the other two mechanisms, manifests itself in higher endorsement of the positive option, regardless of the order of presentation. Therefore, when measuring ARS, we would expect a higher level of ARS when the positive option is presented first and the negative option is presented last. This is due to the effect being reinforced by scale direction effect, which is likely to be caused by satisficing and the anchoring-and-adjustment process.

Given that more national surveys are moving toward web data collection, this study also compares the impact of scale direction on response styles between web and face-to-face surveys. Liu et al. (2016) showed that face-to-face survey respondents reveal more ARS and ERS than web survey respondents. They attributes the mode effect on response styles to the impression management concerns of respondents in the face-to-face survey. Another goal of this study is to evaluate and compare the ARS and ERS bias between two scale directions in the two modes of data collection separately. Since the primacy effect is typically stronger in a self-administered survey than in an interviewer-administered survey, it is possible that we will observe a larger scale direction effect on ARS in a web survey than in a face-to-face survey. Also, web surveys use only visual display, hence respondents are more likely to use the first visually displayed option as the anchoring point when making the judgment than face-to-face respondents. Interviewers in a face-to-face survey utilize both visual (show cards) and aural channels of communication and may facilitate a more full interpretation of the rating scales. This consideration also leads to the prediction of a larger scale direction effect on ARS in web surveys than in face-to-face surveys.

To sum up, this study utilizes an experiment embedded in a national probability survey to explore the impact of scale direction on response styles in face-to-face and web survey data collection. Specifically, this study will answer three research questions. First, do response styles, including ARS and ERS, differ by scale direction in agree-disagree Likert scales? Second, does the impact of scale direction on response style, if any, differ between face-to-face surveys and web surveys? Third, after adjusting for response style, does the scale direction still have an impact on the substantive content latent class variables? Given the previous research in this area, we expect to find more endorsement of positive options when they are presented first than when they are presented last. Also, we expect to see a larger effect of this in web surveys than in face-to-face surveys. As for the scale direction effect on ERS, our study is largely exploratory, and we do not have a clear expectation. However, given that ERS is an important response style and a previous study shows that the rating scales used in this analysis suffer from ERS (Liu et al. 2015), we also examine whether or not the scale direction has any impact on ERS.

3. Data and Measures

The data analyzed in this study come from the 2012 American National Election Studies (ANES). The ANES is a national time series survey on political candidates, parties, politics in general, and other related topics conducted in election years in the United States. In 2012, the ANES, for the first time, conducted surveys using two modes of data collection, namely face-to-face and web. The survey has two independent national representative samples: one for each mode, and one identical questionnaire. The target population for both samples is US citizens aged 18 or older by the 2012 Election Day. The web survey sample came from the GfK KnowledgePanel, a probability online panel of US adults. The online panel members were recruited using two sampling methods: address-based sampling and random-digit dialing. After a household was selected, all members in the household were enumerated, and panel members were selected. Households without Internet access or necessary equipment for participating in web surveys were provided with such equipment. The face-to-face survey involves an address-based, stratified, multi-stage cluster sample. The sample includes a nationally representative main sample and two oversamples for African Americans and Hispanic Americans. (For more information about the ANES sampling design, see http://www.electionstudies.org/studypages/anes_timeseries_2012/anes_timeseries_2012_userguidecodebook.pdf.)

The ANES contains two stages of data collection, with one wave prior to the presidential election and one wave after the election. In 2012, 5,914 pre-election interviews (2,014 of the face-to-face surveys) and 5,510 post-election interviews (1,929 of the face-to-face surveys) were completed. Response rates for the pre-election study were 38% and two percent for face-to-face and web, respectively (AAPOR RR1). The re-interview rates (conditional on the pre-election study response rates) in the post-election stage for the two modes were 94% and 93%, respectively. According to the survey organization, the response rate for the web survey is a function of the recruitment of panel members, the retention of panelists from the time of recruitment to the point at

which they were invited to take the ANES survey, and the response to those survey invitations.

The 2012 ANES pre- and post-election studies included an experiment that randomly assigned respondents to one of two scale direction conditions. Of the 190 ordinal items included in the experiment, three sets of balanced multi-item Likert scales can fully serve the purpose of examining the effect of scale direction on ARS and ERS. Randomization is performed at the respondent level, not at the question level. That is, each respondent received all ordinal rating items labeled in one direction or the other.

For the multi-scale items analyzed in this study, the first scale contains four items about attitudes toward traditionalism. The second scale contains four items about the position of Blacks in society. The third scale contains six items about the attitudes toward egalitarianism (see Appendix for wording of items). All items were measured on the same five-point rating scale. The scales were labeled 'disagree strongly,' 'disagree somewhat,' 'neither agree nor disagree,' 'agree somewhat' or 'agree strongly' (forward condition) or in reversed order (reversed condition) without numeric labels. In the face-to-face survey, show cards were used to visually present the response options to the respondents. (For show cards in the face-to-face survey, see http://www.electionstudies.org/studypages/ anes_timeseries_2012/anes_timeseries_2012_respbooklet_post.pdf.) The response options were displayed vertically in both modes. A "don't know" option was not explicitly provided in either mode but it was accepted in the face-to-face survey. In the web survey, respondents could skip the question if they chose not to answer. Both "don't knows" and "skips" are coded as missing in the analysis. Overall, the design of the questionnaire was kept as unified as possible between these two modes. Also, the fully labeled five-point agree-disagree scale conforms to the best practice in the literature (Revilla et al. 2013; Krosnick and Presser 2010).

In this study, latent class analysis (LCA) was used to examine the experimental data on scale direction. Several analytical models exist in the literature on how to measure response styles. We choose this particular analytical approach because it can simultaneously estimate ARS, ERS, and the substantive content of the rating scales as different latent class variables. Moors (2003) was among the first to adopt the LCA model to examine response styles. Specifically, he treated the rating scales as nominal variables and the latent class variables as equidistant ordered variables in order to estimate ERS. The reason for treating the rating scales as nominal rather than as ordinal is because a U-shape is expected for the ERS. In other words, the coefficients for the two endpoints are positive, and the coefficients for the middle options are negative, which is an indication of the endpoint preference. Morren et al. (2011) simplified the model by treating the rating scales as ordinal variables when measuring the content latent class variables and maintaining the rating scales as nominal variables when measuring ERS latent class variables in order to capture the nonmonotonic (U-shape) relationship. This is a more parsimonious model in that, for each response item, only one coefficient is estimated for the content latent class variables. This modeling approach was further simplified by forcing the style latent class variable coefficients to be equal across all items (Kieruj and Moors 2013; Moors et al. 2014). This model only outputs one set of coefficients for all items when estimating the response style latent class variables. This constraint is theoretically meaningful because, by definition, a response style is content-irrelevant, and

its impact on all items should be equally likely. For the data in this study, the model can be written as:

$$log \frac{P(Y_{ij} = c + 1 | F1_i, F2_i, F3_i, E_i, A_i)}{P(Y_{ij} = c | F1_i, F2_i, F3_i, E_i, A_i)}$$

$$= (\beta_{0jc+1} - \beta_{0jc}) + \beta_{1j_1}F1_i + \beta_{2j_2}F2_i + \beta_{3j_3}F3_i + (\beta_{4jc+1} - \beta_{4jc})E_i + \beta_{5j}A_i$$

Where Y_{ij} denotes the response of respondent i to Likert-type item j, $i = 1, . . .,I$, $j = 1, . . .,14$;
$F1_i$ denotes the "moral traditionalism" latent class variable;
$F2_i$ denotes the "position of Blacks in society" latent class variable;
$F3_i$ denotes the "egalitarianism" latent class variable;
E_i denotes the ERS latent class variable;
A_i denotes the ARS latent class variable;
β_{1j_1} denotes the effects on the adjacent category logit for the "moral traditionalism" latent class variable, $j_1 = 1, 2, 3,$ or 4;
β_{2j_2} denotes the effects on the adjacent category logit for the "position of Blacks in society" latent class variable, $j_2 = 5, 6, 7,$ or 8;
β_{3j_3} denotes the effects on the adjacent category logit for the "egalitarianism" latent class variable, $j_3 = 9, 10, 11, 12, 13,$ or 14;
$\beta_{4jc+1} - \beta_{4jc}$ denotes the nonmonotonic (U-shape) relationship between the ERS latent class variable and the Likert-type items;
β_{5j} denotes the effects on the adjacent category logit for the ARS latent class variable; and
c denotes the response options, $c = 1, 2, 3$ or 4.

This model is illustrated in Figure 1, showing that the individual items only load on their corresponding content latent class variables (F1, F2, F3) with no cross-loadings specified. The three content latent class variables are allowed to be correlated with each other. For ARS and ERS latent class variables, all items are loaded on these two style latent class variables. This is because response styles should affect all items regardless of the specific question content. The two style latent class variables are not correlated with each other, nor do they correlate with the content latent class variables. The rating items are estimated as nominal variables for measuring ERS. Effect coding is used and, thus, the model outputs five coefficients – one for each response option. For the other four latent class variables, the rating items are estimated as ordinal variables and, thus, one coefficient is estimated for each item. The scale direction is introduced into the model as a covariate. This is the key variable of interest in this model. Previous research has shown that, given the data structure, this model should fit the data well (Liu et al. 2015). However, we also test a few other alternative models in order to find the best fitting empirical model. We use the Bayesian information criterion (BIC) to guide our choice of models. A smaller BIC indicates a better model fit. However, although a more complex model tends to have a better model fit based on BIC, we try to present a conceptually meaningful and parsimonious model rather than a purely data-driven model. The model is applied to face-to-face and web survey data separately because a previous study has shown that
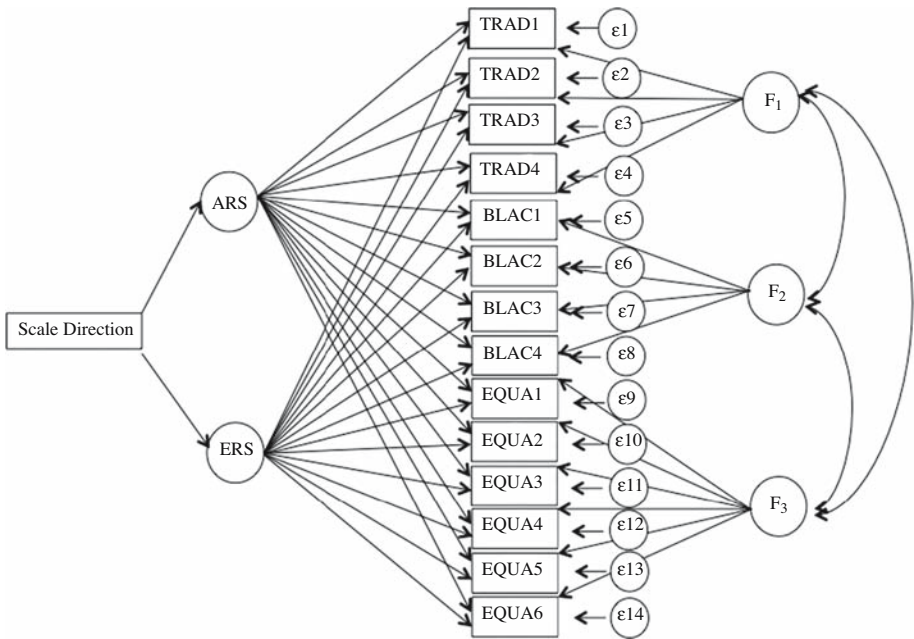
Fig. 1. Latent class analysis model of acquiescent response style (ARS), extreme response style (ERS), and content latent class variables (F_1, F_2, F_3), with covariates (scale direction).

response styles differ in these two modes of data collection (Liu et al. 2015). All analysis is weighted using the weight variable provided by the survey organization, which adjusts for the probability of household selection, the probability of respondent selection within the household, nonresponse, and random sampling error. Based on the survey documentation, the weights are poststratified to produce estimates that match known population proportions for selected characteristics. The weights were created separately for face-to-face and web survey so that the weighted analysis of each survey should, in theory, produce unbiased estimates of the same target population. Missing values are deleted listwise. All analyses are performed in Latent Gold 5.0 (Vermunt and Magidson 2013).

4. Results

Before fitting the LCA models, we first estimated the demographic distributions between the two experimental conditions (forward vs. reversed) in the two modes separately. As Appendix B shows, none of the weighted demographic variables, including gender, age, race/ethnicity, education, marital status, or household income, differed in the two conditions in the two response order conditions and the two modes of data collection. Therefore, it is not likely that the weighted differences observed between the two conditions and two modes are due to the demographic composition differences.

Although the LCA model introduced above is theoretically meaningful, we fit several alternative models, and compare them using BIC in order to find the best fitting empirical model. The first step was to determine whether the data actually reflected response styles. That is, whether adding latent class variables for ARS, ERS, or both to the model in

addition to the content latent class variables could improve the model's fit. According to Table 1, this was the case for both the face-to-face and the web survey. In comparison with the content-only model (Model 1), when ARS (Model 2), ERS (Model 3), or both ARS and ERS (Model 4) were included in the model, the BIC dropped; Model 4 has the smallest BIC among the four models. This indicated that both ARS and ERS were critical latent class variables to be added to the model. In other words, respondents' answers to the rating scales not only reflected their opinions toward the substantive content of the questions, but also their response styles.

The next step was to determine the number of levels for all five latent class variables. As we mentioned above, the latent class variables are equidistant ordered latent class variables with at least two levels. Each of the latent class variables in the aforementioned four models contains two levels. Next, we increased the number of levels to three (Model 5) and four (Model 6). However, the four-level latent class model has small and not meaningful class sizes for several latent class variables for both fact-to-face and web surveys. (For the face-to-face surveys, the class sizes for ARS are 0.19, 0.50, 0.24, and 0.07, and for ERS they are 0.11, 0.63, 0.20, and 0.06. For the web surveys, the class sizes for ARS are 0.20, 0.66, 0.09, and 0.05, and for ERS they are 0.28, 0.56, 0.10, and 0.06.) Furthermore, the model becomes more difficult to interpret and replicate. We hope to identify a simple, easy-to-interpret, and theoretically meaningful model. Considering this, we choose to proceed with the three-level model with ordinal latent classes (Model 5).

The last step in model-building is to examine whether or not adding equality constraints can improve the model fit. In Model 5a, for each latent class variable, the coefficients are set to be equal. The BIC of Model 5a is substantially larger than in Model 5 for data from both the face-to-face and the web survey, indicating such constraints deteriorate the model fit. In Model 5b, the coefficients are set to be equal for ARS and ERS, but the coefficients for the three content latent class variables are free to vary. The BIC increases slightly compared with Model 5, but the model has 65 (158-93) more free parameters, which makes it a much more parsimonious model. In addition, this model is conceptually more meaningful, since response styles are content-free, and they should equally influence responses to items. As a result, we conclude that Model 5b is the best-fit model.

Given this information, we deduced that two response style latent class variables exist. An important subsequent step is to determine whether these two style latent class variables

Table 1. *Model Fit Statistics, 2012 American National Election Studies.*

	BIC		
	Face-to-face	Web	Npar
Model 1: Content only (two-level)	72909	137782	79
Model 2: Content + ARS (two-level)	70639	131711	137
Model 3: Content + ERS (two-level)	71610	133396	95
Model 4: Content + ARS + ERS (two-level)	69581	128620	153
Model 5: Content + ARS + ERS (three-level)	68907	126078	158
Model 5a: Equality on all latent class variables	73643	141588	82
Model 5b: Equality on style latent class variables	68928	126467	93
Model 6: Content + ARS + ERS (four-level)	68791	125355	163

Table 2.   Estimated Regression Coefficients (Log Odds) and Standard Errors of ERS and ARS on the Likert Scale Items by Mode of Data Collection, 2012 American National Election Studies.

Response style	Response option	Face-to-face		Web	
		$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.
ERS	Disagree strongly	2.45	0.16[***]	1.70	0.16[***]
	Disagree somewhat	−0.72	0.13[***]	−1.11	0.15[***]
	Neither agree nor disagree	−3.96	0.30[***]	−1.88	0.25[***]
	Agree somewhat	−0.63	0.12[***]	−0.92	0.10[***]
	Agree strongly	2.86	0.14[***]	2.20	0.16[***]
ARS		1.11	0.07[***]	0.98	0.05[***]

[***]$p < .0001.$

actually represent ARS and ERS. Table 2 shows the estimated regression coefficients for these two latent class variables in the face-to-face and web survey separately. As described above, there are five coefficients for ERS: one for each response option. For ERS, the coefficients for the two endpoints need to be in the opposite direction from the three middle options. This is what we find from both modes. The coefficients for the two endpoints are positive, and the coefficients for the middle three options are negative. This confirms that this latent class variable is indeed ERS. Recall that ERS is a three-level latent class variable. The coefficients in Table 2 suggest that respondents at the higher level of the latent class variable tend to select the endpoint of the scales more frequently than respondents at the lower level of the latent class variable. That is, a lower level of the latent class variable indicates "avoid-ERS" and a higher level of the latent class variable indicates "pro-ERS". For ARS, because the rating items are treated as ordinal variables, there is only one coefficient output from the model. In the analysis, the negatively worded items are recoded so that they go from negative (disagree strongly = 1) to positive (agree strongly = 5), the positive coefficients in both modes mean that a higher level of the ARS variable indicates pro-ARS and a lower level of the ARS variable indicates avoid-ARS.

The scale direction effect on response styles is presented in Table 3. In this case, the latent class variables are dependent variables and scale direction is the predictor.

Table 3.   Estimated Scale Direction Effect (Log Odds) on Response Style Latent Class Variables by Mode of Data Collection, 2012 American National Election Studies.

	Face-to-face					
	ERS			ARS		
	$\hat{\beta}$	S.E.	*p*-value	$\hat{\beta}$	S.E.	*p*-value
Scale direction[a]	−0.17	0.23	0.44	−0.28	0.27	0.31

	Web					
	ERS			ARS		
	$\hat{\beta}$	S.E.	*p*-value	$\hat{\beta}$	S.E.	*p*-value
Scale direction	−0.10	0.18	0.60	1.73	0.35	<.0001

[a]Scale direction (dependent variable): strongly agree first vs. strongly disagree first (reference group).

Table 4.   *Estimated Scale Direction Effect (Log Odds) on Content Latent Class Variables by Mode of Data Collection, 2012 American National Election Studies.*

	Face-to-face			Web		
	$\hat{\beta}$	S.E.	*p*-value	$\hat{\beta}$	S.E.	*p*-value
F1	-0.07	0.33	0.83	0.04	0.19	0.81
F2	-0.2	0.33	0.53	-0.01	0.22	0.96
F3	-0.06	0.4	0.87	-0.06	0.24	0.81

The coefficients for both ERS ($\hat{\beta} = -0.17$, $p = 0.44$) and ARS ($\hat{\beta} = -0.28$, $p = 0.31$) are not significant in the face-to-face survey. This means that, although both types of response styles exist in the face-to-face survey, they do not differ by the direction of the scale. This is possibly due to the presentation channels (both visual and aural) in the face-to-face survey reduced the impact of design features such as scale direction. Also, the presence of interviewers might have also increased the respondents' motivation so that they were more likely to provide a careful answer and less likely to be influenced by the presentation of the scales. In the web survey, ERS also does not show a significant scale direction effect ($\hat{\beta} = -.1$, $p = .6$), while ARS does show a significant scale direction effect ($\hat{\beta} = 1.73$, $p < .0001$). Morren et al. (2011) found in their study that the three levels of ERS variable were not ordinally related to covariate variables. Therefore, we also analyzed the data by treating the ERS latent variable as a nominal variable. The result showed that scale direction was not significantly related to ERS.

When the positive option is presented first on the scale (i.e., agree-disagree scale), the odds of having ARS change are by a factor of 5.6 ($= \exp(1.73)$) compared to the scale where the negative option is presented first (i.e., disagree-agree scale). As was described in the introduction, the examination of the ERS is exploratory and we did not have any expectation on how the scale direction could influence ERS. The significant scale direction effect reflect on ARS is possibly due to a combination of satisficing and anchoring-and-adjustment.

The literature shows that scale direction can affect content latent variables. However, previous studies do not control for response styles when testing scale direction effects on the substantive content latent class variables. Table 4 shows scale direction effects on the three content latent class variables after explicitly controlling for ARS and ERS in the LCA model. The effects on all three variables and in both models are small, and none are significant. Since earlier literature that examined the relationship between scale direction and substantive content variables did not control for response style, we also analyzed the data without adjustment for response styles. We reached the same conclusion: that scale direction was not significantly related to the content variables, without adjusting response styles (results not shown).

5.   Discussion and Conclusion

The direction of rating scales and their influence on responses to attitudinal questions has attracted researchers' attention for more than half a century (Belson 1966). Previous research has primarily focused on the effect of scale direction on the response distribution of individual items, with occasional focus on latent variables. Response styles are one

form of measurement bias that is frequently examined for multi-item rating scales. Some recent studies have examined the different scale design formats and their impacts on response styles (Kieruj and Moors 2013; Moors et al. 2014). However, no such effort has been devoted to testing the relationship between scale direction and response styles.

This study set out to test whether response styles, including ARS and ERS, differ by the direction of scales using an experiment embedded in a national probability survey conducted both face-to-face and on the web. Using latent class analysis, we reached the following conclusions:

1) ARS and ERS exist in both scale directions in both survey modes;
2) ERS is similar for scales presented in both directions in both face-to-face and web surveys;
3) ARS is significantly different between the two scale directions in the web surveys but not in the face-to-face surveys;
4) the scale direction does not impose a significant influence on the substantive content latent class variables, with or without controlling for response styles.

Several factors can explain the higher endorsement of the positive options in the agree-disagree scale than the disagree-agree scale in the web surveys. First, this could be interpreted as the result of respondents satisficing (Krosnick 1991) that leads to a primacy effect due to the more frequent selection of the option presented first, regardless of whether the option is positive or negative. Respondents might not read and process all scale points once they find a good enough response option. Assuming that satisficing might play a bigger role in self-administered survey modes than interviewer administration, this would also explain the nonfindings for the face-to-face survey.

Second, it is possible that respondents use the top-most response option as the anchoring point and then adjust their responses. Since the adjustment process is often insufficient, the anchoring point and the adjacent options are most likely to be endorsed. Both satisficing and anchoring-and-adjustment predict more selection of the position option when it is presented first than when it is presented last. Third, acquiescence predicts more selection of the positive option regardless of the scale direction. As a result, the overall shift of the distribution toward the positive end of the scale is most likely to be the consequences of a combination of the three mechanisms. Future research needs to be done to determine which of these explanations reflects the true cause of the scale direction effect.

The lack of scale direction effect in the face-to-face survey may be attributable to the channel of communication. In the face-to-face survey, as opposed to the web survey, both visual and aural communication is adopted during the interview. Such a design feature can alleviate the impact of scale direction and, hence, no such effect on response style latent class variables is detected from face-to-face surveys. In the web survey, questions and response options are only present visually. Thus, there is a higher likelihood that the respondent's anchoring point changes from one scale direction to another. Consequently, we observe a significant effect on ARS in the web survey. Note that this finding is the opposite of recency effects found in telephone interviews (Krosnick 1991). Furthermore, interviewers in face-to-face surveys can explain the purpose of the survey, answer questions, address concerns, and motivate respondents. Hence, respondent commitment is likely to be higher among face-to-face respondents than among web respondents. This can

also be a reason for the lack of response style difference between the two scale directions, as respondents might just have been more committed to the task and less likely to be impacted by the design difference.

This study also finds that substantive latent class variables do not show a reliable scale direction effect, whether the response style latent class variables were controlled for or not. Previous research reported mixed results on the impact of scale direction on latent traits (Krebs and Hoffmeyer-Zlotnik 2010; Saris and Gallhofer 2007; Chan 1991), and this study provides one more piece of evidence that supports the lack of impact of the scale direction. Although the univariate response distribution shifted by the change of scale direction, it does not necessarily reflect a change in the substantial latent construct. Rather, it is likely to be a reflection of the change of response style latent variables.

This study probably raises more questions than it can answer. First, is it possible to generalize the findings in this study to other scale types? In this study, we only examine the scale direction effect using five-point agree-disagree Likert scales. A previous study has shown that other question types, such as item-specific scales, produce different response style patterns (Liu et al. 2015). Whether the scale direction influences other rating scales similarly should be examined in the future. Second, do other data collection modes exhibit a different scale direction effect? This study demonstrates that the scale direction effects are not identical between face-to-face and web surveys. Future studies should also explore and compare the scale direction effect on response styles among other survey modes (e.g., telephone, mobile web). Third, are the results replicable for other survey questions? Although response style is content-irrelevant, the scale direction effect on response style may differ regarding the question topic (e.g., sensitive vs. nonsensitive topics) or type (e.g., attitudinal vs. behavioral questions). We also find no scale direction effect on content latent class variables after controlling for the response styles. Future studies should attempt to replicate this finding and test whether it is possible to generalize. Last but not least, the scale direction may interact with the wording of the item. A positively worded item may show a different effect from a negatively worded item. We encourage future research to explore this possibility.

Regardless of the unsolved questions, this study demonstrates that the scale direction, a seemingly trivial survey design feature, can influence response style in web surveys. Researchers need to take the scale direction into serious consideration when designing rating scales. As many flagship national and international surveys move toward using the web, this is becoming a particularly relevant issue. The good news is that after controlling for response styles in the analysis model, the scale direction does not exert significant impact on the substantial latent class variables. Since in most cases the substantial, rather than the response style latent variables, are of interest, it is important to control for the response styles in the analysis model.

Appendix A
Question Wordings

The world is always changing and we should adjust our view of moral behavior to those changes. (TRAD1)
The newer lifestyles are contributing to the breakdown of our society. (TRAD2)
We should be more tolerant of people who choose to live according to their own moral standards, even if they are very different from our own. (TRAD3)
This country would have many fewer problems if there were more emphasis on traditional family ties. (TRAD4)

Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors. (BLAC1)
Generations of slavery and discrimination have created conditions that make it difficult for Blacks to work their way out of the lower class. (BLAC2)
Over the past few years, Blacks have gotten less than they deserve. (BLAC3)
It's really a matter of some people not trying hard enough; if Blacks would only try harder they could be just as well off as whites. (BLAC4)

Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed. (EQUA1)
We have gone too far in pushing equal rights in this country. (EQUA2)
One of the big problems in this country is that we don't give everyone an equal chance. (EQUA3)
This country would be better off if we worried less about how equal people are. (EQUA4)
It is not really that big a problem if some people have more of a chance in life than others. (EQUA5)
If people were treated more equally in this country we would have many fewer problems. (EQUA6)

Appendix B
Demographic Distributions (means and standard errors [S.E.]) by Scale Direction in Face-to-face and Web Surveys, 2012 American National Election Studies.

	Face-to-face						Web					
	Forward		Reversed				Forward		Reversed			
	Mean	S.E.	Mean	S.E.	χ^2	p	Mean	S.E.	Mean	S.E.	χ^2	p
Male	47.98	2.18	48.00	2.14	0.00	0.99	47.29	1.52	48.48	1.49	0.31	0.58
Female	52.02	2.18	52.00	2.14			52.71	1.52	51.52	1.49		
<30	20.16	1.81	22.14	1.82	7.49	0.19	20.32	1.42	21.69	1.44	3.32	0.65
30–39	16.47	1.50	14.76	1.45			15.31	1.13	14.66	1.08		
40–49	15.59	1.52	18.80	1.76			16.01	1.08	17.37	1.13		
50–59	19.64	1.70	18.20	1.64			20.13	1.14	20.07	1.10		
60–69	16.59	1.77	12.21	1.38			16.30	1.01	16.24	0.94		
70<	11.56	1.48	13.89	1.58			11.93	0.93	9.96	0.79		
Non-Hispanic white	70.78	1.71	70.96	1.75	0.35	0.95	70.11	1.44	71.73	1.40	4.92	0.18
Non-Hispanic black	12.20	1.15	11.69	1.13			11.21	0.97	12.59	1.11		
Hispanic	11.04	1.00	10.75	1.06			11.76	1.04	10.76	0.92		
Other	5.98	0.88	6.60	1.02			6.92	0.83	4.92	0.64		
Less than high school	9.95	1.15	10.43	1.14	1.01	0.91	11.30	1.08	9.26	0.94	2.70	0.61
High school	29.23	2.05	31.46	2.08			29.33	1.47	30.35	1.43		
Some post-high school, no Bachelor's degree	31.09	1.99	29.25	1.89			30.29	1.34	30.51	1.38		
Bachelor's degree	19.83	1.79	18.74	1.72			18.21	1.13	19.10	1.11		
Graduate degree	9.91	1.29	10.13	1.31			10.88	0.82	10.78	0.83		
Married	53.76	2.15	52.78	2.12	1.02	0.91	53.91	1.52	52.62	1.50	4.51	0.34
Widowed	5.86	0.97	6.27	0.92			5.94	0.69	5.42	0.60		
Divorced	12.19	1.19	12.98	1.32			13.49	1.02	12.69	0.95		
Separated	2.94	0.71	2.22	0.40			2.67	0.51	1.96	0.37		
Never married	25.26	1.83	25.75	1.80			23.98	1.36	27.31	1.42		
0–49999	51.71	2.25	48.60	2.19	3.16	0.37	49.30	1.55	49.66	1.51	0.56	0.90
50000–99999	27.71	2.09	31.96	2.06			31.62	1.44	32.25	1.44		
100000–149999	11.08	1.58	12.19	1.67			11.83	0.98	11.57	0.90		
150000+	9.50	1.49	7.25	1.37			7.25	0.72	6.53	0.66		

7.   References

Baumgartner, H. and J.E.M. Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research* 38: 143–156. Doi: http://dx.doi.org/10.1509/jmkr.38.2.143.18840.

Belson, W.A. 1966. "Effects of Reversing Presentation Order of Verbal Rating Scales." *Journal of Advertising Research* 6: 30–37.

Carp, F.M. 1974. "Position Effects on Interview Responses." *Journal of Gerontology* 29: 581–587. Doi: http://dx.doi.org/10.1093/geronj/29.5.581.

Chan, J.C. 1991. "Response-Order Effect in Likert-Type Scales." *Educational and Psychological Measurement* 51: 531–540. Doi: http://dx.doi.org/10.1177/0013164491513002.

Christian, L.M., D.A. Dillman, and J.D. Smyth. 2008. "The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys." In *Advances in Telephone Survey Methodology*, edited by J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavrakas, M.W. Link, R.L. Sangster, 250–275. New Jersey: John Wiley & Sons, Inc.

Dillman, D.A., T.L. Brown, J.E. Carlson, E.H. Carpenter, F.O. Lorenz, R. Mason, J. Saltiel, and R.L. Songster. 1995. "Effects of Category Order on Answers in Mail and Telephone Surveys." *Rural Sociology* 60: 674–687. Doi: http://dx.doi.org/10.1111/j.1549-0831.1995.tb00600.x.

He, L., T. Yan, F. Keusch and S. Han. 2014. "The impact of question and scale characteristics on scale direction effect." In Proccedings of the AAPOR 69th Annual Conference, Anaheim, California, May 15–18. Available at: http://www.aapor.org/AAPOR_Main/media/AnnualMeetingProceedings/2014/Session-J-6-2-He-Y.pdf (accessed January 2017).

Kalton, G., M. Collins, and L. Brook. 1978. "Experiments in Wording Opinion Questions." *Applied Statistics*, 149–161. Doi: http://dx.doi.org/10.2307/2346942.

Kieruj, N.D. and G. Moors. 2010. "Variations in Response Style Behavior by Response Scale Format in Attitude Research." *International Journal of Public Opinion Research* 22: 320–342. Doi: http://dx.doi.org/10.1093/ijpor/edq001.

Kieruj, N.D. and G. Moors. 2013. "Response Style Behavior: Question Format Dependent or Personal Style?" *Quality & Quantity* 47: 193–211. Doi: http://dx.doi.org/10.1007/s11135-011-9511-4.

Krebs, D. and J.H.P. Hoffmeyer-Zlotnik. 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: 118–127. Doi: http://dx.doi.org/10.1027/1614-2241/a000013.

Krosnick, J.A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5: 213–236. Doi: http://dx.doi.org/10.1002/acp. 2350050305.

Krosnick, J.A. and S. Presser. 2010. "Question and Questionnaire Design." In *Handbook of Survey Research, Second Edition*, edited by P.V. Marsden and J.D. Wright, 263–314. United Kingdom: Emerald Group Publishing Limited.

Liu, M., F.G. Conrad, and S. Lee. 2016. "Comparing Acquiescent and Extreme Response Styles in Face-to-Face and Web Surveys." *Quality & Quantity*. Doi: http://dx.doi.org/10.1007/s11135-016-0320-7.

Liu, M., S. Lee, and F.G. Conrad. 2015. "Comparing Extreme Response Styles between Agree-Disagree and Item-Specific Scales." *Public Opinion Quarterly* 79: 952–975. Doi: http://dx.doi.org/10.1093/poq/nfv034.

Mingay, D.J. and M.T. Greenwell. 1989. "Memory Bias and Response-Order Effects." *Journal of Official Statistics* 5: 253–263.

Moors, G. 2003. "Diagnosing Response Style Behavior by Means of a Latent-Class Factor Approach. Socio-Demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Reexamined." *Quality and Quantity* 37: 277–302. Doi: http://dx.doi.org/10.1023/A:1024472110002.

Moors, G. 2008. "Exploring the Effect of a Middle Response Category on Response Style in Attitude Measurement." *Quality & Quantity* 42: 779–794. Doi: http://dx.doi.org/10.1007/s11135-006-9067-x.

Moors, G., N.D. Kieruj, and J.K. Vermunt. 2014. "The Effect of Labeling and Numbering of Response Scales on the Likelihood of Response Bias." *Sociological Methodology* 44: 369–399. Doi: http://dx.doi.org/10.1177/0081175013516114.

Morren, M., J.P.T.M. Gelissen, and J.K. Vermunt. 2011. "Dealing With Extreme Response Style In Cross-Cultural Research: A Restricted Latent Class Factor Analysis Approach: Extreme Response Style In Cross-Cultural Research." *Sociological Methodology* 41: 13–47. Doi: http://dx.doi.org/10.1111/j.1467-9531.2011.01238.x.

Paulhus, D.L. 1991. "Measurement and Control of Response Bias." *Measures of Personality and Social Psychological Attitudes. Volume 1 in Measures of Social Psychological Attitudes Series*. Available at: http://doi.apa.org/psycinfo/1991-97206-001.

Revilla, M.A., W.E. Saris, and J.A. Krosnick. 2013. "Choosing the Number of Categories in Agree-Disagree Scales." *Sociological Methods & Research* 43: 73–97. Doi: http://dx.doi.org/10.1177/0049124113509605.

Salzberger, T. and M. Koller. 2013. "Towards a New Paradigm of Measurement in Marketing." *Journal of Business Research* 66: 1307–1317.

Saris, W.E. and I.N. Gallhofer. 2007. "Estimation of the Effects of Measurement Characteristics on the Quality of Survey Questions." In *Design, Evaluation, and Analysis of Questionnaires for Survey Research*, edited by W.E. Saris and I.N. Gallhofer, 237–253. New Jersey: John Wiley & Sons, Inc.

Stapleton, C. 2013. "The Smart (phone) Way to Collect Survey Data." *Survey Practice* 6(2).

Tourangeau, R., M.P. Couper, and F.G. Conrad. 2004. "Spacing, Position, and Order Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly* 68: 368–393. Doi: http://dx.doi.org/10.1093/poq/nfh035.

Tourangeau, R., M.P. Couper, and F.G. Conrad. 2007. "Color, Labels, and Interpretive Heuristics for Response Scales." *Public Opinion Quarterly* 71: 91–112. Doi: http://dx.doi.org/10.1093/poq/nfl046.

Tversky, A. and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185: 1124–1131. Doi: http://dx.doi.org/10.1126/science.185.4157.1124.

Vermunt, J.K. and J. Magidson. 2013. Technical guide for Latent GOLD 5.0: Basic, advanced, and syntax. Belmont, MA: Statistical Innovations Inc.

Yan, T. and F. Keusch. 2015. "The Effects of the Direction of Rating Scales on Survey
    Responses in a Telephone Survey." *Public Opinion Quarterly* 79: 145–165. Doi: http://
    dx.doi.org/10.1093/poq/nfu062.