Investigating the Impact of Violations of the "Left and Top Means First" Heuristic on Response Behavior and Data Quality

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

Ting Yan
Westat (USA)

Abstract

Web surveys are an established data collection mode that use written language to provide information. The written language is accompanied by visual elements, such as presentation formats and shapes. However, research has shown that visual elements influence response behavior because respondents sometimes use interpretive heuristics to make sense of the visual elements. One such heuristic is the "left and top means first" (LTMF) heuristic, which suggests that respondents tend to believe that a response scale consistently runs from left to right or from top to bottom. We conducted a web survey experiment to investigate how violations of the LTMF heuristic affect response behavior and data quality. For this purpose, a random half of respondents received response options that followed a consistent order and the other half received response options that followed an inconsistent order. The results reveal significantly different response distributions between the two groups. We also found that inconsistently ordered response options significantly increase response times and decrease data quality in terms of criterion validity. We therefore recommend using options that follow the design strategies of the LTMF heuristic.

Keywords: criterion validity, field experiment, interpretive heuristics, response behavior, response times, web survey

Introduction

Question and response scale design is a long-standing issue in quantitative social research. One early example is the study by Mathews (1929), which investigated the impact of the response option order on respondents' answers. Since then numerous scientific contributions followed. To better understand how designs of response options affect the way respondents comprehend and select them, researchers started to use theoretical frameworks from psychology to explain response behavior in surveys (Schwarz, 2007). The climax of this development was exemplified in 1983 with the establishment of "Cognitive Aspects of Survey Methodology (CASM)" as an interdisciplinary research field that combines survey research and psychology (see Jabine, Straf, Tanur, & Tourangeau, 1984).

Under this paradigm, Schwarz (1996) suggests that respondents are "cooperative communicators" who use any information provided by a questionnaire or survey to answer questions meaningfully. Accordingly, they not only use textual information, such as phrases and words, but also non-textual information in the form of visual elements, such as presentation formats and shapes (Couper, Tourangeau, & Kenyon, 2004; Toepoel & Dillman, 2011; Tourangeau, Couper, & Conrad, 2004). Such non-textual visual elements represent an additional information source for respondents when answering questions.

Particularly, web surveys facilitate the employment of visual elements (Couper et al., 2004; Dillman, Smyth, & Christian, 2014; Toepoel & Dillman, 2011) that, for instance, help respondents to correctly complete surveys or to make survey experience more enjoyable. However, they can also unintendedly affect response behavior and data quality. One reason for the impact of visual elements is that respondents sometimes use interpretive heuristics when answering questions. Expanding beyond the principles of Gestalt psychology, Tourangeau et al. (2004) listed five such heuristics that assign meaning to visual elements: (1) "middle means typical," (2) "left and top means first," (3) "near means related," (4) "up means good," and (5) "like means close" (see also Toepoel & Dillman, 2011).

This study focuses on the "left and top means first" (LTMF) heuristic. The LTMF heuristic posits that the first response option of a scale – either the leftmost one in horizontal scales or the top one in vertical scales – is interpreted as the first one by respondents. The LTMF heuristic is in line with the reading direction in many Western languages, such as English, French, and German (Rayner, 1998; Tourangeau et al., 2004). It states that respondents expect the first option of a list of ordered options to be the starting point (e.g., "very important"). Furthermore, they expect the subsequent options to follow it consistently (e.g., "somewhat important", "neither important nor unimportant" etc.) and that the lowermost option represents the end (e.g., "very unimportant"). Tourangeau et al. (2004) empirically examined the LTMF heuristic by systematically varying the succession of response options. For instance, some respondents received the response options in an order consistent with the LTMF heuristic (i.e., agree strongly, agree, it depends, disagree, disagree strongly) and other respondents received the same response options in an order inconsistent with the LTMF heuristic (i.e., it depends, agree strongly, disagree strongly, agree, disagree). As implied by the LTMF heuristic, respondents presented with response options in an inconsistent order produced significantly longer response times, which indicates that the inconsistent order slowed respondents down (see also Holbrook, Krosnick, Carson, & Mitchell, 2000). The authors also found significantly different response distributions by order. For instance, when the middle option "it depends" was presented in the middle of the scale, more respondents chose this option than when it was presented at the top (Tourangeau et al., 2004, p. 383). The authors interpreted these results as evidence for the application of the LTMF heuristic.

As pointed out by Höhne, Lenzner, Neuert, and Yan (2019), most of the existing literature on visual question design strategies and interpretive heuristics only deals with question processing and response behavior and does not address data quality. For instance, Tourangeau et al. (2004) examined response times and response distributions, but they did not look at data quality, such as the reliability and validity. In this study, we therefore attempt to fill the gap by experimentally manipulating the order of response options presented on a web survey. We then investigate the impact on response behavior and data quality when the presentation of response

options does not comply with the LTMF heuristic. We attempt to answer the following research question: How do violations of the LTMF heuristic affect response behavior (i.e., response distributions and response times) and data quality (i.e., criterion validity)?

Method

Data Source

Data for this study were collected in the German Internet Panel, which is part of the Collaborative Research Center 884 "Political Economy of Reforms" at the University of Mannheim. The German Internet Panel is based on an initial recruitment in 2012 and two refreshing recruitments in 2014 and 2018. While the recruitments in 2012 and 2014 are based on a three-stage stratified probability sample of the German population, the recruitment in 2018 is based on a two-stage stratified probability sample of the German population. For a detailed methodological description of the German Internet Panel, we refer interested readers to Blom, Gathmann, and Krieger (2015).

Sample

For this study, we use data from wave 42 in July 2019. In total, 4,714 respondents participated in wave 42. Of these respondents, 27 broke off before being asked any study-relevant questions. As a result, 4,687 respondents remain for statistical analysis. These respondents had a mean age of 51.0 years and 48.1% of them were female. In terms of education, 13.5% had graduated from a lower secondary school, 31.2% from an intermediate secondary school, and 51.3% from a college preparatory secondary school or university. Further, 1.5% still attended school or had finished without a diploma and 2.5% reported another degree than mentioned above.

Study Design

Respondents were randomly assigned to one of two experimental groups. The first group (n = 2,346) received response options consistent with the LTMF heuristic (the "consistent condition"). The second group (n = 2,341) received response options inconsistent with the LTMF heuristic (the "inconsistent condition"). These two conditions were directly adapted from Tourangeau et al. (2004).

To evaluate the effectiveness of random assignment and the sample composition between the two experimental groups, we conducted chi-square tests. The results showed no significant differences regarding age, gender, and education.

Questions

Target questions: We used four questions on political efficacy that were adapted from Beierlein, Kemper, Kovaleva, and Rammstedt (2012). The questions were written in German and were presented on a separate screen with five-point, vertically aligned scales.

Criterion question: One question on political interest, also adapted from Beierlein et al. (2012), is used as the criterion measure to evaluate criterion validity. This method has been used in previous research (see, for instance, Yeager & Krosnick, 2012). This particular question was chosen as a criterion question because it was shown to be conceptually relevant to the topic of the target questions (see Beierlein, et al. 2012). In addition, it correlated significantly with all experimentally manipulated target questions in the full sample. To determine criterion

validity, we investigate which one of the two conditions (consistent and inconsistent order) produces higher correlations between the target questions and the criterion question. The criterion question is also written in German and used a five-point, vertically aligned scale (the Appendix displays the English translations of all questions and response options).

Results

Response Distributions

In line with our research question, we first compared response distributions of the four target questions between the two experimental groups that differ regarding the consistency of the order of the response options. Chi-square tests were conducted to determine whether differences between the response distributions are statistically significant. Table 1 displays the results. For three out of the four target questions on political efficacy, the overall response distributions significantly differ by scale condition. Respondents tend to select "disagree strongly" more often when it was presented as the third option in the inconsistent condition than when it was presented as the last option in the consistent condition.

We then specifically looked into the endorsement of the middle option ("it depends") by condition. In contrast to Tourangeau et al. (2004, p. 383), we did not find that the middle option was systematically selected more often when it was presented in the middle of the scale in the consistent condition than when it was presented at the top of the scale in the inconsistent condition. For the first political efficacy question, the percentage of selecting "it depends" increased from 46.5% when it was the middle option to 56.9% when it was the top option. The third political efficacy question shows almost no differences in the percentage of respondents selecting "it depends" across the two experimental groups.

Response Times

We investigated whether the two experimental groups differ with respect to response times (in seconds). For this purpose, we calculated non-parametric Mann-Whitney U tests that determine median differences.[1] Table 1 displays the results. The results revealed significantly longer response times for the experimental group with an inconsistent order of the response options. This applies to all four political efficacy questions. These findings are line with the results reported by Holbrook et al. (2000) and Tourangeau et al. (2004), confirming that the inconsistent order that violates the LTMT heuristic slowed respondents down.

Criterion Validity

Lastly, we investigate data quality in terms of criterion validity between the group with a consistent order of response options and the group with an inconsistent order of the response options. Specifically, we looked at the strength of the correlations between the target questions on political efficacy and the criterion question on political interest. The criterion validity analyses were conducted by estimating unstandardized OLS regression coefficients.

---

[1] As robustness check, we also tested for mean differences using the following outlier treatment: excluding respondents with response times below the 5% percentile or above the 95% percentile. The main results did not change. In addition, we conducted all response time analyses with and without log transformation, but there were almost no differences. We report the non-log transformed results in the paper. We did not adjust response times for baseline reading speed.

Table 1. Response distributions (in percentage) and median response times (in seconds) of the target questions on political efficacy by experimental condition

	Agree strongly	Agree	It depends	Disagree	Disagree strongly	Significance level	Response times	Significance level
Political efficacy 1								
Consistent	14.6	32.6	46.5	4.9	1.4	***	12	***
Inconsistent	13.5	22.9	56.9	4.6	2.1		17	
Political efficacy 2								
Consistent	0.4	6.7	39.9	43.8	9.2	***	10	***
Inconsistent	0.6	10.5	36.6	37.9	14.4		13	
Political efficacy 3								
Consistent	12.8	31.8	37.7	13.6	4.1	n.s.	8	***
Inconsistent	14.1	30.8	37.4	14.2	3.5		10	
Political efficacy 4								
Consistent	0.3	5.3	36.4	45.6	12.4	***	9	***
Inconsistent	0.5	8.3	31.7	44.5	15.0		10	

Note. ***$p < 0.001$. n.s. = non-significant. Response distributions are in percentage. The response options with an inconsistent order were recoded to a consistent order. Response times were measured in seconds.

Table 2. OLS regressions predicting political interest with political efficacy (unstandardized coefficients)

Independent variables	Consistent (coefficients)	R^2	n	Inconsistent (coefficients)	R^2	n	Differences between coefficients
Political efficacy 1	0.59*** (0.02)	0.331	2,331	0.48*** (0.02)	0.225	2,341	0.11*** (0.03)
Political efficacy 2	0.13*** (0.02)	0.014	2,332	0.02[n.s.] (0.02)	0.001	2,340	0.11*** (0.03)
Political efficacy 3	0.55*** (0.01)	0.410	2,332	0.50*** (0.01)	0.344	2,341	0.05* (0.02)
Political efficacy 4	0.20*** (0.02)	0.032	2,332	0.11*** (0.02)	0.011	2,340	0.09** (0.03)

Note. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. n.s. = non-significant. Standard errors in parentheses. Dependent variable: political interest (1 = very interested; 2 = fairly interested; 3 = somewhat interested; 4 = hardly interested; 5 = not at all interested). Coding of the independent variables on political efficacy (1 = agree strongly; 2 = agree; 3 = it depends; 4 = disagree; 5 = disagree strongly). The response options with an inconsistent order were recoded to a consistent order.

As shown in Table 2, the correlations between the target questions on political efficacy and the criterion question on political interest significantly differ between the two experimental groups. More specifically, the group with a consistent order produced significantly higher correlations than the group with an inconsistent order. This is suggested by the positive significant differences between the regression coefficients (see last column of Table 2). These results show that violations of the LTMF heuristic reduce criterion validity.

Discussion and Conclusion

This study continues the research on how to best design questions and response scales motivated by CASM combining survey research and psychological approaches. The goal is to better understand how violations of the LTMF heuristic influence response behavior and data quality. For this purpose, we experimentally manipulated the order of response options and compared response distributions, response times, and criterion validity. Regarding response distributions we found that consistently and inconsistently ordered response options result in significantly different distributions. In addition, we found for two out of four questions that the middle option "it depends" was not selected significantly more often when it appeared in the middle than when it appeared at the top. This partially contradicts findings reported by Tourangeau et al. (2004) and indicates that selecting the middle option does not only depend on its respective position in the scale, but also on the wording or content of the question stems. However, this is only an attempted explanation that needs further research.

With respect to response times we found that they increase with order discrepancies, replicating findings reported by Holbrook et al. (2000) and Tourangeau et al. (2004). It seems that respondents need more time to make sense of inconsistently ordered options, which, in turn, increases the overall effort in responding. These findings support the assumption that respondents expect response options to consistently run from left to right or from top to bottom.

The criterion validity analyses have shown that violations of the LTMF heuristic not only affect response distributions and response times, but also the quality of survey data. More specifically, criterion validity is significantly higher for questions with options following a consistent order than for questions with options following an inconsistent order. One explanation is that respondents find it difficult to make sense of the literal meaning of response options when they are inconsistently ordered. Supporting this argument, Höhne et al. (2019) showed in their eye-tracking study that response scales with an inconsistent order of the options produce significantly more and longer fixations as well as more re-fixations on response options, indicating comprehension difficulties.

This study has some limitations. First, we only used questions that dealt with political topics. Future research could therefore employ questions dealing with other topics to increase the generalizability of our findings. Second, we only employed one single criterion measure for evaluating criterion validity. It would be interesting to see if our findings hold for different criterion measures. Third, we only used one single indicator for data quality; namely, criterion validity. We therefore suggest extending the current state of research on question design strategies and interpretive heuristics by using multiple data quality indicators, such as reliability and satisficing response behavior. Finally, we did not randomly display the options, but simply adopted the inconsistent order used by Tourangeau et al. (2004). Therefore, we suggest that

future studies employ a random order of the response options to investigate the implications of the LTMF heuristic for response behavior and data quality.

Considering our empirical findings, it seems that the LTMF heuristic is at work for web survey respondents. Response options that are not presented in a consistent order negatively affect response behavior in terms of response distributions and response times and data quality in terms of criterion validity. This is important because many surveys do not present the response options in a linear (e.g., in one column from top to bottom) but in a non-linear way (e.g., in two columns downwards from left to right), which constitutes a violation of the LTMF heuristic (see Toepoel & Dillman, 2011). As a result, we highly recommend presenting response options in a linear order consistent with the LTMF heuristic to improve response behavior and data quality.

References

Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012). Ein Messinstrument zur Erfassung politischer Kompetenz- und Einflussüberzeugungen: Political Efficacy Kurzskala (PEKS). In GESIS – Leibniz-Institut für die Sozialwissenschaften (Ed.), *GESIS-Working Papers* (pp. 5–24). Cologne, Germany.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting up an online panel representative of the general population: The German Internet Panel. *Field Methods, 27*, 391–408.

Couper, M. P., Tourangeau, R., & Kenyon, K. (2004). Picture this! Exploring visual effects in web surveys. *Public Opinion Quarterly, 68*, 255–266.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley.

Holbrook, A., Krosnick, J. A., Carson, R. T., & Mitchell, R. C. (2000). Violating conversational conventions disrupts cognitive processing of attitude questions. *Journal of Experimental Social Psychology, 36*, 465–494.

Höhne, J. K., Lenzner, T., Neuert, C. E., & Yan, T. (2019). Re-examining the middle means typical and the left and top means first heuristics using eye-tracking methodology. *Journal of Survey Statistics and Methodology*. DOI:10.1093/jssam/smz028

Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (1984). *Cognitive Aspects of Survey Methodology: Building a Bridge between Discipline*. Washington D.C.: National Academy Press.

Mathews, C.O. (1929). The effect of the order of printed response words on an interest questionnaire. *Journal of Educational Psychology, 20*, 128–134.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21*, 277–287.

Toepoel, V., & Dillman, D. A. (2011). How visual design affects the interpretability of survey questions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research*

and The Internet. Advances in Applied Methods and Research Strategies (pp. 165–190). New York, NY: Routledge.

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

Yeager, D. S., & Krosnick, J. A. (2012), Does mentioning "some people" and "other people" in an opinion question improve measurement quality? *Public Opinion Quarterly, 76*, 131–141.

Appendix

English translations of the target questions on political efficacy and the criterion question on political interest.

Table A1. Questions and response options used in this study

Political efficacy 1	I am good at understanding and assessing important political issues.
Political efficacy 2	Politicians strive to keep in close touch with the people.
Political efficacy 3	I have the confidence to take active part in a discussion about political issues.
Political efficacy 4	Politicians care about what ordinary people think.
Response options (consistent order)	agree strongly, agree, it depends, disagree, disagree strongly
Response options (inconsistent order)	it depends, agree strongly, disagree strongly, agree, disagree
Political interest (criterion measure)	In general, how interested would you say you are in politics?
Response options	very interested, fairly interested, somewhat interested, hardly interested, not at all interested

Note. The order of the questions in the web survey corresponds to the order displayed in Table A1. The original German wordings of the questions and response options are available from the first author on request.