

# Different Order, Different Results? The Effects of Dimension Order in Factorial Survey Experiments

Sabine Düval<sup>1</sup> and Thomas Hinz<sup>2</sup>

Field Methods

2020, Vol. 32(1) 23-37

© The Author(s) 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/1525822X19886827

[journals.sagepub.com/home/fmx](http://journals.sagepub.com/home/fmx)



## Abstract

Factorial surveys are widely used in the social sciences to measure respondents' attitudes, beliefs, or behavioral intentions. In such surveys, respondents evaluate short descriptions of hypothetical situations, persons, or objects that vary across several dimensions. An important prerequisite of the method's validity is that respondents are able to deal with the highly complex task created by the need to consider several variable dimensions within one coherent judgment. We analyze the effects of the order in which dimensions are presented in running text vignettes. An experimental setup with four order treatments was randomly allocated to 787 respondents (based on a random sample of register data), yielding 3,119 vignette evaluations. The analyses compare respondent groups across age, education, and response speed. Overall, there is no strong evidence for order effects. However, we find a slight tendency for fast responders to be more prone to recency effects.

<sup>1</sup> Department of Sociology, University of Munich, Germany

<sup>2</sup> Department of Sociology, University of Konstanz, Germany

## Corresponding Author:

Thomas Hinz, Department of Sociology, University of Konstanz, 78464 Konstanz, Germany.

Email: [thomas.hinz@uni-konstanz.de](mailto:thomas.hinz@uni-konstanz.de)

Factorial surveys have been widely used in the social sciences to measure respondents' attitudes, beliefs, or decisions. In such surveys, respondents are asked about one or several vignettes (scenarios) describing hypothetical situations, objects, or persons. According to the experimental design, across the vignettes, attributes (dimensions) vary in their values (levels). This allows the estimation of the impact of each dimension's value on the respondents' judgments. The method is used in many disciplines, such as sociology, political science, criminology, and consumer research. While the method is widespread (for a review, see Wallander 2009), research on methodological issues still lags behind (Auspurg and Jäckle 2015). The evaluation task, based on a multifactorial design, is relatively complex, and this could call into question substantive conclusions drawn from such studies.

We focus on one prominent issue in survey research, namely order effects, and analyze whether the order in which the vignette dimensions are presented affects results. So far, there have been only two studies on this using samples consisting of university students (Auspurg and Jäckle 2015; Robbins and Kiser 2018). We use a more heterogeneous respondent sample, which allows the study of possible variations according to respondents' cognitive abilities (measured by proxies of age and degree of education). Respondents were asked to evaluate a factorial survey module on the substantive issue of potential financial support of university students. In four experimental splits, we varied the order of the seven different vignette dimensions. In general, we found no evidence of order effects. This was true for several subgroups of respondents (higher vs. lower cognitive ability, different speeds of response). A closer look revealed a slight tendency toward recency effects (i.e., vignette dimensions presented at the very end of a vignette having a comparably higher impact). To document our analytical strategy, we provide comprehensive Online Appendices, to which we refer interested readers.

## **Overview of Current Research**

### *Order Effects in Survey Research*

The cognitive revolution in survey research of the 1980s and 1990s (e.g., Beniger and Gusek 1995) raised awareness of the fact that responsive behavior is likely to be affected by the cognitive burden of taking a survey. In general, order effects arise when the order of response alternatives affects results. In complex question formats, survey questions demand the

completion of different cognitive tasks: Respondents need to remember the question and the different response alternatives, evaluate the consequences of each alternative, and choose the one they most agree with (Schwarz 2007; Schwarz et al. 1991). Given complex tasks, respondents might be unable to actively remember all presented alternatives at the same time (Ayidiya and McClendon 1990). While alternatives presented first in a list enter long-term memory, alternatives presented at the end are more likely to enter short-term memory. When there is no delay between the presentation of the alternatives and the answering of the question, information from the short-term memory (i.e., alternatives presented at the end of a list) is more easily recalled than information from the long-term memory (Schwarz et al. 1991). This leads to recency effects. Primacy effects are the opposite of recency effects.

Heterogeneity of respondents might moderate order effects. Recalling information should be easier for individuals who are more motivated, who have more experience solving complex cognitive problems, or who even enjoy solving them (Krosnick 1989). Also, respondents with greater knowledge have more skills available to help them choose the optimal response (McClendon 1991). At the same time, people with less working memory capacity are expected to be more prone to order effects (Knäuper 1999).

Most studies have measured working memory capacity by using proxies like education or age (e.g., Holbrook et al. 2007; Krosnick and Alwin 1987). As there is a strong correlation between education and scores on tests of cognitive skills, education should be a good proxy of cognitive skills (Ceci 1991). Also, age-related reductions in cognitive resources starting around the age of 60 make it more difficult to hold large amounts of information in short-term memory (Schaie 1996; Schwarz and Knäuper 2000:248).

Respondents who are cognitively challenged when answering survey questions might follow a satisficing strategy, instead of providing an optimal answer, to decrease their cognitive effort (Krosnick 1991). Instead of thinking about all the different options simultaneously and trying to arrive at a single answer based on all of them, respondents might adopt the first reasonable alternative and base their answer disproportionately on this item. Such a satisficing approach takes much less memory capacity and therefore can be detected through lower response times.

### *Order Effects in Factorial Surveys*

In factorial survey experiments, respondents read the vignettes, which are displayed in a paper questionnaire or on a screen in web surveys, in a self-

administered mode. Factorial survey experiments are considered cognitively challenging because respondents have to take a number of (varying) dimensions into account before making a judgment. To our knowledge, there have been only two methodological articles that have focused on the order of vignette dimensions (Auspurg and Jäckle 2015; Robbins and Kiser 2018). Auspurg and Jäckle (2015) find some indications of order effects but only when the survey task becomes more complex (in terms of dimensions and number of vignettes to evaluate). Moreover, they do not find any signs of primacy or recency effects. Robbins and Kiser (2018) do not find any evidence of order effects in a similar experimental setting.

Both studies demonstrate this with vignettes in a tabular format (rarely used in the standard applications of factorial surveys where running text dominates) and with data from homogeneous samples of university students and hence samples with high levels of cognitive capacity. We expand the methodological research on order effects to the presentation of running text vignettes and to a general population sample with respondents aged between 18 and 83 years and with different levels of education. This allows us to examine to what extent possible order effects are moderated by respondents' cognitive ability.

As indicated above, it is worth considering one specific order effect, namely whether the last vignette attribute evokes recency effects. When answering a question after reading a vignette description spanning several questionnaire lines, information in a person's short-term memory might be much more easily recalled, which would favor such recency effects.

When it comes to respondents' moderators, we expect older respondents, and those with less formal education, to be more prone to order effects than younger respondents or respondents with a higher degree of education. Finally, we expect fast responders to be more prone to order effects because of satisficing strategies.

## Data and Methods

### *Data and Sample*

To answer our questions, we rely on a general population sample. The *Konstanzer Bürgerbefragung* is an annual panel survey of the citizens of the German city of Konstanz that has been conducted since 2008 (Hinz et al. 2012), with different main survey topics such as neighborhood quality, life satisfaction, and political participation. We use data collected in 2011 via computer-assisted self-interview, when the thematic focus was the

educational system. All respondents were randomly drawn from the population register, with an overall return rate of 37.1%.

The data collection included a factorial survey to evaluate opinions on who should be supported by state-financed scholarships when applying for university admission. Each respondent was asked to evaluate four vignettes describing hypothetical applicants. The total number of observations was 3,640 vignette evaluations by 910 respondents. We excluded five respondents with missing values on the vignette evaluation. Furthermore, following standard procedures, only plausible response time measures were used: All time values above the 99th percentile and under the 1st percentile were dropped; in a second step, all values were excluded that were two standard deviations above or under the mean (i.e., 80 respondents; Mayerl et al. 2005). Additionally, 38 respondents with missing values on the proxy measures of cognitive ability (age and educational level) were dropped. Excluded respondents did not differ with regard to education and response time from nonexcluded participants (there was a small difference in age—excluded respondents were 1.63 years younger on average). Finally, we use an analysis sample of 787 respondents, who evaluated 3,119 vignettes.

### *Factorial Survey Module*

All hypothetical situations describe students at the University of Konstanz applying for scholarships by the state to support the students with €500 per month (e.g., see Appendix A1). To ensure that the respondents differentiated between the students, they were told that only a third of the applicants could be supported, and it was up to the respondents to decide whether the award of a scholarship was reasonable or not. The question asked at the end of each vignette was as follows: In your opinion, should this person rightfully receive a scholarship or not? using an 11-point rating scale ranging from  $-5$  not at all to  $+5$  by all means.

The students who applied for scholarships were characterized by seven variable dimensions that are known to influence allocation decisions in education: (1) the name of the student (clearly indicating both gender and ethnic background), (2) place of birth, (3) average grade at high school graduation (grade point average), (4) parents' income, (5) number of siblings, (6) students' work experience, and (7) career orientation. Every dimension varied on two or three levels (e.g., parents' income was either low, middle, or high; for all dimension levels, see Appendix A2).

Of the set of 2,916 possible vignette scenarios, we generated a sample of 200 vignettes using a D-efficient sampling technique that reflects both

**Table 1.** Information on Treatment Split.

Voluntary social year	Treatment Split: Four Order Groups			
	1	2	3	4
Position of dimension in vignette	1 Name	Name	Name	Name
	2 Place of birth	GPA	GPA	Place of birth
	3 GPA	Place of birth	Place of birth	GPA
	4 Parents' income	Parents' income	Work experience	Work experience
	5 Siblings	Siblings	Career orientation	Parents' income
	6 Work experience	Career orientation	Parents' income	Siblings
	7 Career orientation	Work experience	Siblings	Career orientation

Note: GPA = grade point average.

orthogonality and level balance (Auspurg and Hinz 2015; Kuhfeld et al. 1994; for information on D-efficiency, see Appendix E1). With a D-efficiency value of 97.7, our design has enough statistical power for our research. We split the sample of 200 vignettes into 50 questionnaire versions (sets) of four vignettes, again using a D-efficient sampling technique that maximizes orthogonality and level balance.

To mirror standard practices in factorial survey research (Wallander 2009), we decided to present vignettes in running text (and not in a tabular order). Thus, a complete randomization of the order of all dimensions was not feasible while keeping a logical sentence structure in a running text. For instance, introducing a vignette person requires information on their names at the beginning of the text, followed by information that would be expected (e.g., place of birth and school achievement). It would be odd to start running texts describing students with dimensions of career orientations or the number of siblings and providing names and place of birth at the end of the text. Of 720 possible arrangements of dimension orders (keeping names at the beginning), we decided on four selected, different orders of dimensions as treatment groups (Table 1). The four order treatments each represent smooth vignette texts, given the structure of the German language, and enable us to test for possible overall order effects and order effects across respondent groups. Each of the 50 sets of vignettes was assigned at least once to all four treatment groups. Additionally, the order of the

vignettes within each set was randomized for each respondent. The respondents were randomly allocated to one of the sets and treatment groups.

For each of the four treatments (dimension order), we have at least 709 vignette evaluations from at least 180 respondents (for detailed information, see Appendix A3). Respondents' characteristics, such as age, educational level, response time, gender, and migration background, did not systematically vary across the four treatment groups.

### *Operationalization of Respondents' Cognitive Ability and Response Time*

We measured cognitive ability via the proxies of respondents' age and education level. Following Auspurg and Hinz (2015), respondents who are 60 years and older are classified as one subgroup, while the reference is under 60 years. As Narayan and Krosnick (1996) found that respondents with low education levels (those who did not complete high school) showed stronger order effects than respondents with a medium or high level of education (who did not differ from each other), we decided to split respondents accordingly: Respondents who at least graduated from high school are said to have a high education level, and all others are in the reference group of lower education level. Further analyses revealed that the results are robust to a different criterion (e.g., academic degree). In all estimations of age effects, education is controlled and vice versa.

The composition of the subgroups of age and education shows that the mean age of the analysis sample is 45.98 years ( $SD = 14.97$ ); 2,478 vignettes were evaluated by respondents under the age of 60 years, while older respondents evaluated 641 vignettes. More than three-fourths (75.22%) of the sample at least graduated from high school, while 24.78% only graduated from junior high school or below (see Appendix A3).

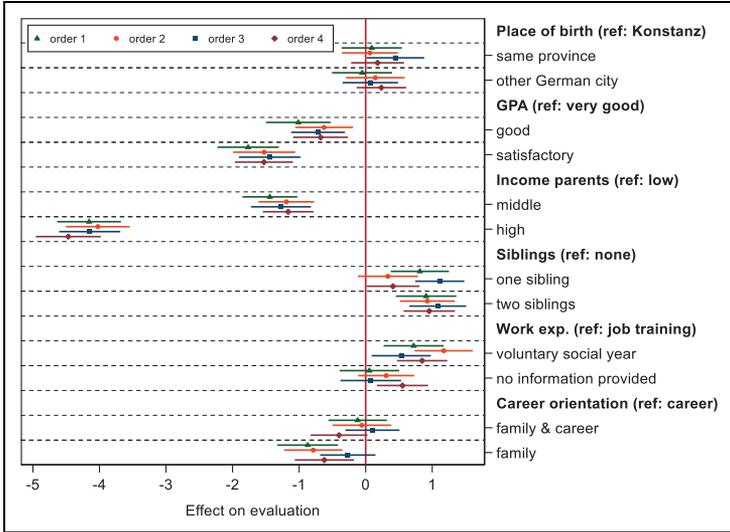
To measure response time, we used time stamps to calculate the time span respondents needed to read and evaluate the vignettes. When analyzing response time, it must be taken into account that the baseline speed for answering survey questions differs between respondents because of cognitive factors but also because of technical aspects relevant to an online survey, such as differing Internet speed or device effects (Mayerl et al. 2005). Only if the raw time needed for answering the four vignettes is calibrated by each respondent's individual baseline speed can a meaningful comparison of response time across different respondents be made. In short, this measure reveals whether the respondents needed a particularly short or long time to answer the vignettes, in comparison with the time they needed

to answer a different part of the survey (Sauer et al. 2011). In this study, baseline speed is estimated through four questions on life satisfaction that represent a medium cognitive burden. This baseline speed is used to adjust the raw time span by using a residual index approach (Urban and Mayerl 2007). Higher values on the residual index indicate that respondents needed more time to answer the vignettes than their baseline speed would indicate; accordingly, lower values indicate a faster response time. By means of a median split, the respondents were classified into the two subgroups fast and slow responders.

### *Analytical Strategy*

In all estimations, vignette evaluations are regressed on all seven vignette dimensions by using ordinary least squares (OLS) regressions. To adjust for the unequal variances of error terms caused by the clustering of evaluations within respondents, we use clustered standard errors (Rogers 1993). Detailed information on all models and further robustness checks can be found in extensive tables in the Appendices (e.g., Appendix E).

We test for overall order effects within the full sample and within each of the six subgroups (under 60 years, 60 years and older, higher education level, lower education level, fast responders, and slow responders). All tests are estimated for the full sample and for each of the six subgroups, separately. In a first step, we estimate OLS regressions for all four order treatments. Differences in regression coefficients between the four order splits would give us first indications of order effects. Second, we estimate a joint model of all four order splits, including two-way interaction terms of each dimension with an indicator variable for the order split. Here, a joint Wald test (also known as a “Chow test,” see Wooldridge 2009) allows us to test whether the interaction terms are all jointly zero, meaning there are no significant differences between the order splits caused by the order of the dimensions. Third, Wald tests for each individual vignette dimension allow an even closer look at dimension order effects: Now we compare each individual dimension’s coefficient across all four treatment groups. Fourth, with “moderator tests” (Auspurg and Jäckle 2015), we test whether possible differences in the extent of order effects between the relevant subgroups are significant. The test logic is comparable to that described above, but instead of two-way interaction terms of the relevant dimension and the order indicator, we introduce three-way interaction terms, with the respective moderator variable (age, education, or speed) as an additional variable.



**Figure 1.** Coefficient plot of regression of vignette evaluation by treatment splits. The figure shows regression coefficients of the vignette dimensions by the four treatment splits (see Appendix B1 for the regression table). Only dimensions whose positions change between treatment splits are shown. Standard errors with 95% confidence intervals are attached.

To focus on the occurrence of recency effects, we apply a similar strategy, again separately for the full sample and for all subgroups. For each of the three dimensions that are placed on the last position of the vignette in at least one order split (i.e., career orientation, work experience, and siblings), we again estimate separate OLS regressions including a two-way interaction term of the relevant dimension with an indicator as to whether the dimension was in the last position. A significant interaction term indicates a recency effect. Again, we use joint Wald tests to test whether the interaction terms are zero.

## Results

### *Is There an Overall Order Effect?*

For the full sample, Figure 1 depicts the influence of each dimension on the vignette evaluation for all four order splits. With the exception of the dimension level “one sibling,” there is no clearly visible difference between

**Table 2.** Wald Tests and Moderator Tests for Full Sample and Subgroups.

	Wald Test		Moderator Test	
	<i>N</i>	<i>F</i> ( <i>p</i> )	<i>N</i>	<i>F</i> ( <i>p</i> )
Full sample	2,030	0.86 (.697)		
Under 60 years	1,615	1.30 (.124)	2,030	1.17 (.235)
60 years and older	415	0.80 (.774)		
Higher education	1,515	1.06 (.380)	2,030	0.82 (.755)
Lower education	515	0.72 (.868)		
Fast responders	1,014	1.31 (.123)	2,030	0.85 (.704)
Slow responders	1,016	0.61 (.956)		

**Table 3.** Individual Wald Tests (Full Sample).

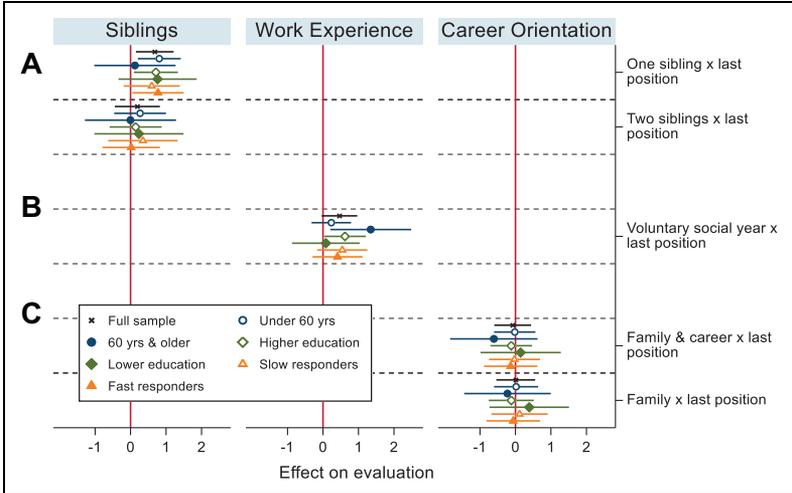
Comparison	Dimension	$\chi^2$ ( <i>p</i> )
Order 2 versus 3	Siblings (ref.: no siblings)	
	One sibling	6.987 (.008)
	Work experience (ref.: job training)	
Order 3 versus 4	Voluntary social year	4.175 (.041)
	Siblings (ref.: no siblings)	
	One sibling	4.349 (.037)

Note: We compare the dimensions' coefficients between two order groups. See Appendix C1 for nonsignificant combinations and further information.

the four treatment groups. The result of the joint Wald test supports the conclusion that there is no overall dimension order effect in the full sample ( $F = .86, p = .697$ ; see Table 2).

In Table 3, Wald tests for each individual vignette dimension show that only 3 of all 66 possible comparisons of the dimensions' positions indicate significant order effects. Those 4.5% significant interaction terms are expected to be significant by chance only. However, it is worth noting that all three effects indicate recency effects (i.e., their impact on the vignette evaluations is higher if the dimension is the last one posed in the vignette).

Within the six respondent subgroups, the results of the OLS regressions do not point toward order effects (see Appendix B2–B4). Moreover, Wald tests reveal no significant order effects (Table 2). Also, additional Wald tests for each interaction term individually do not point toward subgroups that are particularly prone to dimension order effects (not displayed, see Appendix C2–C7). The only exception are fast responders, where



**Figure 2.** Coefficient plot of recency effects for the full sample and subgroups. This figure shows regression coefficients of two-way interaction terms of the relevant dimension level ([A] siblings, [B] work experience, [C] career orientation), with an indicator that the dimension was in the last position. Standard errors with 95% confidence intervals are attached. See Appendix D2–D5 for regression tables.

5 of 66 possible dimensions show significant order effects. Those 7.6% are slightly more than could be expected to occur by chance (for a detailed interpretation, see Appendix E3).

The moderator tests do not support the idea that respondents with lower cognitive abilities are more prone to order effects compared with respondents with higher cognitive abilities (see Table 2). In addition, the two speed groups do not differ in their susceptibility to order effects.

### Is There a Recency Effect?

In Figure 2A, the point estimates for the full sample suggest that the dimension siblings show significant recency effects for one of the dimension levels: The effect size increases when the dimension level “one sibling” is in the last position ( $p = .012$ ). The point estimates for work experience (Figure 2B) and career orientation (Figure 2C) do not show any significant recency effects. All results are supported by joint Wald tests (see Appendix D1).

The results in Figure 2 indicate recency effects in the dimension work experience ( $p = .021$ ; Figure 2B) for people over 60, while younger people show recency effects in one level of the dimension siblings ( $p = .009$ ; Figure 2A). Concerning education, if any effect is visible, it is that more educated respondents show more recency effects (for a detailed interpretation, see Appendix E4). In sum, the results do not clearly suggest that respondents with lower cognitive abilities are consistently more prone to recency effects than respondents with higher cognitive abilities. Concerning response time, *fast responders* are, as expected, slightly more prone to recency effects than *slow responders*. This difference is, however, based only on one significant recency effect for fast responders (see Figure 2A, dimension level “one sibling,”  $p = .036$ ).

## Discussion and Conclusion

Our study adds a comprehensive experimental test of effects of dimension order in factorial survey experiments to the methodological literature on survey research. We asked whether the position of vignette dimensions in running text systematically influences the respondents’ evaluations. We tested for *overall order effects* and for *recency effects* (i.e., whether, because of the complex structure of all the information to be retained and because of their proximity to the eventual evaluation task, respondents focus specifically on dimensions at the *end* of the running text). Moreover, in contrast to prior research that was based on student samples only, by using a general population sample, we were able to identify heterogeneous order effects across subgroups.

Following standard procedures, age, education, and response speed (as grouped variables) served as indicators of respondents’ cognitive abilities. Overall, for our vignette module with four evaluations per respondent and seven different vignette dimensions, using a standard 11-point rating response scale, we found almost no systematic overall order effects. At best, there was weak evidence of recency effects within the group of fast responders. By age and education, only few and inconclusive effects can be reported, again with a slight tendency toward recency effects. In sum, the study supports the overall validity of factorial survey experiments with a medium degree of complexity (seven dimensions are a typical design in this research; Wallander 2009) in regard to the order in which vignette dimensions are presented.

Evidently, the study also has its limitations. First, the setup did not allow us to test for *primacy* effects because the first dimension was actually

constant across all cases: All vignettes started with names indicating gender and ethnic background, for logical reasons. We indeed started variation at the second dimension, which is less salient in running texts than the first dimension. Second, we tested across seven dimensions, which is a typical number of dimensions used in applications. There are, however, applications using a higher number of dimensions, and there are good arguments that order effects most probably emerge only with more complex vignettes (Auspurg and Jäckle 2015). Third, even more importantly, for obvious practical reasons, we tested only a subsample of possible vignette dimension orders. In all four treatment conditions, the attributes of the hypothetical applicants in the final position of the vignette texts only marginally influenced the substantive judgments in the evaluation task. One could argue that vignette dimensions with a higher substantive relevance, such as parental income and achievements at high school, as compared to number of siblings, career orientation, and work experience, could change the conclusions. This should be tested in future research.

Nevertheless, the study has its specific strengths in its rigorous tests based on data from a heterogeneous respondent population. The good news for practitioners of factorial survey experiments is that we did not find strong order effects of vignette dimensions at all. However, we also learned from the tests that vignette evaluations are obviously prone to some (small) recency effects. In designing factorial survey experiments, and particularly dimension order, researchers are well advised to consider a logical, realistic structure for their scenarios—with a specific focus on the dimension placed last. We suggest avoiding having dimensions with the highest supposed influence at the very end of running text.

The most challenging result of our study seems to be the sensitiveness of order effects to the differing speeds of responders. Those who take questionnaires quickly (who are probably “satisficers”) are slightly more prone to possible bias caused by order effects. The use of online access panels is increasing, and respondents to these ubiquitous surveys are probably more likely to provide quick and imprecise responses, which might also invalidate results in other ways. Thus, another conclusion from our study points to the relevance of para-data when applying factorial survey experiments—and survey research in general. Such data might promote the elimination of possible bias caused by order effects and other effects related to satisficing behavior.

## Acknowledgments

We are very grateful to Katrin Auspurg and two anonymous reviewers for their valuable comments on an earlier version of this article.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Supplemental Material

Supplemental material for this article is available online.

## References

- Auspurg, K., and T. Hinz. 2015. *Factorial survey experiments*. Thousand Oaks, CA: Sage.
- Auspurg, K., and A. Jäckle. 2015. First equals most important? Order effects in vignette-based measurement. *Sociological Methods & Research* 46:490–539.
- Ayidiya, S. A., and M. J. McClendon. 1990. Response effects in mail surveys. *Public Opinion Quarterly* 54:229–47.
- Beniger, J. R., and J. A. Gusek. 1995. The cognitive revolution in public opinion and communication research. In *Public opinion and the communication of consent*, edited by T. L. Glasser and C. T. Salmon, 217–48. New York: Guilford.
- Ceci, S. J. 1991. How much does schooling influence general intelligence and its cognitive components? A reassessment of evidence. *Developmental Psychology* 27:703–22.
- Hinz, T., A. Joos, and K. Schleicher. 2012. Bildung und Lebenszufriedenheit. Ergebnisse der Konstanzer Bürgerbefragung 2011–4. Welle. Statistik Bericht 5. Konstanz, Germany: University of Konstanz.
- Holbrook, A. L., J. A. Krosnick, D. Moore, and R. Tourangeau. 2007. Response order effects in dichotomous categorical questions presented orally. *Public Opinion Quarterly* 71:325–48.
- Knäuper, B. 1999. The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly* 63:347–70.
- Krosnick, J. A. 1989. Attitude importance and attitude accessibility. *Personality and Social Psychology Bulletin* 15:297–308.

- Krosnick, J. A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5:213–36.
- Krosnick, J. A., and D. F. Alwin. 1987. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly* 51: 201–19.
- Kuhfeld, W. F., R. D. Tobias, and M. Garratt. 1994. Efficient experimental design with marketing research applications. *Journal of Marketing Research* 31:545–57.
- Mayerl, J., P. Sellke, and D. Urban. 2005. Analyzing cognitive processes in CATI-surveys with response latencies: An empirical evaluation of the consequences of using different baseline speed measures. *Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart* 2:1–18.
- McClendon, M. J. 1991. Acquiescence and recency response-order effects in interview surveys. *Sociological Methods & Research* 20:60–103.
- Narayan, S., and J. A. Krosnick. 1996. Education moderates some response effects in attitude measurement. *Public Opinion Quarterly* 60:58–88.
- Robbins, B., and E. Kiser. 2018. Legitimate authorities and rational taxpayers: An investigation of voluntary compliance and method effects in a survey experiment of income tax evasion. *Rationality and Society* 30:247–301.
- Rogers, W. H. 1993. Regression standard errors in clustered samples. *Stata Technical Bulletin* 3:19–23.
- Sauer, C., K. Auspurg, T. Hinz, and S. Liebig. 2011. The application of factorial surveys in general population samples: The effects of respondent age and education on response times and response consistency. *Survey Research Methods* 5:89–102.
- Schaie, K. W. 1996. *Intellectual development in adulthood. The Seattle Longitudinal Study*. Cambridge: Cambridge University Press.
- Schwarz, N. 2007. Cognitive aspects of survey methodology. *Applied Cognitive Psychology* 21:277–87.
- Schwarz, N., and B. Knäuper. 2000. Cognition, aging, and self-reports. In *Cognitive aging: A primer*, edited by D. Park and N. Schwarz, 233–52. Philadelphia: Psychology Press.
- Schwarz, N., F. Strack, H.-J. Hippler, and G. Bishop. 1991. The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology* 5:193–212.
- Urban, D., and J. Mayerl. 2007. Antwortlatenzzeiten in der survey-basierten Verhaltensforschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59: 692–713.
- Wallander, L. 2009. 25 years of factorial surveys in sociology: A review. *Social Science Research* 38:505–20.
- Wooldridge, J. M. 2009. *Introductory econometrics. A modern approach*, 4th ed. Mason, OH: South Western.