

LONGER INTERVIEWS MAY NOT AFFECT SUBSEQUENT SURVEY PARTICIPATION PROPENSITY

PETER LYNN*

Abstract Survey researchers often assume that respondent burden is an important determinant of survey participation propensity and that interview length is a good indicator of burden. However, there is surprisingly little evidence of the effect of length of a completed interview on subsequent participation propensity, particularly in the case of face-to-face surveys. This article presents results from a large-scale randomized experiment in which respondents experienced interviews of different lengths at wave 1 of a panel survey. Subsequently, respondents were asked to complete a self-completion questionnaire and then to take part in further waves of the survey. For each of these subsequent tasks, the study compares completion rates between those administered the shorter and those administered the longer version of the wave 1 interview. No evidence is found that wave 1 interview length affects subsequent participation propensity.

Introduction

Survey designers must decide how much content should be included in a questionnaire. There is often pressure to include additional questions, as this can widen the analytical potential of a survey data set for a relatively modest cost. But additional questions can also have negative implications. The additional time that it consequently takes a respondent to complete the interview or questionnaire may impose greater cognitive burden, greater discomfort, or greater disruption to their other activities. These difficulties are characterized by survey researchers as components of *respondent burden* (Bradburn 1978;

PETER LYNN is a professor of survey methodology at the Institute for Social and Economic Research, University of Essex, Colchester, UK. This study was funded by an Economic and Social Research Council Award [ES/H029745/1, “Understanding Society and the UK Longitudinal Studies Centre,” to Nick Buck]. Important contributions to the design, management, and data preparation of the waves of the Innovation Panel used in this study were made by Noah Uhrig, Jon Burton, Alita Nandi, Gundi Knies, and other ISER colleagues. Data collection was carried out by NatCen Social Research. *Address correspondence to Peter Lynn, ISER, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK; e-mail: plynn@essex.ac.uk.

Sharp and Frankel 1983), a phenomenon that may affect the quality of answers given to survey questions, particularly later in an interview (Krosnick 1999; Holbrook, Green, and Krosnick 2003; Galesic 2006; Roberts et al. 2010), and may affect the willingness of the respondent to undertake subsequent additional survey tasks. The latter consideration is particularly pertinent in the case of longitudinal surveys, where value comes from having the same respondents participate repeatedly. Thus, to help survey designers make decisions about instrument length, this study presents evidence of the effect of additional content on respondent propensity to agree to participate in subsequent survey tasks. Data come from a large-scale randomized experiment with a nationally representative sample in which respondents experienced interview instruments of different lengths at wave 1 of a panel survey. The study presents estimates of the effect of interview length on completion rates of a self-completion questionnaire administered immediately after the interview and on requests for subsequent interviews at one-year intervals.

Background

As a concept, questionnaire/interview length is broad and has been defined and operationalized in different ways by different researchers.

In the context of studying the effect of questionnaire length on participation propensity in mail surveys, researchers have tended to focus on the number of questionnaire pages or, less frequently, the number of items (Yammarino, Skinner, and Childers 1991; Dillman, Sinclair, and Clark 1993; Dillman, Smyth, and Christian 2008). This approach makes sense, as in the mail survey context the entire questionnaire is visible to the sample member before he or she decides whether or not to complete it. It is likely that a sample member will form an idea of the amount of effort required to complete the questionnaire based on a cursory impression of the number of pages and questions. Longer questionnaires have generally been found to reduce response rates on mail surveys (Heberlein and Baumgartner 1978; Yammarino, Skinner, and Childers 1991; Dillman, Smyth, and Christian 2008), though Champion and Sear (1969) found that, holding the number of items constant, a three-page questionnaire obtained a lower response rate than either six-page or nine-page versions.

In other survey modes, sample members are unable to form their own impression of the amount of effort required to participate, but instead rely on information communicated by the researcher. The communication may be written (a prenotification letter, an invitation to participate in a web survey) or verbal (via an interviewer). Studies of the effect of length on participation propensity have therefore tended to focus on the statement made by the researcher/ interviewer about the anticipated time required to participate (e.g., Collins et al. 1988; Groves et al. 1999; Crawford, Couper, and Lamias 2001;

Galesic and Bosnjak 2009). It has been found that increasing the time that respondents are told the survey will take reduces response rates both in web surveys (Crawford, Couper, and Lamias 2001; Marcus et al. 2007; Galesic and Bosnjak 2009) and in telephone surveys (Collins et al. 1988; Roberts et al. 2010). Groves et al. (1999) report a similar finding for face-to-face surveys, though this is based on a laboratory study with low external validity.

Information provided to a sample member prior to starting the interview/questionnaire is the only aspect of length that can affect an initial decision to participate. But this information is not necessarily a good predictor of the actual time that participation will require. Even if the researcher in good faith communicates an accurate mean interview length, the time that the interview will take for any individual respondent can vary substantially from the mean due to variation in circumstances, cognitive ability, response styles, and so on. Thus, the actual completion time, rather than the anticipated time, is likely to influence propensity to continue responding, and willingness to respond to additional requests. This is especially pertinent in the web survey context, where the study of breakoffs has paid attention to the role of elapsed time (Haraldsen 2002; Galesic 2006; Peytchev 2009). In interviewer-administered modes, breakoff is far less common and sample members who commence an interview usually complete it, regardless of the correspondence between anticipated and actual completion time. Furthermore, it has been argued that the respondent's perception of completion time may be more important than actual completion time (Bradburn 1978; Holbrook, Green, and Krosnick 2003), and that tolerance of longer interviews may be mode-dependent: Holbrook, Green, and Krosnick (2003) found that telephone respondents were more likely than face-to-face respondents to express dissatisfaction with the length of interview, even though their interviews were in fact shorter.

In the context of a longitudinal survey, or any other survey in which additional participation requests are made after the initial interview, there is scope for completion time to influence the respondent's disposition to future requests. There have been only two prior experimental studies of this issue on interviewer-administered surveys. With face-to-face interviewing, Sharp and Frankel (1983) found no significant difference in wave 2 response rates between respondents administered a 25-minute interview and those administered a 75-minute interview at wave 1. With telephone interviewing, Fricker et al. (2012) found a lower wave 3 attrition rate among respondents administered a 21-minute wave 2 interview than among those administered a 29-minute interview. Almost two decades ago, Bogen (1996) concluded that there is very little evidence of the effect of different-length instruments on subsequent survey participation. That state of affairs has hardly been altered.

Other studies have looked at the effect of variation in administration time of a single instrument. For example, Branden, Gritz, and Pergamit (1995) found that a longer interview was associated with less attrition in a telephone survey and concluded that this was likely due to greater salience of the questionnaire

content. But this addresses a different issue, the variation between respondents in the experience of the same instrument.

McCarthy, Beckler, and Qualey (2006) examined a different aspect of prior participation experience, namely the effect on subsequent participation of the number of previous survey requests received. They found that larger numbers of requests were associated with greater participation propensity, but their study was not based on an experimental design, so the association may not be causal.

Data

Data are from an experiment mounted on the UK Household Longitudinal Study Innovation Panel (UKHLS-IP). The UKHLS-IP is set up for the purpose of methodological development and testing (Uhrig 2011), primarily to service the main UKHLS, an academically led study designed to provide a multi-disciplinary research resource (Buck and McFall 2012; Hobcraft and Sacker 2012). The UKHLS involves annual interviews with around 70,000 adult sample members and all other adult members of their current household, while the UKHLS-IP is based on a much smaller sample—around 2,500 adults at wave 1—but follows broadly the same design as the main survey.

The UKHLS-IP has a clustered, stratified, probability-sampling design (Lynn 2009). In summary, 120 postcode sectors were selected with probability proportional to size from a list of all sectors in Great Britain,¹ stratified by geographic region, socioeconomic classification, and population density. In each sector, an equal-probability sample of 23 residential addresses was selected. Each address was visited by an interviewer whose task was to identify all resident persons, all of whom became UKHLS-IP sample members. At wave 1 of the survey, all sample members aged 16 or over were eligible for an individual interview.

The experiment that forms the basis of the study reported here was incorporated into wave 1 of the UKHLS-IP, for which data collection took place from January to April 2008. Sample members were randomly assigned to two equal-sized groups, each of which were administered one of two versions of the individual interview, referred to hereafter as the “short interview” and the “long interview.” The short interview consisted of 26 modules of questions, administered face-to-face by a trained survey interviewer in the respondent’s home, using a CAPI instrument programmed in Blaise. Topics included demographics, family background, education, health and disability, employment and labor-market activity, job satisfaction, ethnicity, national identity, religion, and income. The long interview consisted of exactly the same instrument, but

1. Excluding that part of Scotland that lies north of the Caledonian Canal, a remote area that accounts for approximately 0.1 percent of the population of Great Britain.

with two additions. First, a module on partnership history was extended to include a complete lifetime fertility history. Second, an additional module of questions on attitudes to environmental issues was added.

Interview completion times were obtained from electronic time stamps incorporated into the CAPI script. Time was recorded at the start of the interview, at the start and end of each of the 27 modules of questions that constituted the interview, and at the end of the interview. In the analysis, total interview completion time is derived as the difference between the time at the end of the interview and the time at the start,² while completion time for a specific module is analogously derived as the difference between the times at the end and start of that module.

Mean interview completion time was 20 percent longer ($P < 0.0005$) with the long interview (31 minutes 3 seconds) than with the short interview (25 minutes 50 seconds), though there was considerable variation between respondents in interview length, partly reflecting the heavily routed nature of the questionnaire (figure 1). The additional interview time for the fertility and partnership history module was greater for women than for men (table 1), resulting in a significant difference between men and women in the time taken to complete the long-interview version of that section (mean 2 minutes 45 seconds for men, 3 minutes 9 seconds for women, $P = 0.03$). There was no difference between men and women in the mean completion time for the environmental attitudes module (mean 2 minutes 4 seconds for men, 2 minutes 10 seconds for women, $P = 0.18$). On average, women took 1 minute 25 seconds longer than men to complete the short interview ($P = 0.04$), but women took 3 minutes 12 seconds longer than men to complete the long interview ($P < 0.0005$).

Once an interview was completed, the interviewer requested the respondent to additionally complete a 22-item self-completion paper questionnaire with questions on the quantity and quality of sleep, mental health, neighborhood attachment, life satisfaction, attitudes toward risk, and friendship networks. The questionnaire was a 12-page booklet with instructions on the cover page,

2. Among the 2,399 respondents, there were six for whom time-stamp data was unavailable and one for whom the time-stamp values were implausible. Analysis of interview completion time is restricted to the remaining 2,392 respondents. Edit checks for internal consistency and external validity were applied and resulted in total interview time being edited in 193 cases. These included 96 cases for whom total interview time was missing and was replaced by the sum of the times for the 27 modules. For the other 97 cases, total interview time was implausibly long (over 2 hours). For 48 of these cases, total interview time was replaced by the sum of the times for the 27 modules. For the other 49, this would still have resulted in an implausibly long interview time. Inspection revealed that these were all cases with excessively long times for one or more of the last three modules. These modules asked, respectively, consent to linkage of administrative data to the survey responses, respondent's contact details, and contact details for stable contacts. For these 49 cases, the total interview time was reduced by the difference between the recorded time for these three modules and the mean time among other respondents for these three modules.

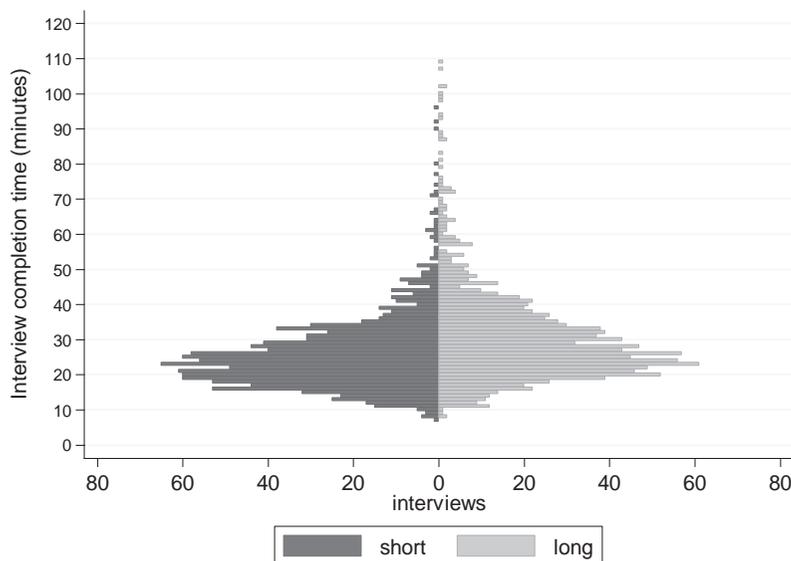


Figure 1. Distribution of Interview Length, for Short- and Long-Interview Treatment Groups.

nine pages of questions, and two blank pages. Where possible, the respondent handed the completed questionnaire back to the interviewer before he or she left the house (often, respondents would complete the questionnaire while the interviewer was interviewing another household member); in other cases, the questionnaire was collected on a later visit or mailed back to the survey organization.

Table 1. Completion Time for the Fertility History and Environmental Attitudes Modules, Means and Standard Deviations by Treatment Group and Gender

	Long interview			Short interview		
	Men	Women	Total	Men	Women	Total
Fertility and partnership history – Mean	2:45	3:09	2:58	0:57	0:53	0:55
(Standard deviation)	(2:34)	(3:42)	(3:15)	(0:35)	(0:37)	(0:36)
Environmental attitudes – Mean	2:04	2:10	2:07	–	–	–
(Standard deviation)	(1:08)	(1:34)	(1:23)	–	–	–
<i>N</i>	544	645	1,189	534	669	1,203

NOTE.—All cell entries are minutes:seconds.

Wave 1 respondents were subsequently reapproached and asked to take part in further interviews at approximately one-year intervals. Wave 2 took place between April and June 2009, wave 3 between April and June 2010, wave 4 between March and May 2011, and wave 5 between March and July 2012.

Results

Table 2 presents completion rates for each of the five subsequent survey requests that were made of the wave 1 respondents. No significant difference is observed in any of the completion rates between sample members administered the short interview at wave 1 and those administered the long interview. In all cases, the rates are remarkably similar. The same is true when the analysis is restricted to women, for whom the difference in wave 1 interview length was greater (table 3).

To test whether these results might depend on actual interview completion time, logistic regression analysis of each of the five completion rates was carried out, in which the independent variables were interview version (long or short), quartile of the interview length distribution,³ and the interaction between these two factors. None of the interaction terms were significant, indicating no significant effect of interview version on outcome for any of the quartiles (results not shown⁴).

Table 2. Outcome Rates for Subsequent Survey Requests, Full Sample

Outcome	Treatment		<i>P</i>
	Short wave 1 interview	Long wave 1 interview	
Self-completion questionnaire	90.0% (<i>n</i> = 1,205)	89.7% (<i>n</i> = 1,194)	0.78
Wave 2 interview	70.1% (<i>n</i> = 1,201)	70.5% (<i>n</i> = 1,187)	0.83
Wave 3 interview	63.0% (<i>n</i> = 1,190)	64.5% (<i>n</i> = 1,157)	0.46
Wave 4 interview	57.8% (<i>n</i> = 1,182)	57.0% (<i>n</i> = 1,147)	0.71
Wave 5 interview	49.9% (<i>n</i> = 1,171)	49.0% (<i>n</i> = 1,141)	0.67

NOTE.—The completion rate for each of the five survey tasks is conditional on participation at wave 1. The wave 1 response rate was 53.7 percent (AAPOR RR1). Additionally, for the wave 2 to 5 interview rates, sample members known to have died prior to that wave, and therefore ineligible to be interviewed, have been excluded from the base. *P*-values are based on independent chi-square tests for each of the five survey requests.

3. To avoid endogeneity, for the purpose of this analysis interview length was defined as the time taken for all modules excluding the two modules in which the content differed between the two treatment groups. Thus, respondents are assigned to quartiles based on the time taken for identical content.
4. The results of this analysis are available in the supplementary materials online.

Table 3. Outcome Rates for Subsequent Survey Requests, Women

Outcome	Treatment		<i>P</i>
	Short wave 1 interview	Long wave 1 interview	
Self-completion questionnaire	90.8% (<i>n</i> = 671)	90.6% (<i>n</i> = 648)	0.91
Wave 2 interview	70.6% (<i>n</i> = 670)	70.5% (<i>n</i> = 647)	0.96
Wave 3 interview	63.2% (<i>n</i> = 663)	65.9% (<i>n</i> = 633)	0.31
Wave 4 interview	57.9% (<i>n</i> = 658)	57.7% (<i>n</i> = 627)	0.95
Wave 5 interview	50.0% (<i>n</i> = 654)	49.0% (<i>n</i> = 624)	0.73

Discussion

This study has found no effect of the length of the wave 1 survey instrument on subsequent participation propensity. This finding holds both for a request immediately following the interview (to complete a self-completion questionnaire) and for later requests to take part in interviews similar to the first interview. The finding also holds both for respondents whose wave 1 interview was relatively short (due to questionnaire routing) and for those whose interview was relatively long. The mean difference in length of the wave 1 interview between the two instrument versions was modest, with the longer version being just over five minutes longer—a 20 percent increase in completion time. However, a difference of this magnitude represents a real decision that a survey designer may face. It is not unusual for survey designers to have to balance the analytical advantages of adding a few extra minutes of questioning against the perceived disadvantages of increasing the respondent burden by making the interview longer. The evidence presented here suggests that the disadvantage in terms of subsequent participation propensity may be negligible or nonexistent. This conclusion should be of particular interest to longitudinal survey managers.

It is reasonable to suppose that any effect of interview length might depend both on the absolute difference in completion time (five minutes in the present case) and on the relative difference (20 percent). Thus, we cannot rule out that a larger difference than the five minutes studied here could have an effect. Similarly, the finding of no effect for an absolute difference of five minutes might not hold if this corresponded to a larger relative difference; in other words, a shorter baseline completion time (though see [Sharp and Frankel \[1983\]](#)). As suggested by [Hansen \(2007\)](#), further experimental research could usefully shed light on this.

To put the findings of this study in context, it should be noted that increased respondent burden may have other effects. Though a participant may be willing to continue responding, the quality of provided responses could suffer when

burden is increased (Sharp and Frankel 1983; Holbrook, Green, and Krosnick 2003). Furthermore, we have no evidence on the relationship between interview completion time and perceived burden. It is plausible that the association is weak, with perceived burden being influenced at least as much by other factors, such as interest in the survey topic(s) and likeableness of the interviewer. In the case of this experiment, the additional questions in the long interview may have had a different level of salience to respondents, on average, than the questions in the rest of the interview.

Aside from potentially affecting respondent burden, increasing the interview completion time affects survey costs. Interviewers must be paid for longer hours, and a greater volume of data must be processed and managed. These factors should all be considered when deciding whether additional interview content is worthwhile.

Supplementary Data

Supplementary data are freely available online at <http://poq.oxfordjournals.org/>.

References

- Bogen, Karen. 1996. "The Effect of Questionnaire Length on Response Rates—A Review of the Literature." Proceedings of the Survey Research Methods Section of the American Statistical Association, 1020–25. Available at http://www.amstat.org/sections/srms/Proceedings/papers/1996_177.pdf.
- Bradburn, Norman. 1978. "Respondent Burden." Proceedings of the Survey Research Methods Section of the American Statistical Association, 35–40. Available at http://www.amstat.org/sections/srms/Proceedings/papers/1978_007.pdf.
- Branden, Laura, R. Mark Gritz, and Michael R. Pergamit. 1995. "The Effect of Interview Length on Nonresponse in the National Longitudinal Survey of Youth." Proceedings of the 1995 Census Bureau Annual Research Conference, 129–54. Available at <http://www.bls.gov/osmr/pdf/nl950030.pdf>.
- Buck, Nick, and Stephanie McFall. 2012. "Understanding Society: Design Overview." *Longitudinal and Life Course Studies* 3:5–17.
- Champion, Dean J., and Alan M. Sear. 1969. "Questionnaire Response Rate: A Methodological Analysis." *Social Forces* 47:335–39.
- Collins, Martin, Wendy Sykes, Paul Wilson, and Nora Blackshaw. 1988. "Nonresponse: The UK Experience." In *Telephone Survey Methodology*, edited by Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls II, and Joseph Waksberg, 213–31. New York: Wiley.
- Crawford, Scott D., Mick P. Couper, and Mark J. Lamias. 2001. "Web Surveys: Perceptions of Burden." *Social Science Computer Review* 19:146–62.
- Dillman, Don A., Michael D. Sinclair, and Jon R. Clark. 1993. "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys." *Public Opinion Quarterly* 57:289–304.
- Dillman, Don A., Jolene Smyth, and Leah Melani Christian. 2008. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 3rd ed. Hoboken, NJ: Wiley.
- Fricker, Scott, Brett Creech, Jeanette Davis, Jeffrey Gonzalez, Lucilla Tan, and Nhien To. 2012. "Does Length Really Matter? Exploring the Effects of a Shorter Interview on Data Quality,

- Nonresponse, and Respondent Burden." Paper presented at the Federal Committee on Statistical Methodology 2012 Research Conference, Washington, DC, USA. Available at http://www.fcsm.gov/12papers/Fricker_2012FCSM_IX-B.pdf.
- Galesic, Mirta. 2006. "Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey." *Journal of Official Statistics* 22:313–28.
- Galesic, Mirta, and Michael Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73:349–60.
- Groves, Robert M., Eleanor Singer, Amy D. Corning, and Ashley Bowers. 1999. "A Laboratory Approach to Measuring the Effects on Survey Participation of Interview Length, Incentives, Differential Incentives, and Refusal Conversion." *Journal of Official Statistics* 15:251–68.
- Hansen, Kasper M. 2007. "The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study." *International Journal of Public Opinion Research* 19:112–21.
- Haraldsen, Gustav. 2002. "Identifying and Reducing the Response Burden in Internet Business Surveys." Paper presented at the International Conference on Questionnaire Development, Evaluation and Testing Methods (QDET), Charleston, South Carolina, USA.
- Heberlein, Thomas A., and Robert Baumgartner. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 43:447–62.
- Hobcraft, John, and Amanda Sacker. 2012. "Guest Editorial: The Origins of Understanding Society." *Longitudinal and Life Course Studies* 3:1–4.
- Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67:79–125.
- Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50:537–67.
- Lynn, Peter. 2009. "Sample Design for Understanding Society." Understanding Society Working Paper 2009-01. Colchester, UK: University of Essex. <https://www.iser.essex.ac.uk/publications/working-papers/understanding-society/2009-01>.
- Marcus, Bernd, Michael Bosnjak, Steffen Lindner, Stanislav Pilishchenko, and Astrid Schuetz. 2007. "Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Nonresponse in Web Surveys." *Social Science Computer Review* 25:372–83.
- McCarthy, Jaki Stanley, Daniel G. Beckler, and Suzette M. Qualey. 2006. "An Analysis of the Relationship Between Survey Burden and Nonresponse: If We Bother Them More, Are They Less Cooperative?" *Journal of Official Statistics* 22:97–112.
- Peytchev, Andy. 2009. "Survey Breakoff." *Public Opinion Quarterly* 73:74–97.
- Roberts, Caroline, Gillian Eva, Nick Allum, and Peter Lynn. 2010. "Data Quality in Telephone Surveys and the Effect of Questionnaire Length: A Cross-National Experiment." Institute for Social and Economic Research Working Paper 2010-36. Colchester, UK: University of Essex. Available at <https://www.iser.essex.ac.uk/publications/working-papers/iser/2010-36>.
- Sharp, Laure M. and Joanne Frankel. 1983. "Respondent Burden: A Test of Some Common Assumptions." *Public Opinion Quarterly* 47:36–53.
- Uhlig, S. C. Noah. 2011. "Using Experiments to Guide Decision Making in Understanding Society: Introducing the Innovation Panel." In *Understanding Society: Early Findings from the First Wave of the UK's Household Longitudinal Study*, edited by Stephanie L. McFall and Christine Garrington, chapter 13. Colchester, UK: University of Essex. Available at <http://research.understandingsociety.org.uk/findings/early-findings>.
- Yammarino, Francis J., Steven J. Skinner, and Terry L. Childers. 1991. "Understanding Mail Survey Response Behavior: A Meta-Analysis." *Public Opinion Quarterly* 55:613–39.