

A Laboratory Approach to Measuring the Effects on Survey Participation of Interview Length, Incentives, Differential Incentives, and Refusal Conversion

Robert M. Groves¹, Eleanor Singer², Amy D. Corning², and Ashley Bowers³

A laboratory-based experimental method is used to study influences on survey participation. The experiment tested specific hypotheses concerning the combined effects of person-level attributes and survey design features on cooperation with a survey request. The laboratory protocol presented subjects with videotaped vignettes portraying interviewer persuasion attempts, and then elicited their reactions to them.

The vignettes were constructed so that three factors were systematically varied in a factorial design: (1) disclosure of differential payment to other sample households by the interviewer; (2) refusal conversion efforts by the interviewer; and (3) burden, as reflected in the length of interview. These factors were completely crossed in the design. In addition, two other factors were partially crossed: offering a rationale for the differential payment versus no rationale; and offering an incentive prior to the interview versus offering no incentive.

This article presents a description of the methodology as well as analyses testing the effects of the experimental factors.

Key words: Nonresponse; burden; laboratory; experiment; vignettes; incentives.

1. Introduction

There is some evidence that the difficulty of obtaining cooperation with sample households in surveys of the United States and other developed countries is growing over time (De Heer and Israëls 1992). Even though overall response rates on some surveys may be relatively constant (Smith 1995), it appears that the component of nonresponse due to refusals will grow unless large increases in expenditures are made (Couper and Groves 1996b). In this environment the search for combinations of survey design features that increase participation is intense.

Progress in evaluating different combinations of design features has been slow, however. Previous studies of survey design features are usually based on field experiments, in which random subsamples are assigned different treatments. The vast majority of these experiments are single factor designs, measuring the effect of varying one design feature (e.g., offers of incentives). The studies have yielded diverse findings, suggesting lack of control over all confounding variables (e.g., Dillman, Gallegos, and Frey 1976; O'Neill, Cannell, and Groves 1978). Thus, the relative value of various features can be measured

¹ Survey Research Center, University of Michigan, and Joint Program in Survey Methodology, University of Maryland, 1218 Lefrak Hall, College Park, MD 20742, U.S.A. E-mail: bgroves@survey.umd.edu

² Survey Research Center, University of Michigan, PO Box 1248, Ann Arbor, MI 48106, U.S.A.

³ Survey Research Unit, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, U.S.A.

Acknowledgments: Revision of a paper prepared for presentation at the 1997 AAPOR conference, Norfolk, Virginia, U.S.A. May, 1997. Research supported by the National Science Foundation and internal funds of the University of Michigan Survey Research Center and the Department of Sociology.

only by meta-analytic studies (e.g., Heberlein and Baumgartner 1978; Yu and Cooper 1983; Church 1993; Singer, Gebler, Raghunathan, Van Hoewyk, and McGonagle 1999). Further, most theoretical treatments suggest that the influences on survey participation are multivariate in nature (e.g., Goyder 1987; Groves and Couper 1998). Given such a phenomenon, it is essential that researchers find techniques to examine the combined effects of distinct independent variables on participation.

This article reports on an experiment to investigate the separate and joint effects of four design features on survey participation: the offer of monetary incentives to respondents (incentives), a disclosure that others were paid for participating although the subject was not (differential incentives), repeated attempts to persuade reluctant persons to participate (refusal conversion), and the burden of the interview request (interview length). The first three factors are alternative means to reduce nonresponse rates in surveys. The last factor has been hypothesized as being important in modifying the effects of the former (Singer et al. 1999). Because we were especially interested in measuring interactions among the independent variables, and because of the prohibitive costs of a field experiment, we designed a laboratory experiment using videotaped vignettes of doorstep interactions between interviewers and respondents in order to investigate the problem.

The use of incentives has a long history in mail surveys, and their effectiveness has been well documented there (see, for example, qualitative reviews by Armstrong (1975), Houston and Ford (1976), Kanuk and Berenson (1975), and Linsky (1975), and quantitative meta-analyses by Church (1993), Fox, Crask, and Kim (1988), Heberlein and Baumgartner (1978), Yammarino, Skinner, and Childers (1991), and Yu and Cooper (1983)). Church (1993) notes that prepaid incentives are more effective than promises of rewards conditional upon survey completion, and that monetary incentives are more effective than nonmonetary gifts. He points out, furthermore, that response rates increase with the cash amount of an incentive, although some studies suggest they do so at a decreasing rate. A meta-analysis of some 37 experimental studies indicates that monetary incentives can also be used effectively in telephone and face-to-face surveys and in surveys involving exceptional effort on the part of respondents (e.g., keeping a diary, or long-term panel participation (Singer et al. 1999)), with effects very similar to those in mail surveys.

One question that arises with respect to incentives is whether they should be offered to all respondents, or only to those who initially refuse. In the context of mail surveys Kulka (1994) reviewed the relative effectiveness of sending an incentive with the first mailing versus offering an incentive only to nonrespondents to the first mailing. Most of the experimental studies he cites find that a monetary incentive is equally effective (i.e., results in similar final response rates) regardless of when it is offered. However, the strategy of offering incentives only to nonrespondents, while apparently cost-effective, does raise concerns about respondent reactions. If such incentives are perceived as unfair by respondents, they may have the unintended consequence of reducing willingness to cooperate.

At the same time, offering refusal conversion payments may be the only alternative to repeated attempts by interviewers to persuade reluctant respondents to reconsider the survey request. At this writing, it is common for U.S. survey organizations to do this after the first refusal, after the second refusal, and sometimes, even after the third refusal. It is our experience that between 25% and 40% of the initial refusals are converted to interviews through such callback protocols. From the householder's perspective, such

efforts might be interpreted as harassment. Indeed, under these designs the only guaranteed outcome of saying “no” is another request for the same interview. An important research question therefore is how householders compare such repeated requests to protocols involving incentives, especially when these are paid only to reluctant respondents.

To gain initial insights into respondent thinking on this issue, we held two exploratory focus groups in the fall of 1994. One, in Detroit, Michigan, included black low-income participants, all but two of whom were women. The others in Livonia, Michigan, consisted of middle-income, mostly white, men and women. Participants in both groups were paid \$15 for an hour and a half of their time.

We asked participants for their opinions about the use of incentives in surveys, and also about the use of differential incentives to convert refusals. In the black, low-income group, attitudes toward incentives were mixed, but in general participants agreed that a) if the study is important for the social good, no incentives should be offered; but b) if someone stands to benefit financially from the survey (e.g., a marketing study), incentives are proper. When asked how they felt about differential incentives, participants varied in their responses, but none raised the issue of fairness – their negative reactions centered on the possibility that offering an incentive might bias responses.

Opinions concerning the desirability of offering incentives also varied within the middle-income group. In general, people felt that special effort warranted payment; also, that commercial surveys should pay respondents whereas socially useful surveys should not. Several people were concerned about the possible contaminating effects of payment. None had heard about the practice of paying only refusers, but when specifically queried about their reactions, only two expressed any sense that the practice was unfair. Some said that if those who were reluctant were especially important to the study, it would be appropriate to pay them. Everyone, on the other hand, objected to survey organizations continuing to call after a person had declined to participate, though they modified their objections when the moderator suggested different rationales that the interviewer might use for the callback (e.g., “You refused earlier; let me tell you a little bit more about the survey”).

These group discussions offered little evidence for perceptions of inequity in survey organizations’ payment practices, but they did provide support for some of our speculations concerning the bases for perceptions of inequity. Of course, these were verbal responses, offered under conditions of group influence; they do not tell us how the focus group participants would react individually to a survey that used a differential incentive strategy, nor how they would react to repeated persuasion attempts by interviewers in the face of persistent refusals. To address these questions, we mounted a multi-factor experiment, permitting examination of the separate and joint effects of several influences on survey participation. The experiment was designed to answer the following questions:

1. Can the positive effect of incentives neutralize the negative effect of longer interviews? Are the effects of incentives themselves a function of the length of the interview requested?⁴

⁴ There is a small literature, nicely summarized by Bogen (1996), on the effects on participation rates of the length of questionnaire, generally measured in number of pages, and achieved response rates. In mailed self-administered surveys, the findings from experimental studies are mixed, with a tendency to show higher response rates with shorter questionnaires. However, the effect seems to be sensitive to the format of the questionnaire (see Champion and Sear 1969), and to efforts at following up initial nonrespondents. Similarly mixed results have been obtained in interviewer-mediated studies (Sharp and Frankel 1983; Collins, Sykes, Wilson, and Blackshaw 1988; Sobal 1982).

2. If subjects are informed about incentives given to others when they themselves are not offered such payment, does that reduce their own likelihood of participation? If so, can providing a rationale for the payment mitigate these negative effects?
3. How do subjects react to the two principal methods of converting reluctant persons – differential incentives and refusal conversion efforts? Do they prefer one to the other?

To answer these questions, we presented subjects with videotaped vignettes portraying interviewers requesting a survey interview. The technique of vignettes has been used in studies as diverse as attributions of rape victimization attitudes and judgments of social status (Alexander and Becker 1978; Rossi 1979; Rossi and Nock 1982; Rossi, Sampson, Bose, Jasso, Passel 1974; Alves and Rossi 1978; Jasso and Rossi 1977). The technique permits the researcher to obtain from each respondent their reactions to a controlled, but randomized, set of stimuli (i.e., the hypothesized influences on the dependent variable). From the statistical design viewpoint each subject is exposed to several different combinations of influences so that rich, within-subject analyses can be performed.

There are three potential weaknesses of the vignette method – uncertainty about the power of the experimental stimuli, confounding effects of the order of presentation of the vignettes on the dependent variable, and low external validity (i.e., the vignette response is of low predictive value for real-world behavior). Some evaluation of each of these potential weaknesses is needed. Manipulation checks are the standard measure of whether the experimental stimuli performed as expected. If we find that subjects do not perceive the experimental manipulations as intended, then we will suspect the subjects do not use them as anticipated to make decisions about survey participation. For the potential order effects in the vignette presentation, using a randomized order of presentation allows us to examine order effects. Thus, if we find that the response of subjects to one vignette is dependent on the contents of the prior vignette, we will be alerted to contamination in the dependent variable due to order. Finally, choosing experimental factors that have long track records in the survey methodological literature helps assess the external validity of the method. If the vignette experiment does not replicate the previously found main effects of the experimental factors, then it is less likely that the vignette findings of joint effects of the factors are informative about real behaviors.

2. Research Design

2.1. Experimental factors

The vignettes were constructed so that three factors were systematically varied in a factorial design: (1) disclosure by the interviewer of payment (25 USD) to other sample households, but not to the respondent; (2) refusal conversion efforts by the interviewer following an initial refusal; and (3) burden, reflected by the length of the interview requested. These factors were completely crossed in the design. In addition, we introduced two factors which were only partially crossed: offering a rationale for the differential payment versus

offering no rationale;⁵ and offering a 40 USD incentive prior to the interview versus offering no incentive. We hypothesized that offering subjects a convincing explanation for differential payments to respondents would reduce the negative effect of disclosing such payments, and that offering an initial incentive would increase the motivation to participate.

The result was 12 distinct vignettes, having the characteristics shown in Table 1. These vignettes were videotaped, using one interviewer throughout, a second interviewer when a refusal conversion was involved, a female respondent's voice, and a male respondent's voice. Scripts from three vignettes are included in the Appendix.

2.2. Administration design

In order to measure both individual and joint effects of factors, each subject was shown a randomly selected set of six vignettes. Each set of six vignettes was balanced in predetermined ways. The randomization ensured that each vignette would be viewed the same number of times, minimized combinations of vignettes likely to produce contrast effects, and counterbalanced such combinations when they were unavoidable. In developing the design, we adhered to the following principles: (a) each vignette should be administered to the same number of subjects; (b) no vignette should be administered more than once to the same subject; (c) adjacent vignettes should ideally differ on only one experimental factor, when possible; (d) the order of initial incentive/no initial incentive and rationale/no rationale vignettes should be balanced across the sequences; that is, the number of sequences where an initial incentive vignette precedes a no initial incentive vignette should be equal to the number of sequences where an incentive vignette follows a no incentive vignette, and the same for rationale/no rationale vignettes.

The final experiment administered 6 vignettes to each of 96 subjects, for 576 observations. Each of the 12 vignettes was viewed a total of 48 times. Of the 96 sequences, 60 contained incentive/no incentive pairs, and 60 contained rationale/no rationale pairs. Both the incentive/no incentive and rationale/no rationale pairs were balanced in terms of the order in which the two vignettes in the pair were presented.

2.3. Subject recruitment

Subjects were recruited through the Parent-Teacher Organizations (PTOs) at two public elementary schools in Ann Arbor, Michigan, to which we offered to donate \$200 in return for their assistance. At one school, subjects were recruited by telephoning parents using a directory provided by the PTO. At the other school, a packet of materials was distributed

⁵ The rationale read as follows: "Unfortunately, we aren't able to pay everyone, because that would make it too expensive for us to do our research at all. Sometimes we do offer money to people to encourage them to participate – we really do need to get the opinions of everyone who's selected. Otherwise our conclusions could be wrong or misleading. Our results are used to plan important social programs that in the long run affect everyone. So, getting everyone's answer is extremely important even if we do have to pay some people. Most people enjoy the interview and find the questions interesting. Why don't we get started and see how it goes?"

By contrast, the norationale condition read as follows: "Unfortunately, we aren't able to pay everyone, because that would make it too expensive for us to do our research at all. But we really need to get everyone's opinions. Why don't we get started and see how it goes?"

Limited research funds forced the unbalanced design, not assurance that the investigators knew the effects of the missing cells on participation judgments.

Table 1. Observation counts by experimental factors^a

	No incentive				Incentive			
	Low burden		High burden		Low burden		High burden	
	No refusal conversion attempt	Refusal conversion attempt	No refusal conversion attempt	Refusal conversion attempt	No refusal conversion attempt	Refusal conversion attempt	No refusal conversion attempt	Refusal conversion attempt
No Disclosure	48	48	48	48	48	X	48	X
Disclosure, No Rationale	48	48	48	48	X	X	X	X
Disclosure, Rationale	48	X	48	X	X	X	X	X

^aCells with “X” were omitted from the design.

to parents; the packet included a letter describing the project, a letter of endorsement from a PTO officer, and a form for interested volunteers to return. Forty-nine subjects were recruited from one school and 47 from the other.

2.4. Administration

A week or so before viewing the videos the subjects were interviewed by telephone about their political attitudes and community involvement, as well as about their attitudes toward surveys and their sensitivity to equity issues.⁶ Approximately one week later, each subject was brought into the laboratory (a room set aside in each school) to view six vignettes, selected as above, and completed a brief self-administered questionnaire after seeing each one. Subjects were matched by sex with the respondent's voice on the videotapes (e.g., female subjects were shown videos where the respondent's voice was female).

We examined different modes of administering the vignettes. Instead of watching videotapes, the first 23 subjects at one of the schools were assigned to respond to a written version of the vignettes. A comparison with 24 subjects who had viewed the videotaped vignettes at the other school showed no differences on the key dependent variables at that time, and we decided to use only the videotaped vignettes with the remaining subjects. Analyses that follow include a statistical control on mode of vignette administration.

The questionnaires administered after each vignette included the main dependent variable: "How likely would you be to agree to participate in the interview?" as well as some manipulation checks. Following the last video, subjects also completed a final "debriefing" questionnaire, which repeated a few of the questions from the intake interview, primarily those about their perception of the fairness of differential payments to survey respondents, and also asked whether they would be willing to take part in a future interview.

Since the external validity of these experiments is a function of the ability to describe reactions of cross-sectional household samples in a field setting, we must first be sensitive to any mismatches between characteristics of the subjects and the general household population. The subject group is quite homogeneous on several sociodemographic characteristics. Almost all are married (94%), white (96%), and female (86%). All have children. Their median income is about 90,000 USD. Their median age is about 41 years. The literature has shown higher response rates among females (Smith 1979) and among

⁶ In all contacts with subjects, the study was described as "a research project about interviewing." In calls to potential subjects, the research assistant said:

"Hello, my name is ___ and I'm calling from the Institute for Social Research at the University of Michigan. We are conducting a research project about interviewing, and we're asking parents from the ___ School to help us out. The study involves answering a few questions on the telephone and then making an appointment to come to the school and watch six short videos and then answer some questions about them."

The interviewer also mentioned the donation to the PTO, the length of time required from the subject (about half an hour altogether), administered the phone intake interview if the subject agreed to participate, and made an appointment for the subject to come to the school to view the videos. At the school, subjects were greeted by the research assistant and told:

"I'm going to ask you to watch six short videos that show interviewers approaching people for a survey interview. After each video, you'll get a questionnaire to fill out. At the end, you'll get a few more questions. As you watch the videos, pretend that the interviewer is speaking to *you* and asking you to participate. Try to imagine how you would react to the things the interviewer says, and try to answer the questions after each video based on what your own reactions would be. The researchers are interested in *whatever* reactions you have, even if they're negative or critical. Your responses are completely confidential. After you've completed all the questionnaires, put them in this envelope and seal it. Please be sure you read the instructions in the questionnaires carefully, and answer all the questions except those you are asked to skip."

households with children (Couper and Groves 1996a). Income and race show mixed effects in prior studies. We expect that the subject group is best viewed as an unusually cooperative group, one that would tend to be among the respondents (versus nonrespondents) in most surveys. We speculate that this may decrease their sensitivity to design features manipulated in the experiment, thus attenuating the chances of finding treatment effects.

3. Results

Before the key hypotheses were investigated, we examined evidence for the three potential weaknesses of the vignettes mentioned above; that is, a) we examined built-in manipulation checks to see if the experimental stimuli were perceived as intended by the subjects, b) we examined whether the results replicated the prior research on main effects of influences on survey participation, and c) we examined whether there were confounding effects of order and mode of presentation for the experimental factors.

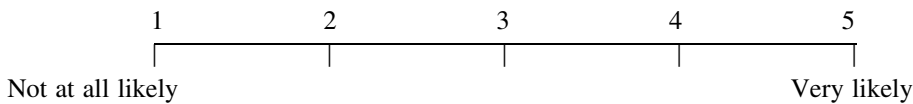
3.1. Experimental manipulation checks

The first analysis is a check that the experimental manipulations were perceived by subjects as intended. After the subjects had completed the participation question for each vignette, they were asked how long the interview in the vignette was to last, whether the respondent was offered money, whether anyone else was offered money, and if so, how much they were offered. For all these questions, 94% or more of the subjects provided the correct answer, and incorrect answers were not associated with any one vignette.

3.2. Testing whether the vignettes replicate past results in the survey methodological literature

While survey participation is a dichotomous phenomenon (response versus nonresponse), the ordinal scale used for the key dependent variable permitted an expression of levels of uncertainty:

“If you were the respondent in the vignette you just saw, how likely would you be to agree to participate in the interview? Circle a number from 1 to 5, where 1 means you would be *not at all likely* to participate, and 5 means you would be *very likely* to participate.”



Responses

Percentage	21%	22%	23%	22%	12%
<i>n</i>	123	127	133	125	68

The distribution above reflects a pooling of all 576 observations. Our analyses address the variation over vignettes in the likelihood of participation and were conducted both

on the 1–5 scale and using a binary variable (contrasting (1,2,3) versus (4,5)).⁷ Since the same conclusions were reached using both analytic strategies, we limit our presentation to the binary variable. Table 2 presents percentages answering “4” or “5” (labeled “willing” respondents) on this scale for each of the experimental cells. We view the data records as 576 vignette measurements clustered by 96 different subjects. Standard errors and hypothesis tests employ a random groups estimator through SUDAAN.

Analyses based on Table 2 can be used to determine whether the experiment replicated the findings in the previous literature. We found that the request for a 1.5-hour interview generated lower willingness to participate (28%) than the .5-hour requests (39%) ($t = 2.22, p = .03$). We found less willingness to respond to survey requests not involving initial incentives (41%) than to those with incentives (59%) ($t = 3.15, p = .002$). Vignettes portraying refusal conversion attempts generated lower cooperation (22%) than did those that did not portray conversion attempts (32%), albeit at weaker levels of statistical significance ($t = 1.69, p = .09$). These results are consistent with the cumulative findings of the field experimental literature on survey participation.

3.3. Tests of order and mode effects in the presentation of vignettes

The last test of whether the vignette design was subject to fatal flaws was the examination of order effects. That is, if we found that the likelihood of a subject agreeing to participate in the survey portrayed in a vignette was itself a function of what vignette preceded it, then the order of presentation would be a confounding variable.⁸

We postulated that the largest order effects would likely involve adjacent pairs of vignettes. For example, assume that for subjects in group A the first vignette portrayed a request involving an incentive, but for subjects in group B it did not. If the first group reported consistently different likelihoods of participation than the second group, we would have evidence that the order of presentation affected the dependent variable. We performed a large number of such order analyses, involving each of the experimental factors. In none of the cases was there a significant main effect of order of presentation.

Finally, as noted earlier, a preliminary comparison of responses to the written and videotaped vignettes suggested that there were no differences between them, and we therefore included responses to the written vignettes in the final analysis. Controls for mode indicated that respondents to the written vignettes were significantly less likely to say they would be willing to participate in the interview, but there were no significant interactions between mode and any of the experimental variables.

In sum, we constructed three analytic hurdles for the vignette design to jump, prior to examining the key hypotheses of the study. The data indicate a) the experimental manipulations were perceived as intended, b) the experiment replicates the main effects of

⁷ That collapsing was used because alternatives were either statistically or substantively less attractive. Contrasting the “5” responses versus others would have limited statistical power to detect differences, given only 68 responses in the cell (spread over the 96 subjects). Combining 3, 4, and 5 would have treated the logical neutral category as part of the positive end of the scale.

⁸ Even if order effects were found, the design might yield useful results about survey participation. Since the order of presentation was randomized (within some constraints), the order of presentation could be used as a covariate in the analysis of likelihood to participate. This would introduce statistical control of the confounding variable of order.

Table 2. Percentages of “willing” respondents by experimental factors^a

	No incentive				Incentive			
	Low burden		High burden		Low burden		High burden	
	No refusal conversion attempt	Refusal conversion attempt	No refusal conversion attempt	Refusal conversion attempt	No refusal conversion attempt	Refusal conversion attempt	No refusal conversion attempt	Refusal conversion attempt
No Disclosure	47.9% (7.3)	33.3% (6.8)	33.3% (6.8)	16.7% (5.4)	64.6% (6.9)	X	54.2% (7.2)	X
Disclosure, No Rationale	27.1% (6.4)	25.0% (6.3)	20.8% (5.9)	14.6% (5.1)	X	X	X	X
Disclosure, Rationale	37.5% (7.0)	X	27.1% (6.4)	X	X	X	X	X

^aStandard errors reflecting clustering of vignette observations into subjects are presented in parentheses.

experimental factors as found in years of methodological research, and c) there were no measured order effects of the experimental vignettes. Thus, with these initial tests passed, we examined the combined effects of experimental factors relevant to our key hypotheses.

3.4. The combined effects of incentives, burden, and refusal conversion attempts on survey participation

We measured the joint influences of all the experimental factors within a logistic regression framework, using the collapsing of the cooperation measure described in Section 3.2. Two covariate controls were introduced based on separate analyses – mode of administration (written versus video) and the subject’s rating of whether survey respondents should be compensated. Table 3 presents the logit model coefficients in a model containing all the experimental factors and the covariates. Since the design was not balanced across all experimental factors, all analyses were conducted separately on the balanced and the full sample. The results in all cases yield the same conclusions; hence, the full sample results are presented below.

1. Can the positive effect of incentives neutralize the negative effect of longer interviews? Are the effects of incentives themselves a function of the length of the interview being requested?

One practical question survey designers face is whether offering an incentive for a longer interview can attain the same or better response rate as a shorter interview without an incentive. The vignettes allow us to test whether a 1.5-hour interview request with an incentive generates less willingness to participate than a .5-hour interview request without an incentive. Calculations based on the logit model in Table 3 suggest that this is not the case ($p = .46$).

We were also interested theoretically in the hypothesis that the effect of an incentive would itself be a function of the burden of the interview. Incentives were expected to have larger effects on participation in high-burden surveys. This was tested by fitting an interaction term to the logistic model, but there was no support for this hypothesis in

Table 3. Logistic Regression Coefficients for Model Predicting High Likelihood of Participation^a

	Coefficient	Standard Error
Intercept	.11	.28
Predictors		
Incentive	.94**	.27
Burden	-.58*	.26
Disclosure, Rationale Conditions:		
No Disclosure ^b	–	–
Disclosure, No Rationale	-.59**	.18
Disclosure, Rationale	-.25	.23
Refusal Conversion Attempt	-.49	.32
Mode	-.75*	.37
Compensation	-.85*	.40

^aStandard errors reflect the clustering of observations within subjects. The dependent variable is coded 0 (for responses 1, 2, 3) and 1 (for responses 4, 5). “Compensation” is coded “1” for “strongly agree” or “somewhat agree” and “0” otherwise. Asterisks denote p -values: * $p < .05$, ** $p < .01$.

^bOmitted category in dummy variable regression.

the data. A meta-analysis by Singer et al. (1999) similarly fails to find an interaction between burden and incentive in their effect on response rate differences.

2. *If subjects are alerted to the fact that others have been paid for participation, does that reduce their own likelihood of participation? Can providing a rationale for the payment mitigate these effects?*

Table 3 shows that, controlling for all other variables, the disclosure of payment to others without a rationale for the payment suppresses the reported willingness to participate ($p < .01$). Providing a rationale for the differential incentive substantially dampens its effects. Is there a difference in expected likelihoods between a disclosure of differential incentives with and without giving the respondent a rationale for the incentive? We cannot detect a large enough difference between disclosure with and without rationale to achieve statistical significance at traditional p -levels (instead, $p < .094$). The raw data corresponding to these conditions are shown in Table 2. If we average across rows to obtain total percentages by each category of disclosure, we see that with no disclosure of payment to others, 42% express willingness to participate; with such a disclosure, but no rationale for the differential payment, 22% are willing to participate. With a rationale for the differential payment, 32% express a willingness to participate. Finally, contrary to our expectations, a variable measuring subjects' perceptions of the fairness of this practice had no significant main or interaction effect on their expressed willingness to participate in the interview (data not shown).

3. *How do people react to the two principal methods of converting reluctant persons – differential incentives and refusal conversion efforts? Do they prefer one to the other?*

The vignettes exposed subjects to the undesirable features of both methods: payment to others who were reluctant and not to themselves (differential incentives), and a repeat visit by an interviewer after an initial refusal (refusal conversions). The question above is addressed by a comparison of two coefficients in the logistic regression model, one for a refusal conversion effort and one for differential incentives (with no rationale provided). The difference between these two coefficients (-0.59 for the differential incentives and -0.49 for the refusal conversion attempts) is not statistically significant. Thus, our experiment cannot detect a preference for either of the methods commonly used to convert refusals.

4. Summary and Conclusions

Because this study was a first test of the vignette approach, we constructed three tests of the method prior to using it to examine combined effects of influences on survey participation. The data passed the tests of sensitivity to experimental manipulation, replication of main effects of experimental factors based on prior research, and the lack of order effects. This permitted us to examine the combined effects of the experimental factors.

All of the experimental factors had individual effects on participation likelihoods as the previous literature implied. Incentives encourage participation; long interviews depress it; observing an initial refusal and an interviewer's conversion efforts leads to less willingness to participate.

The unique advantage of the within-subject design, however, was that it gave us the ability to assess the joint effects of survey design features. For example, our findings suggest that incentives for a long survey can make it achieve the same response rates as a shorter survey without incentives. This finding deserves replication in a field experiment. A useful next research step would be to alter length and incentive amounts to examine how robust the finding is to such changes. In addition, variation in the sponsorship of the survey could be introduced to examine whether the role of incentives and interview length vary by whether a commercial, academic, or government agency conducts the survey.

All the measured effects are no doubt a function of how sharp a contrast was drawn between treatment groups on any one dimension. For example, the experiment did not support the hypothesis of larger incentive effects for higher-burden surveys; the explanation for this may lie in too small a contrast between the two interview lengths examined. On the other hand, there was support for hypotheses concerning the effects of revealing payment to others, with and without a rationale for the payments. Informing subjects about differential incentives significantly reduced their willingness to participate in the survey, but when a rationale was provided for this practice, willingness to participate was not significantly reduced. One fruitful use of the vignette approach would be to pretest particular design options (e.g., providing different rationales for the use of incentives) before launching a larger and more expensive field study.

Thus, our reaction to this first use of vignettes to study survey participation is one of cautious optimism. There are some aspects of the procedure that we would now change. For example, we chose to focus the camera on the interviewer and to have the householder represented only by a disembodied voice. We suspect that this may have focused the attention of the experimental subject too fully on the interviewer, leading to greater empathy with the interviewer than to identification with the householder, which was our goal. The fact that written presentation of the vignettes resulted in significantly lower levels of expressed willingness to participate supports this interpretation.

We are concerned about limits on the external validity of the experimental method. We attracted a homogeneous group of subjects with attributes commonly found to lead to high survey participation rates. As we have already noted, we suspect that this may result in an overestimation of participation willingness in general. At the same time, presenting the vignettes to subjects in the laboratory can be expected to sharpen the effects of the experimental variables, compared to the effects that might be observed in the field (Boruch and Gomez 1977). To some extent, the problem of external validity might be repaired by attempting different recruiting mechanisms – unemployed persons from state or local unemployment offices, and workers at specific companies. There will never be a full repair of this problem, however, without probability sampling procedures. In this regard, there is nothing to stop the implementation of video vignettes in a CAPI household survey, addressed to a probability sample of householders.

Use of probability sampling would attack limits on external validity from the population side. Remaining, however, is the question of how well responses by subjects to a question that begins ‘‘If you were the respondent in the vignette you just saw . . .’’ would predict their actual behavior. On this issue, we believe that field testing of the findings from such vignette studies is the logical next step after discoveries in the laboratory. That is, we believe that the real value of the vignette approach may be to permit rather inexpensive

investigations of many multivariate hypotheses in the laboratory; then, for those showing promise, related experiments imbedded in real surveys might be mounted in order to replicate the laboratory results.

Appendix: Three Illustrative Vignettes

A. Factors: No incentive, no refusal conversion attempt, low burden, disclosure, no rationale

(Interviewer approaches the house, knocks. Door opens.)

Interviewer: Hello, my name is (points to badge) and I'm from the Survey Research Center at the University of Michigan. We're conducting research on how people use their free time. You might remember that we sent you a letter telling you a bit about our research, explaining that an interviewer would be contacting you –

Respondent: I don't remember getting any letter.

Interviewer: Well, let me tell you a little bit about what was in the letter. We're interested in how people spend their leisure time, and I'd like to ask you some general questions about what hobbies you enjoy, what television programs and videos you like to watch – those kinds of things.

Respondent: Oh, I'm sorry, I don't think I can help you then – we don't have a TV.

Interviewer: Oh, that's OK if you don't have a TV – it's important for us to get information on how everyone spends their leisure time. I'm not asking questions only about TV – I've also got some questions about hobbies, going to movies and other activities.

Respondent: Is this for a TV company?

Interviewer: Oh no, this isn't for any *company* – we don't do market research. I'm from the Survey Research Center at the University of Michigan, and we do studies on issues that are important for *everyone*.

Respondent: I don't know ... how did you get my name?

Interviewer: In fact, I don't have your name, all I have is your address, and this address was selected to be in our study.

Respondent: And how long did you say it would take?

Interviewer: Well, that depends – the time it takes varies from person to person, but on average, it's about half an hour.

Respondent: You know, I think my friend across the street got interviewed for that study. They paid her \$25! Am I going to be paid?

Interviewer: No, unfortunately, we aren't able to pay everyone, because that would make it too expensive for us to do our research at all. But we really need to get *everyone's* opinions. Why don't we get started and see how it goes?

B. Factors: No incentive, high burden, refusal conversion attempt, disclosure, no rationale

(Interviewer approaches the house, knocks. Door opens.)

Interviewer: Hello, my name is [*points to badge*], and I'm from the Survey Research Center at the University of Michigan –

Respondent: Yes?

Interviewer: We're conducting research on how people spend their leisure time, and you have been selected to be in our study –

Respondent: Oh, no, I'm not interested. (*Respondent closes door.*)

(*A few days later... A different interviewer approaches the house, knocks. Respondent opens door.*)

Interviewer: Hello, my name is (*points to badge*) – I'm from the University of Michigan's Survey Research Center. One of my colleagues came by a few days ago, and I'm here to explain a little more about our research and to answer any questions that you might have.

Respondent: Well, what's the study about?

Interviewer: We're interested in how people spend their leisure time, and I'd like to ask you some general questions about what hobbies you enjoy, what television programs and videos you like to watch – those kinds of things.

Respondent: Well, we have a TV, but I never watch anything but the news. I'm sorry I can't help you. (*Respondent closes door.*)

(*One week later... The same interviewer approaches the house, knocks. Respondent opens door.*)

Interviewer: Hello, I'm . . . , and I'm glad I found you at home. I came back because I wanted to tell you how important it is that we speak to people like you – people who don't watch a lot of TV. I'm not asking questions only about TV – I've also got some questions about hobbies, going to movies, and other activities. It's really important that we include you in our research.

Respondent: Is this for a TV company?

Interviewer: Oh no, this isn't for any *company* – we don't do market research. I'm from the Survey Research Center at the University of Michigan, and we do studies on issues that are important for *everyone*.

Respondent: I don't know . . . how did you get my name?

Interviewer: In fact, I don't have your name, all I have is your address, and this address was selected to be in our study.

Respondent: And how long did you say it would take?

Interviewer: Well, that depends – the time it takes varies from person to person, but on average, it’s about one and a half hours.

Respondent: You know, I think my friend across the street got interviewed for that study. They paid him \$25! Am I going to get paid?

Interviewer: No, unfortunately, we aren’t able to pay everyone, because that would make it too expensive for us to do our research at all. But we really need to get *everyone’s* opinions. Why don’t we get started and see how it goes?

C. Factors: Incentive, no refusal conversion attempt, low burden, no disclosure

(Interviewer approaches the house, knocks. Door opens.)

Interviewer: Hello, my name is (points to badge), and I’m from the Survey Research Center at the University of Michigan. We are conducting research in the tri-county area about how people use their free time, and you have been selected to be in our study.

Respondent: Oh, yes, you sent me a letter with a cheque for \$40! I was wondering what this is all about.

Interviewer: Well, let me tell you a little bit about it. We’re interested in how people spend their leisure time, and I’d like to ask you some general questions about what hobbies you enjoy, what television programs and videos you like to watch – those kinds of things.

Respondent: Oh, I’m sorry, I don’t think I can help you then. We have a TV, but I never watch anything but the news.

Interviewer: That’s OK if you only watch the news – it’s important for us to get information on the kind of programs everyone watches, even if they’re not entertainment programs. And I’m not asking questions only about TV – I’ve also got some questions about hobbies, going to movies, and other activities.

Respondent: Is this for a TV company?

Interviewer: Oh no, this isn’t for any *company* – we don’t do market research. I’m from the Survey Research Center at the University of Michigan, and we do studies on issues that are important for *everyone*.

Respondent: I don’t know . . . how did you get my name?

Interviewer: In fact, I don’t have your name, all I have is your address, and this address was selected to be in our study.

Respondent: And how long did you say it would take?

Interviewer: Well, that depends – the time it takes varies from person to person, but on average, it’s about half an hour. Why don’t we get started and see how it goes?

5. References

Alexander, C.S. and Becker, H.J. (1978). The Use of Vignettes in Survey Research. *Public Opinion Quarterly*, 42, 93–104.

- Alves, W. and Rossi, P. (1978). Who Should Get What? Fairness Judgments of the Distribution of Earnings. *American Journal of Sociology*, 84, 541–564.
- Armstrong, J.S. (1975). Monetary Incentives in Mail Surveys. *Public Opinion Quarterly*, 39, 223–250.
- Bogen, K. (1996). The Effect of Questionnaire Length on Response Rates – A Review of the Literature. *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Boruch, R.F. and Gomez, H. (1977). Sensitivity, Bias, and Theory in Impact Evaluation. *Professional Psychology*, November, 411–434.
- Champion, D.J. and Sear, A.M. (1969). Questionnaire Response Rate: A Methodological Analysis. *Social Forces*, 47, 335–339.
- Church, A.H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*, 57, 62–79.
- Collins, M., Sykes, W., Wilson, P., and Blackshaw, N. (1988). Nonresponse: The UK Experience. In *Telephone Survey Methodology*, eds. Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls, W.L., and Waksberg, J., Chapter 13, New York: Wiley.
- Couper, M.P. and Groves, R.M. (1996a). Household Level Determinants of Survey Nonresponse. In *Advances in Survey Research*, eds. Braverman, M.T. and Slater, J.K., 70, Summer, 63–80.
- Couper, M.P. and Groves, R.M. (1996b). Social Environmental Impacts on Survey Cooperation. *Quality and Quantity*, 30, 173–188.
- De Heer, W.F. and Israëls, A.Z. (1992). Nonresponse Trends in Europe. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Boston.
- Dillman, D.A., Gallegos, J.G., and Frey, J.H. (1976). Reducing Refusal Rates for Telephone Interviews. *Public Opinion Quarterly*, 40, 66–78.
- Fox, R.J., Crask, M.R., and Kim, J. (1988). Mail Survey Response Rate: A Meta Analysis of Selected Techniques for Inducing Response. *Public Opinion Quarterly*, 52, 467–491.
- Goyder, J. (1987). *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, CO: Westview Press.
- Groves, R.M. (1992). Respondent Incentives, Minimum Mean Square Error per Unit Cost, and the Resident's Responsibility to the Polity. Discussion, COPAFS Symposium on Providing Incentives to Survey Respondents, John F. Kennedy School of Government, Harvard University, Cambridge, MA.
- Groves, R.M. and Couper, M.P. (1996). Contact-level Influences on Cooperation in Face-to-face Surveys. *Journal of Official Statistics*, 12, 63–83.
- Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Heberlein, T.A. and Baumgartner, R. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. *American Sociological Review*, 3, 447–462.
- Houston, M.J. and Ford, N.M. (1976). Broadening the Scope of Methodological Research on Mail Surveys. *Journal of Marketing Research*, 13, 397–403.
- Jasso, G. and Rossi, P. (1977). Distributive Justice and Earned Income. *American Sociological Review*, 42, 639–651.

- Kanuk, L. and Berenson, C. (1975). Mail Surveys and Response Rates: A Literature Review. *Journal of Marketing Research*, 12, 440–453.
- Kulka, R.A. (1994). The Use of Incentives to Survey “Hard-to-Reach” Respondents: A Brief Review of Empirical Research and Current Practice. Paper prepared for a seminar on new directions in statistical methodology, Bethesda, MD, May 25–26.
- Linsky, A.S. (1975). Stimulating Responses to Mailed Questionnaires: A Review. *Public Opinion Quarterly*, 39, 82–101.
- O’Neil, M., Groves, R.M., and Cannell, C.C. (1979). Telephone Interview Introduction and Refusal Rates: Experiment in Increasing Respondent Cooperation. Proceedings of the Section on Survey Research Methods, American Statistical Association.
- Rossi, P. (1979). Vignette Analysis: Uncovering the Normative Structure of Complex Judgments. In *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld*, eds. Merton, R.K., Coleman, J.S., and Rossi, P., 176–186, New York: Free Press.
- Rossi, P., Sampson, W.A., Bose, C.E., Jasso, G., and Passel, J. (1974). Measuring Household Social Standing. *Social Science Research*, 3, 169–190.
- Rossi, P. and Nock, S. (1982). *Measuring Social Judgments: The Factorial Survey Approach*. Beverly Hills: Sage.
- Sharp, L.M. and Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. *Public Opinion Quarterly*, 47, 36–53.
- Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The Effect of Incentives in Interviewer-mediated Surveys. *Journal of Official Statistics*, 15, 2, 217–230.
- Smith, T.W. (1979). Sex and the GSS. General Social Survey Technical Report No. 17. Chicago: National Opinion Research Center.
- Smith, T.W. (1995). Trends in Nonresponse Rates. *International Journal of Public Opinion Research*, 7(2), 157–171.
- Sobal, J. (1982). Disclosing Information in Interviewer Introductions: Methodological Consequences of Informed Consent. *Sociology and Social Research*, 66, 348–361.
- Yammarino, F., Skinner, S.J., and Childers, T.L. (1991). Understanding Mail Survey Response Behavior: A Meta Analysis. *Public Opinion Quarterly*, 55, 613–639.
- Yu, J. and Cooper, H. (1983). A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *Journal of Marketing Research*, 20, 36–44.

Received January 1998

Revised May 1999