STRAIGHTENING THE SEAM EFFECT IN PANEL SURVEYS

LANCE J. RIPS
FREDERICK G. CONRAD
SCOTT S. FRICKER

Abstract Panel surveys, such as the Survey of Income and Program Participation and the Consumer Expenditure Survey, interview respondents every 3 or 4 months, but ask the respondents for monthly data. A typical finding in such surveys is that changes in responses to a question are relatively small for adjacent months within a reference period but much more abrupt for adjacent months across reference periods. Previous studies have attributed this "seam effect" either to underreporting of changes within the periods or to overreporting of changes across them. In the present studies, we attempt to distinguish these possibilities, using an experimental method that allows us to gauge respondents' accuracy as well as the number of times they change their answers. The studies produced seam effects and accompanying evidence for forgetting of queried information and bias toward constant responses within the reference period. In general, seam effects appear to increase as a function of the demands on memory. We also find that separating questions with the same content in the survey instrument decreases the seam effect. To account for these data, we propose a model in which respondents' answers are initially based on attempted memory retrieval. Inability to recall leads to (possibly biased) guessing or subsequent repetition of an earlier answer.

A main advantage of longitudinal surveys over cross-sectional ones is the accuracy with which longitudinal surveys measure change. By following the same individual or household over time, survey researchers can pinpoint fluc-

LANCE J. RIPS is a professor in the Department of Psychology at Northwestern University (e-mail: rips@northwestern.edu); FREDERICK G. CONRAD is Associate Research Professor at the Institute for Social Research, University of Michigan; and SCOTT S. FRICKER is a graduate student in the Joint Program in Survey Methodology, University of Maryland. We thank Jami Barnett, Jennifer Behr, Pam Douglas, and Marissa Pishel for their help with this project. We also thank Catheryn Dippo, Paul Hsen, Elizabeth Martin, Jeffrey Moore, Adrianna Silberstein, Monroe Sirken, Judith Tanur, Roger Tourangeau, Clyde Tucker, and Gordon Willis for comments on earlier presentations. NSF grant SES-9907414 supported the research we report here.

Public Opinion Quarterly Volume 67:522–554 0033-362X/2003/6704-0002$10.00

tuations in respondents' behaviors or attitudes that might be difficult to detect across different samples in cross-sectional surveys (see, e.g., Menard 1991; Plewis 1985). But these advantages can come with corresponding costs. Longitudinal surveys sometimes produce their own spurious changes that accompany or obscure true differences. This article investigates one of these survey-induced artifacts called the "seam effect," and it reports a series of experimental studies that attempt to trace the causes of this type of response error. The perspective we take on the seam effect focuses on the respondents' cognitive resources (see, e.g., Bradburn, Rips, and Shevell 1987). Questions that produce this artifact make demands on respondents' memory. If memory cannot provide the basis for a reasonable answer, respondents are forced to rely on alternative means to infer or estimate the relevant information. The size of the seam effect, we believe, depends on this interaction between recall difficulty and the accuracy of the inference strategies that take over when memory fails.

The Seam Effect

Some national panel surveys interview their respondents every 3 or 4 months, but ask these respondents to provide data for each month within the reference period. For example, the Survey of Income and Program Participation (SIPP) schedules its interviews every 4 months and, at each interview, asks respondents items like this one:

(1) Did . . . receive food stamps in (*Read each month*)
(Last month)?
(2 months ago)?
(3 months ago)?
(4 months ago)?

Similarly, the Consumer Expenditure Survey (CE) interviews its panels every 3 months, but asks some of its questions about 1-month intervals within the longer reference period:

(2) What was the total for all labor, materials, appliances, or equipment THEY [i.e., builders and contractors] PROVIDED IN—
(month, 3 months ago)?
(month, 2 months ago)?
(month, 1 month ago)?

Questions like these provide data from each respondent (or household) and for each month in which the respondent participates in the survey. For example, if SIPP interviews a respondent in May, the question in (1) will provide data about receipt of food stamps for the months of January ("4 months ago"), February ("3 months ago"), March ("2 months ago"), and April ("last month").

The same respondent (or household) will be reinterviewed in September and at that time provide food stamp data for May, June, July, and August. This pattern will repeat for as long as the respondent participates in the survey (usually about 2.5 years for SIPP). The top half of Figure 1 depicts an interviewing schedule of this type for a SIPP panel. In our example, month 1 in the figure corresponds to January, month 2 to February, and so on.

What is important about this interviewing schedule, for our purposes, is its effects on month-to-month change in responses to an individual item, such as (1). Figure 1 shows that data from months 1 and 2 come from the same interview; for example, the January and February data for our hypothetical panel come from the May interview. This is also true for months 2–4. However, the data from months 4 and 5 (April and May in the example) come from two different interviews (the April data from the May interview and the May data from the September interview). The transition between months 4 and 5 (April to May) is across the “seam” between two reference periods, as is the transition between months 8 and 9. But the transitions between the remaining pairs of months (1 and 2, 2 and 3, 3 and 4, 5 and 6, 6 and 7, 7 and 8, 9 and 10, 10 and 11, and 11 and 12) are all “off-seam”: The data from each of these pairs of months come from the same interview.

The “seam effect” is the finding that month-to-month changes in responses tend to be larger for the seam months than for adjacent months off the seam. The lower half of Figure 1 illustrates this difference in SIPP data (from Burkhead and Coder 1985; see also Jabine, King, and Petroni 1990). The curves in this figure come from reports of receiving food stamps (solid line) and reports of receiving social security benefits (dashed line) over a 12-month period. The y-axis records the number of respondents who changed their answer (from “yes” to “no” or from “no” to “yes”) from one month to the next. (Respondents do not report these changes directly; rather, the changes are calculated from their month-to-month answers about receiving these benefits.) Off the seam, the number of changes is modest and relatively constant across pairs of months. For the seam months, however, there is an abrupt increase in the number of changed responses. Because different groups of respondents enter SIPP each month, it is not likely that the effect is due to seasonal or other trends tied to the calendar. Nor is the effect due to proxy reporting, since the same effect appears when the data are restricted to respondents reporting their own income sources (Burkhead and Coder 1985). The effect is not limited to SIPP but also occurs in several different surveys, including the Panel Study of Income Dynamics (Hill 1987) and the Income Survey Development Program (Moore and Kasprzyk 1984).

As we have noted, the usual goal of a longitudinal survey is to measure change over time. If the seam effect produces spurious changes, it can create serious interpretive problems for the results of such surveys. This difficulty is compounded by the fact that the size of the seam effect can be quite large for some items. In a study comparing respondents’ answers to administrative

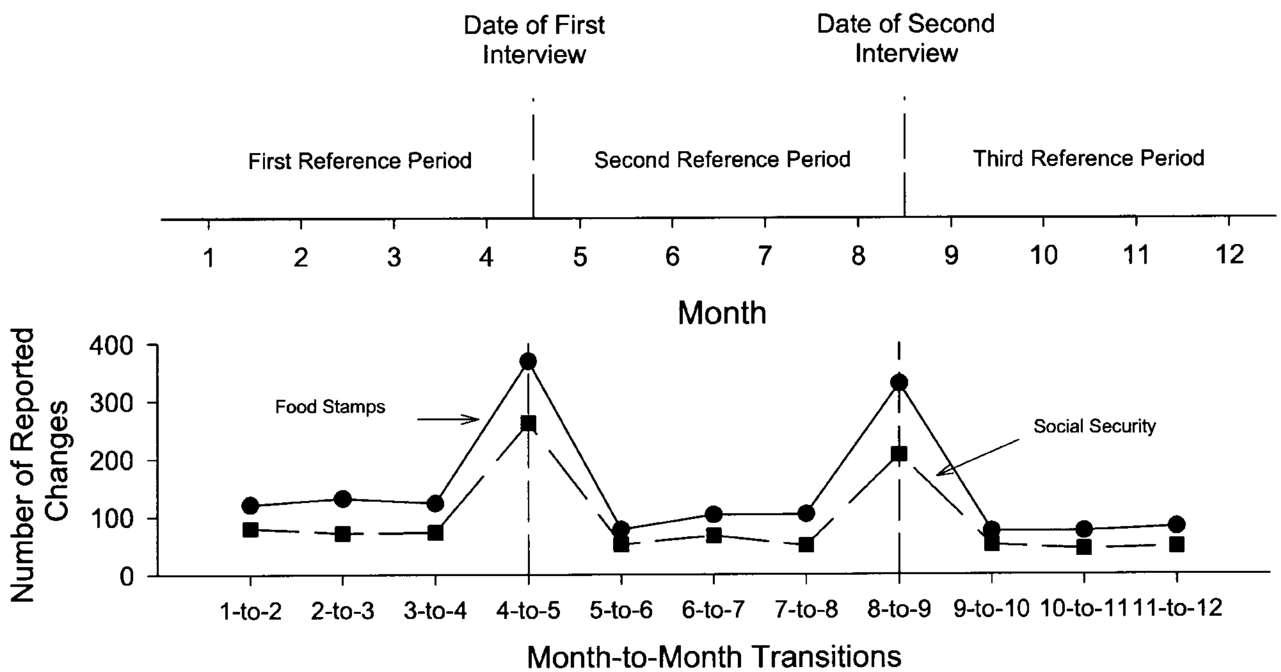


Figure 1. Time course of a hypothetical panel survey, and number of changes in reports of receiving food stamps (solid line) and social security benefits (dashed line) from SIPP data (Burkhead and Coder 1985).

records, Marquis and Moore (1990) found that SIPP overestimates the number of changes in program participation by 20 percent to 135 percent (across different programs) for seam months and underestimates changes by up to 64 percent at nonseam months (see also Moore and Marquis 1989). Seam effects appear in both discrete variables (e.g., food stamp receipt or employment status) and continuous ones (personal earnings, total family income, or amount of social security received; see Kalton and Miller 1991; Young 1989). The effect also seems to be a stubborn one. An experimental version of SIPP that included procedures specifically designed to reduce the bias at the seam failed to eliminate the effect (Moore, Marquis, and Bogen 1996).[1] Unless overestimates of change at the seam months are canceled by underestimates at off-seam months, the seam effect introduces systematic and potentially large amounts of error.

Possible Causes of the Seam Effect

What creates the seam effect? This bias almost certainly depends on the fact that the data for off-seam transitions come from the same interview, whereas the data for seam transitions come from different interviews. However, in order to correct for the seam effect statistically or to revise survey procedures to eliminate it, we need to know more about its origin. As mentioned earlier, Marquis and Moore's (1989, 1990) record-check study suggests that the seam effect reflects both overestimates of changes at the seam and underestimates of changes elsewhere. Although overestimates at the seam tended to be larger in absolute value than underestimates off-seam, both types of biases appeared for reports of participation in most of the examined programs (e.g., Aid to Families with Dependent Children, Food Stamps, and Unemployment Insurance). We therefore need an explanation that predicts both types of error.

THE ROLE OF MEMORY

One class of hypotheses about the seam effect is that it results from respondents forgetting relevant information during the reference period. As the elapsed time increases between a critical event and the interview, the more likely respondents are to forget the event (see, e.g., Cannell [1965] for evidence that respondents are less able to recall hospitalizations as the time since these incidents increases).

There are a number of findings in the literature on autobiographical memory

1. The experimental version of SIPP included probe questions for income sources that respondents reported in one interview but not the next, probes for discrepancies in sources that occurred during an overlapping period included in both interviews, and pressure on respondents to use records of payments. These procedures may have reduced underestimates of changes at off-seam months, but may have increased overestimates of changes at the seam months (Moore, Marquis, and Bogen 1996).

that may be at least indirectly relevant to the seam difference (see Belli [1998] and Shum and Rips [1999] for reviews of the role of autobiographical memory in survey responses). First, when people must recall what happened to them over a several-month period, their narratives sometimes take the form of streams of incidents associated with a particular theme, such as events that occurred at work or events connected with an ongoing romantic relationship (Barsalou 1988). People may describe one of these streams from beginning to end and then double back to describe another. Some investigators have taken such results to suggest that long-term memory encodes events in the form of thematically and chronologically structured histories (Barsalou 1988; Conway 1996). Survey questions, such as question (2), the Consumer Expenditure Survey item, that give rise to seam effects may tap into these histories, for example, evoking recall of a particular home improvement project and facilitating memory for relevant events within the reference period. It is unclear, however, why such histories should produce the characteristic scallops in transitions between reference periods. For example, unless the CE question elicits memories of different histories in different interviews, it is hard to explain why changes in responses should be greater across the seam than within it.[2]

Second, there is evidence that people's ability to recall events depends on the structure of the calendar they live under (Kurbat, Shevell, and Rips 1998; Pillemer, Rhinehart, and White 1986; Robinson 1986). For example, when students recall events from the last year or two, they tend to recall more incidents from the period at the beginning or end of semesters, especially if these transition points are accompanied by a change in activity or locale. It seems possible that a panel survey interview schedule like that in Figure 1 may impose its own calendar on respondents' lives, leading to better memory for events that take place at the start or end of the reference periods. But although a calendar effect of this type might boost recall at the juncture points, its impact on transitions is uncertain. Respondents may find it easier to remember incidents that took place in the final month of the first reference period and the initial month of the second reference period; however, this may make them less likely, rather than more likely, to change their answers across these seam months.

A more promising account of the seam effect might begin instead with the simpler observation that, other things being equal, recall accuracy decreases with elapsed time. Suppose that respondents have fairly accurate memories for payments they received in the most recent month of the reference period, but decreasing memory for payments in earlier months. If respondents who have actually received payments throughout the reference period report par-

2. Tapping these histories, however, may improve respondents' ability to recall certain events and, perhaps, eliminate seam transitions. We discuss attempts of this kind in connection with event history calendars at the end of this article.

ticipating only when they have a clear memory of the payments, this will produce relatively many correct "yes" responses for the most recent months (months 4, 8, and 12 in Fig. 1) but fewer correct "yes" responses for the earliest month in a reference period (months 1, 5, and 9). This difference in the rate of positive answers will produce spurious changes across the seam (between months 4 and 5 or months 8 and 9). Forgetting could also be responsible for underestimates off the seam if memory is poor for most of the early months within each reference period (months 1–3, 5–7, 9–11). To account for the exact shape of the SIPP function in Figure 1, however, a forgetting hypothesis may well require some additional assumptions.

CONSTANT WAVE RESPONDING

A second class of theories is that the seam effect is due to response strategies that respondents adopt to simplify the reporting task. One possibility along these lines is that respondents simply provide the same answer for each month within a reference period. For example, a respondent might report receiving social security benefits in months 1, 2, 3, and 4 (and report the same amount received), whether or not he or she actually obtained benefits in each of these months. This tendency—called "constant wave" responding (e.g., Martini 1989; Young 1989)—means that respondents do not have to calculate separate answers for each month but can give a single summary answer instead. If this summary answer differs from one interview to the next (e.g., reporting receipt of benefits in months 1–4 but no benefits in months 5–8), this tendency will give rise to changes at the seam but flat responding elsewhere.

PREVIOUS STUDIES OF COMPONENTS OF THE SEAM EFFECT

Earlier assessments of these explanations have proved mixed. For example, positive evidence for the role of forgetting comes from Kalton and Miller's (1991) study of SIPP data on social security payments. The size of individual payments increased by 3.5 percent in January 1984, and Kalton and Miller examined reports of the increase as a function of the amount of time between January and the respondents' interview. The results showed a drop in reports with elapsed time: 68 percent of those interviewed in February reported such an increase, but only 53 percent of those interviewed in April. (Some respondents seemed to have confused the month in which the increase occurred, but most never reported the increase at all; see Kalton and Miller 1991, Table 6.) Negative evidence for forgetting, however, comes from the study by Marquis and Moore (1989, 1990), who found that underestimates of participation significantly decreased with elapsed time for only one of the eight programs they studied (unemployment insurance).

Of course, researchers do not design surveys like SIPP in order to test hypotheses about forgetting or response bias. Hence, conclusions about the causes of the seam effect have been indirect. A complex survey can mask potentially important effects if different sources of error compete. Effects of forgetting, for example, may fail to appear if respondents compensate for poor memory by estimating or inferring the answer to a survey question (see Sudman, Bradburn, and Schwarz [1996] and Tourangeau, Rips, and Rasinski [2000] for reviews of the role of estimation in survey responding). For this reason, identifying the underlying causes of the seam effect may require experimental procedures specifically directed toward uncovering these factors. In the remainder of this article, we describe some hypothesis-testing experiments of this sort.

Our aim in these studies is primarily theoretical, in the sense that we are concerned with the underlying causes of the seam effect, and we propose a model of these causes and their interactions in the final section of this article. We do not systematically consider practical methods for eliminating the seam effect, although we hope that the model and some of the findings (especially in Study 3) will suggest new ideas for attacking this problem.

Experiments on the Seam Effect

The experiments that we have conducted aim to create seam effects under conditions in which we can monitor the accuracy of the respondents' answers. These experiments are, in some sense, a cross between traditional laboratory studies of memory and field studies of survey methods. As in laboratory studies, we use experimental manipulation of variables that might contribute to a seam effect and random assignment of respondents to conditions.[3] However, we use a time scale of 6 to 8 weeks, which is much longer than traditional laboratory experiments (though, of course, shorter than the duration of a national panel survey), and our respondents come from the community rather than from the usual pool of college students. The goal is to create an experimental version of a panel survey that might provide a tool for evaluating theories such as those we discussed in the preceding section, but without the expense (and lack of control) of a full field test. We do not intend our method to eliminate the need for field testing, record checks, cognitive interviews, or other current methods, but we hope it will provide a new procedure that may prove useful to survey methodologists, perhaps in combination with these other techniques.

3. Previous memory experiments in psychology, unfortunately, do not bear directly on the seam effect. Many of these experiments present lists of items (e.g., single words or word pairs) to participants on multiple occasions, testing recall after each presentation. However, these studies do not assess changes in response to the same items both within and across lists in a way that is analogous to longitudinal surveys.

GENERAL METHOD

Because our studies all employ a similar technique, we summarize the general procedure first and then describe the variations we included in each experiment. Respondents participated for a period of 6 or 8 weeks (depending on the study), and Figure 2 illustrates the main events that took place in the case of the 8-week experiments. (The 6-week study contained one fewer week in each of the two reference periods in Fig. 2.) In each week we mailed a questionnaire to the respondents that asked them about the events of that week. The questions were "yes/no" items about events that may have happened to them, as in (3) below:

(3) During the last week (that is, from [date] to [date]), did you take a day off work due to illness?

Respondents were to mail the questionnaires back within 24 hours. The answers to these questionnaires are not of central concern in these experiments (though we will use them later to clarify aspects of the findings); however, the questionnaire items themselves provided information to the respondents that we probed during later phases of the experiment.

Midway through the study and again at the end, respondents came to the laboratory and answered a sequence of test questions.[4] The approximate temporal positions of these two test sessions appear in Figure 2. In most of the experiments, the questions in these sessions were about the content of the questionnaires that respondents had filled out during the preceding weeks. For example, in the first of these test sessions, we asked questions like those in (4) below:

(4) *a*. In the fourth week's questionnaire (that is, the one you received on [date]), was there an item about taking a day off work due to illness?
b. In the third week's questionnaire . . . , was there an item about taking a day off work . . . ?
c. In the second week's questionnaire . . . , was there an item about taking a day off work . . . ?
d. In the first week's questionnaire . . . , was there an item about taking a day off work . . . ?

4. We asked respondents to come to the lab for the test sessions (rather than mailing them the test items) so that we could administer the critical questions under conditions that are uniform across respondents. Although this difference in surroundings distinguishes the test sessions from the mailed questionnaires, our intent was not to compare responses to these two sets of items. We note that an analogous difference in setting often occurs in national panel surveys between the context in which the interviewer administers the questionnaire (face-to-face or telephone interviews) and the context in which the queried events took place (e.g., doctor's office or hospital in the case of a health survey, crime scene in the case of a crime survey, and department or grocery store in the case of a consumer survey).

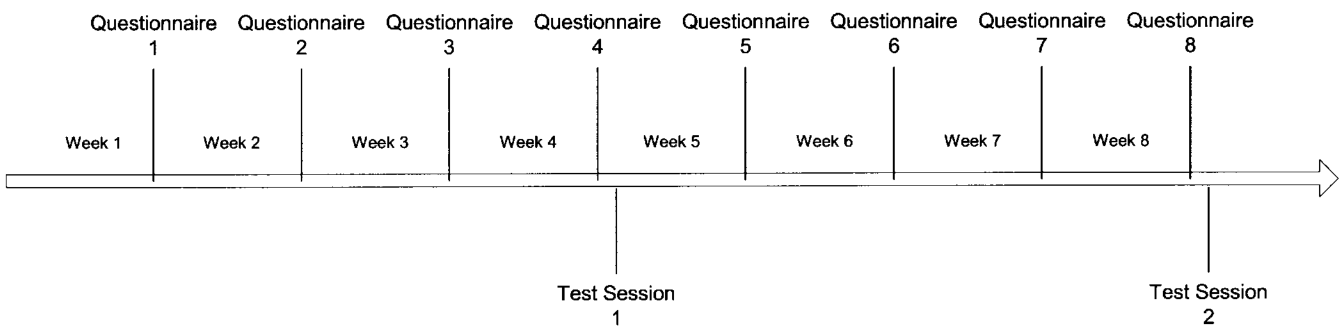


Figure 2. Time course of an 8-week study, such as Study 1

We refer to these items as “test questions” to distinguish them from questionnaire items, such as (3), that we had sent to the respondents earlier. During the first test session, the questions concerned items from the first 4 weeks of the study (3 weeks in the case of the 6-week experiment); during the final test session, the questions concerned items from the last 3 or 4 weeks. Respondents’ answers to test questions such as (4*a*)–(4*d*), provide the data we focus on here.

The two test sessions divide the study into two intervals that are analogous to the reference periods of a panel study (compare Figs. 1 and 2). We can therefore use changes in respondents’ answers to the test questions to detect a seam effect. To see this, consider the sequence of events in Figure 2. During the first test session we asked questions like those in (4) for each of the weeks of the first reference period. During the second test session we asked a similar set of test questions for each of the weeks (5–8) of the second reference period (e.g., “In the fifth week’s questionnaire . . . , was there an item about taking a day off work . . . ?” “In the sixth week’s questionnaire . . . , was there an item about taking a day off work . . . ?” and so on). Differences in a respondent’s answers to these questions give us week-to-week changes in response. For example, if a respondent answers (4*a*) and (4*b*) differently, this produces an off-seam change between weeks 3 and 4. If the respondent gives different answers to (4*a*) and the relevant question from the second test session (“In the fifth week’s questionnaire . . . , was there an item about taking a day off work due to illness?”), this produces a seam change. If the data from the experiments mimic those of the panel surveys, we should expect more changes on seam than off seam.

Test questions such as those in (4) are not typical survey items. Although it would not be unusual for a consumer survey to ask respondents whether they had seen an advertisement or other information, most survey items are not about written sources. Our purpose in using these items was to gain experimental control over the content and timing of the facts that we asked respondents to report. This allows us to pinpoint respondents’ accuracy as a function of elapsed time within the reference period and to study the impact of accuracy (or inaccuracy) on the seam effect. Some trade-off between control and naturalness seems inevitable in an investigation of this kind. However, we can begin to determine whether the findings generalize on the basis of the second study we are about to report, which employs items closer to those in usual surveys.

SPECIFIC STUDIES

The experiments all used questionnaire items about discrete everyday activities, similar to question (3) about work loss due to illness. The differences among these experiments concerned the design of the questionnaires, the type of test question, and the order in which we posed the test items. Table 1 summarizes these differences.

Study 1: Effects of true change versus no change. The purpose of the first of these studies was to find out whether we could induce a seam effect in the absence of any objective change at the seam. Such a result would show that the effect here is entirely due to response error rather than to any actual temporal changes. The study also collects evidence about forgetting and constant wave responding that we can use to evaluate hypotheses about the cause (or causes) of the seam effect.

To this end, we sent one questionnaire to respondents in weeks 1, 2, 7, and 8, and a second questionnaire in weeks 3, 4, 5, and 6. These two types of questionnaire differed only in the specific items they contained. A particular respondent, for example, would have answered one set of questions about their everyday activities (e.g., "During the last week, did you attend a training session for work? . . . did you mail a package at the post office?" and so on) in weeks 1–2 and 7–8, and a separate set of questions (e.g., "During the last week, did you go to visit a doctor or dentist? . . . did you have the oil changed in your car?" and so on) in weeks 3–6. What is important about this setup is that each respondent saw exactly the same set of questions during weeks 4 and 5, the "seam weeks" of the study (see Fig. 2). The only change in the items occurred between the nonseam weeks 2 and 3 and 6 and 7.[5] In each test session we asked respondents to decide whether or not they had seen each of the items from the questionnaires for each of the preceding 4 weeks. During the first test session, for example, we presented the full set of items one at a time, and respondents indicated whether they had seen each item in the questionnaire for week 4; we then presented the items again (in a new random order) and asked respondents whether they had seen each item in the questionnaire for week 3, and so on for weeks 2 and 1. (We refer to this order of presentation as "backward by week" in Table 1.) During the second test session, we repeated the same procedure for the questionnaires of weeks 8, 7, 6, and 5 (in that order). Respondents answered "yes" or "no"; there was no "don't know" option. (The national panel surveys with which we are concerned usually do not offer explicit "don't know" options for questions about monthly activities; see questions (1) and (2) for examples. In

5. To ensure compatibility between the two types of questionnaires, we balanced their content over respondents. For example, approximately half the respondents received the items about attending a training session and mailing a package as part of the questionnaire for weeks 1–2 and 7–8, and they received the items about visiting a doctor and changing the oil as part of the questionnaire for weeks 3–6. The remaining respondents received the reverse assignment. We also equated the two sets of items on the basis of a preliminary normative study (with a different set of respondents) for mean rated importance of the events, degree of affect associated with the events, frequency of participating in the event, and duration of the event. Each questionnaire contained 50 items, and we randomized the order of the items anew at each presentation.

Table 1. Summary of Individual Studies

Study	Length (Weeks)	Questionnaire Design	Test Question Topic	Question Order in Test Session	Response Type	Respondents (n)
1	8	Questionnaire A in weeks 1–2, 7–8; Questionnaire B in weeks 3–6	Questionnaire item	Backward by week	Yes/no recognition	58
2	8	Same questionnaire in weeks 1–8	Personal events	Backward by week	Yes/no recognition	56
3	6	Items related by topic across questionnaires	Questionnaire item	Forward by item Forward by week Backward by item Backward by week	Item recall	65

particular, none of the SIPP questions on income have "don't know" response alternatives.)[6]

We recruited the respondents in all these studies through advertisements in local newspapers and on-line job-bank listings. The number of respondents in each study appears in the last column of Table 1. Respondents in Study 1 were from the Washington, DC, metropolitan area. The ads sought "reliable individuals of varied background to participate in research studies for the Bureau of Labor Statistics" and indicated that individuals would receive $50 for their participation. Of the 58 respondents in Study 1, 40 were female and 18 male. Forty-three were white, 11 African-American, and 4 Asian. Their average age was 38, and their average number of years of education was 15. There were 64 respondents who began the study, but six failed to complete all questionnaire booklets or failed to attend both test sessions, for an attrition rate of 9 percent.

Study 2: Autobiographical questions. The second study attempted to generalize our findings to a situation that is closer in some respects to those of actual panel surveys. We have already noted that the test questions in Study 1 differ from typical survey questions in targeting respondents' memory for written information sources. It is therefore of interest to know whether the same effects (seam effects, forgetting, and constant wave responding) occur in a study that employs that same methodology but whose test questions focus on everyday events in the respondents' lives. We can anticipate some overall differences between the present study and the previous one. It would not be surprising, for example, if respondents were better able to recall autobiographical events than to recall the contents of a questionnaire. Nevertheless, if the same processes are at work when respondents answer autobiographical questions, we should again see evidence both for the seam effects themselves and for any underlying causes of these effects, such as constant wave answers.

This study followed the procedure of Study 1, except that we asked respondents during the test sessions about the events of the previous 4 weeks, rather than about the questionnaire items. During the first test session, for example, instead of asking the questions about the presence of survey items in a questionnaire, as in (4), we asked those in (5) below:

(5) *a*. During the fourth week . . . , did you take a day off work . . . ?
b. During the third week . . . , did you take a day off work . . . ?
c. During the second week . . . , did you take a day off work . . . ?
d. During the first week . . . , did you take a day off work . . . ?

By examining the changes in respondents' answers to series like these, we can again check for a seam effect. Of course, for questions like those in (5), we have no control over the facts that respondents are reporting, and we do

6. For the exact wording of these questionnaires, see http://www.sipp.census.gov/sipp/pubs.html.

not have certain knowledge about the accuracy of their answers. However, we can compare the answers for test questions such as (5) to those the respondent had supplied on the weekly questionnaires (e.g., the item on work loss in [3]). We assume that the questionnaire answers are more likely to be correct than the test answers, since respondents filled out the questionnaires closer to the time of the occurrence of the events. The questionnaires thus provide a partial check on respondents' accuracy. In order to obtain this accuracy information for each week of the study, we also asked the same items in each of the questionnaires, rather than switching questionnaires as we had in Study 1.

The respondents were 56 adults who responded to advertisements placed in Chicago area newspapers. The average age of the respondents was 51 years. Nine respondents were male and 47 were female. (No information about level of education was collected in this study.) Fifty respondents were white, five were African American, and one was Asian. A total of 70 respondents began the study, but 14 were unable to complete all the questionnaires and test sessions. Thus, the attrition rate was 20 percent.

Study 3: Effects of question order on seam effects. Panel surveys such as SIPP and CE typically group questions about a particular topic. The SIPP item in (1), for example, asks all the questions about food stamps one after the other (for the last month, 2 months ago, 3 months ago, and so on). The chronological sequence of the questions, however, differs from survey to survey (and sometimes differs even within the same survey). Whereas the SIPP item in (1), for example, asks respondents to recall information in reverse chronological order (i.e., beginning with the most recent month), the CE item in (2) asks for forward chronological recall (beginning with the earliest month of the reference period). Other CE items ask for backward recall. Previous research on recall order in surveys has found mixed effects on accuracy, with some studies showing an advantage for reverse chronological order and others finding no difference (see Jobe et al. 1990; Loftus et al. 1992).

Our goal in the present study was to determine whether a difference in question ordering and question grouping could impact the seam effect through their effects on memory or constant wave responding. Grouping all questions about a single topic (e.g., receiving food stamps) could encourage respondents to give the same answer to each question. Having just been asked whether they received food stamps in one month, respondents may find it tempting to give the same response if they are then asked about receiving food stamps in a second month. Placing multiple questions about a single topic in different parts of the survey instrument (as we had in Studies 1 and 2) may discourage this type of constant wave responding and reduce the seam effect. Study 3 therefore compared these two grouping strategies explicitly. We also varied the order in which we asked the questions (forward vs. backward chronological order) to check whether the potential advantage of backward recall would surface in this study and, if so, whether it would also translate into a small seam difference.

To investigate these issues, Study 3 employed four recall conditions. In two of these, we asked the test items in backward order as in question (4), and in the other two we asked the test items in forward order. Within each of these two groups, approximately half the respondents answered the questions in a sequence that we blocked by week, with all the questions about one week appearing before the questions about the next. This is the same procedure that we had followed in Studies 1 and 2. The remaining respondents answered the same questions in a sequence that we blocked by item. For these respondents, all the questions about (for example) making a phone call appeared together, as is standard in survey instruments (see, e.g., questions [1] and [2]). Table 2 illustrates these four ordering conditions, using two sample items from the study.

As Table 2 indicates, respondents in all four conditions answered the same test questions, though in different orders. These questions were about items that had appeared in the questionnaires during the previous weeks, as in Study 1. In this study, however, each questionnaire item was related in content to an item in each of the others. For example, the questionnaire for week 1 contained the item, "During the last week, did you make a phone call to a friend?"; the questionnaire for week 2 contained the item, "During the last week, did you make a phone call to a relative?"; and the questionnaire for week 3 contained the item, "During the last week, did you make a phone call to a co-worker?" and so on. Similar sets of items concerned events that occur during a business trip, purchases made at a hardware store, and activities that occur in visiting a doctor, among others. During the test sessions we asked respondents whether they had seen an item about each of these topics for each of the questionnaires (e.g., "Did you see an item about making a phone call on the questionnaire for week 1?"; "Did you see an item about making a phone call on the questionnaire for week 2?"; and so on). If the respondents indicated that they had seen such an item, we asked them to write down the item as it had appeared in the questionnaire.[7]

The 65 adults (48 white, 13 African American, 4 Asian; 24 males, 41 females) in Study 3 had a mean age of 47 years and a mean education level

7. Some of the sequences of items in Study 3 were related by virtue of sharing a common topic (e.g., *making a phone call* in the earlier example). Others were related by occurring within a common causal sequence (e.g., *agreeing to go on a business trip*, *contacting a travel agent for a business trip*, *purchasing tickets for a business trip*, etc.). Our rationale for varying the sequences in this way was to test the possibility that forward order would be more effective for items that have an inherent causal structure, whereas backward order might be more effective for items that merely shared a common topic (see Loftus et al. [1992] for a hypothesis of this kind). In the latter case, memory for recent items might serve as a cue for earlier ones. In this study, however, although respondents tended to remember causally related items better than topically related ones, there was no interaction between item type and order of recall. Hence, for purposes of this article, we consider the two types of items together.

Table 2. Example of the Recall Orders in Study 3

Question Order	Recall Conditions			
	Forward by Item	Forward by Week	Backward by Item	Backward by Week
1	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 1**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 1**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 3**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 3**?
2	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 2**?	Did you see an item about **making a phone call** on the questionnaire for **week 1**? [and so on, for all items from week 1]	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 2**?	Did you see an item about **making a phone call** on the questionnaire for **week 3**? [and so on, for all items from week 3]
3	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 3**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 2**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 1**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 2**?
4	Did you see an item about **making a phone call** on the questionnaire for **week 1**?	Did you see an item about **making a phone call** on the questionnaire for **week 2**? [and so on, for all items from week 2]	Did you see an item about **making a phone call** on the questionnaire for **week 3**?	Did you see an item about **making a phone call** on the questionnaire for **week 2**? [and so on, for all items from week 2]

5	Did you see an item about **making a phone call** on the questionnaire for **week 2**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 3**?	Did you see an item about **making a phone call** on the questionnaire for **week 2**?	Did you see an item about **purchasing something at the hardware store** on the questionnaire for **week 1**?
6	Did you see an item about **making a phone call** on the questionnaire for **week 3**?	Did you see an item about **making a phone call** on the questionnaire for **week 3**?	Did you see an item about **making a phone call** on the questionnaire for **week 1**?	Did you see an item about **making a phone call** on the questionnaire for **week 1**?

of 15 years. The attrition rate was 10 percent. The respondents were from the same pool as those of Study 1, but they had not participated in the earlier study.

Results

In examining the results of these studies, we first check to see whether our procedures produced a seam effect. In the case of Studies 1 and 2 (the 8-week studies), such effects would amount to greater changes in responses between weeks 4 and 5 than between the other pairs of adjacent weeks. In the remaining 6-week study, seam effects imply greater changes between weeks 3 and 4 than between other weeks. We then look at evidence that forgetting and constant wave responding (or other biases) may have contributed to potential seam effects.

SEAM EFFECTS

All of the studies showed evidence of seam effects. The clearest example comes from Study 1, in which there was no objective change at the seam, but where respondents nevertheless switched their responses more frequently across seam weeks than across nonseam weeks. In this study the test questions concerned whether an item had appeared on one of the earlier questionnaires (see the sample questions in [4]), and respondents made a "yes" or "no" response. We can therefore look at the average number of changed answers (from "yes" to "no" or from "no" to "yes") between seam weeks 4 and 5 versus adjacent nonseam weeks.

Table 3, columns labeled "Changed Responses," reports the percentage of changed responses at the seam week and at nonseam weeks. The data from Study 1 show that even when there were no true changes at the seam weeks, changes at the seam outnumber the average changes at the nonseam positions. In fact, the number of seam transitions was greater than that at any of the individual nonseam locations. This is especially noteworthy since there were true changes between nonseam weeks 2–3 and 6–7. To evaluate the statistical reliability of the differences, we use analyses of repeated measures for categorical data (Koch et al. 1977), and we report the Wald statistic (Q_W) for the difference between the number of changed responses at the seam weeks versus the average number of changes at nonseam weeks. For the data from Study 1 in Table 3, $Q_W(1) = 196.99$, $p < .0001$.

Studies 2 and 3 also produced trends in the direction of greater changes at seam weeks, as Table 3 illustrates. In the test sessions of Study 2, respondents were reporting on whether they had participated in specific events (e.g., taking a day off work, as in the sample questions in [5]); so we had no control over the actual number of changes from one week to the next. Perhaps for this

Table 3. Seam, Accuracy, and Constant Wave Data for Studies 1–3

	Changed Responses (%)		Correct Responses (%)		Constant Wave Responses (%)
Study	Seam Weeks	Nonseam Weeks	Earliest Week	Most Recent Week	
1	44.0	31.4	51.7	63.6	34.7
	(50.4)[a]	(33.7)	(47.8)	(66.7)	(37.2)
2	15.2	13.8	82.4	91.2	72.5
3	95.4	67.2	3.8	10.6	11.9

[a] Items in parentheses are predictions from the model described in the final section of this article.

reason, the difference between seam and nonseam weeks is much smaller than for the other studies in the table, but is nevertheless marginally significant, $Q_W(1) = 2.71$, $p < .10$. In Study 3, the test questions concerned sequences of items about a common topic, such as making a phone call (see Table 2), but the specific items to be recalled always differed from one week to the next. For example, respondents were to recall *make a phone call to a friend* for week 1, *make a phone call to a relative* for week 2, and so on. Thus, if the respondents were correct on each test item, the percentage of changes should be 100 percent for both seam and nonseam weeks. Table 3, however, shows more changes at seam weeks, in agreement with the other studies, $Q_W(1) = 349.14$, $p < .0001$. (This should not be taken to mean, however, that respondents were necessarily more accurate for the seam weeks. A respondent could give one response in week 3 and a different response in week 4—the two seam weeks—with neither response being correct. We discuss accuracy in the following section.)

One pattern discernible in Table 3 is that the size of the seam effect appears to increase with the difficulty of the respondent's task. The effect is largest when the respondents had to recall a specific questionnaire item (Study 3). The effect decreases if respondents simply had to indicate whether or not they had seen a questionnaire item (Study 1) and is smallest when they had to decide whether or not a personal event had happened to them (Study 2). The fact that the seam effect depends on the difficulty of the recall task may indicate that memory is a factor in producing the effect. We turn to this possibility next by examining evidence for forgetting during the reference period.

EFFECTS OF FORGETTING

Respondents tended to give more accurate answers for the most recent week of the reference periods, in agreement with a forgetting hypothesis. (For Studies 1 and 2, weeks 4 and 8 are the most recent and weeks 1 and 5 the earliest; see Fig. 2. For Study 3, weeks 3 and 6 are the most recent and weeks 1 and 4 the earliest.) The columns labeled "Correct Responses" in Table 3 display this difference between the earliest and most recent weeks. In each study, respondents were more accurate for the most recent week ($Q_W(1) \geqslant 10.19$, $p < .01$ for all studies). In Studies 1 and 2 respondents indicated whether or not they recognized an item (there were no "don't know" responses). We scored a response as correct in Study 1 depending on whether the corresponding item had appeared on the relevant questionnaire; in Study 2 we scored a response as correct depending on whether the respondent had reported the item in the earlier questionnaire for the relevant week. Study 3 responses were correct if they conveyed the gist of the item to be recalled. All other responses in Study 3, including nonresponses, were scored as incorrect for purposes of this accuracy analysis.

The study-to-study variation in these accuracy differences and in the absolute size of the accuracy rates probably depends on several factors. First, Studies 1 and 2 asked respondents for "yes/no" answers to questions about having seen a questionnaire item or participating in an activity, so respondents could achieve an accuracy of 50 percent by guessing alone. Accuracy is low in Study 3, where respondents had to recall an entire item (e.g., they had to recall the item "During the last week, did you make a phone call to a coworker?"). Second, respondents in Study 2 were answering questions about their own activities rather than about the content of a questionnaire, and this likely had a positive effect on the percentage of correct answers. In addition, because we lacked the objectively correct answers in Study 2, we estimated accuracy indirectly. As we mentioned earlier, we compared respondents' answers in the test sessions to their answers to the same items in the relevant weekly questionnaires. For example, we counted as correct a response to the test question, "During the fourth week . . . , did you take a day off work . . . ?" if the respondent's answer matched the one he or she had given in the fourth week's questionnaire ("During the past week, did you take a day off work?"). This may be responsible for higher levels of recall if the respondents were able to remember their earlier answers. In the other studies, we determined accuracy more directly.[8]

Some studies of SIPP have failed to detect evidence of forgetting over the survey's 4-month response period, as we noted earlier. The same studies demonstrate, however, that respondents' answers are far from perfectly accurate. Thus, one way to reconcile the earlier evidence with the data from Table 3 is to suggest that forgetting of program participation and other SIPP

8. Overall accuracy in Study 3 also depends, of course, on the criteria we used for correct responding. We scored responses in terms of whether respondents preserved the gist of the correct answer rather than in terms of an exact (verbatim) answer, since there were few correct verbatim responses. But although there are uncertainties associated with this criterion, we believe that similar evidence for forgetting would appear for other reasonable scoring policies.

data may sometimes take place "off stage." Forgetting may occur relatively quickly and may reach asymptote by the time of the interview. In our own studies, the average elapsed time between the test sessions and the relevant events was much shorter than in SIPP, which may make it easier to detect forgetting. In addition, SIPP respondents may use estimation or inference strategies that compensate for retrieval difficulties (or that short-circuit retrieval entirely). Although similar strategies may be at work in our own studies, they are less likely to obscure forgetting for the type of information we used here. For example, respondents in SIPP may be able to infer with some accuracy whether they received a social security check in the past few months, but our own respondents would have less to go on when deciding whether a particular item had appeared in an earlier questionnaire.

CONSTANT WAVE RESPONSES

Ideal respondents should have given no constant wave answers in Studies 1 and 3, since the items changed at least once during each of the reference periods. However, Table 3 shows that respondents gave constant answers for 12–35 percent of these items. In Study 3 we calculated the percentages conservatively, omitting cases in which a respondent gave no response to an item for each week in the reference period (see n. 7). Partly for this reason, the percentage of constant wave responses is relatively low in Study 3, reflecting the more complex type of answer this study required.[9]

One factor that may encourage constant wave responding in surveys like SIPP and CE is the way in which they group questions. In SIPP item (1), for example, a respondent first answers the question about receiving food stamps last month, then immediately answers the same question about the periods 2 months ago, 3 months ago, and 4 months ago. This format makes it a simple matter for the respondent to give the same answer to each question. To investigate this possibility, we varied the grouping of items in the test sessions of Study 3 so that some respondents answered questions that we blocked by item (as in SIPP or CE), whereas others answered questions that we blocked by week (see Table 2). Table 4 summarizes the results of this manipulation.[10] The last column of this table shows that grouping by items does in fact increase

9. Table 3 also shows high constant wave percentages for Study 2, which may seem odd in view of the high accuracy rates for this study. However, respondents in this study were recalling their own activities, and some of these activities may actually have been constant from one week to the next. Respondents gave constant answers across each of the weekly questionnaires within a reference period for 65.0 percent of the items (82.5 percent of these were negative cases in which the respondent did not engage in the queried activity in any of the three week of the reference period). This compares with 72.5 percent constant wave answers to the test questions (see Table 3). So the number of incorrect constant wave responses may have been relatively modest.

10. Because of an unequal distribution of missing data, the means of the percentages in Table 4 are not always equal to the overall means for this study in Table 3.

Table 4. Seam, Accuracy, and Constant Wave Data for Study 3, as a Function of Recall Order

	Changed Responses (%)		Correct Responses (%)		Constant Wave Responses (%)
Recall Condition	Seam Weeks	Nonseam Weeks	Earliest Week	Most Recent Week	
Backward by week	94.6	76.5	4.2	16.8	5.6
Forward by week	92.0	75.0	4.1	8.3	6.9
Backward by item	100.0	51.2	3.4	7.9	16.9
Forward by item	97.9	52.3	3.3	10.2	21.6

the likelihood of constant wave answers by 13 percentage points ($Q_w(3) = 76.72$, $p < .0001$). Because constant wave responses tend to decrease the number of changes at off-seam months, grouping questions by item should also inflate the size of the seam effect. The columns labeled "Changed Responses" in Table 4 demonstrate that this effect is also present in our data. In this study, seam effects more than doubled when respondents answered in sequence all questions about a common topic (e.g., the difference in changed responses between seam and nonseam weeks is 18.1 percentage points when respondents recalled items backward by week, but 48.8 points when they recalled backward by item).

Although item blocking clearly affects both constant wave responses and seam effects in these data, its influence on accuracy is more equivocal. In Study 3, respondents who answered questions in backward (i.e., reverse chronological) order with items blocked by week not only produced fewest constant wave responses and small seam effects but also highest accuracy levels. (Correct recall is low overall in this study for reasons we have discussed in the section of the results on "Effects of Forgetting.") This boost in accuracy occurred mainly in the most recent week of the reference period, as Table 4 indicates. Because respondents in this condition (backward by week) answered all questions about the most recent week before the questions about earlier weeks, their accuracy advantage may be due to the fact that they were able to deal with the most recent and accessible items at the beginning of the test session when motivation was probably highest and fatigue lowest. However, in an additional study (not reported here), we have found slightly lower accuracy (by 6 percentage points) when respondents recalled backward by week versus backward by item. (Like Study 3, though, the data also showed fewer constant wave responses and smaller seam effects when respondents recalled backward by week.) The correct answers in Study 3 always varied from week to week, whereas half the correct answers in the additional study were constant across weeks. This suggests that grouping by week may be of most benefit when the experiences are likely to change across weeks. Such grouping may keep respondents from automatically giving the same answer to an item for each week in the reference period and may encourage them to make another retrieval effort.

A Theory of the Seam Effect

The results from these studies suggest a direct relationship between difficulty of the respondents' retrieval task and the size of the seam effect: The harder it is for respondents to recall the queried information, the larger the effect. The relation between constant wave responses and the seam effect is not quite so straightforward in these data, perhaps because of differences in the nature of the responses across studies ("yes/no" vs. recall). Nevertheless, the data from Table 4, which we have just reviewed, show that, at least within an individual study, conditions that favor constant wave responses also produce larger seam effects. One general explanation of these relations is that as information becomes more difficult for respondents to remember, they tend to substitute other response strategies that compensate for their inability to recall. As Cannell (1965, p. 2) observed, "In order to report accurately the respondent must relive or review carefully his experience, constantly checking his own memory, or he must resort to records of the event. The farther away the event in time or the less importance it has, the greater the energy required to recall it. Frequently respondents give inaccurate information merely to avoid the work required to respond accurately." In the present case, one strategy for avoiding work is simply giving the same answer for all intervals—constant wave responding—especially when all questions about the same item appear together, as they do in SIPP and CE or in the by-item conditions of Study 3. These constant wave answers may then further exacerbate the seam effect.

To make this proposal more concrete, we can consider a model in which responses to an individual item depend on the state of a respondent's information about that item at different points in the reference period. Figure 3*a* illustrates our assumptions. According to this model, respondents initially attempt to retrieve from memory facts about the most recent interval of the reference period (e.g., week 4 in Studies 1 or 2). With some probability (p_{mem}), the respondent succeeds in this task and gives a correct answer. The respondent may, however, fail to remember the relevant fact (with probability $1 - p_{mem}$) and instead guess at an answer. In the context of "yes/no" questions, such as that of Studies 1 and 2, we assume that the respondent will guess "yes" with probability p_{guess} and "no" with probability $1 - p_{guess}$. (We treat the probability of guessing "yes" as a parameter, rather than arbitrarily setting it at .5, since we suspect that this likelihood can be influenced by factors such as the overall familiarity of the information.) The respondent next attempts to determine a response for the next-to-most-recent interval in the reference period (e.g., week 3 in Studies 1 or 2). If the respondent's first response was memory based, he or she may again be

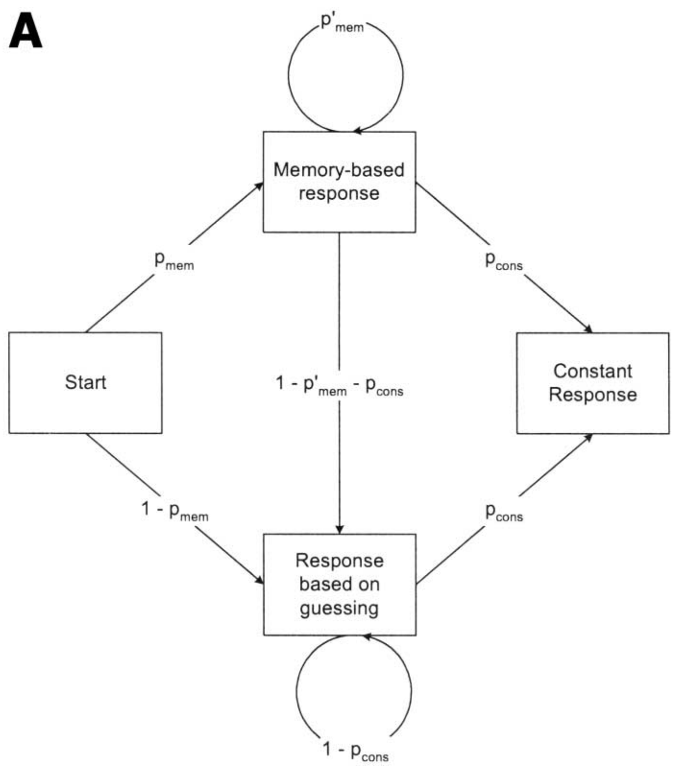
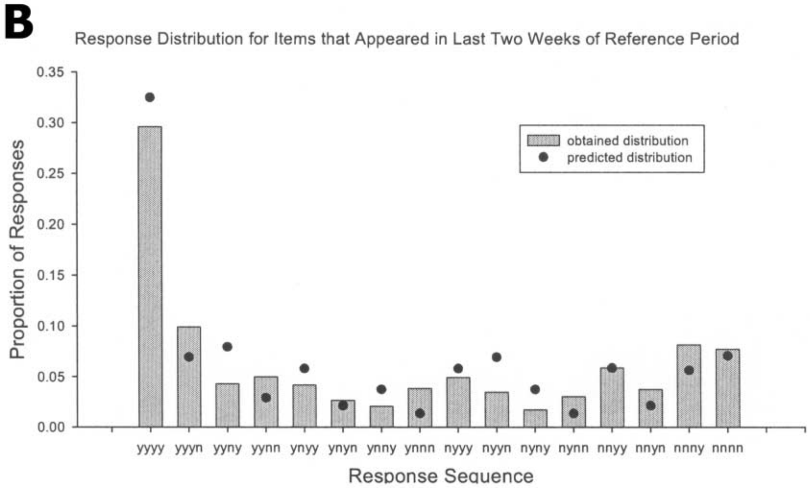
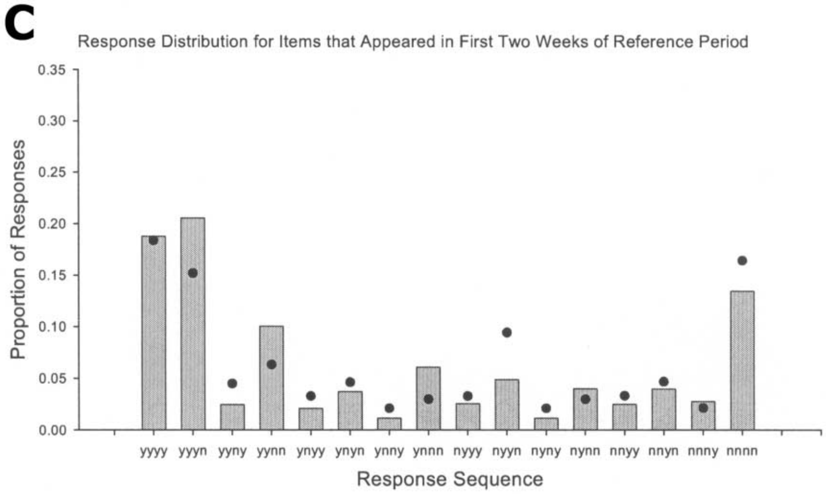


Figure 3. *a*, A model for seam effects. Respondents begin by making either a memory-based response for the most recent interval of a reference period or a response based on guessing. Subsequent responses for earlier intervals in the reference period can be based on memory, guessing, or constant wave responding. Arrows indicate the probability of a transition from one response state to another. *b*, Predicted (black points) and observed (grey bars) response distributions for items that appeared during the last 2 weeks of the reference period in Study 1. The x-axis indicates the sequence of responses about the 4 weeks in each reference period; "y" means the respondent answered "yes" that the item appeared on the questionnaire, and "n" means the respondent answered "no." Thus, "ynyn" represents a sequence in which a respondent said that a particular item appeared on the questionnaire for the first week, did not appear on the questionnaire for the second week, did appear on the questionnaire for the third week, and did not appear on the questionnaire for the fourth week of the reference period. Sequences of the form "yyyy" and "nnnn" are constant wave responses. The sequence "nnyy" is the correct response. *c*, Predicted and observed response distributions for items that appeared during the first 2 weeks of the reference period in Study 1. The sequence "yynn" is the correct response.

B


Response Distribution for Items that Appeared in Last Two Weeks of Reference Period
obtained distribution
predicted distribution
Proportion of Responses
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
yyyy yyyn yyny yynn ynyy ynyn ynny ynnn nyyy nyyn nyny nynn nnyy nnyn nnny nnnn
Response Sequence

C


Response Distribution for Items that Appeared in First Two Weeks of Reference Period
Proportion of Responses
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
yyyy yyyn yyny yynn ynyy ynyn ynny ynnn nyyy nyyn nyny nynn nnyy nnyn nnny nnnn
Response Sequence

able to retrieve information about this second interval, but if recall is too difficult at this point, the respondent can either repeat the answer just given (i.e., make a constant response with probability p_{cons}) or make an attempt to guess the answer. If the respondent's first response was a guess, then the respondent can also repeat that response or make a new guess. This same process continues for the remaining intervals in the reference period.[11]

To demonstrate the usefulness of this model, we have fit it to the data from Study 1. The model predicts the probability of a sequence of responses to a particular item within a reference period (e.g., "yes" for week 1, "no" for week 2, "no" for week 3, and "yes" for week 4); so the output from the model is a probability distribution across the 16 possible sequences. From these distributions, we can then derive predictions for all the statistics in Table 3: percentage of changed responses, correct responses, and constant wave responses. In Study 1 there were two types of correct sequences, since some items appeared in the questionnaires for the first 2 weeks of the reference period, while others appeared in the questionnaires for the last 2 weeks. The model generated separate predicted distributions for these two types of items (but using the same parameter values), since the parameters affect these distributions in different ways. Preliminary model fitting suggested very little memory for information beyond the most recent week of the reference period, so we set the likelihood of subsequent memory-based responses (p'_{mem} in Fig. 3*a*) to zero. The model therefore predicts 32 data points (16 response sequences for each of the two types of items—those that appeared in the first 2 weeks and those that appeared in the last 2 weeks of the reference period) with three parameters (p_{mem}, p_{guess}, and p_{cons}).

To fit the model, we first constructed equations for the probability of each data point in terms of the model parameters just mentioned. These equations were based on the pathways through the diagram in Figure 3*a*. As an illustration of how the model works, consider the situation in which a particular item (e.g., the item about taking a day off work in Question [3]) appeared on the questionnaires in weeks 1 and 2 but did not appear in weeks 3 and 4. Correct responses to the test questions about this item would therefore be "no" to Questions (4*a*) and (4*b*) and "yes" to Questions (4*c*) and (4*d*). Figure 3*a* shows that one way this correct response sequence can occur is if a respondent retrieves from memory the fact that there was no such item on the

11. This model obviously simplifies the range of respondent strategies. For example, the model assumes respondents either remember an item and make a correct response, or fail to remember it and guess (or rely on constant wave answers). But it is possible that a respondent could retrieve partial information about an item and make a (correct or incorrect) response based on inference. Similarly, it is easy to imagine that respondents have heuristics other than constant wave responding in their repertoire. The aim of the modeling effort here is to capture the main features of the data as far as they bear on the seam effect, and in line with previous research and our own findings, we concentrate on the interaction between forgetting and constant wave responding. It is an open question, of course, whether a more fine-grained approach could produce an increase in predictive accuracy that would compensate for its increase in complexity.

questionnaire in week 4 (with probability p_{mem}), then correctly guesses that the item was not on the questionnaire in week 3 (with probability $(1 - p_{cons})$ * $(1 - p_{guess})$), and finally guesses twice more that the item was on the questionnaires for weeks 1 and 2 (each with probability $(1 - p_{cons}) * p_{guess}$). (Recall that $p'_{mem} = 0$, simplifying the expressions just given.) The overall probability of this route is then:

$$p_{mem} * (1 - p_{cons}) * (1 - p_{guess}) * [(1 - p_{cons}) * p_{guess}]^2 =$$
$$p_{mem} * (1 - p_{cons})^3 * (1 - p_{guess}) * {p_{guess}}^2.$$

Figure 3*a* also shows three alternative routes to this same response sequence. For example, a respondent could again correctly retrieve the fact that the item was not on the questionnaire for week 4, guess that it was not on the questionnaire for week 3, guess that it was on the questionnaire for week 2, but then repeat this last response (i.e., make a constant response) for week 1 (with probability p_{cons}). This route therefore has probability:

$$p_{mem} * (1 - p_{cons})^2 * (1 - p_{guess}) * p_{guess} * p_{cons}.$$

The remaining routes to the correct response consist of one in which the respondent guesses correctly on all four occasions and one in which the respondent guesses correctly on the first three occasions and then makes a constant response on the fourth. The total probability for making a correct response (“yes” for week 1, “yes” for week 2, “no” for week 3, and “no” for week 4) is the sum of the probabilities for these four possible routes:

$$\text{Prob (“yes,” “yes,” “no,” “no”)} = p_{mem} * (1 - p_{cons})^3 * (1 - p_{guess}) * {p_{guess}}^2$$
$$+ p_{mem} * (1 - p_{cons})^2 * (1 - p_{guess}) * p_{guess} * p_{cons}$$
$$+ (1 - p_{mem}) * (1 - p_{cons})^2 * (1 - p_{guess})^2 * p_{guess} * p_{cons}$$
$$+ (1 - p_{mem}) * (1 - p_{cons})^3 * (1 - p_{guess})^2 * {p_{guess}}^2.$$

We obtained equations like this for each of the 16 possible response sequences for each of the two item types. We then used nonlinear regression based on the Newton method (see, e.g., Thisted 1988) to estimate the values of the parameters in fitting the equations to the observed data.

The predicted response distributions follow the data quite closely. Figure 3*b* shows the predicted and obtained distribution of the responses for test questions about items that had appeared on questionnaires in the last 2 weeks of the reference period, and Figure 3*c* gives comparable distributions for items from the first 2 weeks. Labels on the x-axis indicate the individual response sequences, with y for “yes” responses and n for “no.” For example, the sequence "nynn" means that the respondent made a "no" response when asked whether an item was on the questionnaire for the first week of the reference period, made a "yes" response that the same item was on the second week's questionnaire, and so on. Constant wave responses are those indicated by "yyyy" and "nnnn." Correct responses for items that appeared in the last 2 weeks are those labeled "nnyy"; correct response for items that appeared in the first 2 weeks (as in the earlier example) are those labeled "yynn." For the regression, $F(3, 29) = 132.34$, $p < .0001$, with a residual mean square of only 0.0006. The estimated values of the parameters are: $p_{mem} = .32$ (SE = .05), $p_{guess} = .61$ ($SE = .02$), and $p_{cons} = .18$ (SE = .02). The value of p_{guess} reflects a bias toward "yes" responses in this study. This is probably the result of the fact that respondents had seen all items prior to the test sessions (though, of course, not in each week); so every item may have seemed familiar to them.[12] Table 3 shows in parentheses the predicted percentage of changed responses, correct responses, and constant wave responses. Except for the predictions for changes during seam weeks, these figures come directly from the predicted response distributions. To get the seam-week predictions, we also assumed that the response sequences for the first reference period were independent of those from the second reference period. We then calculated by hand the likelihood that a "yes" response for week 4 would be followed by a "no" response from week 5 (or the reverse change), using the predictions for the reference periods in Figures 3*b* and 3*c*. The quantitative match to the data in Table 3 seems reasonably good, even though the model was not optimized to fit these statistics, and demonstrates that the same model is consistent with the presence of seam effects, forgetting, and constant wave responding.[13]

Concluding Comments

The studies we have reported attempt to scrutinize the seam effect by examining it under experimental conditions. The results show that this method

12. This bias accounts for the ability of the model to predict the difference in the shape of the distributions in Figure 3*b* and 3*c*. Figure 3*b* shows many more cases in which respondents answered that an item had appeared in all four questionnaires ("yyyy" responses) than that an item had appeared in none ("nnnn" responses). Figure 3*c* shows about the same number of cases for each of these two response types. For items that appeared in the questionnaires for the last 2 weeks of a reference period (Fig. 3*b*), both memory and guessing tend to produce a "yes" response for the most recent week. This response can then become the basis of a constant wave answer and, thus, a "yyyy" series. However, for an item that did not appear in the last 2 weeks (Fig. 3*c*), memory yields a "no" response, while guessing again tends to produce a "yes" answer. Extending these through constant wave responding gives a more even balance of "yyyy" and "nnnn" sequences.

13. A possible qualification is that the model overpredicts by 6 percentage points the proportion of seam transitions. This may be a consequence of our independence assumption. An alternative would be to build into the model further assumptions about the carryover from the first to the second test sessions, but it is unclear whether the improvement in fit would be large enough to warrant this change.

does produce a seam effect—more changes in response across neighboring intervals between reference periods than within those periods—even in the absence of any true changes at the seam (Study 1). The effect also occurred across a variety of memory conditions, including recognition of previous items or events (Studies 1–2) and recall of previous items (Study 3). The effect is robust over different types of question grouping and question order. Some of these variables, however, alter the size of the seam effect. The effect is larger when questions about the same content appear together. The size of the effect also appears to depend on the difficulty of the respondents' memory task: The greater the demands on memory, the larger the seam difference.

Our studies also document the joint influence of forgetting and constant wave responding in the seam effect. Even when respondents are attempting to recognize an event as one that happened to them, their accuracy decreases with time in the reference period (Study 2). Perhaps as a reaction to forgetting, respondents often give the same response to each interval in the reference period, even when the true answers change from interval to interval (Studies 1 and 3). To account for these data, we propose that respondents attempt to retrieve relevant information in answering a question but resort to (possibly biased) guessing or to repeating an earlier response when memory fails. A model based on this hypothesis provides a close fit to the response sequences in Study 1 and is consistent with the presence of seam effects, forgetting, and constant wave responding.

Of course, the methods we employed here differ from the conduct of actual surveys, and we should be cautious about generalizing to survey contexts. The time span of our studies is much shorter than the usual schedules for panel surveys, and this may affect the degree to which the seam difference depends on forgetting and other factors. In addition, the items that appear in surveys are sometimes more memorable than those we employed here and may allow respondents to use different strategies to recall or to estimate their answers. We focused on items in Studies 1 and 3 that were deliberately difficult to reconstruct or to estimate in order to get a clearer view of the role of memory.[14] We acknowledge, however, that questions in longitudinal surveys are unlimited in their variety, and respondents' methods for answering them may also vary in potentially idiosyncratic ways. As far as we can tell from earlier studies, all survey items can give rise to seam effects, but the route to these effects may sometimes differ from the account that appears in the model of Figure 3.

In Study 3, we looked at a potential remedy for the seam effect that comes from placing repeated items in different parts of the questionnaire. When

14. It is relevant that Study 2, which used items closer to those of typical surveys, obtained smaller seam effects and greater accuracy. This does not imply, however, that questions about more naturalistic events will tend to produce minimal seam effects. As we have already noticed in our discussion of SIPP, seam effects can be quite large even for questions about events such as receipt of social security or food stamps.

coupled with backward recall, this blocking by week led to both a decrease in the seam effect and an improvement in accuracy in one study; in a further study that we have not reported in this article, it again decreased the seam effect but with a slight decrease in accuracy. This procedure probably discourages respondents from relying on constant wave answers and, at least for the type of items we employed, may encourage them to try again to recall the queried information. It is possible, of course, that respondents would find their task unnatural in this situation because they are unable to give at one time the full (several-month) account about an individual item, such as receiving food stamps. But we are not yet convinced that it is necessarily more natural to step back and forth in time with each new item, as in current surveys. Respondents may prefer to fill in the details of a given time period completely before moving to the next. Although blocking by time period may require more effort on the part of both interviewers and respondents, its merits and demerits are an empirical matter—one that we think bears closer examination in actual surveys.

Other methods for reducing the seam effect are also worth pursuing. One that we have examined in further experiments is the use of dependent or bounded interviewing (e.g., Neter and Waksberg 1964). In this procedure, interviewers remind respondents of the answer they gave in the previous interview before asking the respondents the same question about the current reference period (e.g., during the second interview in the Fig. 1 scheme, the interviewer would tell respondents what their answer had been to the question, "Did you receive food stamps during month 4?"; the respondents would then answer the questions about whether they had received food stamps during months 5–8). In our experimental analog of dependent interviewing, we used a method similar to that of Studies 1 and 3, but during the second test session, we told respondents the answer they had given for the final week of the first response period before asking about the second response period (see Fig. 2). We compared results from this group to a control group that received the same questions but no dependent information. The results showed that dependent interviewing did indeed reduce the size of the seam effect compared to the control, but with no improvement in overall accuracy. Thus, although the respondents' errors were smoothed across the seam, they did not disappear.

Dependent interviewing is inherently limited because the feedback to the respondent can itself contain errors. Other techniques for reducing the seam effect might be more successful if they directly target forgetting, perhaps by giving respondents more retrieval cues than are available in a single conventional question. One such technique is the event history calendar in which respondents are encouraged to think about events from several contemporaneous life themes (e.g., Conway 1996), such as jobs, relationships, and household composition. With event history calendars, interviewers can press respondents to align events across themes, and this can lead to additional recall. Belli, Shay, and Stafford (2001) found better recall with these calendars than with conventional questioning, where these investigators determined accuracy by comparing responses to those from a conventional interview several years earlier.

It would be premature at this point to draw definitive conclusions about practical attempts to eliminate the seam effect. The results of our studies bear on these attempts, but they are far enough removed from the circumstances of actual surveys to make their implications tentative. What we believe the data suggest, however, is that the seam effect depends on several different but interrelated cognitive factors. Remedies that focus on just one factor can sometimes reduce the seam effect without necessarily increasing overall response accuracy. Because survey researchers are mostly concerned with accuracy, they are likely to see merely shrinking the seam effect as of secondary importance. We suspect that the most successful measures—measures that reduce erroneous responses on and off seam—will be ones that make it easier for respondents to retrieve the facts of interest while making it harder for them to glide through a series of repeated questions with a series of repeated answers.

References

Barsalou, Lawrence W. 1988. "The Content and Organization of Autobiographical Memories." In *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*, ed. Ulric Neisser and Eugene Winograd, pp. 193–243. Cambridge: Cambridge University Press.

Belli, Robert F. 1998. "The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys." *Memory* 6:383–406.

Belli, Robert F., William L. Shay, and Frank P. Stafford. 2001. "Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods." *Public Opinion Quarterly* 65:45–74.

Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell. 1987. "Answering Autobiographical Questions: The Influence of Memory and Inference on Surveys." *Science* 236:157–61.

Burkhead, Dan, and John Coder. 1985. "Gross Changes in Income Recipiency from the Survey of Income and Program Participation." In *Proceedings of the Social Statistics Section*. Washington, DC: American Statistical Association.

Cannell, Charles F. 1965. *Reporting of Hospitalization in the Health Interview Survey*. Vital and Health Statistics, Ser. 2, No. 6. Hyattsville, MD: National Center for Health Statistics.

Conway, Martin A. 1996. "Autobiographical Knowledge and Autobiographical Memories." In *Remembering Our Past: Studies in Autobiographical Memory*, ed. David C. Rubin, pp. 67–93. Cambridge: Cambridge University Press.

Hill, Daniel H. 1987. "Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods." In *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association.

Jabine, Thomas B., Karen E. King, and Rita J. Petroni. 1990. *Quarterly Profile, Survey of Income and Program Participation.* Washington, DC: U.S. Census Bureau.

Jobe, Jared B., Andrew A. White, Catherine L. Kelley, David J. Mingay, Marcus J. Sanchez, and Elizabeth F. Loftus. 1990. "Recall Strategies and Memory for Health-Care Visits." *Milbank Quarterly* 68:171–89.

Kalton, Graham, and Michael E. Miller. 1991. "The Seam Effect with Social Security Income in the Survey of Income and Program Participation." *Journal of Official Statistics* 7:235–45.

Koch, Gary G., J. Richard Landis, Jean L. Freeman, Daniel H. Freeman, Jr., and Robert G.

Lehnen. 1977. "A General Methodology for the Analysis of Experiments with Repeated Measurement of Categorical Data." *Biometrics* 33:133–58.

Kurbat, Matthew A., Steven K. Shevell, and Lance J. Rips. 1998. "A Year's Memories: The Calendar Effect in Autobiographical Recall." *Memory and Cognition* 26:532–52.

Loftus, Elizabeth F., Kyle D. Smith, Mark R. Klinger, and Judith Fiedler. 1992. "Memory and Mismemory for Health Events." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, ed. Judith M. Tanur, pp. 102–37. New York: Russell Sage.

Marquis, Kent H., and Jeffrey C. Moore. 1989. "Response Errors in SIPP: Preliminary Results." In *Proceedings of the Annual Research Conference.* Washington, DC: U.S. Census Bureau.

———. 1990. "Measurement Errors in the Survey of Income and Program Participation (SIPP) Program Reports." In *Proceedings of the Annual Research Conference.* Washington, DC: U.S. Census Bureau.

Martini, Alberto. 1989. "Seam Effect, Recall Bias, and the Estimation of Labor Force Transition Rates from SIPP." In *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association.

Menard, Scott. 1991. *Longitudinal Research*. Newbury Park, CA: Sage.

Moore, Jeffrey C., and Daniel Kasprzyk. 1984. "Month-to-Month Recipiency Turnover in the ISDP." In *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association.

Moore, Jeffery C., and Kent H. Marquis. 1989. "Using Administrative Record Data to Evaluate the Quality of Survey Estimates." *Survey Methodology* 15:129–43.

Moore, Jeffery C., Kent H. Marquis, and Karen Bogen. 1996. "The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation." SIPP Working Paper 212. Washington, DC: U.S. Census Bureau.

Neter, John, and Joseph Waksberg. 1964. "A Study of Response Errors in Expenditures Data from Household Interviews." *Journal of the American Statistical Association* 59:18–55.

Pillemer, David B., Elizabeth D. Rhinehart, and Sheldon H. White. 1986. "Memory of Life Transitions: The First Year in College." *Human Learning* 5:109–23.

Plewis, Ian. 1985. *Analyzing Change: Measurement and Explanation Using Longitudinal Data.* Chichester: Wiley.

Robinson, John A. 1986. "Temporal Reference Systems and Autobiographical Memory." In *Autobiographical Memory*, ed. David C. Rubin, pp. 159–88. Cambridge: Cambridge University Press.

Shum, Michael S., and Lance J. Rips. 1999. "The Respondent's Confession: Autobiographical Memory in the Context of Surveys." In *Cognition and Survey Research*, ed. Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau, pp. 95–109. New York: Wiley.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco: Jossey-Bass.

Thisted, Ronald A. 1988. *Elements of Statistical Computing.* New York: Chapman and Hall.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *Psychology of Survey Responses.* Cambridge: Cambridge University Press.

Young, Nathan. 1989. "Wave-Seam Effects in the SIPP." In *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association.