Session 2

Cognitive Basis for Seam Effect in Panel Surveys

SEAM EFFECTS FOR QUANTITATIVE AND QUALITATIVE FACTS

Lance J. Rips
Northwestern University

Frederick G. Conrad
Bureau of Labor Statistics

Scott S. Fricker
Bureau of Labor Statistics

Jennifer Behr
Yale University

SEAM EFFECTS FOR QUANTITATIVE AND QUALITATIVE FACTS[1]

The accuracy of answers to factual questions degrades over time. People's memory for an event becomes less accurate with the time since the event took place, so it's natural to expect the accuracy of survey responses that depend on such memories to decrease in the same way. If you ask about our income sources, health histories, or other biographical facts, you can probably expect better answers for last month's information than for that of the month before. Many studies of autobiographical memory document this decrease, though the rates of forgetting vary widely from one type of material to another (see Shum & Rips, 1999, and Tourangeau, Rips, & Rasinski, 2000, chaps. 3 and 4, for reviews of the memory literature as it bears on surveys). Of course, people aren't entirely at the mercy of memory, since they usually have ways of estimating or inferring information when memory fails. Response accuracy over time will then depend on the interplay of forgetting and its compensating strategies. Survey researchers face the job of understanding this interplay in order to estimate the true values of the information they seek.

The studies we report here explore a well-documented type of response error called the seam effect that occurs in panel surveys. The seam effect is time-dependent, since it exhibits a clear temporal profile, but its form is more complicated than a simple increase in errors over time. Our goal is to try to understand this effect by examining some of its components. The panel surveys in which the seam effect appears typically take place over a period of several years, which makes it difficult to study efficiently. We've therefore made use of a laboratory analog that produces seam effects and allows us to vary factors that might contribute to them. In this paper, we first describe the nature of the seam effect in actual surveys and the analog of the effect with which we've been working. Next, we briefly review earlier results using this method and then report two new experiments that extend these findings. Finally, we summarize our conclusions about the seam effect and possible ways to eliminate it.

Seam Effects in Panel Surveys

Seam effects occur in panel surveys that ask respondents about events from each of a series of subintervals within the survey's larger response periods. For example, the Survey of Income and Program Participation (SIPP) interviews respondents three times a year, but during an interview the respondents must report about income and employment for each of the past four months. We show this type of schedule in the upper panel of Figure 1. A respondent might be interviewed in May, for example, and provide answers during that interview about whether he or she received social security benefits during each of the months of January, February, March, and April. The same respondent would be re-interviewed in September and would then provide information about receiving social security benefits in each of May, June, July, and August; and so on.

The seam effect appears in plots of changes in the individual respondent's answers from one month to the next in this series. For example, we can count the number of times respondents changed their answer from "yes, I received social security benefits in January" to "no, I did not

receive social security benefits in February" or the reverse change from "no" to "yes." If we then graph the total number of such month-to-month changes across the entire period of the survey, we get the type of scalloped profile that appears in the lower panel of Figure 1, which is taken from SIPP (Jabine, King, & Petroni, 1990). These data are month-to-month changes in reports of receiving social security and food stamps. For the respondent who is interviewed in May, September, and January, the first point on the x-axis corresponds to the change in response between January and February, the second point to the change between February and March, and so on for all pairs of adjacent months.

It's crucial that the change between months 1 and 2 depends on answers that come from the same interview (the May interview in our example), as does the change between months 2 and 3, and months 3 and 4. The change between months 4 and 5, however, is based on data from two separate interviews: Month 4 answers come from the first interview in this series (e.g., the May interview), whereas the month 5 answers come from the second interview (e.g., the September interview). Months 4 and 5 are on the "seam" between the response periods for the first two interviews, and these and other seam transitions appear in the figure at the positions of the dashed vertical lines. The seam effect is the finding that the number of changes at these seams is much greater than the number of changes between other pairs of adjacent months.[2]

A Laboratory Model of the Seam Effect

There is little doubt that the seam effect is due to the fact that data from the seam months come from two separate interviews while data from the nonseam months come from the same interview. We might say that the response period for the interviews (e.g., four months in SIPP) differs from the response period for the questions (one month), and this difference is responsible for the pattern in Figure 1. Similar differences in response periods appear in other surveys, such as the Consumer Expenditure Survey (CE). The issue for survey methodologists and cognitive psychologists is which factors associated with the change in interviews increase or decrease the size of the effect.

Two general explanations of the seam effect are possible, and the literature on the seam effect implicates both. On one hand, the effect might be due to factors that enhance the difference at the seam months. Suppose, for example, that respondents gradually forget information during the period between interviews, as seems likely. Then answers to questions about month 4, the most recent month queried during the first interview, will draw on respondents' relatively rich memory for the events; but answers to questions about month 5, the earliest month queried during the second interview, will draw on relatively impoverished memory. Forgetting for the incidents in question could therefore contribute to the size of the seam difference. On the other hand, the effect could also be due to factors that minimize differences across nonseam months. Respondents might be biased, for example, to give the same answer about each month during an interview. When asked whether they received social security benefits during each of January, February, March, and April, they answer "no" to all four questions (or "yes" to all four) as a way of simplifying their task. These constant-wave responses reduce the changes for nonseam months, making the changes at the seam stand out (Kalton & Miller, 1991; Young, 1989).

To analyze the seam effect, we need to know the facts that respondents are trying to report. Unless we have access to the correct answers, it's difficult to know whether the seam effect is due to exaggerated changes at the seam months, to suppressed changes at the nonseam months, or both. For this reason, we've designed a new procedure that is in some ways a cross between a field study and a laboratory task. Figure 2 illustrates the basic schedule for the experiment. In this procedure, we mail information to respondents each week for eight consecutive weeks. This information is embedded in a questionnaire that they fill out and mail back to us within 24 hours. Respondents come into the lab at the end of the fourth week and again at the end of the eighth week, and during these test sessions, we ask them to report on the information they had seen in the questionnaires during each of the preceding four weeks. These two test sessions are our analogs to the survey interviews in SIPP, CE, and other panel surveys, dividing the interval into two response periods. The questions that we ask during these sessions provide the week-to-week data that we need in order to study the seam differences. Changes in respondents' answers between weeks 4 and 5 are the seam changes, coming from two different test sessions. Changes in answers between the other pairs of neighboring weeks (1-2, 2-3, 3-4, 5-6, 6-7, and 7-8) are nonseam transitions, coming from the same test session. The time scale of the design is in weeks rather than months to allow us to study seam effects more efficiently.

Previous Results

Our earlier experiments demonstrated that we could use this design to produce seam effects and to alter their size. In one such experiment (Rips, Conrad, & Fricker, 2000, Experiment 1), the questionnaires we sent to respondents during weeks 1-8 asked them yes-or-no questions about common activities they might have participated in that week. The questionnaire for week 1, for example, asked, During the last week, did you have the oil changed in your car?, During the last week, did you order a pizza for home delivery?, along with similar items. Each of the eight questionnaires contained a total of 50 questions about the occurrence of everyday events, which we had selected on the basis of norms of rated frequency of occurrence, duration, importance, and affective impact from an earlier study. We composed the questionnaires so that one group of items appeared during weeks 1, 2, 7, and 8, and a separate group of items appeared during weeks 3, 4, 5, and 6. Thus, the questionnaires in seam weeks 4 and 5 were identical, apart from the random order of the items in the questionnaires.

During the two test sessions, we asked respondents to think back to the questionnaires they had filled out in the past four weeks and to decide whether certain items had appeared on those questionnaires. In the first test session, for example, we gave respondents a list of questions and asked them whether each question had appeared in the questionnaire for week 4 (e.g., On the questionnaire for week 4, did you see: [the item about having] oil changed in your car?, ...did you see: [the item about ordering] a pizza for home delivery?, etc.). We next re-presented the same questions in a new random order and asked the respondents whether each item had appeared in the questionnaire for week 3. We then repeated this procedure for week 2 and week 1. The procedure for the second test session was identical, except that we asked respondents about the content of the questionnaires for weeks 8, 7, 6, and 5 (in that order).

The main data from this experiment come from the test sessions: respondents' answers about whether they remembered seeing individual items on the weekly questionnaires. We can look directly at the changes in their answers to these items to see whether they exhibit a seam effect. But because we also know which items actually appeared on the questionnaires, we can also evaluate their responses for accuracy. In fact, the data from this study produced a seam effect: The largest percentage of changes occurred between seam weeks 4 and 5. (These are changes from "yes, I saw that item on the week 4 questionnaire" to "no, I didn't see that item on the week 5 questionnaire" or the reverse change.) The actual items that respondents had seen during these two weeks were exactly the same; so the increase in changes at this point is entirely due to response error. By contrast, true changes in the questionnaire items had occurred between weeks 2 and 3 and again between weeks 6 and 7; however, the data showed no increase in the percentage of changed responses at these two points. Thus, the results showed an increase in changed responses where there were no objective changes (between weeks 4 and 5), but no change in responses where there were objective changes (between weeks 2 and 3 and between weeks 6 and 7).

What is responsible for the form of these data? It seems likely that forgetting contributes to the effect. Respondents were reliably above chance in their ability to recognize items from the questionnaires they had seen just before the test sessions in weeks 4 and 8 (63.5% correct), but fell to near chance accuracy for earlier weeks (e.g., 52.6% correct for weeks 3 and 7, where 50% is chance recognition). There is some evidence, however, that constant-wave responding also contributed to the effect. In 19.8% of cases during the first test session, respondents made positive constant-wave responses, saying that they had seen an item in all four preceding weeks. Similarly, 10.5% made negative constant-wave responses, saying that they had not seen an item in any of the four preceding weeks. During the second test session, the comparable statistics are 27.4% and 9.9%. Thus, on about 30% of occasions in each test session, respondents were making constant-wave responses. Because each test item appeared in exactly two of the four weeks during a response period, these constant-wave responses were incorrect for two of these weeks. Forgetting could be responsible for the negative constant-wave cases: Respondents may simply have been unable to remember an item on any of the last four questionnaires. It's more difficult, however, to account for the more numerous positive constant-wave responses. Some additional bias in favor of "yes" responses must be at work here.[3] We show in an earlier paper (Rips, et al., 2000) that a theory based on a combination of forgetting and positive constant-wave responding can account for the detailed results from this experiment.

Study 1: Seam Effects for Biographical Material

The results from our preliminary study depend on a rather unusual type of question. We asked respondents whether they recalled items from questionnaires – for example, whether they remembered seeing an item on the questionnaire for week 3 about having the oil changed in their car. We asked questions like these because they gave us experimental control over the to-be-recalled information and allowed us to determine the accuracy of the respondents' answers. But, of course, it is also important to know whether the results generalize to items closer to those of actual surveys. We want our conclusions to apply to survey questions about personal information, not just to questions about questions about such information.

To extend our study in a more naturalistic direction, we've modified the basic procedure in Figure 2 to ask respondents during the test sessions about personal incidents. In this new experiment, we again sent respondents weekly questionnaires about ordinary activities, such as whether they had the oil changed in their car or whether they had checked out books from the library last week. As in the study just described, respondents received eight of these questionnaires in consecutive weeks. They filled out the questionnaires, checking off "yes" or "no" to each item, and mailed the questionnaires back to us within 24 hours. The questionnaires themselves were similar to those of the preliminary experiment, except that there was no change of items from one questionnaire to the next: Respondents saw the same set of items (e.g., During the last week, did you take a day off work due to illness?) on each of the questionnaires they filled out. The respondents were 56 adults (average age 50.6) whom we had recruited through advertisements in local newspapers.

The respondents also took part in two test sessions, again following the pattern of Figure 2. This time, however, we asked respondents about the actual incidents they had described earlier. In the first test session, we asked respondents, for example, whether they had taken a day off work due to illness during week 4, whether they had checked out a book from the library during week 4, and so on. There were 60 questions in all. Half these questions the respondents had answered in their earlier questionnaires, and half were new. We then asked the same set of questions about week 3, week 2, and finally week 1. We conducted the second test session in the same way, asking about weeks 8, 7, 6, and 5, in that order. The test sessions in this experiment, then, asked respondents directly about their own individual activities rather than about whether they had seen a questionnaire item about the activity. In this procedure, of course, we have no absolute knowledge of whether their answers to these questions were correct or incorrect, but data from the questionnaires can provide a partial check on accuracy. Since respondents filled out the questionnaires near the time the target events took place, answers on the questionnaires should be more accurate than answers to the same questions during the test sessions.

The design of this experiment gives us two ways to look at the week-to-week changes in respondents' answers. First, we can examine the transitions as they appear in the test sessions: The percentage of times that respondents said, during these sessions, that they had participated in an activity during week k but not during week $k + 1$ (or the opposite change). These data are analogous to those of the preliminary study and to those of the panel surveys, in that there are separate seam transitions (weeks 4-5, where the data come from different test sessions) and nonseam transitions (weeks 1-2, 2-3, 3-4, 5-6, 6-7, and 7-8, where the data come from the same test session). We plot these data as filled circles in Figure 3, and they show a modest seam effect, with a reliable difference in the number of changed responses from week to week. The second perspective on the changes comes from responses to exactly the same questions on the weekly questionnaires. These data appear in Figure 3 as open circles, and as we might expect, they show no increase in the percentage of changed responses for seam weeks. The procedure here is similar to a hypothetical panel survey that interviewed respondents monthly and asked during each interview about the preceding month alone. In the terms we introduced earlier, the results from the test sessions (filled circles) have different response periods for the interview (four weeks) and for the questions (one week), whereas the results from the questionnaires (open circles) have the same response period for both (one week

in each case). The difference between the two curves in Figure 3 illustrates the effect of separating these response periods.

Once again, part of the seam effect is probably due to memory. If we gauge accuracy by the difference between a respondent's answer in the test session and the answer to the same question on the relevant weekly questionnaire, then we find that accuracy is generally quite high – overall, 86.2% of answers matched. Accuracy in the test sessions, however, decreased with time in a regular way. Accuracy was 91.2% for the events of weeks 4 and 8 (the most recent weeks in the two response periods) and declined to 82.4% for events from weeks 1 and 5 (the earliest weeks in the two periods). There is also evidence for constant-wave responding, but the number of incorrect constant-wave responses appears to be smaller than in our preliminary study. In the first test session, respondents made positive constant-wave responses for 16.2% of the items and made negative constant-wave responses for 54.8%. These figures are high, but they may simply reflect the true proportion of times the respondents had taken part in each of the activities during all four of the preceding weeks or had taken part during none of the preceding weeks. To check for bias, we can again compare these percentages to those for the weekly questionnaires. These data show that respondents had answered "yes" in all four questionnaires for 14.0% of the items and had answered "no" in all four questionnaires for 47.2% of items. Thus, constant-wave responding was only slightly (though reliably) more common in the test session than in the original questionnaires. Results were similar for the second test session: Respondents made positive constant-wave responses for 15.2% of items and negative constant-wave responses for 58.8%, only somewhat higher than the 13.3% positive and 53.8% negative constant-wave responses in the questionnaires.

This experiment suggests that we can detect seam effects for personal events using our procedure. The effect is smaller, though, than those of the preliminary experiment in which we used more artificial items. Respondents' memory for the personal events is much better than for the artificial ones, at least if memory is evaluated relative to answers on the earlier questionnaires. Even after four weeks, accuracy is quite good for the everyday events we used, and this may have decreased respondents' tendency to rely on constant-wave responding and other strategies that could increase the size of the seam effect.[4] This, of course, does not imply that seam effects will also be small in surveys that ask about personal events. The longer response periods of actual panel surveys may decrease memory, and the structure of the survey interview may increase constant-wave responding, as we are about to see. Moreover, some of the questions in panel surveys seek quantitative information rather than the sort of qualitative (yes/no) answers that we have looked at so far. Seam effects may be different when respondents have to come up with a number (e.g, the amount of a purchase or the amount received from a source of income) rather than simply deciding whether or not an event happened. We report one further study of quantitative responses before returning to implications for survey methods.

Study 2: Seam Effects for Quantitative Information

To study seam effects for quantitative information, we used a second variation on our standard procedure. In the new experiment, we again sent respondents weekly questionnaires, but for a period of six rather than eight weeks. Test sessions occurred at the end of weeks 3 and 6. The

schedule was similar to that of Figure 2, then, but condensed by two weeks. The more important difference between the studies concerns the questions we asked in the questionnaires and test sessions. The questionnaire items were all of the type: During the last week..., did you (or someone in your household) spend more or less than $X on Y? Please circle either "More" or "Less" or "Did not purchase," (e.g., During the last week..., did you ...spend more or less than $2 on milk and cream from the grocery or convenience store? During the last week..., did you... spend more or less than $17 on electricity for your home?) We based these questionnaire items on ones that appear in CE. We asked an individual respondent about the same items (e.g., milk and cream, electricity, etc.) on each questionnaire. The specific amounts, however, varied for some items. For half the questionnaire items, we asked about the same dollar amount each week (e.g., respondents might be asked on each questionnaire whether they spent more or less than $2 for milk and cream that week); for the remaining questionnaire items, the amount changed from week to week.[0]

During the test sessions, we asked respondents to recall the dollar amounts they had seen on the questionnaires. For example, one item in the first test session was: On the third week's questionnaire, which you filled out on ..., when you were asked about milk and cream, what was the dollar amount you were asked about? Respondents' answers to these questions provided that data that we used to analyze the seam effects. In addition, we varied the way in which we grouped the questions during the tests. In earlier research (Rips et al., 2000), we had found larger seam effects when respondents had to answer all the question about a given topic one after another, and we were interested in determining whether the same would be true for the quantitative questions in this study. For this reason, half the respondents answered the test questions in an order blocked by item: In the first test session, these respondents answered the question about milk-and-cream for week 3, week 2, and week 1; then they were asked the question about electricity for weeks 3, 2, and 1, and so on. In the second test session, they answered the question about milk-and-cream for weeks 6, 5, and 4; then the question about electricity for weeks 6, 5, and 4; and so on through the full set of items. The remaining respondents answered the test questions in an order blocked by week: During the first test session, these respondents answered all the questions about week 3, then all the questions about week 2, then all the questions about week 1; in the second test session, they answered all the questions about week 6, then week 5, then week 4. Fifty-four adults participated in this study. We recruited them in the same way as before, but none had been in the earlier experiment.

The questions about quantitative information produced clear seam effects. Figure 4 plots these new data. The y-axis of this figure shows the mean absolute value of the change in the dollar amounts from week to week. For example, if a respondent said that the questionnaire for week 1 had asked whether s/he had spent $1 for milk and cream and that the questionnaire for week 2 had asked whether s/he had spent $5 for milk and cream, then the change for this item would be $|1-5| = 4$. Figure 4 shows these absolute changes, both for respondents whose questions were blocked by item (filled circles) and for those whose questions were blocked by week (open circles). It's easy to see that while both ways of grouping the questions produced seam effects, the effect was larger for blocking by item. This agrees with our earlier results for qualitative responses (Rips et al., 2000).

Respondents' accuracy for the amounts was quite low overall: They recalled the correct dollar value for only 12.1% of items. Accuracy also decreased significantly over the response periods, although this decrease was relatively small, probably because of a floor effect. On average,

respondents were correct for 14.4% of items during the most recent weeks of the response periods (weeks 3 and 6) and for 9.2% for the earliest ones (weeks 1 and 4). This decrease was about the same whether the questions were blocked by item or blocked by week during the test sessions.[6]

The more interesting findings, however, concerned constant-wave responding. We counted a respondent as making a constant-wave response during a test session if he or she gave the same answer (dollar amount) each time we asked about an item. For example, if a respondent said during the first test session that he or she was asked about spending more or less than $2 for milk and cream on the week 1 questionnaire, $2 on the week 2 questionnaire, and $2 on the week 3 questionnaire, this was scored as a constant-wave response. (A nonresponse on all three questionnaires was not scored as constant wave.) In these terms, respondents made constant-wave responses to 36.0% of the items during the test sessions. As we noted earlier, the correct amount was actually constant from week to week for half the items and varied for half; so a constant response was appropriate for the former items and incorrect for the latter. Respondents' answers, however, were not greatly different for these two item classes. When a constant response was the correct answer, respondents gave constant answers for 44.2% of items; when a constant response was incorrect, they gave constant answers for 33.8% of items. The number of constant-wave responses did, however, depend on whether the questions were blocked by week or blocked by item. When blocked by item (e.g., all questions about milk-and-cream appeared together in the test), 51.4% of responses were constant wave. But constant-wave responses decreased to only 18.5% when the items were blocked by week (e.g., all questions about week 3 appeared together).

These results suggest that grouping questions about the same topic encourages respondents to give the same answer to each item. If consecutive questions ask respondents about the amount for milk-and-cream in week 3, milk-and-cream in week 2, and milk-and-cream in week 1, it's tempting for these respondents to give exactly the same answer each time. Placing these questions in different parts of the test, as we did when questions were grouped by week, greatly decreases this tendency. This difference clearly contributes to the larger seam effect when questions were grouped by item rather than by week, as seen in Figure 4. In earlier research (Rips et al., 2000), we had also obtained greater accuracy when questions were grouped by week than when the same questions were grouped by item. This was not true in the present study: Respondents were correct on 10.9% of questions when these questions were blocked by week and on 13.1% when blocked by item (a small but marginally significant reversal). This difference between experiments may be due to the fact that in the earlier study none of the items had correct answers that were constant across weeks, while in this study half had correct constant answers.

Simulations of the Seam Effect

The exact form of the seam effect differs in these studies, probably as the result of the relative contributions of memory, constant-wave responding, and other factors. To see why, consider respondents in a SIPP-like survey who are faced with yes/no questions about whether they received some benefit. Respondents' memory for the event will be most accurate for the periods just preceding the interview, declining through the response period, probably at a decreasing rate. If the tendency toward constant-wave responding is moderate, this will create sizeable changes in

responses for the months just preceding the interview (when memory is fading most rapidly). It will also produce a sizeable seam effect, since the seam data come from the most recent month of the first interview (when memory is strongest) and the earliest month of the second interview (when memory is weakest). The result will be asymmetric curves, such as the ones in Figure 5a, which come from simulations based on the assumptions just outlined.[0] The figure shows that the degree of asymmetry in the curves – the amount by which the right-most point on the curve is higher than the left-most point – depends on the strength of memory for the target events. If memory for the event is initially weak (front part of the figure), then the asymmetry is relatively mild, whereas if memory for the event is initially strong (rear part of the figure), the curve is much more clearly asymmetrical.

Respondents' tendencies to make constant-wave responses can also affect the global shape of the function. As an extreme case, if memory is negligible for the events in question and if respondents always make a constant wave response, then the function will be perfectly symmetrical, as shown at the front of Figure 5b. The only opportunity for changing a response in this situation occurs for cases in which a respondent makes one constant-wave response (e.g., "yes") during the first interview, and a different constant wave response ("no") for the second. If respondents are less willing to make constant-wave responses, the asymmetry will increase accordingly, as Figure 5b also shows.

These simulations suggest that the asymmetrical curve from our first study (filled circles in Figure 3) may have been the result of the respondents' fairly accurate memory for the everyday, personal events we asked about. The curve shows the rise in the middle and end that we see in Figure 5a. Asymmetries also appeared in our second study when we grouped questions by week (open circles in Figure 4). These asymmetries largely disappeared, however, when we grouped questions by item (filled circles in Figure 4). Grouping by item probably encourages constant-wave responding, increasing the symmetry of the curve. The differences between the two curves in Figure 4 are similar to the differences in the curves of Figure 5b, where we have deliberately varied the underlying rate of constant-wave responding. SIPP also groups questions by item – for example, asking all the questions about receiving one type of benefit before asking about other types – and this may help account for the symmetric curves in the SIPP data that we glimpsed in Figure 1.

Summary and Implications for Surveys

All the studies we have conducted to date have obtained seam effects – larger changes in responses when the data come from two different interviews than from the same interview. In most of these studies, the key questions that produced the seam effect concerned information that we had supplied. However, the first experiment we reported here extends our finding to naturalistic events, similar to those in actual surveys. This study also compares a situation in which the response period of the interviews is the same as that of the questions to one in which the interviews' response period is longer than that of the questions. As Figure 3 illustrates, only the second type of schedule produced a seam effect. The seam effect is the result of economizing on the number of interviews: By interviewing every four months and asking questions about each month in the preceding interval, these surveys produce response errors that would probably not have occurred if the interviews had been conducted every month.

The studies we described here also show that seam effects appear both for questions about quantities (amounts paid for goods in this case), as well as questions about the occurrence or nonoccurrence of events. In addition, the size of the seam effect for both quantitative and qualitative information depends on question order. When respondents receive questions about the same content for one temporal interval after another, it's easy for them to give the same answer to each item in the series. These constant-wave responses, in turn, increase the seam effect.

Our data show that separating questions about the same topic can reduce the size of the seam effect. The results are not so clear about the effect of this manipulation on accuracy. As we mentioned earlier, the outcome on accuracy may depend on the pattern of events that the survey questions target: Separating questions about the same topic may be helpful when the true answers vary from one response interval to another. It may be of less help when true answers are in fact constant across intervals. We believe similar caution is probably warranted for other methods for reducing seam effects. We can probably reduce seam effects by counteracting biases in respondents' answers, such as the constant-wave tendency, but we need to be careful that in doing so we don't also introduce other sources of error.[8]

For example, SIPP has begun dependent interviewing to help reduce the size of the seam effect. In the second and later interviews, respondents are told about the information they provided in the previous interview before they answer related questions about the current response period (e.g., Last time you said you had job X. Do you still hold that job?). It seems likely that dependent interviewing can smooth seam transitions by reminding respondents of their previous answers. In some cases, this could also provide a memory prompt for information that respondents might not otherwise remember. In other cases, though, giving respondents their own earlier answers may simply bias them toward giving the same answer in the current round of questions, providing an anchor for their judgments (Wilson, Houston, Etling, & Brekke, 1996). Although this would minimize the seam effect, it could lead to equally inaccurate responses. We need to check empirically in each case to see whether reducing the seam effect does more harm than good.

References

Jabine, T. B., King, K. E., & Petroni, R. J. (1990). Quality Profile, Survey of Income and Program Participation. Washington, D. C.: Bureau of the Census.

Kalton, G., and Miller, M. E. (1991). The seam effect with social security income in the Survey of Income and Program Participation. Journal of Official Statistics, 7, 235-245.

Marquis, K. H., & Moore, J. C. (1989). Response errors in SIPP: Preliminary results. Proceedings, 1989 Annual Research Conference, U. S. Bureau of the Census. (Arlington, VA, March 19-22, 1989). Washington D.C.: U. S. Bureau of the Census.

Rips, L. J., Conrad, F. G., & Fricker, S. S. (2000). Unraveling the seam effect. Proceedings of the Section on Survey Research Methods, American Statistical Association. Alexandria, VA: American Statistical Association.

Ruben, D. C., & Wetzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. Psychological Review, 103, 734-760.

Shum, M. S., & Rips, L. J. (1999). The respondent's confession: Autobiographical memory in the context of surveys. In M. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), Cognition and survey research (pp. 95-109). New York: Wiley.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response. Cambridge, England: Cambridge University Press.

Wilson, T. D., Houston, D. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. Journal of Experimental Psychology: General, 125, 387-402.

Young, N. (1989). Wave-seam effects in the SIPP. Proceedings of the Section on Survey Research Methods, American Statistical Association. Alexandria, VA: American Statistical Association.

Footnotes

1. NSF Grant SES-9907414 supported the research reported here. We are grateful to Jami Barnett for her help with these studies. We also thank audiences at the American Statistical Association and the Washington Statistical Society for comments on previous reports about this project. Gordon Willis and Elizabeth Martin commented on this paper, and their remarks will appear along with it in the published version. In fairness to these discussants, we have left the text of the paper in the form that they read it; however, we have added several footnotes addressing a few of the problems the discussants raised. The new footnotes appear in angle brackets. Correspondence about this paper should be sent to Lance Rips, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208.

2. <We note that seam effects can be quite large. For example, Kalton and Miller (1991) present SIPP data showing that 98.3% of respondents report no change in social security benefits from one month to the next when the months are off-seam; however, only 34.4% report no change across seam months. Seam effects also appear in a wide range of variables, including such important characteristics as employment status and total family income (Young, 1989). Of course, whether survey researchers need to worry about these differences depends on their purposes. But those who use panel surveys to make inferences about changes (e.g., changes in social security or food stamp benefits) need to be cautious, unless overestimates of changes between seam months exactly balance underestimates between nonseam months.>

3. Memory may still have a role to play, however, in explaining the positive bias. Respondents had seen all items prior to the test sessions – although, of course, not on each week – so the items may have seemed familiar to them. "Yes" responses based on familiarity could explain positive constant-wave answers.

4. < Figure 3 shows that the test-session data underestimate the number of changes both across seam and nonseam months. This finding contrasts with the results of our preliminary study (Rips et al., 2000) and with results from SIPP (Marquis & Moore, 1989). In the earlier work, seam months produced overestimates of the number of changes, whereas nonseam months produced underestimates. The difference between the present study and the earlier ones is probably due to the particular distribution of the personal events we tested here. For example, there were 49 cases in which an event occurred to a respondent in both weeks 4 and 5, but the respondent failed to report it for week 5. This type of error might be due to forgetting and would serve to inflate the seam change. However, there were also 61 cases in which an event did not occur in week 4, did occur in week 5, but was not reported for either week. This could again be due to forgetting, but it would serve to deflate the seam change. Because of the larger number of events of the second type, incorrect reporting tended to produce too few changes at the seam. Underestimates were less severe, however, for seam months than for nonseam months.>

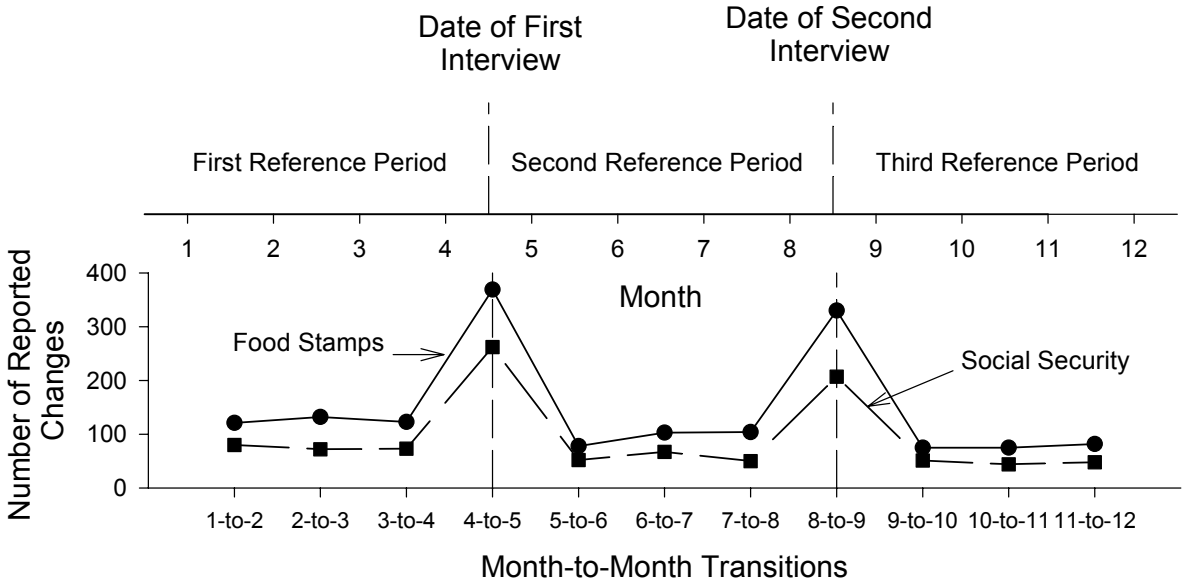
. 5. The amounts for these variable items changed according to four patterns. One group of items increased in amount for weeks 1-3 and increased again for weeks 4-6; a second group increased for weeks 1-3 and then decreased for week 4-6; a third group decreased, then increased; and a fourth group decreased then decreased. For example, if the item about milk and cream was in

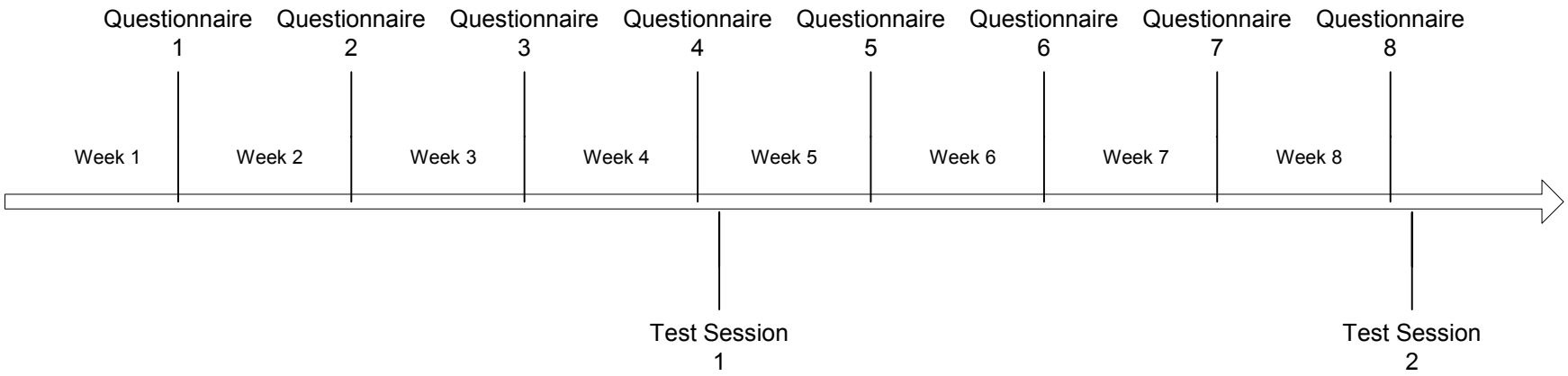
the increase-increase group for a specific respondent, that respondent was asked on the first questionnaire, Last week, did you spend more or less than $1 for milk or cream...?, on the second questionnaire, Last week, did you spend more or less than $2 for milk or cream...?, and on the third questionnaire, Last week, did you spend more or less than $4 for milk or cream...?. This same sequence was then repeated for weeks 4-6. A respondent saw an equal number of items in each of the four groups. Individual items were rotated through the groups across respondents.

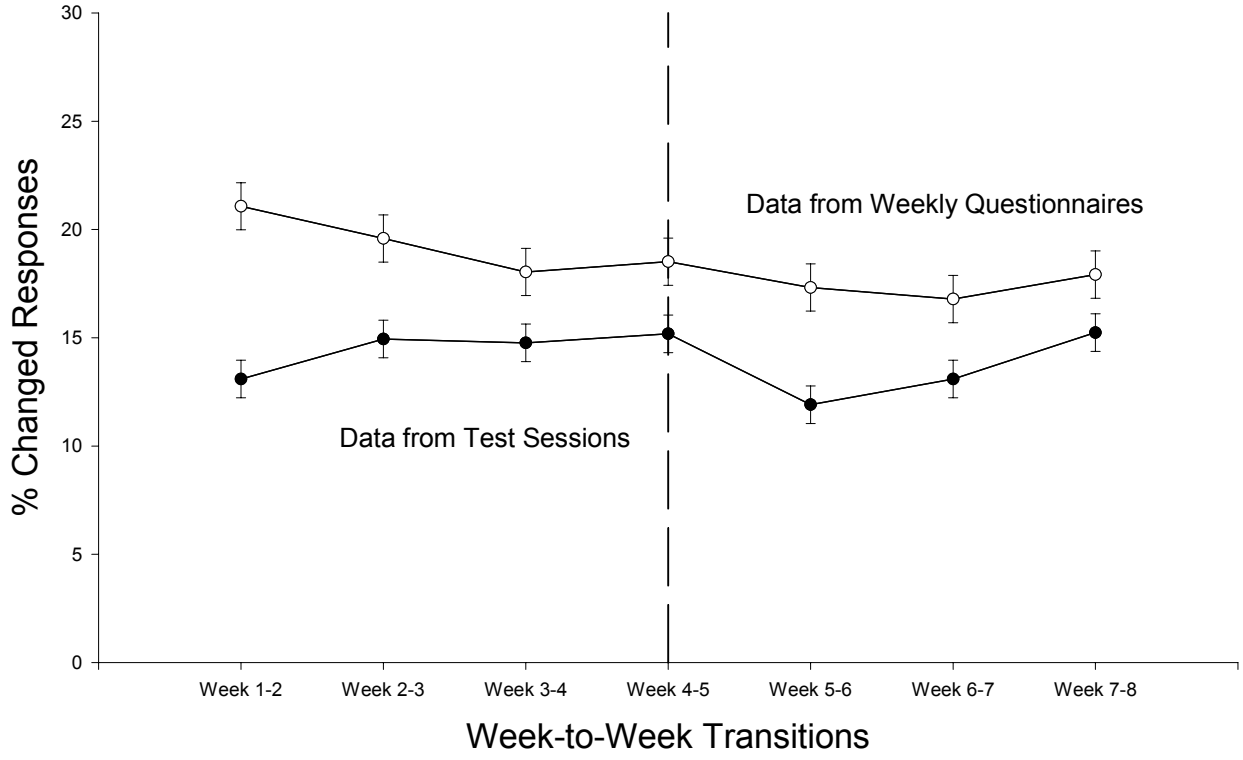
6. <These accuracy figures indicate whether the reported amount exactly matched the amount on the questionnaire. We can construct a more sensitive measure of accuracy by calculating the average absolute deviation (in dollars) between the recalled amount and the true amount. For weeks 3 and 6, the weeks just before the test sessions, the average deviation was $5.88. For weeks 1 and 4, the earliest weeks, the average deviation increased to $6.53. This difference again supports the hypothesis that respondents were forgetting the correct amounts over the reference period. Of course, when respondents no longer remember the amount, they can use a variety of strategies to produce an answer, including constant-wave responding, estimating a usual value for the product or service, or even sheer guessing. The point of the present paper is that the seam effect depends both on forgetting and on the nature of these alternative strategies, as we attempt to show in the following section of this paper.>

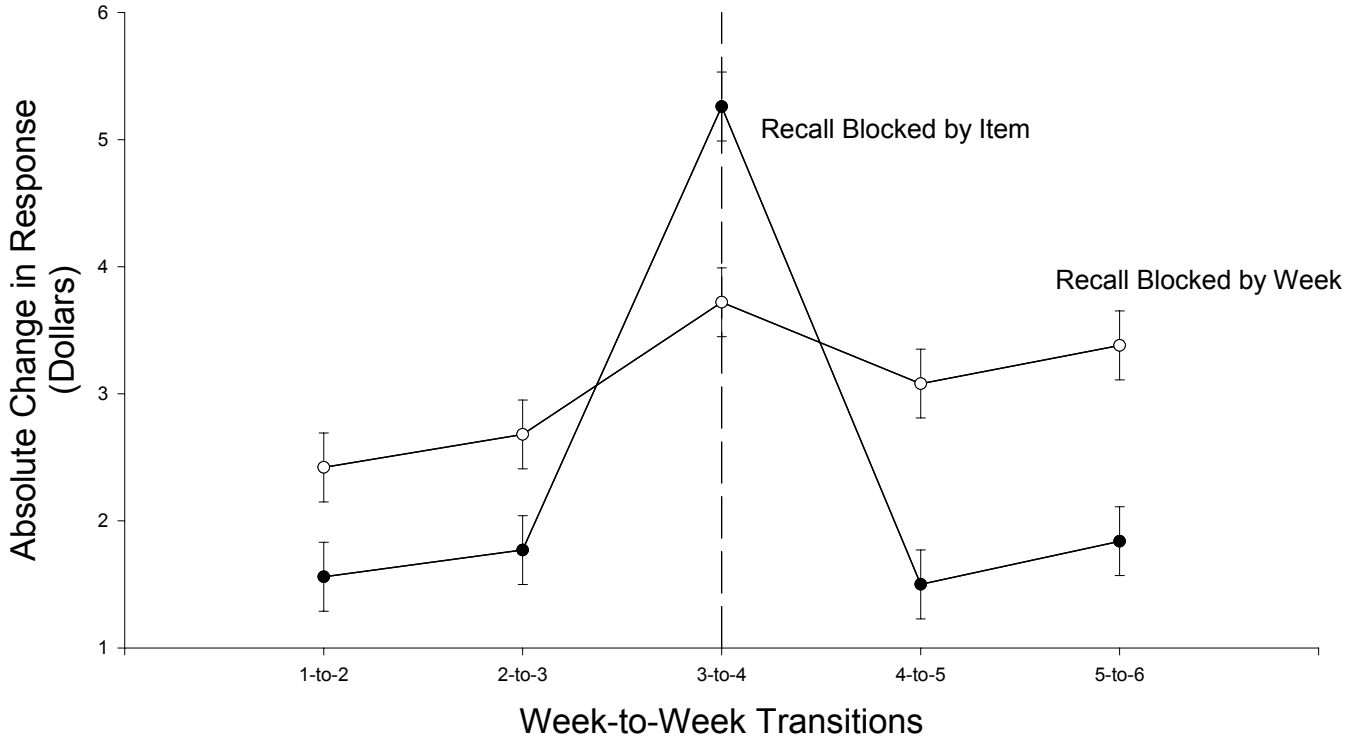
0. 7. For purposes of these simulations, we assumed that forgetting followed a negative logarithmic function of time ($y = a - b*\ln(t)$), in line with earlier work on long-term forgetting (Rubin & Wetzel, 1996). The exact form of the function is probably not crucial, however, as long as memory decreases at a steep rate at first and more gradually thereafter. We also assumed that when a particular piece of information is forgotten, a respondent can rely on one of two strategies. First, the respondent can interpret forgetting as negative evidence (i.e., failure to remember indicates that the event in question didn't occur), and answer "no." Second, the respondent can make a constant-wave response, answering "yes" for all earlier intervals or "no" for all earlier intervals in the response period. The functions in Figure 5 then depend on three parameters: the forgetting parameters (a and b in the equation above), and the probability of a constant wave response. Figure 5a varies the first of these parameters, and Figure 5b varies the third.

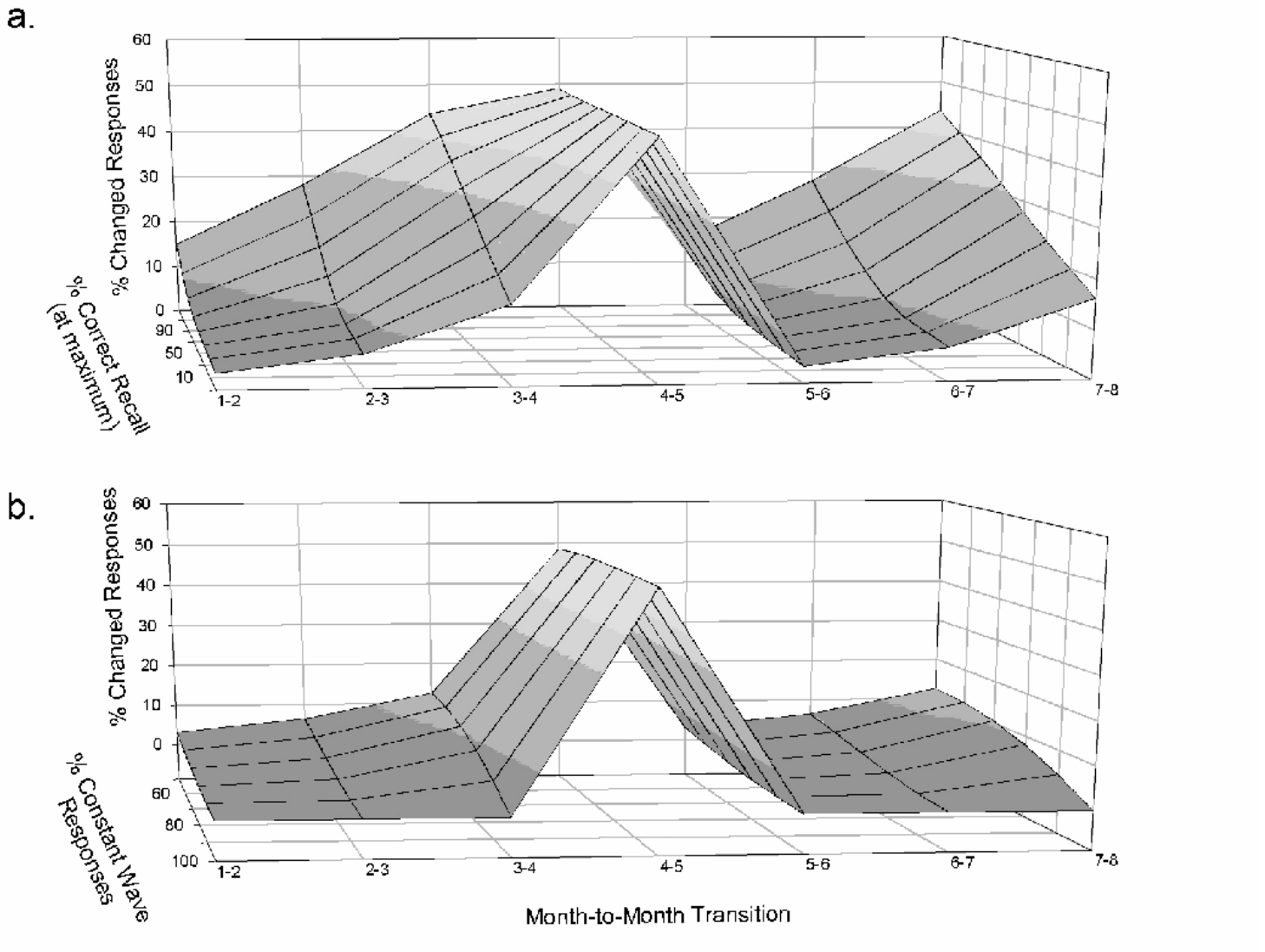
8. <It may seem disappointing that the size of the seam effect isn't a clear indicator of overall data accuracy. However, the seam effect depends on the variability of responses from month to month, and we shouldn't expect variability (or stability) to correlate perfectly with the responses' correctness. We hope that the technique we are developing here can serve as a useful way of finding out how proposed methods for reducing the seam effect will impact other aspects of data quality. As such, methodologists could use the technique alongside other procedures, such as cognitive interviewing, for anticipating problems in actual surveys.>


Date of First Interview
Date of Second Interview
First Reference Period
Second Reference Period
Third Reference Period
1
2
3
4
5
6
7
8
9
10
11
12
Month
Number of Reported Changes
400
300
200
100
0
Food Stamps
Social Security
1-to-2
2-to-3
3-to-4
4-to-5
5-to-6
6-to-7
7-to-8
8-to-9
9-to-10
10-to-11
11-to-12
Month-to-Month Transitions


Questionnaire 1
Questionnaire 2
Questionnaire 3
Questionnaire 4
Questionnaire 5
Questionnaire 6
Questionnaire 7
Questionnaire 8
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Test Session 1
Test Session 2


30
25
20
15
10
5
0
% Changed Responses
Data from Weekly Questionnaires
Data from Test Sessions
Week 1-2
Week 2-3
Week 3-4
Week 4-5
Week 5-6
Week 6-7
Week 7-8
Week-to-Week Transitions


Absolute Change in Response
(Dollars)
6
5
4
3
2
1
Recall Blocked by Item
Recall Blocked by Week
1-to-2
2-to-3
3-to-4
4-to-5
5-to-6
Week-to-Week Transitions


a.
% Changed Responses
60
50
40
30
20
10
0
% Correct Recall (at maximum)
90
50
10
1-2
2-3
3-4
4-5
5-6
6-7
7-8
b.
% Changed Responses
60
50
40
30
20
10
0
% Constant Wave Responses
60
80
100
1-2
2-3
3-4
4-5
5-6
6-7
7-8
Month-to-Month Transition

Rips, Conrad, Fricker, and Behr
Seam Effects for Quantitative and Qualitative Facts

Discussion

Gordon Willis, National Cancer Institute

In general terms, Rips, Conrad, Fricker, and Behr are to be commended for their careful description and analysis of the Seam Effect in panel surveys, and for the way in which they tie this conceptually to "Constant Wave Effects" operating at individual interview points. At a more specific level, my comments begin with a basic question: Is the Seam Effect a worthy protagonist in a research study that endeavors to point the way toward reduction in response error through investigation of the cognitive aspects of survey responding? Potentially, such a study represents an attempt to examine a problem observed in a particular survey environment (here, the Survey of Income and Program Participation, or SIPP), to determine whether there are consistent and modifiable sources of measurement error that reflect the operation of comprehension, recall, and decision processes. Further, to the extent that the Seam phenomenon is pervasive, such research could possibly elucidate cognitive processes that extend to a variety of surveys. Therefore, a vital consideration I pose in evaluating the Rips, et al. investigation is whether the study investigates a phenomenon that a) is non-trivial, b) extends to a range of surveys, and c) is similar in nature to related problems of a cognitive nature that afflict surveys.

Is the Seam Effect an "interesting" phenomenon?

As to whether the Seam Effect is typically of sufficient magnitude to be considered a problem, as opposed to simply a minor source of noise in an otherwise meaningful data distribution, the answer is provided in a separate review by Tourangeau, Rips, and Rasinski (2000), who convincingly portray the sometimes large magnitude of the Seam phenomenon – in particular, the number of changes in status reported between seam months may be as much as 12%, as opposed to a base level of 1-2% between non-seam months.

The Seam Phenomenon also appears to extend beyond the SIPP, as Seam Effects have emerged in the Panel Survey of Income Dynamics, and the Income Development Survey Program. The general paradigm studied conceivably has application to surveys other than those involving income and program participation. For example, health surveys involving cancer risk factors sometimes ask, at one interview point, for reports of status for a series of previous sub-intervals, such as usual weight during a ten-year period, or self-report of male sexual function following treatment for a number of months following prostate cancer. Such procedures do satisfy the procedural requirements necessary, in theory, to produce Constant Wave Effects within interviews, or Seam Effects over repeated interviews. I would therefore propose that a solution to the Seam Effect puzzle is potentially important enough to warrant a systematic examination, especially to the extent that the results are generalizable.

Attempts to generalize the Seam Effect

However, the Rips et al. attempt to directly generalize the Seam Effect to domains beyond that of income and program participation is risky, if this extension is carried out without sufficient consideration of how the cognitive aspects of different behaviors may vary. In particular, the attempt to cover everyday activities such as ordering pizza and changing the oil in one's car, under the Seam Effect umbrella, is problematic. The Seam Effect is hypothesized by the authors to be mediated by Constant Wave Effects, in which the report for an initial sub-interval serves as an anchor that in turn diminishes respondents' tendencies to vary their reports concerning other time intervals. However, Constant Wave effects seem much more likely to occur for some behaviors than others, in fairly predictable ways. For example, if someone changed the oil in a vehicle in one month, it is very unlikely he or she did so the next month as well, so that one might expect, in anything, a "negative wave effect" for that item (so that a positive report in one month would *decrease* reporting of the same behavior in surrounding months). I suggest that the authors need to consider the nature of the behavior studied, and how this may impact the potential for Seam Effects, as they attempt to extend its reach beyond reporting on items such as the receipt of social security benefits and employment.

Does reducing the Seam Effect translate into lower response error?

A key rationale for studying the Seam Effect seems to be that, if this effect is a symptom of cognitive problems, then factors that reduce the magnitude of the Effect will also be likely to improve data quality. For example, in a previous paper reporting on an initial experiment done as part of this research study, Rips, Conrad, and Fricker (2000) found that the use of backward-to-forward temporal ordering of recalled events reduced the magnitude of the Seam Effect, relative to forward ordering, and this finding suggested that the backward order alleviated error. Such a result is important from a methodological perspective, because it potentially provides a proxy measure of survey data quality (the size of the Seam Effect), obviating the need for direct measurement through more expensive procedures such as individual level response validation.

However, the status of Seam Effect-as-proxy-for-quality has become very murky, given the further research that Rips et al. have carried out as part of the current investigation. In particular, the authors find that in some cases manipulations that reduce the size of the Seam Effect may also reduce rather than improve response accuracy. For example, asking a series of survey questions in which related behaviors were organized by time, rather than by topic, did "break up" the tendency for Seam Effects to occur for each topic, but also adversely affected respondents' reporting. As an extreme example, it may be possible to eliminate a Constant-Wave-based Seam phenomenon through the use of random ordering of questions, preventing one response from becoming an anchor for others on similar topics. However, it is doubtful that such a practice would actually result in improved reporting, as this would serve only to treat a possible symptom (the Seam) but not the underlying disorder (response error).

So, it seems that the Seam Effect is not in itself necessarily a direct measure of data quality, and the authors conclude that: "We need to check empirically in each case whether reducing the Seam Effect does more harm than good." However, the need to check in each case (that is, for each new survey or new question environment) effectively nullifies the primary advantage we have for assessing the presence of a Seam Effect in the first place; if its magnitude cannot be mapped

directly to data quality, then the fact that it may be modifiable is perhaps interesting, but of very limited practical utility to survey researchers. In order to move this research forward, one would perhaps need to determine the types of situations in which Seam Effects indicate good versus poor quality data. However, pursuing this road further risks a type of reductionism in which the phenomenon under study is increasingly narrowed in scope so that it retains only academic interest, and is found to be too complex and multivariate to have the type of generality that made it appear attractive initially.

Implications of the Rips et al. study

Despite the limitations of the research, I believe that there are a number of interesting implications that are not directly stated, but that stem from the overall findings:

1. The study of general cognitive processes. The fact that the Seam Effect at the outset appeared to be somewhat straightforward in basis and amenable to study appeared to render it a good candidate as an application of CASM (Cognitive Aspects of Survey Methodology) in which a general law of cognitive functioning as related to questionnaire design could be explicated as the result of the research effort. That is, the intent is to demonstrate that Seam Effects are produced by cognitive mechanisms that can be modified in predictable ways by particular design modifications, that employing these design rules will ameliorate the Seam Effect, and that as a byproduct, survey data quality improvements will be realized. The fact that studying this effect did not lead to such a generalizeable rule, such as "reverse temporal ordering of recall is superior to forward ordering" is consistent with other (somewhat frustrating) failures to produce generally applicable rules of questionnaire design through experimental cognitive study. Thus, questionnaire design and evaluation practice continues to be largely an empirical issue, where the factors that impact on design decisions related to a particular survey instrument are complex, and represent the mutual effects of a number of opposing considerations. This is perhaps why practitioners continue to evaluate questionnaires through empirical techniques such as cognitive interviewing, as opposed to simply relying on a bible of design rules. I do not argue that design rules are useless – simply that they must be regarded as a general starting point, and are insufficient in themselves when we partake of questionnaire design "down in the trenches."

2. Survey responding is problem solving. The point has been made many times that survey respondents do not simply directly output information from memory, as these memories are queried by our survey questions. Rather, in the face of partial and difficult-to-retrieve information, respondents make use of processes such as complex estimation, background knowledge of probability, and other heuristics, in order to produce responses that they feel are reasonable. The detailed study of the Seam Effect by Rips and colleagues further demonstrates this effect. When deciding on how to answer a string of questions over sub-intervals related to the same behavior, respondents consider issues such as the likelihood of an event or behavior occurring in month N, given that it had (or had not) occurred in month N-1. In addition to direct retrieval of specific memories, processes such as knowledge of regularity, frequency, and patterning of particular behaviors (whether ordering pizza, changing the oil, or receiving social security benefits) drives the process by which the respondent attempts to reach a suitable level of accuracy under conditions in which memory itself is insufficient. We must continue to be reminded that answers to survey questions are often not so much reported from storage as they are synthesized on the spot from a variety of information sources.

3. Respondent consistency. The finding of a Constant Wave Effect within interview, but sometimes virtually random perturbation in response between interviews, also has two important implications:

a) Consistency of responses within interview is no assurance of response accuracy. Survey designers sometimes are led to believe that their survey questions "work" because no obvious problems arise, and the responses to related items do not illustrate gross inconsistency. However, a generalization of Constant Wave Effects may be "Consistent Answer Effects" in which respondents strive to maintain a coherent picture ("Because I said X to the previous question, now I better answer Y"). So, especially during pretesting and evaluation, it makes sense to delve into the basis for each answer, rather than simply accepting a seemingly solid and consistent facade.

b) On the other hand, we should not necessarily expect great consistency *between* interviews, even for information which should not have changed between interview. This particular issue arises perennially when we conduct reinterview studies to assess question reliability, or where a longitudinal study requires an answer to the same question at multiple time points. A consistent concern is that at time T2, the respondent is simply recalling the answer he/she gave at time T1, rather than recalling the answer anew. However, if respondents' behavioral tendencies with respect to the Seam Effect can be used as a guide, then perhaps those worries are unfounded; if respondents are not even consistent when we *do* want them to be, then perhaps they also are not attempting (or able) to be consistent when we *don't* want them to be.

To return to the initial question posed – Why study the Seam Effect under the rubric of CASM research? – Perhaps the answer is not that this will provide a means for reducing error by finding ways in which to ameliorate this effect, but rather, that it provides a rich source of data indicating how respondents make decisions as they answer survey questions, specifically about a series of past sub-intervals. The fact that they may be inclined to engage in Constant Wave behavior, when the behavior is viewed as likely to be constant in nature, sensitizes us to the need to emphasize the veracity of the initial response reported, and leads us to consider whether it is even advisable to request information from the respondent that may be severely tainted by other reports they have just given. As a means for investigating this phenomenon further, I advocate additional research which attempts to directly determine the effects of economizing survey reporting by obtaining monthly (or other periodic) information on a less-than-monthly basis. In particular, by comparing the results of a procedure in which some respondents actually are asked the repeated questions (on program participation, ordering pizza, etc.) monthly, and others are asked for the same information, but periodically (e.g., quarterly), we can determine the direct effects of the use of the latter procedure.

References

Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge Cambridge University Press.

Rips, L.J., Conrad, F.G., & Fricker, S.S. (2000). Unraveling the Seam Effect. *American Statistical Association, Proceedings of the Section on Survey Research Methods*

Remarks on "Seam Effects for Quantitative and Qualitative Facts"

Elizabeth Martin
U. S. Census Bureau

It may be useful to begin with a brief recapitulation of the research you've just heard reported. Rips and his colleagues are attempting to reproduce or simulate the seam effect in the lab. To do this, they in effect miniaturize everything--months are transformed into weeks, and the four month wave is transformed into a four (or three) week wave. They ask respondents about two different kinds of events. First, every week they send questionnaires to be filled out, then at the end of four weeks they interview respondents about what questions appeared in the questionnaires. Let's call this "questionnaire recall." Second, they also ask respondents--both in the questionnaires, and in the end-of-wave interview sessions--about ordinary events that may have happened during each week of the "wave." Let's call this "ordinary event recall." "Transitions" are measured as week-to-week changes in respondents' reports. So, if a respondent said a particular question appeared in one of the weekly questionnaires but not the next one, that counts as a transition. For ordinary events, if a respondent said a particular event happened one week and not the next, that counts as a transition. For both types of events, they have measures of truth. For the first, they know which questionnaires they sent out, so they know which questions were in fact asked each week. For the second type of event, they have the responses to the weekly questionnaires as a check on the accuracy of respondents' reports about the same events at the end of the four week "wave."

I think there are three questions we need to ask about the research.

- First, have they produced seam effects in the lab?
- Second, does their laboratory version of the seam effect reproduce or match what we know about essential features of the survey phenomenon?
- Third, if we are satisfied that their laboratory simulation reproduces the survey reporting phenomenon in critical ways, what light does their research shed on the cognitive underpinnings of seam bias?

I would answer the first question with a skeptical "maybe." Despite the authors' interpretations, I do not see evidence of a seam effect for the second type of "ordinary event recall." Compared to the weekly questionnaires which serve as the measure of truth, the test sessions produce consistently lower estimates of week-to-week changes in responses. But we know from record check studies of income reporting that, compared to truth, the number of transitions at the seam is too high, and the number of within-wave transitions is too low (Moore and Marquis, 1989). Hence we should observe the lower line spike up above the top "truth" line at the seam, but it doesn't.

For the other type of "questionnaire recall" the results do seem to show a seam effect which is very much affected by the structure of questioning during the test session interview. For example, respondents were asked in the test session to recall the dollar amounts asked about in questionnaire items for each of the weeks of the "wave." When the recall questions were organized by week, the seam effect is slight; when organized by topic (i.e., respondents were asked to recall the dollar amount referred to in a particular question in each of the weekly questionnaires) it is very large.

The authors theorize that the seam effects in their lab studies, and in surveys such as SIPP, are

produced by the combined effect of recall for events in relatively recent time periods and estimation of events in distant time periods, which respondents do not remember well.

But are Rips and his colleagues measuring recall in their laboratory simulation? The events for which they find seam effects--recall of questions asked in questionnaires--are very ephemeral and inconsequential, in contrast to receipt of income, say. The authors do not provide evidence to support the premise that any recall at all is involved in this reporting task. It seems plausible that the task of "recalling" which of eight or four weekly questionnaires contained a particular item is pure guesswork, and that the seam effects they produce are a consequence of artificial constraints upon the consistency of guesses across the weekly time periods. Answers to the following questions would shed light on whether respondents are engaged in recall or guessing:

1.) What fraction of correct responses should be expected by chance? The test sessions only asked about events (questionnaire items) that really had appeared in the weekly questionnaires, and the results suggest respondents were biased toward positive answers to the questionnaire recall questions. It might be useful to include in the test sessions questions that ask respondents to "recall" items that never appeared in any questionnaire, to learn whether respondents are as likely to say they saw a questionnaire item that never appeared in any questionnaire, as one that did. If so, then it's difficult to interpret the results as being about something other than guessing.

2.) Were respondents given the option of responding "don't know" to the questionnaire recall questions, and what fraction did so (or did not respond)?

3.) Did the researchers conduct any debriefings or think-alouds with respondents to learn how they attempted to solve the questionnaire recall task?

4.) What is the correlation between respondents' reports and truth? (I suspect it is close to zero.)

If the questionnaire recall task does not involve recall, then the answer to the second bulleted question above is "no," because the laboratory version of the income reporting task does not match what is known about the survey phenomenon. Income reports may be characterized by a good deal of error, but no one doubts that income receipt is memorable and that reporting income involves recall.

In seeking to reproduce the seam bias phenomenon, it would be useful to review what is known or suspected about the seam bias as it affects income reporting. I was surprised the authors had not done this. We have a good deal of evidence about the seam bias from the record check studies of the Survey of Income and Program Participation (SIPP), conducted by Kent Marquis and Jeff Moore. That research suggests that:

1.) The seam bias appears to involve both underreporting of true changes within a wave, and overreporting of changes between waves (Moore and Marquis, 1989).

2.) Reporting accuracy (i.e. low underreporting error) varies a good deal by program, but the seam effect turns up even for very accurately reported events, such as Social Security income (see Marquis and Moore, 1990).

3.) It is not clear that more recent events (e.g., income receipt one month ago) are more accurately reported than more distant events (e.g., income receipt four months ago) (see Marquis and Moore, 1990; table 4.2). This finding is counterintuitive, and other research suggests that recall for income receipt does deteriorate over time. For example, Kalton and Miller (1991) find that a one-time

increase in Social Security payments was less likely to be recalled and reported the longer the time interval between the occurrence of the increase and the interview date. The possible role of memory decay in producing the seam bias is an important question for research. The authors beg the question by assuming that memory decay produces better recall for recent events than for more distant ones, taking the evidence of a seam effect as support for this explanation. However, any factor or process that increases the consistency of reporting across weeks within a wave, and/or that reduces the consistency of reporting between waves, could produce a seam effect. It would be useful to contrast their hypothesis with alternative hypotheses about the underlying cognitive processes that may account for the seam bias phenomenon. Existing research has implications for theorizing about the cognitive processes underlying seam effects, and should be taken into account in research on the topic. Results of methodological studies which have attempted to correct the seam bias (see, e.g. Moore, Marquis, and Bogen, 1996) are also pertinent and should be considered in developing cognitive theories of the seam effect and proposing solutions for it.

The authors need to reexamine the task they are using in their laboratory simulation, which should more closely mimic the survey task that gives rise to the seam bias phenomenon. The character of the events being reported about in surveys such as the SIPP is quite different from the questionnaire events for which the authors find seam effects in their laboratory simulation, and the differences almost certainly affect the recall strategies employed by respondents. Income receipt is, for most of us, pretty memorable and consequential. For most of us, it is temporally regular, and patterned in some way. It depends on external, continuing sets of conditions and life circumstances--having a job, being eligible and enrolled for food stamps or social security, and so on. When these conditions change, then income receipt changes--one loses a job, stops receiving wage income, becomes eligible for unemployment compensation, loses eligibility after a certain number of weeks, and so on. Behind the month-to-month changes in recipiency are real transitions in life circumstances which are meaningful to respondents. The fact that income receipt is associated with meaningful transitions influences the response strategies available to respondents as they report income. They can try to reconstruct the timing of a change in income source or amount using associated life events and changes as anchors and landmarks. (Of course, the fact that such recall strategies are available to respondents does not necessarily mean they employ them, or that they produce accurate reports if they do.) Such strategies are unavailable to respondents in this study, because the events (recall of questionnaire items) are meaningless and the "transition" from one item to another in different weekly questionnaires is completely artificial and arbitrary, from the respondent's point of view. (I find it difficult to imagine strategies for answering these questions other than constant wave responding or random guessing.) Thus, the cognitive processes respondents engage in during the questionnaire recall task seem considerably different from the cognitive processes involved in reporting income. For this reason, the answer to the third bulleted question in my opinion is "no," this research has not (yet) shed light on the cognitive underpinnings of the seam bias phenomenon in surveys. However, this is an interim assessment; by better integrating knowledge from the existing methodological literature, and by reexamining and corroborating their assumptions about the nature of the response task and the cognitive processes that respondents engage in, and by exploring respondents' response strategies more directly, the authors would make useful contributions to our understanding of this difficult survey problem.

References

Kalton, G. and Miller, M. E. (1991) "The seam effect with Social Security income in the Survey of Income and Program Participation." Journal of Official Statistics 7(2):235-245.

Marquis, K. H., and Moore, J., (1990), "Measurement Errors in SIPP Program Reports." Proceedings of the 1990 Annual Research Conference, pp. 721-745.

Moore, J., and Marquis, K., (1989), "Using Administrative Record Data to Evaluate the Quality of Survey Estimates." Survey Methodology, Vol. 15, pp 129-143.

Moore, J. C., Marquis, K. H., and Bogen, K. (1996), The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation. Washington DC: Census Bureau.