UNRAVELING THE SEAM EFFECT

Lance J. Rips, Northwestern University, Frederick G. Conrad, Scott S. Fricker, Bureau of Labor Statistics
Lance J. Rips, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston IL 60208

Key Words: Seam effect, Response error, Memory

Respondents in panel surveys sometimes have to answer questions about each of a set of intervals within a longer response period. For example, the Consumer Expenditure Survey (CE) interviews its respondents quarterly, but some questions ask about one-month intervals within the quarterly period. Similarly, the Survey of Income and Program Participation (SIPP) interviews every four months, targeting one-month intervals for some questions. The items in (1) and (2) are examples, the first from CE and the second from SIPP:

(1) 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)?

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

The motive for including micro-questions like these might be that monthly figures are more natural than quarterly figures for some questions or that month-by-month data are needed for comparison with other economic indicators. In any case, questions like these lead to a typical pattern of results called the *seam effect*.

To understand the seam effect, imagine an interviewing schedule, such as that in the top half of Figure 1. This is part of the schedule for one hypothetical panel of respondents who are interviewed just after Month 4 and asked about income received in each of Months 1-4. They are interviewed again just after Month 8 about income in Months 5-8, and so on.

The seam effect appears when we plot the changes in responses from one month to the next. For sample question (2), we can find the number of respondents who change their answers, reporting receipt of food stamps in Month 1 but not in Month 2, or in Month 2 but not in Month 1. The same change can be computed for other pairs of adjacent months. These differences look like those in the lower panel of Figure 1. The month-to-month differences are typically small between Months 1 and 2, 2 and 3, and 3 and 4, where the data come from the same interview. But they increase dramatically between Months 4 and 5, where the data come from two different interviews. The Month 4-5 transition is on the "seam" between the two response periods, and the seam effect is the finding that differences are larger at this point than for

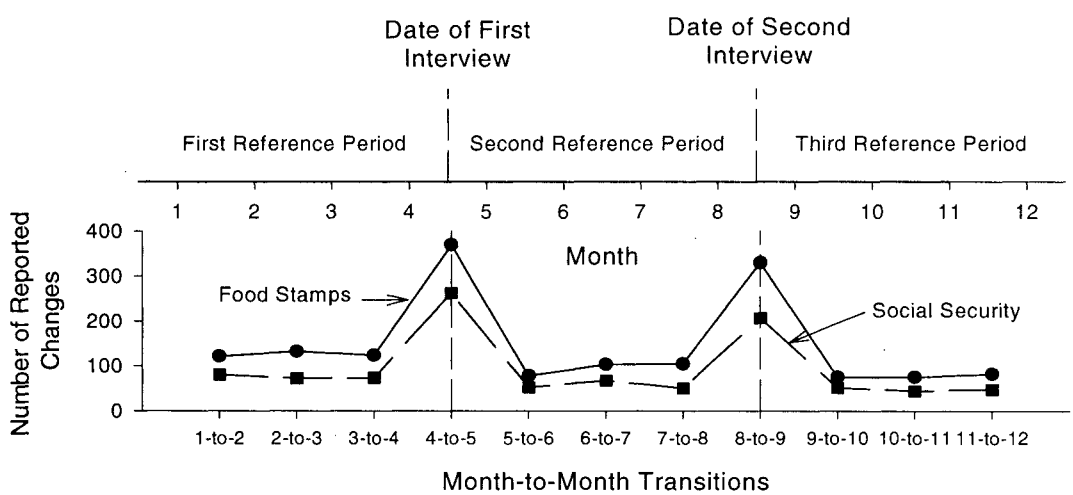

Figure 1. Sample interview schedule for panel survey (upper panel), and changes in reports of receiving food stamps (solid line) and social security (dashed line) from SIPP (Jabine, King, & Petroni, 1990).

the other off-seam months.

The data in the lower part of the figure come from SIPP (Jabine, King, & Petroni, 1990). The solid line shows respondents' reports of receiving food stamps and the dashed line reports of receiving social security benefits over a 12 month period. The number of times respondents changed their answers is clearly much greater between the seam months than between other pairs of adjacent months off the seam.

It's obvious that the seam effect is due to the fact that the data from the seam months come from different interviews while the data from the other pairs of months come from the same interview. The real issue has to do with which factors connected with the different interviews are responsible. In this paper, we report an initial pair of studies intended to address this question. Our method is a hybrid between laboratory studies of memory and survey field tests – a simulated panel survey in which we can control some of the factors that might contribute to the seam effect.

POTENTIAL CAUSES OF SEAM EFFECTS

If respondents have any memory for the queried information, their recall will be best for events that occurred during the most recent portion of the reference periods – for example, the events of Months 4 and 8 in the schedule of Figure 1. Respondents will have more difficulty recalling events that occurred in the earlier parts of the reference period, and we should expect them to base their responses for these early months (e.g., Months 1 and 5) less on memory and more on estimates, guesses, inferences, and other error-prone methods. Let's call this second set of factors *estimates* as a cover term for all such strategies. (See Tourangeau, Rips, & Rasinski, 2000, chap. 3, for a review of memory effects on survey responses, and Smith & Jobe, 1994, for a similar model for dietary surveys.)

Responses at the seam months will then typically depend on different cognitive processes. Responses for Month 4 come from the most recent part of the first response period and may reflect respondents' memory for the events in question. But responses for Month 5 come from the earliest part of the second reference period, and they may depend more on estimates. This means, on one hand, that variables that influence memory, such as salience and distinctiveness, may affect responses for Month 4 but have much less effect on Month 5. On the other hand, variables that affect estimates, such as biases associated with rating scales, are likely to impact Month 5 but not Month 4. The implication of this difference for the seam effect is that when memory and estimates point toward opposite answers, seam effects should be large; but when memory and estimates point to the same answer, seam effects should decrease. This isn't, by any means, a complete theory of the seam effect, but it provides a point from which we can begin to study the effect experimentally.

A METHOD FOR STUDYING SEAM EFFECTS

Our method attempts to model a panel survey using a condensed procedure that takes place over several weeks, rather than several months. In our first experiment, we mailed questionnaires to respondents in each of eight weeks that asked them whether they had engaged in certain activities during the past week. For example, we asked them questions like, "During the last week, did you check out a book from the library?" or "During the last week, did you call a friend or relative long distance?" The respondents filled out the questionnaires and mailed them back to us within 24 hours. At the end of Week 4 and again at the end of Week 8, the respondents also came into the lab, and we tested them on the content of the questionnaires that they had filled out during the preceding four weeks. For example, we asked them during the first test session, "On the 4th week's questionnaire...,was there an item about checking out books?" "On the 3rd week's questionnaire..., was there an item about checking out books?" and so on. The test in Week 8 was the same, except that we asked the respondents about the content of the questionnaires for Weeks 5-8. Thus, the schedule for the experiment was similar to that in the upper panel of Figure 1, but with weeks rather than months as the temporal unit.

In this procedure, the two *test* sessions serve as the analogues of the survey interviews in the panel studies. The test sessions give us data about whether respondents recall seeing a set of items during each of the eight weeks, and we can therefore look at how individual respondents' answers changed from week to week. By looking at the changes between each adjacent pair of weeks, we get the same sort of data that we looked at in connection with the SIPP questions in Figure 1. If we're successful in producing a seam effect, then the number of changed responses should be biggest between Weeks 4 and 5, which are the seam weeks in this experiment.

The critical data come from the test sessions and not from the original questionnaires that we mailed to respondents. The questionnaires function simply as a way to control the information that respondents later report, so that we can check their accuracy. The data we present in the following sections come only from the test sessions. (Of course, it is possible that respondents' answers on the questionnaires influenced their memory for items during the tests. As it turned out, however, respondents were no more likely to say they remembered a particular question if they had previously reported engaging in the named activity than if they reported they had not.)

THE SEAM EFFECT WITHOUT SEAM CHANGES

In our first experiment, the questionnaires that we mailed to respondents contained one set of items during Weeks 1, 2, 7, and 8 and a different set of items during Weeks 3, 4, 5, and 6. Table 1 shows this schedule. For example, the questionnaire for Weeks 1, 2, 7, and 8 – which we will call the *Outside* questionnaire – might have asked the respondents about checking out books from the library, making long-distance calls, and so on, while the questionnaire for Weeks 3-6 – the *Inside* questionnaire – asked about mailing a package at the post office and taking clothes to the laundromat. (The actual content of the questionnaires was balanced across respondents, so that half the respondents saw the item about checking out books on the Outside questionnaire, while the remaining respondents saw it on the Inside questionnaire.)

What's important about this schedule is that, for all respondents, the change in items occurred between Weeks 2 and 3 and between Weeks 6 and 7. There was *no* change to the questionnaire items between seam Weeks 4 and 5. So any difference that we find at the seam is not due to an objective change in the questions. Each questionnaire contained 30 critical items of the sort just described, plus a set of 20 fillers. Respondents simply answered "yes" or "no" for each item on the questionnaire and mailed them back.

During the two test sessions, the respondents came to the lab, and we tested them in a computer-controlled procedure. In the first test session, the computer presented them with a list of all the critical questions they had seen in the last four questionnaires, and they first decided whether they had seen each item on the questionnaire for Week 4 (see Table 1). They pushed one button on a keyboard to indicate "yes, I saw it on the questionnaire for Week 4" and another to indicate "no, I didn't see it on the questionnaire for Week 4." We next presented exactly the same list of items (in a new random order) and asked them to decide whether they had seen each item on the questionnaire for Week 3. They then went through this same procedure for the questionnaires for Weeks 2 and 1. The second test session quizzed respondents on the items from Weeks 5-8; but the two test sessions were otherwise alike.

Table 1. A Schedule for the Events of Experiment 1.

Week	Events
1	Outside Questionnaire
2	Outside Questionnaire
3	Inside Questionnaire
4	Inside Questionnaire First Test Session
5	Inside Questionnaire
6	Inside Questionnaire
7	Outside Questionnaire
8	Outside Questionnaire Second Test Session

We had recruited the respondents from newspaper ads in Washington, D.C. There were 58 respondents in all, with an average age of 40. About half were male and half were female, and they had an average of between two and three years of college.

We expected respondents to rely more on memory in answering questions about the more recent weeks than about the earlier weeks of the response periods. By contrast, they should depend more on estimates in answering questions about the earlier weeks than about more recent ones. In this experiment, respondents had seen all the items prior to the test sessions (though not, of course, in each week); so all items were somewhat familiar to them. For that reason, we expected respondents to be biased toward guessing "yes" (that the questionnaire contained this item on the specified week), when their memory for the questionnaire was weak.

This leads to a prediction about the size of the seam effect. Consider an item from the Outside questionnaire in Table 1. When we ask respondents during the first test session whether they had seen this item on the Week 4 questionnaire, their memory for the questionnaire should be fresh, and they should tend to give a correct "no" answer. When we ask them during the second test session whether they had seen the same item during Week 5, their memory for that questionnaire should be weak, and they should tend to guess "yes" incorrectly on the basis of the item's familiarity. So for items on the Outside questionnaire, seam effects should be relatively large, since respondents should tend to answer "no" for Week 4 and "yes" for Week 5. On the other hand, for items from the Inside questionnaire (see Table 1), we should see a different pattern of results. During the first test, respondents should give a correct "yes" response when we ask if they had seen the item during Week 4. During the second test, familiarity should again lead them to guess "yes" about whether they had seen the item during Week 5. In this case, both memory and guessing yield the same response, and the seam effect should be relatively small. We should therefore predict bigger seam effects for the Outside items than for the Inside items.

The main results from this study appear in Figure 2. The y-axis shows the percentage of times respondents changed their response from one week to the next. This

was either a change from "yes, I saw that item on the Week *k* questionnaire" to "no, I didn't see that item on the Week *k*+1 questionnaire" or the reverse change. The results show a clear seam effect – greater changes between seam Weeks 4 and 5 than between other pairs of weeks off seam. Recall that there was no actual change in the items during the seam weeks, so the effect is entirely due to response error. In the opposite direction, there is no noticeable increase in number of changes between Weeks 2 and 3 or between Weeks 6 and 7, where the real change in items occurred. Respondents' overall accuracy for these weeks was quite low, so it was apparently difficult for them to remember the objective shift in the questions.

The second result that appears in the figure is that the seam effect is larger for items from the Outside questionnaire than for those from the Inside questionnaire. This is the result we predicted on the basis of respondents' memory and estimation strategies. For the Outside questionnaire, memory for Week 4 and estimation for Week 5 produce different responses, and so larger seam effects. In fact, 70% of respondents correctly said that the Outside items were not on the questionnaire during Week 4. However, only 35% correctly said that these items were not on the questionnaire for Week 5. For the Inside questionnaire, memory for Week 4 and estimation for Week 5 produce the same "yes" answer, leading to smaller seam effects. For these items, respondents gave 60% correct "yes" responses for Week 4 and 66% correct "yes" responses for Week 5.

EFFECTS OF RECALL CONDITION

The results of Experiment 1 show that we can reproduce seam effects with our method, so it seems reasonable to use it to explore other factors that might alter the size of the effect. In the study just described, respondents had to answer all the test questions about Week 4 before the questions about Week 3, Week 3 before Week 2, and so on. But this isn't standard practice in panel surveys like SIPP or CE. These surveys more often group questions by topic. SIPP, for example, asks all the questions about food stamps (last month, 2 months ago, etc.) before questions about other income sources. It seemed to us that grouping questions by topic in this way might increase the size of the seam effect by encouraging respondents to give the same response for each question in the group. Once you've decided that you received food stamps last month, it is tempting to give the same answer if you're immediately asked whether you received food stamps two months ago. This can minimize changes for off-seam months, exaggerating changes at the seam. This is a phenomenon called *constant wave response* in previous studies of the seam effect (Kalton & Miller, 1991; Young, 1989).

Although panel surveys group questions by topic rather than by temporal period, they're not always consistent about the temporal order of the questions. CE asks about household repairs and alterations in forward temporal order, as in (1), but it asks about the use of owned property in backward temporal order, beginning with one month ago and ending three months ago. Some cognitive studies have reported better recall in backward temporal order than forward order, which would give an advantage to the questions about owned property (e.g., Whitten & Leonard, 1981, but see Jobe, White, Kelley, Mingay, Sanchez, & Loftus, 1990).

To look at these issues, we used the method of Experiment 1, but with several variations. We mailed all respondents one questionnaire per week for a total of six consecutive weeks, and we tested them for recall of the questionnaire items after Week 3 and again after Week 6.

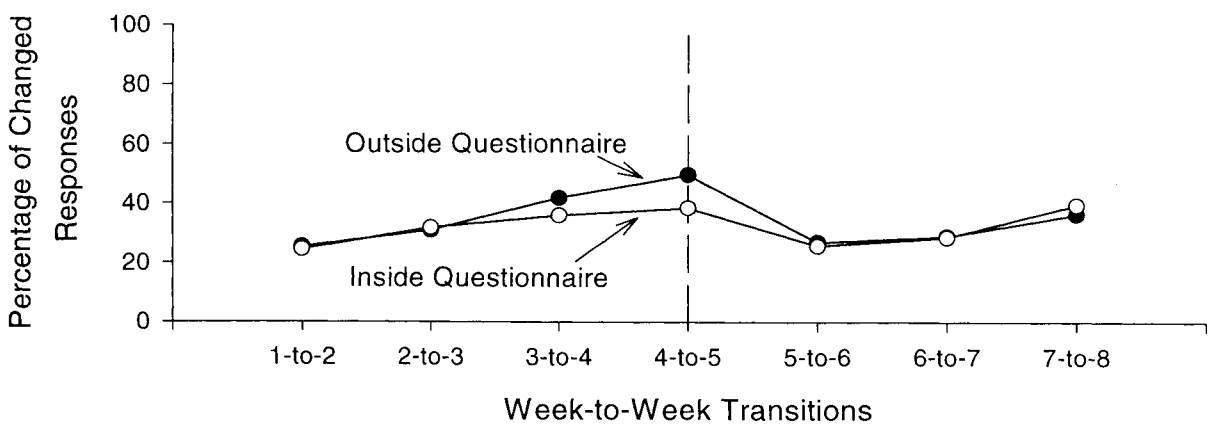

Figure 2. Percentage of times respondents gave different answers for Week *k* and Week *k*+1, Experiment 1. Filled circles represent items from the Outside questionnaire, open circles item from the Inside questionnaire (see Table 1). Dashed line indicates seam weeks.

(We used a slightly shorter schedule here, since we found very little evidence in the earlier study for memory of the questionnaire four weeks back.) The bigger change, though, was that the items were topically related across questionnaires. For example, on the first questionnaire, we asked whether respondents had made a phone call to a friend in the last week; on the second, we asked whether they had made a phone call to a relative in the last week; and on the third, whether they had made a phone call to a co-worker. These items are all related to the same topic of making a phone call. Similar sets of items concerned events on a business trip and shopping at a mall, among others.

We also tested four different groups of respondents who differed in the way they recalled the items. Two of these groups saw test questions grouped by topic: We asked all the questions about making a phone call, for example, before asking questions about the other topics. One group saw these questions in forward temporal order (with questions about Week 1 before Week 2 and so on); a second group saw them in backward temporal order. The tests for these two groups were closest to the procedure of the panel studies. The other two groups of respondents saw the test items blocked by week. We asked about all the topics for one week, then about all the topics for the next. Again, one group of respondents saw the items in forward temporal order (Week 1 to Week 3) and the other in backward order (Week 3 to Week 1).

Each test question asked respondents whether they had seen an item on one of their questionnaires – for example, "Did you see an item about making a phone call on the questionnaire for Week 1?" If they said "yes," we asked them to write down the full item as it appeared on that questionnaire. A correct response would be, for example, "Did you make a phone call to a friend?" We included the recall measure in this experiment because many survey items ask for recall rather than for the simple yes/no judgments of Experiment 1.

The results from this experiment indicate that respondents who recalled the items backward by week were more accurate than those in the other three conditions. Figure 3a shows that this advantage comes mainly from Weeks 3 and 6, the most recent weeks. Respondents' memory is best for these two weeks, and respondents who recall backward-by-week are able to report these most recent memories at the beginning of the test session. Respondents in other conditions have to recall some or all of the more difficult early memories at the beginning of the test session, and this may have put them at a disadvantage. For example, fatigue may have set in before they got to the easier items.

Figure 3b indicates that the seam effect reappears in this experiment. The y-axis in the graph is the percentage of times respondents changed their answer about an item within a specific topical series. For instance, if a respondent said that she saw an item on the Week 1 questionnaire about making a phone call to a friend and an item on the Week 2 questionnaire about making a phone call to a relative, then she would be scored as changing her response from Week 1 to Week 2. In this case, the respondent's answers would be correct. However, we also counted as a change any other shift in wording or a change from response to nonresponse. In this experiment true changes occurred in all weeks; so the difference between seam and off-seam weeks is again due to response error.

Figure 3b also shows that the seam effect depends on recall condition. Although the total number of changes is larger when respondents recall forward or backward by week, the seam effect is actually smaller for recall by week. Respondents who recalled forward or backward by topic produced fewer total changes and larger seam effects. Recalling by topic means answering one-after-another questions like: "What was the item about making a phone call during the third week?" "What was the item about making a phone call during the second week?" "What was the item about making a phone call during the first week?" This sequence may invite constant-wave responses for these items. Since the correct answers differ for each week, constant-wave responses can lead to both inaccuracies and larger seam effects.

IMPLICATIONS

We know of no national surveys that ask respondents to report all the last month's data before reporting on previous months. But the analogous condition in Experiment 2 is the one that produced best performance in terms of accuracy. And although the seam effect hasn't entirely vanished, this condition reduced it to a lower level than that of the more usual procedure of grouping items by topic. Reporting backward by temporal interval seems to take best advantage of respondents' better memory for the most recent time period, and by breaking up the items with the same topic, it seems to reduce the correlation between them.

The results are also consistent with the picture about memory and estimation that we started with. The data from both studies suggest that respondents were able to remember little about the earliest weeks in the response period, and they probably based their answers about these earlier intervals on estimates. Under these conditions, one reasonable estimation strategy is simply to carry over the answer from the most recent interval to earlier ones. This will lead to different constant responses for each of the response periods. The combination of constant answers within the response periods and different answers between response periods will produce a seam effect. This type of estimation strategy is probably more likely when recall is forward or backward by topic; so these recall conditions should increase the size of the seam effect, as they did in

Experiment 2.

On this view, the seam effect depends on the contrast between memory for the most recent portion of one response period and estimates for the earliest portion of the next response period. The underlying problem is that respondents simply can't recall much about the early portion of the period and that forces them to rely on different, less reliable cognitive strategies. We need to recognize that the seam effect is a symptom and that reducing its size does not necessarily eliminate its cause. Nevertheless, some of the conditions in this study had the double advantage of decreasing the seam effect and increasing accuracy.

REFERENCES

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

Jobe, J. B., White, A. A., Kelley, C. L., Mingay, D. J., Sanchez, M. J., and Loftus, E. F. (1990), "Recall Strategies and Memory for Health-care Visits," *Milbank Memorial Fund Quarterly,* 68, 171-189.

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.

Smith, A. F., and Jobe, J. B. (1994), "Validity of Reports of Long-term Dietary Memories: Data and a Model," in *Autobiographical Memory and the Validity of Retrospective Reports,* eds. N. Schwarz and S. Sudman, New York: Springer-Verlag, pp. 121-140.

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

Whitten, W. B., and Leonard, J. M. (1981), "Directed Search through Autobiographical Memory," *Memory & Cognition,* 9, 566-579.

Young, N. (1989), "Wave-seam Effects in the SIPP," in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Alexandria, VA: American Statistical Association, pp. 393-398.

ACKNOWLEDGMENTS

We thank Catheryn Dippo, Paul Hsen, Adrianna Silberstein, Clyde Tucker, and Gordon Willis for comments on earlier presentations of this work. Credit for the title goes to Roger Tourangeau. NSF Grant SES-9907414 supported the research reported here. It was also facilitated by a fellowship from the American Statistical Association, the National Science Foundation, and the Bureau of Labor Statistics. The views expressed here reflect the opinion of the authors and not those of BLS.

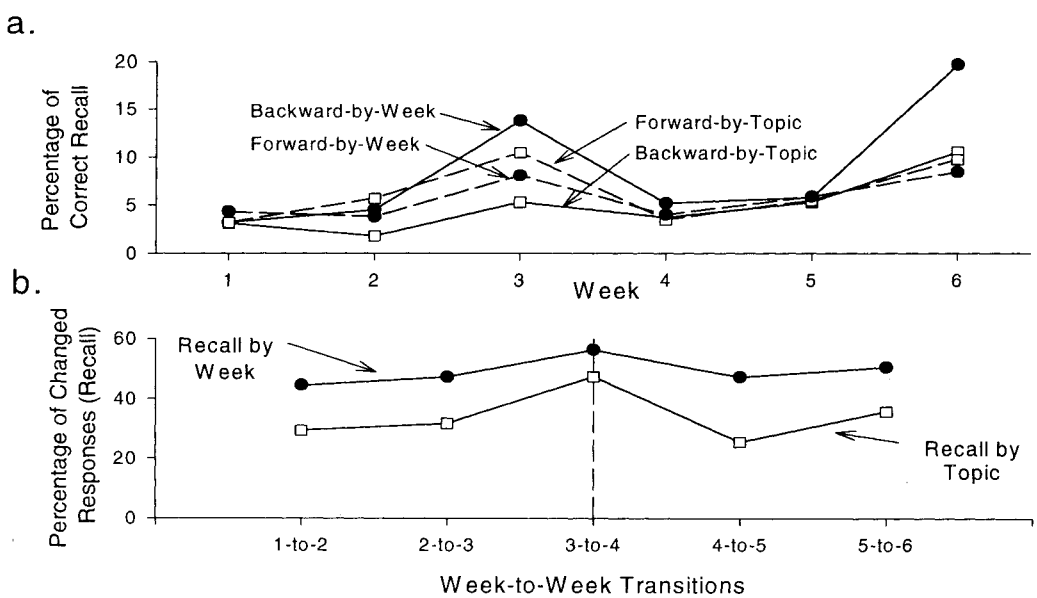

Figure 3. Percentage of items correctly recalled (top panel) and percentage of changed responses from one week to the next (bottom panel), Experiment 2. Vertical dashed line indicates position of the interviewing seam.