

Remembering the Earthquake: Direct Experience vs. Hearing the News

Ulric Neisser, Eugene Winograd, Erik T. Bergman

Emory University, Atlanta, USA

Charles A. Schreiber, Stephen E. Palmer

University of California, Berkeley, USA

Mary Susan Weldon

University of California, Santa Cruz, USA

Three groups of informants—two in California, one in Atlanta—recalled their experiences of the 1989 Loma Prieta earthquake shortly after the event and again 1½ years later. The Californians' recalls of their own earthquake experiences were virtually perfect. Even their recalls of hearing the news of an earthquake-related event were very good: much higher than Atlantan recalls of hearing about the quake itself. Atlantans who had relatives in the affected area remembered significantly more than those who did not.

These data show that personal involvement in the quake led to greatly improved recall, but do not show why. Many Californian informants reported low levels of stress/arousal during the event; arousal ratings were not significantly correlated with recall. The authors suggest that repeated narrative rehearsals may have played an important role.

INTRODUCTION

Most everyday experiences are soon forgotten, but some give rise to vivid, confident memories that last for years or decades. In many cases, such memories are essentially accurate accounts of the original experience. The present study illustrates this point. As we shall see, individuals who had directly experienced the California earthquake in 1989 recalled their experiences confidently and accurately after a delay of a year and a half. But memory is not always so good:

Requests for reprints should be sent to Ulric Neisser, who is now at the Department of Psychology, Cornell University, Ithaca, NY 14853, USA.

A preliminary account of this work was presented at the meetings of the Psychonomic Society, San Francisco, November 1991.

We are grateful to Daniel B. Wright, Steen Folke Larsen, and David C. Rubin for their helpful comments on an earlier draft of this paper.

the subjects of Neisser and Harsch (1992) for example, made gross errors in recalling how they had first heard about the *Challenger* space shuttle disaster. What variables might distinguish experiences that are destined for accurate recall from those doomed to error or oblivion?

At one level, the present findings simply suggest that experiencing events as a participant (rather than as a mere observer) improves the accuracy of recall. This argument is plausible, and has been made before (e.g. Goodman, Rudy, Bottoms, & Aman, 1990). But while participation is obviously important, the mechanisms by which it achieves its effects are less obvious. (This caution applies equally to other claims about the role of the self in memory: cf Greenwald & Banaji, 1989; Hyman & Neisser, 1992.) Most life events and experiences involve "participation" and yet are soon forgotten. What was so special about participating in the earthquake?

The critical variable was apparently not the stress of experiencing the quake itself. Many of our informants were not frightened; they rated their experiences as unexciting, but remembered them anyway. Nor was it the *consequentiality* of what was remembered: the earthquake itself was important, but one individual's experience of its tremor is rarely significant in the same way. Our own interpretation, which we will try to justify in the discussion section, is based on the concepts of *narrative* and *rehearsal*. Each of our California subjects had his or her own "earthquake story", and probably found many occasions to repeat it. Those socially motivated narratives are the most likely basis of the high levels of accuracy documented here.

Historical Introduction

The modern study of such recollections began with Roger Brown and James Kulik's (1977) seminal paper on "flashbulb memories". Brown and Kulik took a critical methodological step: they selected a significant experience that many individuals had shared at more or less the same time, and asked them to recall it after a substantial delay. Most of their target experiences were what Larsen (1988) has called *reception events*: occasions on which one first hears the news of a significant public occurrence. Although they asked their informants about several such events, Brown and Kulik focused on the assassination of John F. Kennedy. Even 12 years later, their informants still claimed to have vivid and detailed memories of how they had heard that news.

Although the concept of flashbulb memory has been very influential, it raises both theoretical and methodological problems. The theoretical puzzle is why one would remember reception contexts (e.g. hearing the Kennedy news) in the first place. The assassination itself is important, but how a given individual hears about it does not seem important in the same way. Perhaps for that reason, more recent studies (Larsen, 1992) have *not* found reception events to be especially memorable. Indeed, most vivid memories are not about news events at all. When

Rubin and Kozin (1984) asked a group of Duke students to describe their clearest memories, most reported events that they had experienced directly: car accidents, sexual episodes, and the like. Only 4 accounts out of 174 were reception events. Similarly, the majority of the college women studied by Pillemer, Koff, Rhinehart, and Rierden (1987) reported having detailed recollections of their first menstruations: again, a personal and potentially narrativisable event. Unfortunately, the accuracy of such memories cannot be assessed.

At the methodological level, Brown and Kulik were perhaps too ready to take their informants' reports at face value. They even suggested that flashbulb memories were essentially immune to forgetting, "... as unchanging as the slumbering Rhinegold" (1977, p.86). This faith may have been misplaced. A long tradition of research on eyewitness testimony, stretching from William Stern (1904/1982) to the present (Wells & Murray, 1984) has often shown that even highly confident witnesses can be wrong. One of us has elsewhere described a personal example of a faulty "flashbulb" memory—of how he himself heard the news about the Japanese attack on Pearl Harbor (Neisser, 1982, 1986; Thompson & Cowan, 1986).

Recent researchers have tried to deal with the problem of accuracy in a new way, by getting reports of the event in question shortly after it happens and using those reports as a baseline to score subsequent recalls. The 1986 explosion of the space shuttle *Challenger* offered just such an opportunity, and Neisser and Harsch (1992) were among those who seized it (see also Bohannon & Symons, 1992; Warren and Swartwood, 1992). On the morning after the disaster, a large group of Emory students were given a set of written questions about how they had heard the news on the preceding day. Three years later, 44 of them were asked the same questions again. Surprisingly, many of these recalls—even those of highly confident informants—were mistaken. Several subjects were sure that they had first learned of the disaster from television when in fact they heard about it from a friend, or in class; one insisted that he had been at home with his parents when in fact he was at Emory. Some of these errors may just have been reports of what Brewer (1988) calls a "wrong time slice", but others surely were not.

To assess these errors quantitatively Neisser and Harsch (1992) devised a "weighted accuracy score", based on the comparison between each subject's first and second reports. These scores ranged from zero (i.e. no aspect of the event was correct in the second recall) to seven (all aspects recalled correctly). In fact, the mean accuracy score was only 2.95. Eleven subjects scored zero; nevertheless, most of them expressed full confidence in their memories.

Further evidence of the fallibility of memory for reception events comes from a diary study by Larsen (1992). For six months in 1986, Larsen recorded two events every day: one was a personal experience, the other was an occasion when he learned about some public news event. As it happened, this period

included two public events of great interest to Larsen, who is Danish: the assassination of Olof Palme (the Prime Minister of Sweden) and the nuclear accident at Chernobyl, which produced a large radioactive cloud threatening Scandinavia. (The *Challenger* explosion also occurred during this period, but Larsen's diary indicates that he did not regard it as a significant event.)

Larsen's memory for these reception events proved to be surprisingly weak. Only five months after hearing the news about Palme, he misremembered both where he had been and who had been with him. (Christianson, 1989, also found substantial forgetting of the reception context for the Palme assassination.) Only seven months after Chernobyl, Larsen misremembered almost everything about the occasion on which he had learned about it: the time, the place, who was with him, what he was doing. What is more, his entirely false reconstruction of that reception context has proved to be surprisingly stable. Years later, Larsen (1992) reported that he still remembers the event that way despite the contrary evidence in his own diary. At least where reception events are concerned, then, vivid and confident memories can be quite mistaken.

What might distinguish events that establish accurate memories from events that do not? At a first level of analysis, it seems likely that participation is a key variable. Although many of Neisser and Harsch's subjects were genuinely upset by the space shuttle disaster—as Larsen was by Palme's murder and by Chernobyl—they were still not personally involved in it. To see whether this variable matters, the present study explicitly compares *participants*—individuals who actually experienced an event—with control subjects who only heard about it. The event in question was an important one: the Loma Prieta earthquake of 17 October, 1989. (Non-Californians often call it the “San Francisco earthquake”, but the actual epicentre was well to the south of that city.) The recall interval was about a year and a half: from October 1989 to mid-spring of 1991.

There is independent evidence that the Loma Prieta earthquake affected those who experienced it. This evidence comes from a study of nightmare content conducted by Wood, Bootzin, Rosenhan, Nolen-Hoeksema, and Jourden (1992). In a design similar to ours, Wood et al. compared Bay area students who had actually experienced the earthquake with Arizona controls who had not. All subjects kept dream logs for 21 nights, starting one or two weeks after the event. About 40% of the California sample reported at least one “slightly to moderately intense” nightmare about an earthquake, compared to only 5% of the comparison group.

Three coordinated teams of investigators—at the University of California campuses in Berkeley and Santa Cruz, and at Emory University in Atlanta—carried out the present study. Our aim was to contrast the memories of California students, for whom the quake was a personally experienced event, with students at Emory for whom it was a reception event. A within-subject comparison of the same two categories is also available: we will contrast the California informants' recollections of their own earthquake experiences with their recalls of how they

first heard about the collapse of the San Francisco Bay Bridge. A number of related variables—especially the subjects' own ratings of their emotional responses—will also be reported.

METHOD

Informants

There were three groups of informants. For most of the Emory students in the Atlanta group ($n=76$), the earthquake was just something that had happened thousands of miles away. Its major consequence for them may have been the postponement of the baseball World Series between the Oakland As and the San Francisco Giants: the earthquake struck just as the third game of that series was scheduled to begin in San Francisco. The subset of Atlanta subjects who had friends or family in the bay area constitute a special case, to be considered in some detail later.

For the Berkeley informants ($n=41$), it was a moderate earthquake experienced at first hand. Berkeley is on the east side of San Francisco Bay, which is some five miles wide at that point. It is connected to San Francisco itself by the Bay Bridge, which was damaged and rendered impassable by the quake. Major damage occurred in San Francisco itself as well as at one freeway at the east shore, but the effects in Berkeley were relatively minor. Although the tremors could be clearly felt, no University buildings were seriously affected. Classes met the next day.

The subjects at the University of California in Santa Cruz ($n=44$) experienced much more severe conditions. The epicentre of the earthquake (which registered 7.1 on the Richter Scale) was relatively close to Santa Cruz, a coastal city some 75 miles south of San Francisco. Its impact there was far greater than anything experienced in Berkeley, or indeed in most of San Francisco itself. There was extensive physical damage to buildings, loss of electric power, and a general disruption of everyday activities. Classes at the University were cancelled for several days.

The timing of our initial questionnaires was as follows:

At Emory: Questionnaires were given to an introductory psychology class ($n=95$) on 19 October, less than 48 hours after the quake. Another introductory class ($n=195$) was given the same questionnaire a week later. These groups did not differ significantly on any outcome measure, and have been combined in all analyses reported here.

At Berkeley: A total of 172 questionnaires were given in various undergraduate courses at Berkeley on 18 October, the day after the quake; another 138 were given on 19–20 October.

At Santa Cruz: Logistical factors prevented us from collecting initial accounts at Santa Cruz until the first week in November, a delay of 15–21 days.

Questionnaires were then obtained from 135 upper- and lower-level students in psychology classes.

Questionnaires

After giving their name, sex, and age, the Emory group first described in their own words ‘how you first learned that such an earthquake had occurred.’ Then they turned the page and answered a number of specific questions:

- What time of day was it when you first heard the news?
- How did you hear it (TV, radio, someone told you, etc)?
- Where were you at the time?
- What were you doing at the time?
- Who told you? (If TV/radio, what programme/newscaster; if person, who?)
- Who else was present?

They then rated their emotional reaction to the news on a 7-point scale (“1 means you have no emotional reaction at all; 7 is the strongest emotion you have ever felt in your life”) and answered several further questions: whether there was a family member or friend—someone they felt close to—in the affected area; whether they had ever experienced an earthquake personally; what they did immediately after hearing the news; how much time they had spent talking about the earthquake and following coverage of it, both on the day of the quake and on the following day.

The Berkeley investigators modified the questionnaire in several ways that seemed appropriate to their situation. (For example, the “informant” category is irrelevant for subjects who experienced the earthquake directly.) Their subjects did not begin with an overall free recall of the experience, but were asked a larger number of specific questions. Besides reporting where they were, what they were doing, what time it was, and who they were with when the earthquake struck, they were asked what went through their mind at the time (and afterwards), how they had reacted, how strong they had thought the quake was at the time, and when they had realised how serious it really was.

After these items, the Berkeley informants were asked a further set of questions about how they had learned about the collapse of the upper deck of the Bay Bridge. (This was a major feature of the earthquake for everyone in the Bay area, widely reported in the news media.) These questions were: When you first heard about the collapse of the upper deck on the Bay Bridge (not to be confused with the Nimitz Freeway), where were you? What time was it? What were you doing? Who were you with? What went through your mind? How did others react?

Three 7-point rating scales concerning reactions to the events concluded the questionnaire:

- How you felt during the earthquake (1 = very calm to 7 = terrified).
- Immediately after the quake, how concerned were you? (1 = not at all concerned to 7 = extremely worried)
- Immediately after you heard the news of the bridge collapse, how concerned were you? (same scale)

The Santa Cruz informants began with a free recall like that of the Emory subjects, but were then asked all the questions used at Berkeley.

Follow-up Survey

The retention interval of a year and a half was dictated by the composition of the California samples, many of whom became graduating seniors during the next academic year. In the spring of 1991, follow-up surveys were administered to all informants who could be reached by telephone and agreed to come to the lab to fill out a questionnaire. (They were not told on the phone that the questionnaire would deal with the earthquake.) These included 76 Emory students (39 from the group that had first responded two days after the quake and 37 from the nine-day group), 41 Berkeley students,¹ and 44 Santa Cruz students. The follow-up questionnaires were essentially identical to the originals except that a confidence rating scale was included with each question: 5-point scales at Emory and Santa Cruz (1 = just guessing, 5 = absolutely certain), 7-point scales at Berkeley. Informants were also asked whether they remembered filling out the earlier questionnaire. In addition, Berkeley subjects were asked how long they had lived in California (75% of the informants indicated 12 years or more of residence), and whether they had ever experienced an earthquake prior to Loma Prieta (of the 39 who answered this question, 36 said “yes”).

Scoring

Accuracy was scored with a modified version of the system used by Neisser and Harsch (1992). For the Atlanta subjects, whose recall concerned hearing the news about the earthquake, judges scored five attributes: *place*, *informant*, *activity* (what they were doing when they heard the news), *others present*, and *time of day*. To maximise reliability, each attribute was scored on a simple 3-point scale. A score of 2 meant that the first and second reports of the attribute were essentially identical, while 0 meant that they were clearly inconsistent (or that the attribute was not mentioned on the second questionnaire). Intermediate degrees of fit were scored 1. The primary source of information for these judgements was the subject's answer to the specific question about the given

¹ Forty-four students were actually tested at Berkeley, but three have been omitted from subsequent analyses because, according to their original reports, they hadn't realised it was an earthquake until subsequently informed by others or the media.

attribute. Judges scoring the Emory and Santa Cruz questionnaires were also free to use the subject's initial free recall (or responses to other items) to resolve any ambiguities.

For many purposes it is convenient to have an index of overall accuracy, summed across relevant attributes. Here we report "Combined Accuracy Percent" (CAP) scores, based on three attributes common to all groups: *place*, *activity before*, and *others present*. No other attribute is suitable for such an index: *informant* does not apply to direct experience, *activity during* does not apply to reception events, *time* was not used because information about it was very widely available. (The media in California often used the impact time, 5:04, as a sort of synonym for the quake itself; it even appeared on T-shirts.)² These three attribute scores, each ranging from zero to two, were summed and converted to a "Combined Accuracy Percent" (CAP) score:

$$\text{CAP} = 100 \times (\{place + activity + others\}/6).$$

Two separate CAP scores were calculated for the California groups, CAP(DIR) for recall of the direct earthquake experience and CAP(BNEWS) for recall of hearing the news about the Bay Bridge. The score for the Emory group may be called CAP(ENEWS) since it concerns news of the earthquake itself.

To assess reliability, two judges independently scored the attribute recalls and converted them into CAP scores which were then correlated. These reliabilities were 0.87 for CAP(ENEWS) at Emory, 0.80 and 0.71 for CAP(BNEWS) in Berkeley and Santa Cruz respectively. Agreement between coders was high for CAP(DIR)—85% perfect agreement in Berkeley, 93% in Santa Cruz—but meaningful correlation coefficients could not be calculated because of ceiling effects. All disagreements were resolved by discussion.

A further measure of scoring reliability was available for the Berkeley group. After they had responded to the 1991 follow-up questionnaire, informants were shown their own 1989 forms. They then rated the consistency of their own current and original reports separately for each attribute, scoring 0 if the two reports were incompatible, 1 if some details were contradictory, 2 if both reports might well be completely correct. These ratings were converted to CAP scores and correlated with the CAPs based on our coding scheme. For CAP(BNEWS) this "reliability" correlation was 0.85. For CAP(DIR), 61% of the scores agreed perfectly.

RESULTS

Almost all informants gave full and complete initial reports. The only exceptions were three informants at Santa Cruz who did not say how they heard the news of

² Friedman (1987) found that individuals who had experienced a 1986 earthquake in Ohio were able to recall the time of impact rather accurately nine months later.

the Bridge collapse, two at Emory who did not answer the question about "others present",³ and three who did not record a time. Most of those who filled out the second questionnaire still remembered having answered the first one: 66/76 at Emory, 32/44 at Santa Cruz, 29/41 at Berkeley. Because there were no significant sex differences, data for men and women have been combined in all analyses reported here.

Accuracy: Global Measures

There are five major sets of CAP scores: CAP(DIR) at Berkeley and Santa Cruz, CAP(BNEWS) at the same two sites, and CAP(ENEWS) at Emory. These means are shown in Fig. 1. Three comparisons are of primary interest here.

Experiencing the Quake vs. Hearing About It. As can be seen, both California samples were essentially at ceiling for recall of direct experience

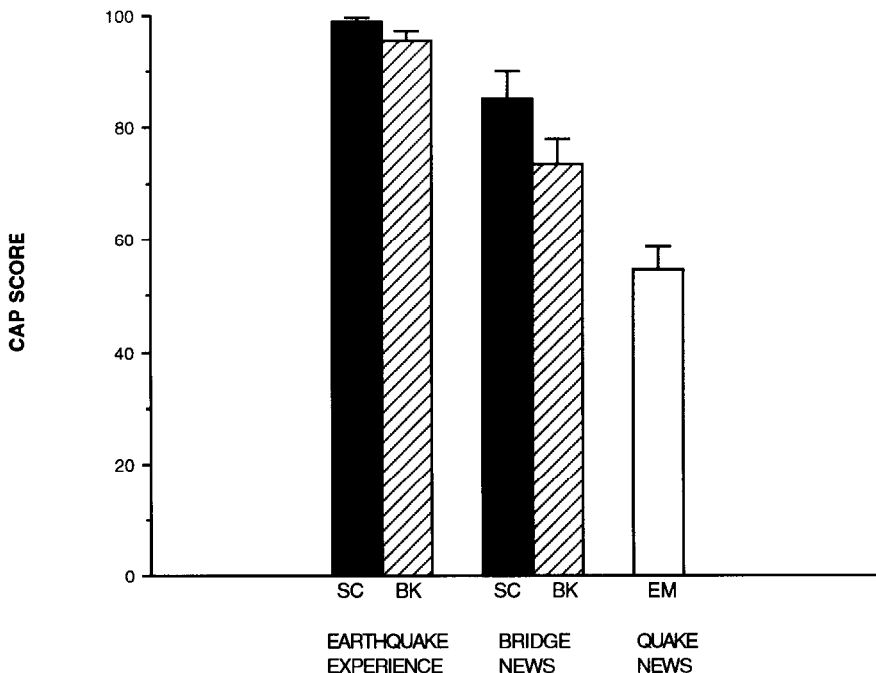


FIG. 1. Mean Combined Accuracy Percent (CAP) scores, with standard error bars, for recalls of: earthquake experiences (Santa Cruz and Berkeley informants), hearing about the Bay Bridge (Santa Cruz and Berkeley informants), and hearing the news of the earthquake (Atlanta informants).

³ These five subjects were eliminated from all computations involving CAP scores.

(first two bars), with mean CAP(DIR)s of 99 and 96 respectively. These scores were very much higher than those at Emory, where the mean CAP(ENEWS) of 55 was significantly lower than either CAP(DIR) by Mann-Whitney U test. (Non-parametric tests are used throughout this section because of the skewed CAP distributions.)

A few of the Emory subjects made gross recall errors, comparable to those described by Neisser and Harsch (1992). Subject LG, for example, was very confident on the follow-up questionnaire: she had first heard about the earthquake on the following morning, listening to the radio while at work, "... sitting on a stool, looking at the radio in amazement at the news." But in fact (according to her own earlier report), her parents had told her about the quake on the telephone on the previous evening: "My dad was going to watch the World Series. The(y) described the destruction of the Bay Bridge for me & the evacuation of the stadium." LG's CAP(ENEWS) score was zero. In contrast, no error of this magnitude was made by any Californian informant in recalling the experience of the earthquake.

Experiencing the Quake vs. Hearing About the Bridge. In both California groups, recall of hearing about the bridge was substantially lower than recall of direct experience. Wilcoxon matched-pair tests show that these differences were significant (for Berkeley $Z=3.74$, $P<0.001$; for Santa Cruz $Z=2.97$, $P<0.01$). Being personally involved in an event is evidently more memorable than just hearing the news of one. But although they were weaker than memories of direct experience, the California recalls (both groups combined) of hearing about the bridge were still more accurate than the Atlanta subjects' recollections of hearing about the earthquake itself ($U=1759$, $Z=4.68$, $P<0.0001$).

Santa Cruz vs. Berkeley. In spite of near-ceiling effects, there are significant differences favouring the Santa Cruz group over the Berkeley group on both CAP(DIR) ($U=766$, $Z=1.98$, $P<0.05$), and CAP(BNEWS) ($U=589$, $Z=2.56$, $P=0.01$). These differences may reflect the greater intensity of the earthquake as experienced in Santa Cruz, but they may also have resulted from procedural differences between the two locations. Two such differences seem especially relevant. First, the Santa Cruz informants (like those at Emory but unlike those at Berkeley) were asked to give free-recall accounts of their experiences before answering specific questions. This format may have aided recall; it may also have affected subsequent scoring, because coders could use the free recalls to clarify answers that might otherwise have seemed inconsistent. Second, the baseline Santa Cruz reports were not obtained until 2–3 weeks after the quake and hence may reflect more solidly established initial narratives. Note, however, that an (admittedly smaller) delay in interviewing some of the Berkeley informants did *not* result in more consistent reports. The mean CAP(DIR) and CAP(BNEWS) scores of the Berkeley subjects first seen on 18

October were almost identical to those of others who were not interviewed until 20 October, two days later.

Accuracy of Individual Attributes

Analysis at the level of individual attributes tells the same story as the global CAP scores. Figure 2 shows each group's mean recall separately for *Activity During*, *Activity Before*, *Place*, *Others Present*, *Time*, and *Informant*. Each attribute shows the same pattern: the California direct-experience scores are

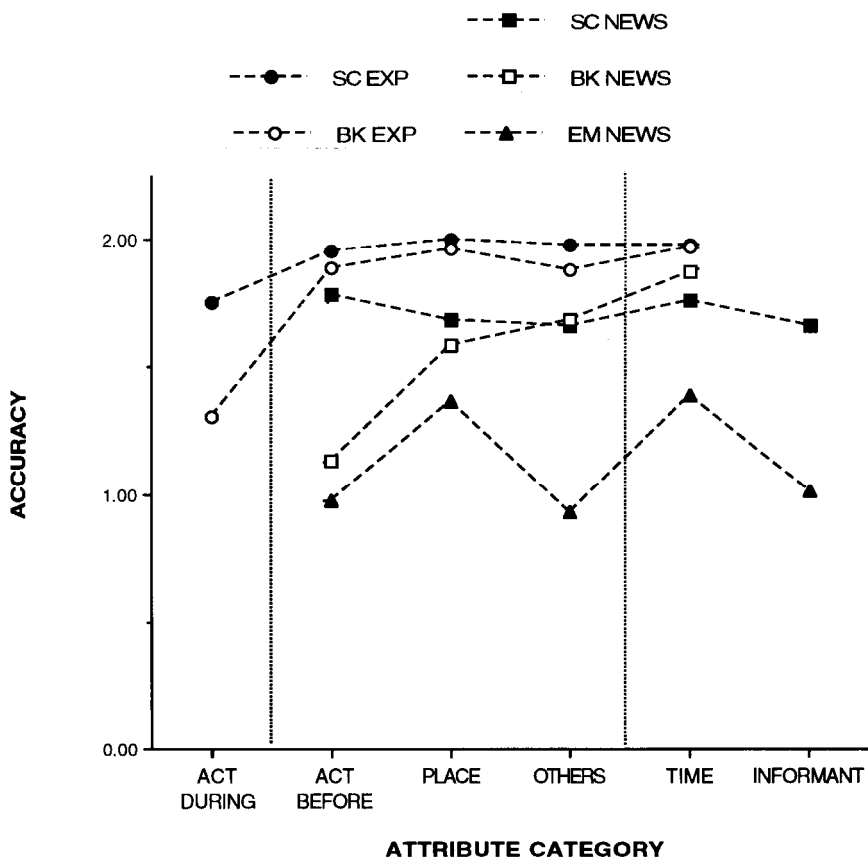


FIG. 2. Mean accuracy scores for individual attribute categories (maximum score 2.00) for recalls of: earthquake experiences (Santa Cruz and Berkeley informants), hearing about the Bay Bridge (Santa Cruz and Berkeley informants), and hearing the news of the earthquake (Atlanta informants). The attributes between the dashed lines (activity before, place, others present) comprise the CAP measure that was graphed in Fig. 1.

highest, the Atlanta hearing-the-news scores lowest, and the Bay-Bridge-news scores in between.

Affect

The subjects' affect ratings can be used to address three distinct issues: the degree of emotional response actually produced by the earthquake, the correlation of reported arousal levels with accuracy of recall, and the subjects' ability to recall their own reported feelings a year and a half later.

Mean Reaction Ratings. How emotionally disturbing was the earthquake, and how disturbing was the subsequent news about the Bay Bridge? Mean initial ratings on three different scales—reaction during the quake, concern after the quake, concern about the Bay Bridge—appear in the first column of Table 1. On all three scales, the mean ratings of the California informants were close to 4.00, the middle of the 7-point scale. (Emory subjects used a differently worded scale and will not be considered here.) On the average, then, the earthquake was *not* a traumatic event. This is congruent with findings of Wood et al. (1992), whose subjects rated their anxiety at only 3.61 (in Stanford) and 3.17 (in San Jose) on a 5-point scale. It is interesting that the Berkeley and Santa Cruz means did not differ significantly on any of these scales, even though Santa Cruz was closer to the epicentre and experienced more damage.

In both cities, the concern engendered by news of the bridge collapse was appreciably *higher* than the concern reported after the earthquake itself. Separate Wilcoxon tests show that the difference was significant in both groups ($Z=4.38$, $P<0.0001$ for Berkeley; $Z=3.45$, $P<0.0001$ for Santa Cruz). Although the two rating scales were worded somewhat differently, we are inclined to take this result at face value. Most of our subjects were not seriously concerned for themselves because, in fact, their local situations were not dangerous. On the other hand, all of them were exposed to dramatic reports of major catastrophes elsewhere.

Affect as a Predictor of Recall. Perhaps the most obvious question to ask of these data is whether higher affect is associated with greater recall. The relevant correlations—between initial affect ratings and subsequent CAP scores—are shown in the second column of Table 1. The answer is clearly negative: none of them is significant. This may be an artifact of ceiling effects in the case of correlations with CAP(DIR), but even without any such constraint the correlations with CAP(BNEWS) in California or CAP(ENEWS) at Emory were still essentially zero.

Recall of Affect Itself. The low correlations between affect and accuracy are not due to any lack of reliability in the affect ratings themselves. We were

TABLE 1
Ratings of Affect

<i>Group</i>	<i>Mean Initial Rating</i>	<i>Correlation with Accuracy (CAP)</i>	<i>Mean Remembered Rating</i>
Berkeley ^a during quake	3.84	-0.02	4.04
Santa Cruz ^a during quake	4.02	-0.05	4.09
Berkeley ^b after quake	3.89	-0.22	3.73
Santa Cruz ^c after quake	4.03	-0.04	4.07
Berkeley ^b after bridge news	5.28	-0.11	5.63
Santa Cruz ^c after bridge news	5.29	-0.14	5.32
Emory ^d after quake news	3.60	-0.03	3.49

^a Scale went from 1 = "very calm" to 7 = "terrified"

^b Scale went from 1 = "not at all concerned" to 7 = "extremely worried and upset"

^c Scale went from 1 = "not at all concerned" to 7 = "extremely worried"

^d Scale went from 1 = "no emotional reaction at all" to 7 = "the strongest emotion you have ever felt in your life"

able to assess that reliability because the same reaction ratings were obtained again on the follow-up survey. The means of these second ratings are shown in column 3 of Table 1; they agree rather well with the ratings given initially. (The only significant difference is for Berkeley ratings of concern over the bridge collapse, where the second ratings are higher.) For a more direct measure of reliability, we looked at within-subject consistency. Correlations between the first and second ratings are shown in the main diagonal of Table 2, separately for the Santa Cruz and Berkeley groups. All of them are substantial and significant. The cross-correlation between *Different* ratings on the two questionnaires (also in Table 2) are generally lower. Even after 18 months, then, our informants could discriminate their feelings during the earthquake from their feelings after it was over.

TABLE 2
Correlations Between Initial and Later Ratings of Reaction to the Events for California Subjects

<i>Initial</i>	<i>Quake During</i>	<i>Recalled Quake After</i>	<i>Bridge News</i>
Quake During	0.81*/0.78*	0.51*/0.63*	0.19 /0.40*
Quake After	0.61*/0.65*	0.50*/0.82*	0.43*/0.48*
Bridge News	0.27*/0.12	0.22 /0.28*	0.53*/0.65*

* $P < 0.05$

Santa Cruz correlations are given first and Berkeley correlations are given second. For Emory, the correlation was 0.52 between the initial and later ratings of affect for hearing the news of the earthquake.

Concern for Friends and Relatives

Individuals who do not experience an event themselves may nevertheless feel concern for others more directly involved. To assess the effect of this concern, we examined the CAP(ENEWS) scores of Emory students who reported having friends or family in the affected area. Table 3 shows that it matters: subjects with friends or family in the Bay area had significantly higher scores than those without ($U=3.86$, $z=2.08$, $P<0.05$). The mean CAP(ENEWS) for those involved in this way was 67.46, only slightly below CAP(BNEWS) in Berkeley. Surprisingly, however, this difference was not reflected in mean affect ratings (see Table 3). These did not differ significantly; in fact, the mean for informants *without* such involvement was slightly higher. Moreover, there was no significant correlation between rated emotional reaction and accuracy in either group.

Confidence

A measure of each informant's overall confidence was obtained by averaging his or her confidence ratings on the same three scales (place, activity, others present) that contributed to the CAP score. For recall of direct experience these confidences were at ceiling, averaging 6.89 on the 7-point scale used in Berkeley, 5.00 on the 5-point scale in Santa Cruz. Subjects were somewhat less confident of their recall of the Bay Bridge news event: 6.33 and 3.96 respectively. At Emory, where a 5-point scale was used, the mean confidence rating was 3.64.

The question of whether subjects' confidence ratings were significantly correlated with their accuracy is meaningless for direct experience, as CAP(DIR) scores and confidence ratings were at ceiling in both California samples. They were not at ceiling for recall of the Bay Bridge news, but both distributions were still so skewed towards high values that the correlations between them ($r=0.05$ in Berkeley, $r=0.32$ in Santa Cruz) cannot easily be tested for statistical significance. The correlation in the Emory group was $r=0.31$, but even there the distribution of confidence was skewed towards high

TABLE 3
Ratings of Affect and Accuracy for Emory Students With and Without Relations and Friends in California

<i>Family or Friends</i>	<i>N</i>	<i>Mean Rating of Affect</i>	<i>CAP</i>	<i>Correlations between Affect and CAP</i>
YES	21	3.43	67.46	-0.32
NO	53	3.72	49.68	0.13
ALL	74	3.61	54.73	0.02

values, while the distribution of CAP(ENEWS) was U-shaped with high frequencies at both 0 and 100.

Comparisons with the Space Shuttle study

For some purposes, it is useful to regard the Emory group tested here as a partial replication of Neisser and Harsch (1992). In both cases a group of Emory students were asked to recall how they had heard the news of a distant disaster: here the 1989 earthquake, then the 1986 explosion of the shuttle *Challenger*. Neisser and Harsch got initial reports from their informants one day after the event and asked for recalls some three years later; we got initial reports two or nine days after the event and recalls only one-and-a-half years later. Given these differences, it is not surprising that our subjects were more accurate. The CAP mean for the Emory earthquake group was 55, while the corresponding mean for the space shuttle subjects was only 42.⁴ The shorter delay may also account for the fact that 87% of our earthquake group remembered filling out the first questionnaire, substantially more than the 25% in the shuttle study.

More surprising, perhaps, was an observed difference in confidence. The mean of our Emory confidence ratings⁵ was 3.64; on the same 5-point scale, the space shuttle subjects had averaged no less than 4.17! (A Mann-Whitney *U*-test marks this difference as significant, $U=1180$, $Z=2.11$, $P<0.05$). Only 16% ($N=11$) of the Emory earthquake subjects showed perfect confidence (ratings of 5 on all CAP attributes), as compared with 30% ($N=13$) of the shuttle subjects. Moreover, most of the perfectly confident earthquake informants were justified in their confidence: 10 of the 11 had CAP scores of 67 or higher. (The other was subject LG, quoted earlier, whose CAP score was zero.) In contrast the median CAP of the 13 fully confident space shuttle informants was only 29, and three of them were at zero. For some reason, asking people about the space shuttle explosion produced memories that were far more confident—though much less accurate—than those of the earthquake.

It is also worth considering the incidence of what Neisser and Harsch called “TV priority”, i.e. the error of believing that one first heard about an event from television when this was not actually the case. This mistake was common in the space shuttle study, where 12 subjects erred by adding TV to their memory while only two erred by dropping it (a significant difference). The Emory earthquake group showed a similar but less striking trend: 14 subjects erred by saying they had learned of the quake from TV (when they had not), while eight made the opposite mistake.

⁴ Scores on the 7-point scale used by Neisser and Harsch (1992) have been converted to an “accuracy percent” comparable to our 6-point-based CAP scores.

⁵ As already noted, two Emory subjects made no response at all on one of the CAP attributes; four others failed to make a confidence rating on such an attribute. These six have been eliminated from the confidence analyses, leaving $N=70$.

DISCUSSION

Two independent comparisons in these data show that experiencing the earthquake directly was more memorable than just hearing about it. First, the subjects in Berkeley and Santa Cruz who had experienced the earthquake themselves remembered much more than those in Atlanta for whom it had just been a reception event. Second, both California groups remembered their direct experience of the quake better than they remembered learning about the Bay Bridge collapse. How can these differences be explained? Several alternative hypotheses present themselves.

The Arousal Hypothesis

Perhaps the most obvious possibility is that the difference results from emotional arousal. Is it not reasonable to assume that earthquakes are frightening, and that the heightened arousal level they produce improves memory (Christianson, 1992; Gold, 1992)? Perhaps surprisingly, our data do not support this interpretation. For one thing, most of our California subjects were simply not frightened: their ratings of affect and concern were by no means high. It is important to realise that earth tremors are common in California. As noted earlier, more than 90% of our Berkeley sample had experienced at least one previous earthquake. Two of the present authors, both Californians (ETB & CAS), are sure they have been through more quakes than they can now recall. For such subjects, feeling the earth move is neither frightening nor intrinsically memorable.

Even more important is the absence of correlation between affect ratings and CAP scores in any group. The most concerned or frightened subjects did *not* establish the strongest memories. This negative finding contradicts any simple arousal hypothesis. Moreover, the subgroup of Atlanta subjects with friends or relatives in the Bay area, who had relatively high CAP scores, rated their feelings as no stronger than did their peers without such attachments.

These data do not make an arousal hypothesis untenable; they just make it more complicated. Perhaps our affect ratings failed to capture some key aspect of emotional response; perhaps arousal itself is unconscious, or at least goes unrecognised. It is also possible that what matters is not the subject's arousal during the quake itself but during some later memory-consolidation phase. At present, however, these hypotheses remain *ad hoc*.

The Consequentiality Hypothesis

Another variable that must be considered is the "consequentiality" of the event, a term first introduced by Brown and Kulik (1977). They asked subjects to assess "what consequences for my life, both direct and indirect, has this event had", and got very high ratings for the John F. Kennedy assassination. In the same vein, Conway et al. (1994) use "consequentiality" to explain why Britons

remember how they heard the news of Margaret Thatcher's resignation better than do nationals of other countries. We find this argument unsatisfying, especially where reception events are concerned. Although the fact that an assassination or resignation has occurred may have real consequences for citizens of a given country (though even this is not obvious: cf Neisser, 1982, p.46), the *occasion on which they heard about it* surely does not. Nevertheless it is just that occasion which they are asked to remember in these experiments. The subjects of our study were not asked to recall facts about the earthquake itself (which indeed had major consequences) but rather to relate their own personal experiences at the time (which probably did not). Why then was their memory so good?

The Distinctiveness Hypothesis

Some experiences are unique and distinctive; others are familiar, commonplace, similar to much that has gone before. In general, the former are remembered better than the latter. In Brewer's (1988) study for randomly sampled personal events, for example, the uniqueness of an event was the best overall predictor of how well it would be recalled later on. Some of our subjects certainly had unique or distinctive experiences during the earthquake, often including a substantial disruption of their ongoing activity. Winograd and Killinger (1983) suggest that such a disruption can be a source of distinctiveness in its own right.

It is possible that recalls of direct experience were more accurate than recalls of hearing about the Bay Bridge simply because subjects' immediate experiences of the quake were more distinctive than the various ways in which they got the Bridge news. (The bridge-news recalls may also have been more vulnerable to what Brewer {1988} has called "wrong time slice" errors: the subject recalls a reception event that really did happen, but was not the occasion on which he or she *first* heard the news.) Unfortunately, this argument depends entirely on post hoc assessments of distinctiveness.

A further issue concerns the scope of the event whose distinctiveness is being assessed. Even when a California subject's momentary experience of the quake was *not* distinctive, the full set of his or her earthquake-related experiences—the first shock, hearing about the Bay Bridge, talking about it all repeatedly, encountering the sequelae of the quake for days afterwards—must have been unique indeed. This brings us to the final and (in our view) most persuasive explanation of the findings: subjects remembered their earthquake experience so well because they had talked about them so much.

The Narrative Hypothesis

One's personal experience of a major earthquake is definitely worth talking about. Some of its narrative demands are essentially social: friends and relatives want to know how you survived, acquaintances want to compare your story with

their own. Others are more ego-related: one achieves a sort of vicarious importance in recounting one's experience of the big quake, even if that experience was unremarkable. (Consider the three Berkeley informants mentioned in Footnote 1, who did not even notice that an earthquake was happening: later, they nevertheless gave good accounts of what they had been doing at that moment!) In any case, most of our Californian informants surely told their earthquake stories many times.⁶ Atlanta subjects, in contrast, had little reason to talk about how they had heard the news of an earthquake on the other side of the continent.

Our assumptions about the frequency of narrative rehearsal receive independent support from the data of Pennebaker and Harber (1993), who conducted a telephone interview study of reactions to the same earthquake. Their informants from the San Francisco Bay area (unlike those from control cities) reported very high rates of "talking about the earthquake" in the first two weeks after the event. In the third week those rates began to diminish, although informants reported that they were still *thinking* about the quake for several weeks more. People eventually got bored: T-shirts reading "Thank you for not sharing your earthquake experience" appeared in the Bay area after about a month.

The narrative rehearsal hypothesis may also explain another aspect of the data. The collapse of the Bay Bridge was the defining moment of this particular earthquake, the feature that established it as "big". For this reason many individuals may have incorporated the occasion of hearing about the bridge into their narratives, and rehearsed it nearly as often as their direct experiences.

The relation between these two memories is particularly intriguing. We have argued that, in Berkeley at least, most people's direct experience of the earthquake was neither dramatic nor unique. It became memorable only in view of what was learned afterwards, including (but not limited to) the news about the Bay Bridge. Brown and Kulik's (1977) "novelty" and "consequentiality" hypotheses make the wrong prediction here. It was the Bay Bridge news, not

⁶ In an attempt to get some measure of "rehearsal", the first Berkeley questionnaire asked informants to estimate how much total time they had spent in "talking about the earthquake". As in previous studies (Neisser & Harsch, 1992), these estimates were unrelated to the accuracy of later recall—e.g. the *r* with CAP(BNEWS) was 0.05. Such estimates are intrinsically flawed: not only is it hard to recall and combine time intervals, but the question did not specifically refer to how long they had spent in relating *their own experiences*. Moreover, an internal contradiction suggests that such estimates cannot be taken at face value: the total times reported by subjects interviewed on 19/20 October were *smaller* (on average) than those from 18 October, although the former had had at least one extra day to talk! This incongruity appeared in two independent groups: (a) 234 informants who completed only the first questionnaire (the 18 October subjects averaged 4.3 hours; the 19/20 October subjects averaged 4.1 hours); (b) 36 informants who—having answered this question initially—later completed the follow-up (4.7 & 3.4 hours respectively). It is easy to think of explanations for this pattern of results; our point is only that such estimates are not valid measures of rehearsal.

their own direct experiences of the earthquake, that was novel and consequential for these subjects. Nevertheless, the direct experiences were remembered much better than the occasions on which they learned about the bridge.

The hypothesis that memory for a given event can be enhanced by later developments (in this case, by learning how “big” the quake had been) has been proposed before (e.g. Linton, 1982). A phone call that asks whether you might be interested in a certain job becomes memorable if you accept that job and it changes your life; otherwise, it will probably be forgotten. Your first date with your future spouse may not have been all that remarkable; you still remember it, but only because it later became a turning point in your life narrative. Some years ago one of us gave a similar interpretation of Brown and Kulik’s (1977) Kennedy flashbulbs: “{Such experiences} are the places where we line up our own lives with the course of history and say ‘I was there’” (Neisser, 1982, p.48). In the same way, experiencing a relatively modest earth tremor becomes memorable if—and only if—it turns out to be part of an historical disaster.

The Atlanta subjects who had friends or relatives in the Bay Area are a particularly interesting group. Although they did not rate themselves as especially aroused or upset, they had substantially higher CAP scores than those without such links to the event itself. This advantage, too, can be explained in terms of narrative and rehearsal. Most of these subjects had surely called up their California friends after the quake, just to see if they were all right. In most cases the resulting conversation must have included an eager exchange of earthquake narratives, one of which was the Atlanta subject’s account of how he or she had first heard the news.

Although we are convinced that it was the subsequently established narratives (rather than the immediate levels of arousal) that made these experiences so memorable, the details of that process are not entirely clear. We do not know whether establishing a narrative strengthens memory because it increases the distinctiveness of the experience as a whole or because it is actually rehearsed on many occasions (or both). Rehearsal and distinctiveness are never easy to separate in naturalistic studies: the distinctive stories are the ones that tend to get rehearsed. Brewer (1992) has recently emphasised the joint roles of rehearsal and distinctiveness while noting the difficulty of teasing them apart.

Sources of Error. We believe it was the narratives that our California informants constructed—after the event—that made their earthquake experiences so memorable. But why were those particular narratives accurate when many other recollections (including those in the Emory group) are not? Several possibilities are worth mentioning.

First, there may have been crucial differences in the frequency and spacing of rehearsal. Given the ubiquity of earthquake talk in the first few hours after the event, all our Californian informants had probably retold their stories before the day was out. Equally important, in view of the efficacy of spaced retrieval

(Bjork, 1988), is that they probably found more opportunities to do so in subsequent days and weeks. In contrast, Emory students probably had few occasions, even on day one, to rehearse the reception events about which we would later question them. (The quake itself may have been a topic of discussion even at Emory, but there was little reason to review how anyone happened to hear about it.) Thus the reception-event narratives would have been less well established and sooner forgotten, leaving more scope for constructive processes and potential error.

Second, differences in how informants are situated with respect to the focal event can affect the weight given to their narratives. In the case of direct experience, each informant has a real link to the big event: he or she had felt the *very same quake* that devastated San Francisco. Narratives can emphasise those links in many ways. One of our favourites is the *how I didn't die* story: "If it wasn't for . . . {some unusual departure from routine}, I would have been crossing the Bay Bridge at the very moment when . . ." Nearness to real danger gives such stories a distinctiveness that few accounts of "hearing the news" can match.

These hypotheses are speculative. Many relevant variables remain uncontrolled, and there may be others that we have not considered. Whatever the proper interpretation of our findings, they may at least offer some counterpoise to the many current demonstrations of the fallibility of memory. Recall can be accurate, even if it takes an earthquake to make it so.

Manuscript received 3 February 1995

Manuscript accepted 21 September 1995

REFERENCES

- Bjork, R.A. (1988). Retrieval practice and the maintenance of knowledge. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp.396–401). Chichester, UK: Wiley.
- Bohannon, J.N.I., III, & Symons, V.L. (1992). Flashbulb memories: Confidence, consistency, and quantity. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp.65–91). New York: Cambridge University Press.
- Brewer, W.F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp.21–90). New York: Cambridge University Press.
- Brewer, W.F. (1992). The theoretical and empirical status of the flashbulb memory hypothesis. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp.274–305). New York: Cambridge University Press.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, 5, 73–99.
- Christianson, S.-A. (1989). Flashbulb memories: Special, but not so special. *Memory and Cognition*, 17, 435–443.
- Christianson, S.-A. (Ed.). (1992). *Handbook of emotion and memory: Research and theory*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Conway, M.A., Anderson, S.J., Larsen, S.F., Donnelly, C.M., McDaniel, M.A., McClelland, A.G.R., Rawles, R.E., & Logie, R.H. (1994). The formation of flashbulb memories. *Memory and Cognition*, 22, 326–343.

- Friedman, W.J. (1987). A follow-up to "Scale effects in memory for the time of events": The earthquake study. *Memory and Cognition*, 15, 518–520.
- Gold, P.E. (1992). A proposed neurobiological basis for regulating memory storage for significant events. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp.141–161). New York: Cambridge University Press.
- Goodman, G.S., Rudy, L., Bottoms, B.L., & Aman, C. (1990). Children's concerns and memory: Issues of ecological validity in the study of children's eyewitness testimony. In R. Fivush & J.A. Hudson (Eds.), *Knowing and remembering in young children* (pp.249–284). New York: Cambridge University Press.
- Greenwald, A.G., & Banaji, M. (1989). The self as a memory system: Powerful but ordinary. *Journal of Personality and Social Psychology*, 57, 41–54.
- Hyman, I.E. Jr., & Neisser, U. (1992). The role of the self in recollections of a seminar. *Journal of Narrative and Life History*, 2, 81–103.
- Larsen, S.F. (1988). Remembering without experiencing: Memory for reported events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp.326–355). New York: Cambridge University Press.
- Larsen, S.F. (1992). Potential flashbulbs: Memory of ordinary news as the baseline. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp.32–64). New York: Cambridge University Press.
- Linton, M. (1982). Transformations of memory in everyday life. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp.77–91). New York: Freeman.
- Neisser, U. (1982). Flashbulbs or benchmarks? In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp.43–48). New York: Freeman.
- Neisser, U. (1986). Remembering Pearl Harbor: Reply to Thompson and Cowan. *Cognition*, 23, 285–286.
- Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp.9–31). New York: Cambridge University Press.
- Pennebaker, J.W., & Harber, K.D. (1993). A social stage model of collective coping: The Loma Prieta earthquake and the Persian Gulf war. *Journal of Social Issues*, 49, 125–145.
- Pillemer, D.B., Koff, E., Rhinehart, E.D., & Rierden, J. (1987). Flashbulb memories of menarche and adult menstrual distress. *Journal of Adolescence*, 10, 187–189.
- Rubin, D.C., & Kozin, M. (1984). Vivid memories. *Cognition*, 16, 81–95.
- Stern, W. (1904/1982). Realistic experiments. *Beitrag zur Psychologie der Aussage*, 2, 1–31. {Translated and reprinted in U. Neisser (Ed.) *Memory observed: Remembering in natural contexts* (pp.95–108). New York: Freeman.}
- Thompson, C.P., & Cowan, T. (1986). Flashbulb memories: A nicer interpretation of a Neisser recollection. *Cognition*, 22, 199–200.
- Warren, A.R., & Swartwood, J.N. (1992). Developmental issues in flashbulb memory research: Children recall the Challenger event. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp.95–120). New York: Cambridge University Press.
- Wells, G.L., & Murray, D.M. (1984). Eyewitness confidence. In G.L. Wells & E.F. Loftus (Eds.), *Eyewitness testimony* (pp.155–170). New York: Cambridge University Press.
- Winograd, E., & Killinger, W.A. Jr. (1983). Relating age at encoding in early childhood to adult recall: Development of flashbulb memories. *Journal of Experimental Psychology: General*, 112, 413–422.
- Wood, J.M., Bootzin, R.R., Rosenhan, D., Nolen-Hoeksema, S., & Jourden, F. (1992). Effects of the 1989 San Francisco earthquake on frequency and content of nightmares. *Journal of Abnormal Psychology*, 101, 219–224.

