Memory & Cognition
2004, 32 (3), 443-454

Autobiographical memories for the September 11th attacks: Reconstructive errors and emotional impairment of memory

STEPHEN R. SCHMIDT
Middle Tennessee State University, Murfreesboro, Tennessee

College students were asked about their personal memories from September 11, 2001. Consistency in reported features over a 2-month period increased as the delay between the initial test and 9/11 increased. Central features (e.g., Where were you?) were reported with greater consistency than were peripheral features (What were you wearing?) but also contained a larger proportion of reconstructive errors. In addition, highly emotional participants demonstrated poor prospective memory and relatively inconsistent memory for peripheral details, when compared with less emotional participants. Highly emotional participants were also more likely to increase the specificity of their responses over time but did not exhibit greater consistency for central details than did less emotional participants. The results demonstrated reconstructive processes in the memory for a highly consequential and emotional event and emotional impairment of memory processing of incidental details.

On September 11, 2001, terrorists crashed planes into the World Trade Center in New York and into the Pentagon in Washington. Millions viewed and listened to the important events of the day as they unfolded in the national media. Travel plans throughout the country were disrupted, the stock markets plunged, and U.S. foreign policy changed dramatically. The psychological toll was equally drastic, inasmuch as 44% of Americans reported substantial stress symptoms (Schuster et al., 2001) and visits to counseling centers (Gallagher, 2003) and reports of insomnia (National Sleep Foundation, 2003) increased. Tragic as they were, the events of 9/11 provided a unique opportunity for investigating autobiographical memories for momentous events.

Brown and Kulik (1977) developed the flashbulb memory hypothesis in an attempt to explain people's vivid recollections of just such momentous events. They argued for a three-stage process: "First comes the recognition of high novelty or unexpectedness; then comes a test for biological meaning for the individual; if this second test is met, there follows the permanent registration not only of the significant novelty, but of all recent brain events" (p. 76). Most researchers interpret the flashbulb memory hypothesis to imply that we record a detailed and indelible record of the experiences surrounding a highly emotional or consequential event (e.g., Christianson, 1989; McCloskey, Wible, & Cohen, 1988). Support for the flashbulb memory hypothesis can be found in numerous studies reporting vivid and detailed memories for consequential events, including Brown and Kulik, Conway (1995), Conway et al. (1994), and Finkenauer et al. (1998). For example, Conway et al. reported that both the surprise and the intensity of an emotional response to Margaret Thatcher's resignation were reliable predictors of consistent flashbulb reports.

Despite this empirical support for the flashbulb memory hypothesis, numerous investigators have argued that all memories, even those for highly emotional and consequential events, are subject to reconstructive errors and are, therefore, not the indelible reproductions suggested by Brown and Kulik (1977). McCloskey et al. (1988) found that memories for the explosion of the space shuttle *Challenger* were often inconsistent over time and subject to errors. Christianson (1989) reported that the narrative substance of flashbulb memories remained intact but that memory for details of a flashbulb experience was often inconsistent. Schmolck, Buffalo, and Squire (2000) demonstrated that autobiographical memories concerning the O. J. Simpson verdict were subject to increasing numbers of distortions with increasing retention intervals. Winningham, Hyman, and Dinnel (2000) reported that consistency in autobiographical reports increased as the delay between the event and an initial report of the memory increased. The high levels of consistency reported in some studies of flashbulb memories may result from the long initial delay between the event and the first autobiographical report, rather than from the veracity of the initial report.

Further doubt concerning the flashbulb memory hypothesis has arisen from laboratory investigations into

I give special thanks to the faculty and students of the psychology department for their help in collecting these data. Several students worked closely with me in the conduct of this research, including Abby Hughes, Heather Klingensmith, Mark Mullen, Mary Catherine Powers, and Amanda Turnbo. I also thank Constance R. Schmidt for her comments on earlier drafts of this article. Correspondence concerning this article should be addressed to S. R. Schmidt, Psychology Department, Middle Tennessee State University, Murfreesboro, TN 37132 (e-mail: sschmidt@mtsu.edu).

the role of strong emotions in the formation of memories. Researchers have argued that strong emotion does not lead to either a complete record of the emotional event or vivid recollection of insignificant peripheral details, as is implied by the flashbulb memory hypothesis. Rather, strong emotion may create a relatively detailed and accurate memory for the central aspects of an event but poor memory for peripheral information. This attention-focusing idea has several guises, including Easterbrook's (1959) *cue-utilization* hypothesis, *tunnel vision* (Mackworth, 1965), the *weapon focus* (Loftus, 1979), and Christianson's (1992a, 1992b) two-stage model. Numerous laboratory experiments, as well as field studies of the weapon focus (Steblay, 1992), have provided support for attention focusing (see Christianson, 1992a, for a review). For example, Burke, Heuer, and Reisberg (1992) had participants view a series of slides depicting the story of a mother and her son visiting the boy's father at work. Memory for central details was generally better in an emotional condition (a view of open heart surgery) than in a neutral condition. In contrast, memory for background details from the emotional slides was worse than memory for the background details from the corresponding neutral slides.

These criticisms of the flashbulb memory hypothesis have several shortcomings as well. First, the events investigated in some tests of the flashbulb hypothesis (e.g., the O. J. Simpson verdict) may not have been particularly significant or consequential for many of the participants (Conway et al., 1994). Second, the memory inaccuracies reported in the flashbulb memory studies do not necessarily provide evidence for reconstructive processes. Such inaccuracies may result from interference, retrieval failures, or variability in the retrieval of information across contexts (see Brewer, 1992, for a discussion). Only systematic memory errors that can be linked to specific reconstructive processes will provide unambiguous evidence for the reconstructive nature of flashbulb memory. Alba and Hasher (1983) have described four principles of schematic processing that provide evidence for the reconstructive nature of memory: selection, interpretation, abstraction, and integration. Similarly, Johnson (1962) has described sharpening and normalization as indices of reconstructive memory. Thus, specific evidence for reconstruction would be found if people reported the central circumstances of the flashbulb event to the detriment of unimportant peripheral details (schematic selection), filling in missing or forgotten central details (interpretations or normalizations). The gist of the events should be retained, whereas the exact form of the events should be forgotten (abstraction). Also, events spread over several hours or days should become integrated into one autobiographical episode.

Criticisms of the flashbulb memory hypothesis are also weakened by the lack of definitive evidence for attention focusing in studies of autobiographical memory. Some weak support for focusing was reported by Christianson (1989). Swedish participants were asked to report memories associated with their discovery in 1986 that their prime minister had been assassinated. Consistency in reported memories was studied across a 1-year retention interval. The participants were more consistent in reporting how they had heard the news, where they were, what they where doing, and who they were with than in reporting what clothes they were wearing. Furthermore, memory consistency for these central details exceeded consistency for a control (nonflashbulb) event over the same retention interval. Similar findings were reported by Weaver (1993) in his study of people's memory for the beginning of the U.S. bombing raids in the 1991 Gulf War. Berntsen (2002) found that autobiographical memories for negative (but not positive) events contained more central than peripheral details.

However, these studies can be interpreted within the framework of schematic selection, wherein information relevant to the gist of an emotional event is selected and schematically irrelevant features are lost. The attention-focusing idea posits the allocation of attentional resources to central features of an emotional event to the detriment of peripheral processing (Christianson, 1992a). Strong support for this idea requires a detailed comparison of memory for central and peripheral details between an emotional (flashbulb) event and an unemotional control event. However, Christianson (1989) and Weaver (1993) reported memory for the control event in one combined score, failing to distinguish between memories for central and peripheral details. Berntsen (2002) looked at the proportion of central features reported for emotional events but did not provide comparable data for nonemotional control events.

Supporting evidence for attention focusing might also be found if memory for central and peripheral details could be compared between high- and low-emotion participants. High-emotion participants should have better memory for central features and worse memory for peripheral details than do low-emotion participants. Unfortunately, even though Christianson (1989), Weaver (1993), and Berntsen (2002) measured the intensity of emotional reactions, intensity was not predictive of difference in memory for central information.

A series of experiments reported by Schmidt (2002) cast further doubt on the attention-focusing hypothesis. Pictures of nudes were employed as emotional stimuli, and they were embedded in a series of pictures of nonnude people. As compared with a control picture of a clothed person, good memory for the gist of the picture of the nude occurred at the expense of memory for background picture details and for pictures following the nude. However, the emotional stimulus (the nude) was also a distinctive stimulus (it was the only picture of a nude in the series). Further research revealed that the gist of a picture containing a clothed person embedded in a series of pictures containing nudes was better remembered than a picture of a nude presented in a series of nudes, illustrating that access to the pictures was greatly influenced by distinctiveness. Once picture access was

equated, either statistically or through list structure, central details were not remembered better from the emotional than from the nonemotional pictures, challenging the attention-focusing idea. Schmidt (2002) concluded that the impact of emotion on memory is that it impairs the processing of peripheral details. Positive effects of emotion on memory in support of the attention-focusing hypothesis may be due to the increased access to the gist of an event when the emotionally significant event is also distinctive.[1]

A definitive test of the attention-focusing hypothesis requires a clear differentiation between central and peripheral information, as well as equal access to the gist of the event across conditions. Although the *focusing* idea invokes a camera metaphor (as does the flashbulb memory hypothesis), attention focusing is often framed within a constructivist framework. For example, Pillemer (1998) argued that our autobiographical memories are active constructions that contain a thematic narrative of the event. As such, our personal event memories should include a "particular time and place" (p. 50). Christianson (1992a) argued that "memory for central and peripheral emotional information refer[s] to a distinction between information that is connected with the source of the emotional arousal (i.e., the gist of the event and its central details) and information that is irrelevant or spatially peripheral to the source of the emotional arousal" (p. 291). In Berntsen's (2002) study, features were classified as central if they were related to the shocking event and could not be omitted without changing the nature of the event. In the present study, four features were chosen as central information because they seemed necessary for a cogent narrative. If one uses the core elements of journalistic reporting, the narrative should include answers to the questions of who, what, when, and where. Peripheral features were chosen that should have been incidental to most people's narratives and yet were still given some attention during the course of the day. For example, one's choice of clothing on 9/11 would require some attention but would probably be irrelevant to the central narrative of the events of the day. Note, however, that lack of clothing might have been a relevant feature of some people's autobiographical memories if, for example, they were dragged out of the shower and shown a building collapsing on TV.

The research reported here had two primary goals: to provide a more definitive evaluation of the reconstructive view of autobiographical memories for a highly significant event and to provide a clear test of the attention-focusing account of autobiographical memories. A large sample of undergraduate students was asked to fill out questionnaires concerning their autobiographical memories for the September 11th attacks over the week following the attacks. A large number of these students then filled out a second questionnaire approximately 2 months later. To test the attention-focusing idea, comparisons were made between memory for central and peripheral details and between students who rated themselves as having a high emotional response to the 9/11 event and students with a lower rating of emotion. The nature of reconstructive errors for central and peripheral information was examined as a function of rated emotion. The relatively short retention interval was employed in part to capture a large number of students. However, the short retention interval should ensure that most of the participants still had access to the gist of the day's events. An evaluation of reconstructive errors over a short retention interval should demonstrate the ubiquitous role of reconstruction in autobiographical memory. A *delayed-test-only* group was included to determine the potential impact of the first memory test on the second test.

METHOD

Participants

A general request to participate in the research was sent to all undergraduate psychology instructors at Middle Tennessee State University. The classes of those instructors who volunteered class time were placed into two approximately equal groups on the basis of enrollment estimates: an initial-testing group and a delayed-testing-only group. Most of the participants were enrolled in introductory psychology courses. The participants were also drawn from other lower division courses, with the constraint that different sections of the same course provided the participants in the initial-testing and in the delayed-testing-only conditions. All the students voluntarily participated without financial reward or class credit given for their participation. The initial testing included 493 students (141 males, 324 females, and 28 students who did not indicate gender). Of these, 211 (64 males, 144 females, and 3 unknowns) could be matched with questionnaires completed in November, forming the November\September group. The delayed-only group included 272 students (92 males, 169 females, and 11 unknowns). Employing enrollment figures to estimate return rate, 78% of the students enrolled in surveyed sections completed the initial questionnaire in September. This figure dropped to 45% in the November-only sections. Of those students who gave identifying information in September, 45% were matched with questionnaires completed in November. Thus, the return rate of the repeated tested group was equivalent to the return rate of the delayed group.

Materials

A questionnaire was developed by adapting instruments used by previous researchers (e.g., Christianson, 1989; Weaver, 1993). Both the questionnaire and the scoring procedure outlined below were pretested on a small sample of students who were asked to report their memories for Princess Diana's death (Hatmaker & Schmidt, 1998). Several questions tapped features thought to be central to autobiographical memory, including "who told you," "what were you doing," "where were you," and "what time did you hear the news." The participants were also asked to report their "first thought" and "most vivid impression" from the events. Several questions assessed the potential level of emotional and personal involvement in the attacks, whereas others addressed the extent to which the attack disrupted the participant's day. Finally, four questions tapped what were thought to be thematically incidental or peripheral information: "What were you wearing," "what did you eat for breakfast," "what did you eat for lunch," and "what was the weather like." Due to the length of the questionnaire, confidence judgments were collected only on the combined *who*, *what*, and *where* questions and on the *when* question. A second questionnaire, which included several minor edits, was administered in November. The complete questionnaires can be found in the Appendix.

Procedure

The questionnaires were first distributed to psychology instructors on September 12, 2001, and the students completed the ques-

tionnaires during one of their regular class meetings. Approximately 2 months later, the second questionnaire was distributed to the same instructors, and questionnaires were given to the second group of instructors, whose students formed the delayed-test-only group.

RESULTS

The presentation of the results is broken down into two sections. In the first section, the participants' reactions to the events, including the number of features reported, their emotional responses, their responses to the prospective memory questions, and their confidence ratings, are summarized. Primary attention is given to the second section, where consistency in reported autobiographical memories is evaluated.

Reactions to the Events

Table 1 includes a summary of some of the reactions to the events of September 11th that were tapped by the questionnaire. Focusing on the September group (the largest sample), the participants tended to rate themselves high on the emotion scale ($M = 6.21$ on a 7-point scale) and report that the events interrupted their activities for the day ($M = 5.47$). The correlation between participants' emotional ratings and their ratings of changed activity was high and statistically significant [$r(476) = .95$]. (An alpha level of .05 was employed in this and all subsequent analyses.) Males and females were equally affected by the events, with $F(1,447) = 2.82$, $MS_e = 1.16$, for emotion and $F(1,447) = 1.33$, $MS_e = 2.32$, for changed activity. In our region of the country, personal involvement in the events was relatively low ($M = 1.99$). Personal involvement was significantly correlated with a change in activity, but the correlation was rather small [$r(474) = .10$].

Some interesting differences emerged from the comparison of responses to the first and the second questionnaires. For example, modest declines were observed in ratings of emotion and changed activity over the 2-month period (see Table 1). These declines were significant both in the between-groups comparison between those tested in September and those tested only in November [$F(1,739) = 4.87$, $MS_e = 1.42$, for emotion and [$F(1,758) = 11.80$, $MS_e = 2.66$, for changed activity] and in the within-groups comparison of participants tested in both September and November [$F(1,210) = 9.36$, $MS_e = 0.51$, for emotion and $F(1,210) = 18.71$, $MS_e = 0.85$, for changed activity].

Table 1
Descriptive Statistics for the Three Samples, Including Mean Ratings of Emotion, Changed Activity, and Personal Involvement, Proportion of Respondents Indicating That They Were Late or Forgot an Appointment, and Confidence Ratings on the Memory Questions

Question	September	November Only	November\ September
M emotion	6.21	6.00	6.05
M changed activity	5.47	5.04	5.00
M personal involvement	1.99	1.97	1.98
Proportion *late*	.34	.30	.30
Proportion *forgot*	.42	.30	.24
M confidence on			
Combined question	2.94	2.94	2.79
When	2.79	2.70	2.51

Further evidence that the events disrupted our students' lives was found in their reports of whether they forgot an important obligation or were late for class or some other appointment. In September, 42% reported some forgetting, and 34% reported that they were late. Interestingly, the rate of reported forgetting significantly declined from the first to the second assessment, to 30% in the November-only group [$\chi^2(1) = 9.79$] and from 40% to 24% in the November\September group [$\chi^2(1) = 21.78$]. Reporting being late declined to 30% in the November-only group and from 31% to 30% in the November/September groups, but these changes were not significant. These results mirror those reported above, in that the participants seemed to retrospectively reduce the immediate impact of the day's events on their lives. Focusing on the September group (the largest sample with the most immediate memory), these results also provided some evidence for attention focusing or an emotional cost. The participants rating themselves as a 7 on the emotional scale were more likely to report being late (39%) than were the participants rating themselves below a 7 [27%; $\chi^2(1) = 7.15$]. Similarly, the participants with a 7 on the emotional scale were more likely to report having forgotten something (46%) than were the participants with a lower emotional rating [37%; $\chi^2(1) = 4.30$].

Analyses of the confidence data of the September and the November groups revealed that the participants were more confident in their answers to the combined question ($M = 2.94$) than in their answers to the *when* question [$M = 2.76$; $F(1,726) = 124.24$, $MS_e = 0.1028$]. Further between-groups comparisons revealed a decline in confidence over the 2-month period [$F(1,726) = 5.19$, $MS_e = 0.1550$]. Similarly, confidence declined over time in the within-subjects analysis for the participants tested in both September and November [$F(1,188) = 61.89$, $MS_e = 0.16$]. However, the interactions between time of test and question were reliable in both the between-groups [$F(1,726) = 6.47$, $MS_e = 0.10$] and the within-groups [$F(1,188) = 9.10$, $MS_e = 0.08$] analyses. Confidence in the *when* question declined more than confidence in the combined question (see Table 1). The November-only group was more confident in their answers to the combined question ($M = 2.94$) than was the November\September group [$M = 2.79$; $F(2,461) = 22.07$, $MS_e = 0.11$] and in their response to the *when* question [*M*s = 2.70 and 2.51, respectively; $F(2,461) = 16.49$, $MS_e = 0.25$]. These results suggest that repeated testing lowers confidence.

The mean numbers of autobiographical features reported on both memory questionnaires are summarized in Table 2. Comparing the first and the second questionnaires, there was a modest decline in the number of features reported, a trend that was observed in both the

Table 2
Proportion of Participants Reporting Each of the 10 Autobiographical Features of the Events

Question	September	November Only	November\ September
Who told you	1.00	1.00	1.00
Where were you	1.00	1.00	1.00
What activity	1.00	1.00	1.00
When	.99	.99	.98
First thought	.99	.99	.99
Image	.99	.99	.99
Clothing	.95	.77	.87
Weather	.98	.94	.97
Breakfast	.95	.87	.83
Lunch	.91	.69	.66

between- and the within-groups comparisons. In the between-groups comparison, the mean number of features reported in September was 9.65 (out of 10 possible features), and in November that mean dropped to 9.10 [$F(1,763) = 55.58$, $MS_e = 0.95$]. In the within-groups comparison, the number of features reported dropped from 9.79 to 9.27 [$F(1,210) = 58.34$, $MS_e = 0.48$]. The decline in the number of features reported occurred only for the peripheral features of clothing, breakfast, and lunch (see Table 2). Only one difference was observed between the November-only and the November\September groups: The November\September group was more likely to report their clothing than was the group tested only in November [$\chi^2(1) = 8.17$]. These results demonstrate a strong tendency on the part of the participants to provide information for the 10 autobiographical features tested in this survey. Furthermore, the reports of these features dropped only modestly over the course of 2 months.

Consistency in Autobiographical Reports

Previously, researchers have scored memory consistency by using strict and lenient criteria (e.g., Christianson, 1989; McCloskey et al., 1988; Weaver, 1993) or by assigning a memory score reflecting overall accuracy (e.g., Conway et al., 1994). However, such scoring reveals only the degree of consistency and is insensitive to specific reconstructive errors. As was noted earlier, reports that do not meet even a lenient criterion may be the result of any number of forgetting processes, including retrieval failures and schematic reconstructions. To provide a more sensitive test for reconstructive processes, answers to each question were placed into one of six categories. A summary of these categories, along with examples, can be found in Table 3.

Questions that were left blank on both surveys were scored as *blanks*. As was noted, most participants provided at least partial answers to all of the questions, so the blanks category was rarely used. Responses that were essentially the same on both questionnaires were scored as *consistent*. The consistent category required nearly an exact match between first and second responses, so that it was equivalent to the *strict* scoring employed by previous researchers. Responses were judged as *more specific* if, on the second test, details were added to the answer. For example, one participant noted that she was "in my room" on the September questionnaire and was "in my bed " on the November questionnaire. The rare question that was left blank of the first test but answered on the second test was also scored as more specific. More specific responses are theoretically interesting because they may provide evidence for schematic processes, such as elaboration or interpretation. A fourth scoring category was used for responses that were *more general*. More general responses are simply the converse of more specific responses. For example, on the first test, one participant responded to the "who told you" question with "Channel 2 news." On the second test, she responded with the more general "on the news." A fifth category was reserved for responses on the second test that were *inconsistent* with, contradicted, or did not share anything in common with, first test responses. For example, one

Table 3
Reconstructive Scoring Criteria With Examples From the Questions "Where Were You?" and "What Was the Weather Like?"

Scoring Category	Description	Examples (First Response/Second Response)
Blank	Left blank or "don't know" on both tests.	don't remember/don't remember
Consistent	The same elements mentioned. Exact wording not required.	at home/at home sunny and warm/warm and sunny
More specific	Additional details provided. Word or adjective added. Left blank first time, filled in second.	at home/at home in bed sunny and warm/sunny and 80
More general	Less specific. Substituted generic term. Word or adjective omitted.	at home in bed/at home sunny and warm/sunny
Inconsistent	No common elements. Contradictions of details.	in Joe's room/in the hall in bed/at home sunny and warm/sunny and hot
Omissions	Completed on first test, blank on second test.	at home/_____ sunny/don't remember

student wrote on the first test that "I saw and heard from T.V." On the second test she wrote "my mom." The final category of memory errors was *omissions*. These are simply questions that were answered on the first test, but not on the second test. Scoring procedures employed in many previous flashbulb studies can be easily derived from the six scoring categories reported here. Strict scoring is the same as *consistent* recall, and lenient scoring is simply the sum of the consistent, more specific, and more general scoring categories.

The "when did you hear the news" question required special consideration in scoring. Memory for the events of September 11th has been firmly associated with an exact date, with 94% of the September respondents identifying the date they heard the news as 9/11. This proportion increased to 98% for both November groups, revealing an interesting normalization in memory for date. Over 98% of the September participants reported that they had heard about it on a Tuesday, and this dropped to 85% and 87% for the November\September and the November-only participants, respectively. The November\September group was 82% consistent in reporting the day of the event, with 11% omissions and only 7% inconsistent responses.

Memory for exact time proved much less reliable and, as such, provided a richer source of reconstructive errors. Several researchers have noted that people employ cyclical anchors, such as *weekend* versus *workday*, to reconstruct when autobiographical events occur (e.g., Friedman, 1987, 1993; Thompson, Skowronski, Larsen, & Betz, 1996). The analysis of time of day revealed similar anchors in the distribution of responses over the landmarks of the standard analog clock. Responses falling on either the hour or the half hour accounted for over 50% of the responses, with approximately 20% of reported times landing on the even quarter hour. Furthermore, the tendency to use these landmarks increased over the short period during which the first questionnaire was administered (see Figure 1). Over the course of just a few days, the reporting of the more general landmark times (e.g., 9:00) increased in frequency, whereas the reporting of more specific times (e.g., 9:10) declined in frequency [$\chi^2(8) = 19.53$].

By using the distributions of reported times and the shift in this distribution during the week following September 11th, it is possible to identify a hierarchy of times ranging from the hour to the minute (e.g., 9:00, 9:30, 9:15, 9:10, 9:05).[2] Consistency in reporting times was defined relative to this hierarchy. Any changes that went up the hierarchy (e.g., 9:15 to 9:00) were scored as more general. Times that went down the hierarchy (e.g., 9:15 to 9:10) were scored as more specific. Times at the same level of the hierarchy and within the same 30-min segment (e.g., 9:10 to 9:20) were judged as consistent. Any other changes (e.g., 9:00 to 9:30) were scored as inconsistent. The forward or backward direction of changes in time was irrelevant to the scoring.

Table 4 contains a summary of the consistency scoring for the 10 memory questions. The summary data reveal a remarkably low level of consistency, especially given the short retention period. However, consistency levels for the central features (excluding *when*) in the combined

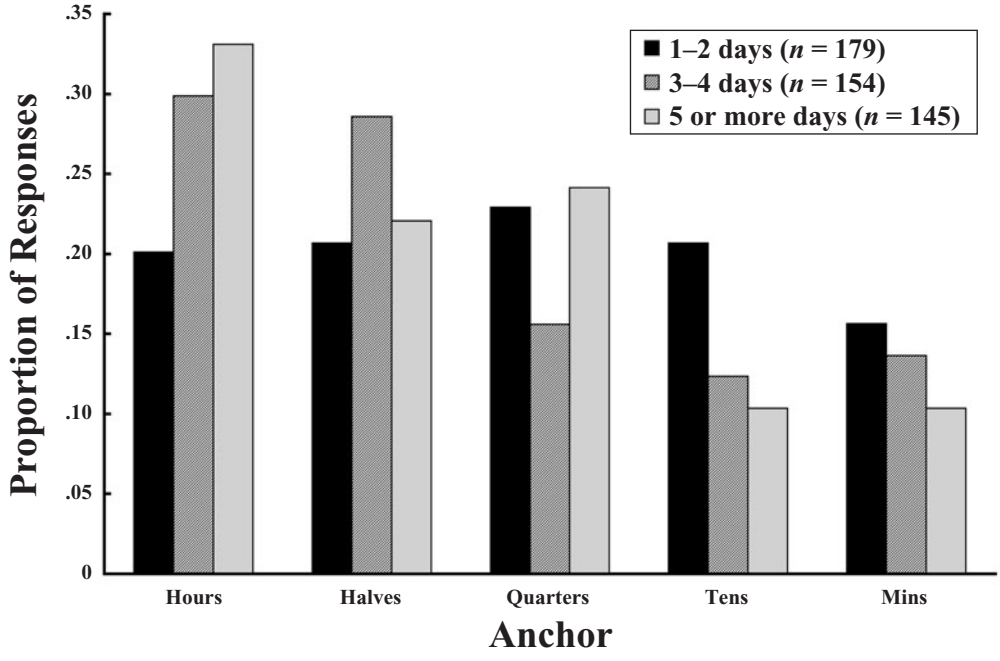

Figure 1. Proportions of participants using various time anchors in reporting *when* they heard the news in the September group as a function of when they completed the questionnaire. The 5-min and 1-min divisions of the clock are combined.

Table 4
Consistent Memories and Memory Errors for the Major Components of Participants' Autobiographical Memories

Component	Consistent	More Specific	More General	Inconsistent	Omissions
Who told you	.63	.07	.15	.15	.00
Where were you	.56	.13	.15	.16	.00
What activity	.50	.09	.19	.22	.00
When	.21	.14	.30	.33	.02
First thought	.29	.05	.21	.44	.00
Image	.22	.15	.27	.35	.01
Clothes	.25	.14	.29	.18	.09
Weather	.26	.06	.26	.40	.01
Breakfast	.40	.04	.05	.34	.14
Lunch	.20	.03	.10	.35	.28

consistent, more general, and more specific categories are similar to the lenient levels of consistency reported in previous flashbulb memory research (e.g., Christianson, 1989; Conway et al., 1994; Weaver, 1993). To determine the impact of emotion on memory, the participants were separated into two emotion groups. The *high-emotion* group contained participants who rated themselves as a 7 on the first test and a 6 or a 7 on the second test (n = 111). The mean emotional rating for this group, collapsed across the two questionnaires, was 6.87. A second group included the remaining participants (n = 100). Because the average emotional ratings for this group were still well above the middle of the rating scale (M = 5.36), I will refer to this as the *moderate-emotion* group. Memory consistency for the 10 memory questions, grouped by emotional reaction, is summarized in Figure 2. The high-emotion group was numerically more consistent than the moderate-emotion group in their answers to three questions: who told you, when did you hear the news, and most memorable image.[3] However, none of these differences approached significance [$\chi^2(1) = 0.75$ for *who*, $\chi^2(1) = 0.94$ for *when*, and $\chi^2(1) = 0.07$ for image]. By contrast, the moderate-emotion group was more consistent than the high-emotion group on six measures, including two that were significant: lunch [$\chi^2(1) = 6.80$] and first thought [$\chi^2(1) = 3.43$, $p < .07$].

The observations underlying the consistency data are essentially binary, and responses are clearly dependent. These properties of the data limit both the types of data analyses that can be performed and the power of those

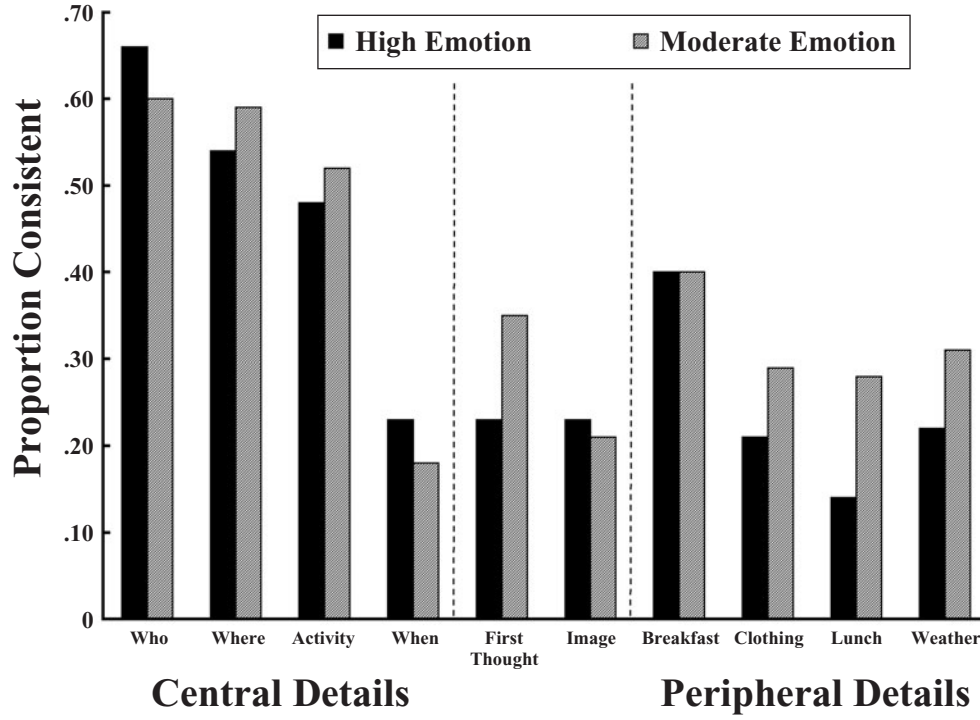

Figure 2. Memory consistency for the 10 autobiographical features as a function of emotional group.

analyses. To provide a strong test of the attention-focusing hypothesis, the memory data were combined into two a priori categories: central information (who told you, what activity, where were you, and when did you hear the news) and peripheral information (clothing, weather, breakfast, and lunch). Each person could earn a score from 1 to 4 for each type of information (central vs. peripheral) for each of five memory types (consistent, more specific, more general, inconsistent, and omissions). A multivariate analysis of variance (MANOVA) was calculated, treating type of information (central vs. peripheral) as a within-subjects factor and emotional group (high vs. moderate) as a between-subjects factor. An additional between-groups factor was created by grouping participants into three date groups on the basis of when they completed the first questionnaire. The groups included participants who completed the questionnaire within 2 days ($n = 85$), 3–4 days ($n = 50$), and 5–9 days ($n = 76$) after the event. The dependent variables in the MANOVA were the proportion of questions answered (out of 4) in each of the five memory categories (consistent, more specific, more general, inconsistent, and omissions). Summaries of these results are reported in Tables 5 and 6.

The MANOVA revealed main effects of type of information [$\Lambda(5,201) = 0.56$] and date group [$\Lambda(10,402) = 0.79$], but no effect of emotional group [$\Lambda(5,201) = 0.96$]. Subsequent univariate ANOVAs revealed that memory was more consistent [$F(1,205) = 71.57$, $MS_e = 0.06$], less inconsistent [$F(1,205) = 18.61$, $MS_e = 0.05$], and included fewer omissions [$F(1,205) = 97.48$, $MS_e = 0.02$] for central than for peripheral information. Alternatively, a greater number of more specific responses were observed for central than for peripheral details [$F(1,205) = 4.61$, $MS_e = 0.02$; see Table 5]. The high-emotion group had a marginally greater number of more specific responses ($M = .10$) than did the moderate-emotion group [$M = .07$; $F(1,205) = 3.38$, $MS_e = 0.02$, $p < .07$]. The ANOVAs also revealed a marginally reliable interaction between type of information and emotional group for inconsistent responses [$F(1,205) = 3.11$, $MS_e = 0.05$, $p < .08$]. The attention-focusing hypothesis made the specific predictions that the high-emotion group should be more consistent on central information and less consistent on peripheral information than the lower emotion group. LSD tests confirmed only the latter half of this prediction: The high-emotion group had a lower number of consistent memories ($M = .24$) for the peripheral information than did the moderate-emotion group [$M = .32$; $t(209) = 2.38$]. Similarly, the high-emotion group had a higher proportion of inconsistent peripheral memories ($M = .34$) than did the moderate-emotion group [$M = .28$; $t(209) = 1.89$]. However, no significant effects of emotion on the consistency of central information were observed.

Table 5
Proportion of Consistent Memories and Memory Errors as a Function of Type of Information and Emotional Group

	Central Information		Peripheral Information	
Measure	High Emotion	Moderate Emotion	High Emotion	Moderate Emotion
Consistent	.48	.49	.24	.32
More specific	.12	.09	.08	.06
More general	.19	.19	.19	.16
Inconsistent	.20	.22	.34	.28
Omissions	.00	.01	.12	.16

One prediction of a reconstructive account of memory is that participants should be more likely to reconstruct central than peripheral details. Support for such reconstructions was found when the analyses above were repeated with the proportion of each of the four types of memory errors (i.e., excluding consistent responses) as the dependent variables. As is predicted by the reconstructive view, the participants were more likely to report more specific information for central ($M = .19$) than for peripheral ($M = .10$) information [$F(1,205) = 9.35$, $MS_e = 0.07$]. The participants were also more likely to report more general information for central ($M = .35$) than for peripheral ($M = .24$) information [$F(1,205) = 10.56$, $MS_e = 0.11$] and more likely to make errors of omission for peripheral ($M = .19$) than for central ($M = .01$) information [$F(1,205) = 84.00$, $MS_e = 0.04$]. Inconsistent responses were not significantly different for central ($M = .34$) and peripheral ($M = .43$) information [$F(1,205) = 1.54$, $MS_e = 0.12$].

The impact of the delay of the initial testing on reconstructive errors was equally informative. Test date had a significant effect on the proportion of consistent responses [$F(2,205) = 8.50$, $MS_e = 0.07$] and on the proportion of more general responses [$F(2,205) = 10.89$, $MS_e = 0.04$]. Consistency increased with the delay of the first test, whereas the proportion of more general responses declined with delay (see Table 6). There was also an effect of delay on the proportion of omissions [$F(2,205) = 5.45$, $MS_e = 0.02$], but this trend was nonmonotonic and may have been contaminated by a floor effect.

Additional analyses were conducted on the consistency data to rule out uninteresting interpretations of the above results. For example, several different methods of scoring time were employed, such as requiring exact recall of time or allowing a 15-min window for time. In one analysis, memory for time was omitted entirely from the measure of central information. In this analysis, central questions led to a larger proportion of more specific [$F(1,159) = 18.07$, $MS_e = 0.10$] and more general [$F(1,159) = 10.77$, $MS_e = 0.12$] responses than did peripheral questions. There was some concern that the memory for peripheral information was unduly affected by memory for lunch. Indeed, removing lunch reduced the magnitude of the emotion effects on memory for peripheral information. Nonetheless, the high-emotion group still had lower levels of consistent peripheral memories ($M = .27$) than did the moderate-emotion group [$M = .33$; $t(209) = 2.63$]. The difference between the high-emotion ($M = .32$) and moderate-emotion ($M = .29$) groups on inconsistent memories was only marginally significant in this subset of the data [$t(209) = 1.52$, $p < .10$]. The general conclusions that central in-

Table 6
Proportion of Consistent Memories and Memory Errors as a Function of the Date That the First Questionnaire Was Completed

	Date Completed		
Measure	9/12–9/13	9/14–9/15	9/16–9/20
Consistent	.32	.38	.44
More specific	.09	.08	.09
More general	.24	.18	.13
Inconsistent	.30	.24	.25
Omissions	.05	.11	.06

formation was more prone to reconstructive errors than was peripheral information and that high-emotion participants had poorer memory for peripheral details than did moderate-emotion participants were not changed in these alternative analyses.

The results reported above do not provide a definitive test of the flashbulb memory hypothesis. A subgroup of our participants may have had true flashbulb experiences and remembered the central details of the event—as well as incidental details not tapped by the questionnaire—with near perfect accuracy and consistency. In fact, 48 participants in the November\September group were perfectly consistent in their memories for *who*, *what*, and *where* (10 of these were also consistent on *when*; no one was consistent on more than eight of the autobiographical features). The 48 participants with consistent memories for central details distinguished themselves from the remainder of the repeated tested group on only two additional measures. Those with consistent memories had significantly lower average ratings for changed activity ($M = 4.94$) than did those with less consistent memories [$M = 5.53$; $t(209) = 2.35$] and marginally lower ratings of emotion [Ms = 6.02 and 6.33 for consistent and inconsistent participants, respectively; $t(209) = 1.64$]. These observations further challenge the idea that increased emotion supports memory consistency.

DISCUSSION

Previous researchers demonstrating inconsistent memories for momentous events have often relied on small sample sizes (Christianson, 1989, $n = 40$; McCloskey et al., 1988, $n = 29$; Neisser & Harsch, 1992, $n = 44$; Weaver, 1993, $n = 22$) and have measured degree of memory consistency without considering types of memory errors across conditions. Very little evidence has existed concerning memory for peripheral information surrounding momentous events. And finally, some of the events studied have been less than monumental. The research reported here was an attempt to expand our knowledge of memory for momentous events beyond these previous studies.

In replication of previous research, the participants reported a large number of autobiographical features, the number of reported features remained high over a 2-month retention interval, and the participants were highly confident in these memories. Thus, a large proportion of our participants had *flashbulb memories* in the most general sense of the term. However, none of the autobiographical reports was completely consistent. Consistency increased as the delay between the attacks and the first test increased, suggesting that high levels of consistency are obtained only after one develops a consistent narrative that is retold many times (Winningham et al., 2000).

There was very little evidence to suggest that September 11th led to highly detailed and veridical autobiographical records of the day. Evidence for elaborations was found in shifts toward more specific responses on the second questionnaire and in inconsistent responses across successive questionnaires. As would be predicted from a constructivist view of autobiographical memory, there were proportionally more errors of commission for central features than for peripheral features and proportionally more errors of omission for peripheral features than for central features. For example, if I do not remember where I was when I heard the 9/11 news, I will generate a likely location, because *where* is a feature central to all autobiographical memories. In contrast, clothing is not an important component of most autobiographical memories. Failures to retrieve such peripheral details should lead to omissions, rather than to incorrect guesses. Nonetheless, a large majority of the participants supplied seemingly incidental details. Evidence for reconstruction was also found in the answers to the *when* question. The participants tended to answer *when* questions with times that fell on the clock's major landmarks, and this tendency increased with the delay of the first test. The time results suggest the use of cyclical anchors in the reconstruction of *when* (Thompson et al., 1996). Whereas *when* might be an important central feature of autobiographical memories, specific clock time may not be. In fact, distribution of answers to the *when* question in Table 4 more closely resembles the distributions for the peripheral features *clothing* and *weather* than those for the central features *who*, *what*, and *where*. Clock time is probably reconstructed from other autobiographical cues, such as *where* and information concerning daily routines (Friedman, 1993).

The results reported here provided only weak support for the attention-focusing hypothesis. Memory consistency was greater for central than for peripheral information, but this may have resulted from selection processes operative for all autobiographical memories, independent of their emotional content. Memory for peripheral information was impaired for the highly emotional participants relative to the participants with less extreme emotional reactions. However, memory for the central information was apparently not enhanced by strong emotion, as is required by the attention-focusing framework. In fact, the high-emotion participants were particularly prone to more specific responses. It is possible that the present investigation was not sensitive enough to detect a positive impact of strong emotion on memory for central details. Perhaps the contrast between the high-emotion and the moderate-emotion groups was attenuated by the strong reactions of both groups to the 9/11 events. But the

attention-focusing hypothesis posits an explicit tradeoff in which the basis of good memory for central details is a reallocation of resources away from peripheral details. The results reported here were sensitive enough to detect poor memory for peripheral details by highly emotional participants. By extension, the results should have been sensitive enough to detect greater consistency in memory for central details. Studies of flashbulb memory have not shown evidence for increased accuracy of central features at the expense of peripheral features. For example, Christianson (1989) measured primarily central features of flashbulb reports and failed to find any differences in memory consistency as a function of rated emotion.

Perhaps strong emotion generally impairs memory for peripheral information without any concomitant increase in memory for central details. Additional support for this idea can be found in Pezdek's (2002) investigation of memory for 9/11. She recruited participants from New York City, California, and Hawaii. Not surprisingly, the New York sample had the highest average rated emotion. However, this group also had the poorest autobiographical memories for the day's events. Previous investigations demonstrating a positive impact of emotion on autobiographical memory may have confounded differences in distinctiveness with differences in emotionality. For example, Conway et al. (1994) demonstrated increased consistency associated with increased affective response, but their measure of affect included both surprise (distinctiveness?) and the intensity of the reaction, and they did not control for differences in access to the gist of the autobiographical events as a function of affect. Emotion may impair processing of the peripheral details of a momentous event while distinctiveness provides access to the memory representation of the event (see Hunt, 2001; Schmidt, 1991).[4] Once comparisons are made across equally distinctive events (as was the case here, in Pezdek, 2002, and in Schmidt, 2002), the negative impacts of emotion on memory become clear.

The events of September 11th, 2001 were truly momentous. The attacks represent an autobiographical landmark (Loftus & Marburger, 1983; Shum, 1998) around which we may organize our memories for the years surrounding the attacks. These landmarks are readily retrieved and, because they demand a great deal of attention and are repeatedly retold, may contain more information than do memories for more mundane or typical days in our lives. However, we may retain little more than schematic *highlights* of landmark events. These accounts include information about how we first heard the news of the terrorist attacks, where we were, and what we were doing. However, our memory for the exact time we learned the news is probably poor, as is our memory for peripheral events from the day. Forgotten details may have been replaced with general statements and information that is consistent with our daily scripts. People who reacted with extremely strong emotion paid a memory cost in the form of further loss of peripheral details. This pattern of memory performance can be explained in terms of a combination of well-established reconstructive processes, the impact of distinctiveness on memory, and the role of emotion in impairing cognitive processes.

REFERENCES

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, **93**, 203-231.

Berntsen, D. (2002). Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Memory & Cognition*, **30**, 1010-1020.

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). Cambridge: Cambridge University Press.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, **5**, 73-99.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition*, **20**, 277-290.

Christianson, S.-Å. (1989). Flashbulb memories: Special, but not so special. *Memory & Cognition*, **17**, 435-443.

Christianson, S.-Å. (1992a). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, **112**, 284-309.

Christianson, S.-Å. (1992b). Remembering emotional events: Potential mechanisms. In S.-Å. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 307-340). Hillsdale, NJ: Erlbaum.

Conway, M. A. (1995). *Flashbulb memories*. Hillsdale, NJ: Erlbaum.

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 & Cognition*, **22**, 326-343.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, **66**, 183-201.

Finkenauer, C., Luminet, O., Gisle, L., El-Ahmadi, A., van der Linden, M., & Philippot, P. (1998). Flashbulb memories and the underlying mechanisms of their formation: Toward an emotional-integrative model. *Memory & Cognition*, **26**, 516-531.

Friedman, W. J. (1987). A follow-up to "Scale Effects in Memory for the Time of Events": The earthquake study. *Memory & Cognition*, **15**, 518-520.

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, **113**, 44-66.

Gallagher, R. P. (2003). *The psychological impact of 9-11 on college students and suggestions for how counseling centers can prepare for war and/or future terroristic attacks* [on line]. Available: http://www.acpa.nche.edu/comms/comm07/Feature3-03.htm.

Hatmaker, M., & Schmidt, S. R. (1998, May). *Flashbulb memories for Princess Di's death.* Paper presented at Middle Tennessee Psychological Association, Nashville.

Hunt, R. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1359-1366.

Johnson, R. W. (1962). The retention of qualitative changes in learning. *Journal of Verbal Learning & Verbal Behavior*, **1**, 218-223.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, **11**, 114-120.

Mackworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic Science*, **3**, 67-68.

McCloskey, M., Wible, C. G., & Cohen, N. J. (1988). Is there a special flashbulb memory mechanism? *Journal of Experimental Psychology: General*, **117**, 171-181.

National Sleep Foundation (2003). *National Sleep Foundation 2002 Sleep in America poll results relating to September 11* [on line]. Available: http://www.sleepfoundation.org/PressArchives/911poll.html.

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). Cambridge: Cambridge University Press.

Pezdek, K. (2002, November). *Event memory and autobiographical memory for the events of September 11, 2001*. Paper presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

Pillemer, D. P. (1998). *Momentous events, vivid memories.* Cambridge, MA: Harvard University Press.
Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, **19**, 523-542.
Schmidt, S. R. (2002). Outstanding memories: The positive and negative effects of nudes on memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 353-361.
Schmolck, H., Buffalo, E. A., & Squire, L. R. (2000). Memory distortions develop over time: Recollections of the O. J. Simpson trial verdict after 15 and 32 months. *Psychological Science*, **11**, 39-45.
Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliott, M. N., Zhou, A. J., Kanouse, D. E., Morrison, J. L., & Berry, S. H. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine*, **345**, 1507-1512.
Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin*, **124**, 423-442.
Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law & Human Behavior*, **16**, 413-424.
Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. L. (1996). *Autobiographical memory: Remembering what and remembering when*. Mahwah, NJ: Erlbaum.
Weaver, C. A. (1993). Do you need a "flash" to form a flashbulb memory? *Journal of Experimental Psychology: General*, **122**, 39-46.
Winningham, R. G., Hyman, I. E., Jr., & Dinnel, D. L. (2000). Flashbulb memories? The effects of when the initial memory report was obtained. *Memory*, **8**, 209-216.

NOTES

1. Distinctiveness may also arise when an emotionally significant stimulus is embedded in a series of nonsignificant stimuli. Such a stimulus is *emotionally distinctive* (see Schmidt, 1991). However, an emotional stimulus embedded in a series of similarly emotionally charged stimuli is not distinctive and, thus, does not benefit from the positive effects of distinctiveness on memory performance.

2. Very few participants reported times to an exact minute that did not fall on one of the larger landmarks. For this reason, the 5-min and the 1-min responses were combined.

3. When the participants were asked to report their most memorable image, they were not asked to report the most memorable initial image or impression. Thus, their answers may reflect changes in their impressions over time, rather than an autobiographical memory for the experience. For this reason, the image responses were dropped from subsequent analyses.

4. Not all emotional events are distinctive and well remembered. For example, Schmidt (2002) demonstrated that a picture of a nude in the context of a series of nudes was not as well remembered as a picture of a clothed person embedded in a series of nudes. A similar phenomenon may have been operative in people's memory for the events of 9/11. The sight of the Pentagon burning would have been quite distinctive on any other news day. However, none of our research participants listed this image as one of their most memorable impressions of the day. The image of the Pentagon was lost in the blur of other emotional events occurring on September 11th.

APPENDIX

Note—Items in italics were added on the second survey.

Public Event Questionnaire

Personal information: Age: ____ Gender (circle one) M / F Today's Date: ____/_____/_____

1. Describe in as much detail as possible how you heard about the news of apparent terrorist attacks on the World Trade Center, and the Pentagon.

Who told you?
Where were you?
What were you doing?
What was your first thought when you heard the news?
How confident are you in these responses: 1) unsure 2) fairly sure 3) sure

2. When exactly did you hear the news?

Day of the Week: _____ Date: _____ Time: __: _ _, AM or PM

How confident are you in this response: 1) unsure 2) fairly sure 3) sure

3. What was the most memorable image or impression you got from the news coverage?

4. Please rate how much the news caused a change in your ongoing activity:

1 2 3 4 5 6 7
not at all I completely stopped what I was doing.

5. As a result of this event, were you late for work or one of your classes? (*Circle one*: yes/no)

6. During the day of the attacks, did you forget to attend any appointments, complete any class assignments, or run any errands that you were supposed to do that day? (*yes*/*no*). Please explain.

APPENDIX (Continued)

7. Please rate how upsetting you found this news:

1	2	3	4	5	6	7
not upsetting						extremely upsetting

8. Do you have family members or close friends who were present, nearby, or otherwise directly involved in one of these incidents? (*yes*/*no*) Please explain.

9. Please rate your personal involvement in the event:

1	2	3	4	5	6	7
no direct involvement						a close friend or relative was injured or killed

10. Do you anticipate further personal involvement as a direct consequence of these attacks? Explain. (*yes*/*no*)

11. Describe in as much detail as possible the clothes you were wearing at the time you heard the news.

12. What was the weather like here in Middle Tennessee that day?

13. What did you eat for breakfast that day (briefly)?

14. What did you eat for lunch (briefly)?

15. About how many times (excluding these surveys) did you discuss with other people the circumstances surrounding when, where, and how you heard about the terrorists' attacks.

_____

Approximate number: _____ **or** *circle one*: *0–50*, *51–100*, *101–150*, *151–200*, *more than 200*

16. Did you complete this questionnaire in September? (yes/no)

We may wish to contact you later to ask you additional questions concerning your memory for this event. In order to match your responses on a later questionnaire to your responses on this questionnaire, please record the last four digits of your Social Security number below. This will be the only identification information connected to your responses in this study.

The above paragraph was replaced by the following on the November questionnaire:

In order to match your responses with the earlier questionnaire, please record the last four digits of your Social Security number below. This will be the only identification information connected to your responses in this study.

Last four digits in your SS# _ _ _ _

(Manuscript received December 3, 2002;
revision accepted for publication October 21, 2003.)