Running head: Ten-Year Follow-up

A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events

William Hirst
New School for Social Research

Elizabeth A. Phelps
New York University

Robert Meksin
New School for Social Research

Chandan J. Vaidya
Georgetown University

Marcia K. Johnson
Yale University

Karen J. Mitchell
West Chester University

Randy L. Buckner
Harvard University

Andrew E. Budson
Boston University School of Medicine

John D. E. Gabrieli
Massachusetts Institute of Technology

Cindy Lustig
University of Michigan

Mara Mather
University of Southern California

Kevin N. Ochsner
Columbia University

Daniel Schacter
Harvard University

Jon S. Simons
Cambridge University

Keith B. Lyle
University of Louisville

Alexandru F. Cuc
Nova Southeastern University

Andreas Olsson
Karolinska Institutet

Word Count: 13,356

Authors' Note

Please address correspondence to: William Hirst, Department of Psychology, New School for Social Research, 80 Fifth Avenue, New York, NY 10011.

We thank the many student coders without whom this research project would have been impossible, as well as Brett Sedgewick, who assisted with the supervision of this project at its earliest stages.

Support from the James S. McDonnell Foundation, and the National Institute of Mental Health, grant R01-MH0066972, is gratefully acknowledged.

Abstract

Within a week of the attack of September 11, 2001, a consortium of researchers from across the United States distributed a survey asking about the circumstances in which respondents learned of the attack (their flashbulb memories) and the facts about the attack itself (their event memories). Follow-up surveys were distributed 11, 25, and 119 months after the attack. The study, therefore, examines retention of flashbulb memories and event memories at a substantially longer retention interval than any previous study employing a test-retest methodology, allowing for the study of such memories over the long-term. There was rapid forgetting of both flashbulb and event memories within the first year, but the forgetting curves leveled off after that, not significantly changing even after a 10-year delay. Despite the initial rapid forgetting, confidence remained high throughout the 10-year period. Five putative factors affecting flashbulb memory consistency and event memory accuracy were examined: (1) attention to media (2) the amount of discussion, (3) residency, (4) personal loss and/or inconvenience, and (5) emotional intensity. After ten years, none of these factors predicted flashbulb memory consistency; media attention and ensuing conversation predicted event memory accuracy. Inconsistent flashbulb memories were more likely to be repeated rather than corrected over the ten-year period; inaccurate event memories, on the other hand, were more likely to be corrected. The findings suggests that even traumatic memories and those implicated in a community's collective identity may be inconsistent over time and these inconsistency can persist without the corrective force of external influences.

Keywords: Long-term memory, flashbulb memories, event memories, September 11, autobiographical memories, collective memories

A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events

The present paper reports on a 10-year longitudinal study of both flashbulb and event memories of the attack in the United States on September 11, 2001. As we use the term, a *flashbulb memory* is a memory for the circumstance in which one learned of a public event. The memory may be reported at any time, but an essential characteristic of a flashbulb memory is that it is long-lasting, that is, that people can report on the reception event long after it occurred. Flashbulb memories are not about the public event itself, nor are they about what is learned about the event from others; rather they are about the reception event in which one hears news about a public event. (The event is, by definition, public, inasmuch as it is the topic of a public discussion.) Although people no doubt form memories of reception events after hearing any piece of news, most of these memories will be quickly forgotten in most instances. Flashbulb memories, on the other hand, are exceptional because people report details of the reception event after extraordinarily long periods of time, as Brown and Kulik (1977) emphasized in their pioneering work. That is, the formation of a flashbulb memory is probably the exception rather the rule when it comes to reception events. In the context of discussions of flashbulb memory, the term *event memory* is reserved for memories for the public event that led to the flashbulb memory (Larsen, 1992). In the case of 9/11, people possess both flashbulb memories e.g., where they were when they learned about the attack, and event memories e.g., that four planes were involved.

Flashbulb memories, and their associated event memory, have garnered a great deal of interest since Brown and Kulik (1977) first brought them to the attention of the psychological

community. There are many reasons for this interest: Researchers treat them as examples of long-term emotional memories and as special cases of the more general class of traumatic memories, for instance (e.g., Berntsen & Rubin, 2006). In addition, flashbulb memories are a distinctive form of autobiographical memory, a topic of great interest to psychologists (Rubin, 2005). Unlike most autobiographical memories, they are built around public events. In the case of flashbulb memories, the public event appears to have personal meaning and consequences for the individual rememberer. That is, the memory refers to events for which the personal and the public intersect. As a result, flashbulb memories are of concern to scholars studying both individual and social identity (Neisser, 1982).

Flashbulb memories and Forgetting

A central question for those interested in flashbulb memories is how to best characterize their long-term retention. When Brown and Kulik (1977) first examined this issue, they initially simply asked participants many years after the event occurred whether they remembered the circumstances in which participants learned of the event. Brown and Kulik noted, however, that "the division [between possessing a flashbulb memory and not possessing one] [is] not so absolute, and, more importantly, within the account there [is] much to interest us" (p. 79). As a result, they moved beyond their dichotomous measure and also examined specific features of participants' reports of reception events, the "who, what, when, and where" of the reception event. They then investigated how many of these canonical features were or were not contained in a reported flashbulb memory. For Brown and Kulik, errors of omission (that is, those instances in which people claimed *not* to remember a specific canonical feature) provided insight into "variations as well as constancies …in the content of the reports" (p. 79).

For many researchers, this focus on errors of omission captures only one aspect of what it means to forget (e.g., see Johnson & Raye, 1981; Winograd & Neisser, 1992). In most tests of memory, people are said to fail to remember previously studied material not only if they state that they do not remember, but also if they remember the material inaccurately. That is, forgetting involves not just errors of omission, but also errors of commission. Brown and Kulik's methods focused on errors of omission, and did not assess errors of commission.

To study both errors of omission and errors of commission, psychologists have often employed a test-retest methodology. Participants are asked as soon as possible after a consequential, often emotionally charged public event for their memory for the circumstances in which they learned of the event and then are given follow-up assessments several months or more afterwards. The follow-up recollections are then compared with the first recollection, resulting in a *consistency* score between the first and subsequent assessments. A "forgetting" curve can then be plotted on the basis of these consistency scores.

Of course, this method does not permit a researcher to assess absolute accuracy, in that the researcher still does not have access to what actually unfolded during the reception event (see Curci & Conway, 2012, for a discussion of this concern, and related issues). Moreover, it creates a situation in which participants might confuse their memory for the reception event with their memory of their first report, though the number of retests does not appear to affect the level of consistency (Kvavilashvili, Mirani, Schlagman, Foley, & Kornbrot, 2009). There is also some concern that the first report is an unreliable index if it is not collected within a day or two after the event, though in most studies collection within the first week or so appears to be adequate (Collucia, Bianco, & Brandimonte, 2006; Julian, Bohannon, & Aue, 2009; Kvavilashvili et al., 2009; Lee & Brown, 2003; but see Schmidt, 2004; Winningham, Hyman, & Dinnel, 2000).

Thus, although the comparison of retest with test offers a good first approximation of reception recall if one assumes that the initial, immediate report is fairly accurate, researchers who employ this methodology cautiously use the term *consistency* rather than *accuracy* when reporting their results.

To date, using either Brown and Kulik's method or the test-retest methodology, researchers have studied a large number of public events, including the deaths of public figures, e.g., John F. Kennedy (Brown & Kulik, 1977), Princess Diana (Kvavilashvili, Mirani, Schlagman, & Kornbrot, 2003), and King Baudouin of Belgium (Finkenauer, Luminet, Gisle, El-Ahmandi, van der Linden et al., 1998); politically salient events, including the resignation of Margaret Thatcher (Wright, Gaskell, & O'Muircheartaigh, 1998) and the election of Barack Obama (Koppel, Brown, Stone, Coman, & Hirst, 2013); and natural or man-made disasters, including an earthquake in Turkey (Er, 2003) and the explosion of the Challenger (Bohannon & Symons, 1992; Neisser & Harsch, 1992). Though mostly emotionally charged and negatively valenced, some studied events have been positively valenced, e.g., the German withdrawal from Denmark at the end of War World II (Berntsen & Thomsen, 2005), the fall of the Berlin Wall (Bohn & Berntsen, 2007), and success at the World Cup (Kopietz & Echterhoff, 2014; Tinti, Schmidt, Testa, & Levine, 2014). Moreover, although most research has involved events salient to the public at large, some have studied emotionally charged public events relevant to only a small group of people, as is the case when family or friends learn of the death of a loved one (Pillemer, 2009; Rubin and Kozin, 1984).

As to the 9/11 attack, there have been over 20 studies of learning to date (e.g., Budson, Simons, Waring, Sullivan, Hussoin, & Schacter, 2007; A. R. Conway, Skitka, Hemmerich, & Kershaw, 2009; Curci & Luminet, 2006; Davidson, Cook, & Glisky, 2006; Greenberg, 2004;

Hirst, Phelps, Buckner, Budson, Cuc et al., 2009; Kvavilashvili. Mirani, Schlagman, Foley, & Kornbrot, 2009; Luminet & Curci, 2008; Luminet, Curci, Marsh, Wessel, Constantin, Genocz, et al., 2004; Paradis, Solomom, Florer, Thompson, 2004; Pezdek, 2003; Qin, Mitchell, Johnson, Krystal, Southwick et al., 2003; Schmidt, 2004; Shapiro, 2006; Sharot, Martorella, Delgado, & Phelps, 2007; Smith, Bibi, & Sheard, 2003; Talarico & Rubin, 2003; Tekcan, Berivan, Gülgöz & Er, 2003; Weaver & Krug, 2004).

Several conclusions have emerged from this flurry of research. First, consequential and emotionally charged public events indeed can lead to long-lasting memories for the reception events. For the retention intervals studied to date, Brown and Kulik's (1977) characterization that "nobody forgets" is correct in that, for the studied public events, few people fail to report that they simply cannot remember the circumstances in which they learned of the event. Second, these memories can contain both errors of omission *and* commission. For some researchers, then, flashbulb memories may be exceptional, in that people continue to report a recollection of the reception event even after a substantial delay, but they are unexceptional in that they are replete with errors of omission and commission, just like "ordinary" autobiographical memories (e.g., Talarico & Rubin, 2003). For us, the mere existence of selective errors of omissions and commission does not disqualify a memory as being classified as "flashbulb," even if it makes it "ordinary" (see Curci & Conway, 2012, for a discussion of this point). Like Brown and Kulik (1977), we also feel that it is necessary to explore the 'variations as well as constancies …in the content of the reports" (p. 79). Consequently, among other things, we will be interested here in the consistency between initial and subsequent recollections.

There is one exceptional feature of flashbulb memories, besides their long-term retention, that is widely accepted and deserves mention. Unlike many "ordinary" autobiographical

memories, people are extremely confident in the accuracy of even their inconsistent flashbulb memories, even after several years have passed (see, for instance, Neisser & Harsch, 1992; Neisser, Winograd, Bergman, Shreider, Palmer, & Weldon, 1996). For example, a day after the 9/11 attack, Talarico and Rubin (2003) asked participants to record both the circumstance in which they learned of the attack, as well as another "important" autobiographical event that had occurred in the last week. In follow-up testing one, six, and 32 weeks after the attack, there were similar rates of forgetting for both types of memories. On the other hand, whereas confidence in ordinary autobiographical memories declined over time as the consistency scores declined, confidence in flashbulb memories remained high, despite the similar rates of decline in consistency. This finding is important in that it suggests that individuals are not simply "filling in" missing details about the reception event with guesses as they recollect. Rather when responding with an inconsistent memory, they truly believe that they are "remembering" it accurately (i.e., they are making reality/source monitoring errors, Johnson, Hashtroudi & Lindsay, 1993).

Factors Affecting Flashbulb Memory Consistency

As more psychologists utilized the test-retest methodology, questions about what makes a flashbulb memory consistent or inconsistent over the long-term naturally arose. If flashbulb memories are like "ordinary autobiographical memories," in that they evidence errors of omission and commission, then psychologists hypothesized that many of the factors affecting the retention of "ordinary" autobiographical memories would also affect flashbulb memories, for example, the emotional intensity of the event, its importance or consequentiality, the degree to which it is rehearsed, its distinctiveness, the level of surprise associated with it (e.g., Conway, 1995; Conway, Anderson, Larsen, Donneley, McDaniel, McClelland, & Rawles, 1994;

McCloskey, Wible, & Cohen, 1988;). Whereas each of these factors has been shown to affect consistency in some studies, other studies fail to find the same effect. Comparison across studies is difficult, inasmuch as it often involves different events, but the diverse set of findings suggests that there may be no single necessary condition for a consistent flashbulb memory (see Talarico & Rubin, 2008, for a review).

Interestingly, researchers have generally not contrasted the factors that might predict consistency with those that might predict confidence ratings or event memory accuracy (for an exception, see Day & Ross, 2014). This is surprising, in that it is well appreciated that the relation between accuracy and confidence is not straightforward (Roediger, Wixted, & DeSoto, 2012). For instance, even in the laboratory, manipulations affecting confidence do not necessarily lead to similar effect on accuracy (e.g., Wells & Bradfield, 1999).

Long Long-term Memory

Surprisingly, inasmuch as an essential defining characteristic of flashbulb memories is their long-term retention, most flashbulb memory studies examine retention intervals of a few months, or a most a few years. The few examining retention intervals greater than a few years, as Brown and Kulik (1977) did, did not use a test-retest methodology, the established standard at present. Some, such as Brown and Kulik and Yarmey and Bull (1978), only examined whether participants possessed a flashbulb memory, without any concerns about its consistency. Others used innovative techniques that only allow for assessing accuracy of a limited range of features of the flashbulb memory (Berntsen & Thomsen, 2005). The present ten-year longitudinal study is, therefore, unique. A few studies of "ordinary" autobiographical or semantic memories have longitudinally tracked memories over a 10-year period, or more (e.g., Burt, Kemp, & Conway, 2001; Catal & Fitzgerald, 2004; see also Linton, 1986; Rubin, 2005; Wagenaar, 1986). Using a

cross-sectional methodology, Bahrick and his colleagues studied forgetting curves for a wide variety of material for intervals stretching up to 50 years or more (Bahrick, 1983; 1984; Bahrick, Bahrick, & Whittlinger, 1975; see also Conway, Cohen, & Stanhope, 1991; Squire & Slater, 1975; Stanhope, Cohen, & Conway, 1993). For instance, they asked college alumni ranging in age from their 20's to their 70's to recollect the Spanish vocabulary they learned while in college or the names of the streets of their college town.

As they pertain to the interests in this investigation, the results of these studies indicate that (1) there is substantial forgetting of both autobiographical and semantic memories in the first few years, both in terms of reported recollection and in terms of accuracy, (2) the forgetting rate asymptotes thereafter, conforming to a power function, (3) there is consistency between what is recollected at one time, even if incorrect, and what is recollected subsequently and (4) if a memory is retained for a certain period of time, it is unlikely to show further forgetting. Bahrick and his colleagues suggested that memories go into a "permastore" after approximately six years (Bahrick, 1984). The conclusions, however, are at best limited to "ordinary" autobiographical memories, or semantic memories. Whether the "permastore" concept is applicable to flashbulb memories is an open question.

The Present 9/11 Study

Can what has been observed in studies of flashbulb memories with retention intervals of a few months to three years be extended to longer retention intervals, such as ten years? And can one extend the studies of long-term memory of "ordinary" autobiographical event and semantic memories to long long-term flashbulb memories and their associated event memories? With questions such as these in mind, within a week of the attack of September 11, 2001, the authors of this paper distributed a survey at locations around the United States and asked about the

circumstances in which participants learned of the attack, facts about the attack (e.g., how many planes were involved), characteristics of the memories participants held (e.g., confidence ratings), and the way participants reacted to the news, among others. We then followed up with similar questions 11 months, 35 months, and finally 119 months after the attack. (We always tested participants one month before an anniversary of the attack.)

Several years ago, we offered a preliminary three-year report (Hirst et al., 2009). The findings fell into three general categories. First, we examined the consistency of flashbulb memories and their associated level of confidence, as well as the accuracy of event memories. We observed a decline in the consistency of the reported flashbulb and event memories over the three-year period, even as confidence ratings remained high. Most of the forgetting occurred in the first year. The dissociation between consistency and confidence ratings is consistent with several studies, including Talarico and Rubin's (2003) 9/11 study. The finding that the level of consistency seemed to flatten out after a year is consistent with Kvavilashvili et al. (2009), who found, again for 9/11, similar levels of consistency between their two-year retest and the final three-year retest. Our results differed from Schmolck, Buffalo, and Squire's (2000) study of flashbulb memories for the O.J. Simpson verdict. These researchers showed little forgetting after 15 months, but more dramatic forgetting after 32 months, suggesting that memory distortions built up over time. They attributed this build-up largely to source monitoring errors.

The second class of results in our previous report focused on predictions of consistency, examining: (1) *residency* at the time of the attack, (2) the level of *emotional intensity*, (3) the *personal loss or inconvenience*, (4) the amount of *media attended* to immediately after the attack and in the ensuing period, and (5) the amount of discussion with others about the attack (referred to as *ensuing conversation*). None of these were correlated with flashbulb memory consistency.

As for event memories, all the studied factors other than emotional intensity were correlated with accuracy. The effects of residency and personal loss and inconvenience could be traced to variations in the level of ensuing conversation. Moreover, the association between media attention and accuracy was reinforced by contrasting the forgetting curve associated with memories for the event of the Challenger explosion with the forgetting curves associated with memories for the event of the 9/11 attack. Not only did the curves differ, but, critically, the difference reflected variations in the level of media coverage of these two events. Both media attention and ensuing conversations probably affected event memory because they encouraged rehearsal of the facts surrounding the attack (see Breslin & Safer, 2011).

The third set of results we examined traced how the memories changed over time. Unlike most investigations to date, we not only assessed changes in the average consistency or accuracy scores, but also what happened to inconsistent or inaccurate memories from the first year. Inconsistent flashbulb memories were more likely to be repeated than corrected, while inaccurate event memories were more likely to be corrected than repeated. Furthermore, the accuracy of event memory was mediated by the level of media attention, suggesting that media might not only reinforce accurate memories, but also correct inaccuracies.

What might we expect ten years after the attack? It is difficult to generalize from Hirst et al (2009), for several reasons. (1) In that study, we based our conclusions on performance at three separate time periods, one of which served as the baseline from which to compare the other two periods. A relation observed with two points might constitute a trend, but does not guarantee a long-term pattern. (2) There is no a priori reason to assume that what holds for three years holds for ten. For instance, the increase in memory distortions over time that Schmolck et al. (2000) observed for the Simpson verdict may not have been observed by Hirst et al. for delays

as short as three years, but may be observed after a ten-year delay. (3) Although forgetting seemed to be slowing in Hirst et al., there was not an additional time interval to assess whether memories had become resistant to forgetting, as Bahrick's (1984) work suggests they eventually should. (4) Declines in confidence that are not detected after a short retention interval may be detected after a longer retention interval. (5) The factors determining the shape of the forgetting curves for the first three years may differ from those associated with forgetting over longer retention intervals. (6) Inconsistencies may be repeated once, but may not continue to be repeated.

With these considerations in mind, we distributed a survey similar to the one we used at Year 3 to previous respondents. We also collected an additional "new" sample of individuals who had never participated in the project.

Method

Because we have described the method as it applied up to Year 3 in detail elsewhere (Hirst et al., 2009), we focus here on information relevant to our ten-year follow-up.

Participants, Recruitment, and Procedure

For the first three surveys, we recruited participants between September 17, 2001 and September 21, 2001 (Survey 1), August 5 and August 20, 2002 (Survey 2), and August 9 and August 20, 2004 (Survey 3). Recruitment for Survey 4 took place between August 1, 2011 and August 15, 2011. The original recruitment was done in seven locations: Boston and Cambridge, MA; New Haven, CT; New York, NY; Washington, DC; St. Louis, MO; Palo Alto, CA; Santa Cruz, CA. As a result, follow-up participants in subsequent surveys came from the same areas, though in many instances they no longer lived in these areas. The one exception is Boston, which used a slightly different procedure when recruiting participants on Survey 1 and hence did

not figure in the follow-up surveys reported here. Over the first three surveys, we had a total of 3,246 participants. For Survey 4, we tried to contact all 3,246 participants, either through postal mail or email. When an email or a postal envelope was returned, we searched through the web to find additional means of reaching a respondent, using, in the main, Facebook, Google+, and LinkedIn. Although we could not associate contact information with a particular survey, codes that participants generated allowed us to connect the responses of a given individual across the surveys they filled out. Some participants claimed that they had previously participated, but they supplied an incorrect ID. Attempts were made to find a match by examining handwriting, demographic information, and so on. The participants in the No Match category reflect those for which a match could not be found at any point in the project. In the end, 202 participants took all four surveys. These participants will be our main focus of interest.

As shown in Table 1, 52% of those who had completed Surveys 1, 2, and 3 returned Survey 4, a good return rate for studies of this kind (Baruch, 1999). Of the 2117 participants who returned Survey 1, 10% of them ended up participating at every stage of the ten-year project. This return rate is reasonable, given the length of the project, the difficulty in keeping track of people over such a long time period, the extensive nature of the survey, and the fact that we did not compensate participants for their efforts. Although it involved participants across the United States, our sampling should not be viewed as representative of the American public.

We undertook several dropout analyses. For the features of age, gender, residence at the time of the attack, student membership at the time of the attack, political affiliation, race/ethnicity, and religion, we found no differences between those who took all four surveys and (1) those who only took Survey 1 and (2) those participants who took Surveys 1 through 3, but not Survey 4 ($ps > .20$). The one dimension on which we did find a dropout effect concerned

personal loss and/or inconvenience. The measure is described in detail in Hirst et al. (2009) and below. Briefly, the percentage of participants reporting personal loss and/or inconvenience was 26.9% for those participants who only took Survey 1, 27% for those who took Surveys 1 and 2; 33.8% for those who took Surveys 1, 2, and 3; 35.6% for those who took Surveys 1, 2, 3, and 4. According to Fisher's exact test, those who took Survey 1 only or Surveys 1 and 2 could not be considered as two groups. Similarly, those who took Surveys 1, 2, and 3 and Surveys 1, 2, 3, and 4 could also not be considered as two groups. We therefore have two sets of participants: those who only took the early surveys and those who took the surveys in the latter part of the study as well. The associations between personal loss and/or inconvenience and these two sets of participants was significant, $p < .01$. Participants were probably more motivated to continue if they had experienced personal loss and/or inconvenience.

At every stage of the study, we recruited new participants, individuals who had not responded to any previous survey. They served as controls for the returning participants. This *New Group* is labeled "Survey 4 only" in Table 1. To obtain this group of participants, we distributed surveys by hand from tables set up on or near the campus of New York University, the New School, and Georgetown University (that is, in New York City and Washington, DC). In addition, we sought new participants through Amazon Mechanical Turk and paid them 50 cents for their efforts.

For all four surveys, participants were told that they had a week to fill out the survey. Surveys could be filled out either on-line or on paper and then mailed back to the experimenters using a stamped return envelope. We combine the two samples as there were no differences in the responses offered on-line as opposed to on paper (for each question in the survey, ps > .50).

Demographic information can be found in Table 2. Although the match was not perfect, the New sample resembled the Four-Sample survey to a substantial degree. Analyses of how responses to the questions in the survey differed across these demographic categories are beyond the scope of the present report.

Surveys

The content of Survey 4 was similar to the content of Surveys 2 and 3, with small changes made to reflect the time Survey 4 was taking place. For example, Survey 2 asked questions that began "In the last year,…."; Survey 3 asked "In the last three years,…"; Survey 4's questions began, "In the last seven years, …." Inasmuch as Survey 4 was administered one and half months before the 10th Anniversary of the attack, for some questions, such as those involving media attention, we also began several questions with "In the last few months, …." Inasmuch as our findings did not differ for the "seven-year" questions and the "last few months" questions, we do not report the results for the "last few months" questions here. Each survey was approximately 17 pages long (when printed) and took about 45 minutes to complete. Copies can be found at http://911memory.nyu.edu. Although the surveys explored a variety of topics, we focus here on those relevant to the formation and retention of flashbulb and event memories.

As Table 3 indicates, questions fell into three categories: (1) *specific memories for the circumstances in which one learned of the attack*, focusing here on six canonical features; (2) *specific memories of the event itself*, focusing on five different "facts" (e.g., number of planes, location of President Bush at the time of the attack); and (3) *ratings related to factors affecting performance*. We examined five potential factors, as in Hirst et al. (2009). With respect to the questions about the canonical features of flashbulb memories, we followed each probe with one of two follow-up questions. Half of the potential participants were sent a questionnaire assessing

their confidence in their response [with the question "How confident (on a 1 – 5 scale) is your recollection?"]. The other half received a questionnaire assessing how well they thought that they could remember the information in the future ["In 10 years, how well (on a 1 – 5 scale) do you think you will remember this?"] Ninety-two of the 202 participants in our Four-Survey sample returned the confidence version; 110, the forecasting version. We also asked questions pertinent to what kinds of relevant media participants had been exposed to over the ten-year period (e.g., did you see the film *Fahrenheit 911*, the film *United 93*?) In what follows, we present data from all 202 participants, except when discussing confidence. For these discussions, we confine ourselves to the 92 participants who received the confidence questions.

Coding

A detailed coding manual was developed for Survey 1 and subsequently adapted for Surveys 2 – 4 (see Hirst et al., 2009, for details; the manuals can be found at http://911memory.nyu.edu). There were two general coding schemes. For some questions, there might be multiple ways to respond, but only a single response was required. For instance, when asked "How did you first learn of the attack?" participants could respond, for example, with "TV," "Radio" or "Telephone," but only one of these devices played a role when they *first* learned of the attack. The coding scheme listed alternatives [specifically, TV, Radio, E-mail/Instant message, Phone call (including Voice messages), Visual sighting, Word of mouth, Sounds/screams, Sirens, Other, Not stated]. Using a number system, the coding indicated which of these options best fit the participant's response. The second coding scheme allowed for multiple responses. For instance, more than one city was the target of the attack. As a result, coders recorded all the responses a participant gave to the question "In the vicinity of which cities did the airplanes end up?" To assess interrater reliability, we randomly selected 10% of

the surveys to be dual coded. We then calculated kappas or Cronbach's alphas, whichever was appropriate, for each question. For Survey 4, as well as the other surveys, interrater reliability was always greater than .80.

Results

We divide our results into three sections: (1) Forgetting and confidence, (2) Factors affecting performance, and (3) Changes in inconsistent or inaccurate memories. In each section, we first discuss flashbulb memories and then event memories. In the main, because we are interested in mnemonic change over time in the same person, we confine our analyses to those individuals who participated in all four surveys. We note any differences arising between the results reported in Hirst et al. (2009) with the larger Three-Survey sample and the results uncovered here for the Four-Survey sample. We examined other combinations of participation, (e.g., participation, e.g., Surveys 1 & 2 only), and did not find any differences that would affect the conclusion in the present report. Details can be obtained from the first author. We present the relevant data from our New participants at various points below when they address specific concerns.

Forgetting and Confidence

Flashbulb Memories. In the Four-Survey sample, in Survey 4, 99.5% of participants responded with a recollection for at least 5 of the six questions probing for canonical features. Using Brown and Kulik's dichotomous measure (did or did not report remembering the reception event, regardless of accuracy and level of elaboration), it would appear that all but one of our participants, if not all of them, still have a flashbulb memory of 9/11 after ten years. But, as noted, there is another sense of forgetting in the literature on flashbulb memories, captured by the question: Even though participants produced recollections, did they nevertheless produce

errors of commission? And given our interest in long-term retention, was the level of forgetting, in this sense, the same, greater, or lesser for Survey 4 as it was for Survey 3?

Consistency. Inasmuch as all the flashbulb memory questions required a single response, a response on Survey 2 thru Survey 4 was considered consistent with the response offered on Survey 1 if the coding matched. For example, for the question "How did you first learn about the attack?", if a participant stated that she learned from the radio on Survey 1 and from TV on Survey 4, her Survey 4 response was scored as a "0," that is, inconsistent. If the Survey 4 response had been "radio," it would have been scored as a "1," that is, consistent. The *total consistency score* was the average of the six questions with which we probed the canonical features of flashbulb memories, producing a value from 0 to 1 (again, see Table 3 for the canonical features). Our scoring method differs slightly from others, such as Neisser and Harsch (1992), who used a three-point (0 – 2) rather than a two-point (0 or 1) classification scheme. For Neisser and Harsch, responses were either absent or inconsistent (0), or consistent with different degrees of specificity (1 or 2). We collapsed their two consistency scores into one single value. Our measure, then, is more likely to emphasize the consistency of a response than would Neisser-Harsch, in that the latter method could elicit a lower score, relative to the entire scale, than ours would if a consistent response was not very specific. In terms of the pattern of results, as opposed to absolute values, Hirst et al. (2009) found no difference between the Neisser-Harsch scoring and the present one for the Three-Survey sample. An examination of 10% of the Four-Survey sample produced equivalent similarities. Figure 1a contains the results across the four surveys for those who completed all four surveys, using our coding scheme.

Given the .63 measure of consistency observed for Survey 2 in Figure 1a (that is, on average, 2.22 of the 6 recalled canonical items in Survey 2 were inconsistent with what was

reported on Survey 1), one can reasonably state that the flashbulb memories of our participants after one year did not reflect to a notable degree what they reported in the first week. This forgetting slowed in the next two years, with a decline of just .07 in the consistency score between Survey 2 and Survey 3. The level of forgetting stabilized between the third and tenth year, with a non-significant improvement of .03 in consistency scores between Survey 3 and Survey 4. We conducted a repeated-measures ANOVA with Time as a within-subject factor and the total consistency score as the dependent measure. There was a main effect for Time, $F(2, 200) = 8.32$, $p < .001$, $n_p^2 = .08$. In follow-up analyses, we found a significant difference between the total consistency scores for responses on Survey 2 and those reported on Survey 3, $t(201) = 4.07$, $p < .001$, $d = .33$, CI[.03,.09], but not between Survey 3 and Survey 4, $t(201) = 1.92$, $p = .06$, $d = .16$, CI[-.07,.00]. Interestingly, the consistency scores on Survey 2 appear to be comparable to those reported by Kvavilashili et al. (2009) and Talarico and Rubin (2003), though exact comparison is difficult because the scoring procedures and probes differ across the three studies. Kvavliashili et al., for instance, used Neisser and Harsch's (1992) weighted average score, which could range from 0 to 7. After three years, participants scored, as read from the graph they present, on average, approximately 5.00, or, if we used the top score as a denominator, a consistency score of approximately .71. One difference between Neisser and Harsch's scoring and ours is that we treat emotional response as equally important as the other canonical factors, but as shown by Hirst et al. (2009) and Koppel, Winkel, and Hirst (2014), consistency scores for emotional response can be quite low. If we exclude emotional responses from the current calculations, we obtain a consistency score of .65 (SD = .24) for Survey 4. It is unlikely that the stabilization between Surveys 3 and 4 probably is the result of a floor effect. People are more likely to forget, rather than remember, most of the events of their lives,

suggesting that a total consistency score around .60 after a ten-year retention interval would be considered high for most "ordinary" autobiographical memories.

In Conway et al. (1994), flashbulb memories were defined as any memory with no more than one inconsistency. Although we used a different measure of consistency than Conway et al., we found that, for the Four-Survey sample, 30.2%, 24.35%, and 23.8% of the participants on Surveys 2, 3, and 4, respectively, had no more than one inconsistency (that is, had a consistency score of .83 or better). We report the distribution of responses for Surveys 2 and 4 in Figure 2. The forgetting observed here is worse than Conway et al. found for their British participants remembering the resignation of Margret Thatcher (85.6% had scores indicating an exact memory for all but one of five attributes). The difference between these two results may reflect different scoring procedures, different probes, or the exceptional nature of the Thatcher resignation. We should note that Conway et al.'s treatment of flashbulb memories as accurate seems counter to both the use advanced by Brown and Kulik (1977) and by most researchers in the field, who are interested in the question of whether or not flashbulb memories are accurate.

Confidence. We calculated the *total confidence rating* as the average of the confidence ratings associated with our six canonical features, as indicated by those who received the confidence-version of the survey (see Figure 1b). Clearly, confidence remained high across the ten years. Because we kept participants' identifying information separate from the responses to the surveys, we could not ensure that we distributed the confidence version of the survey to the same participants on each subsequent round and consequently cannot use an ANOVA to compare confidence ratings across all waves of the survey. For statistical purposes, then, we examined the subset of participants who happened to fill out the confidence version of both Surveys 2 and 4 and then a similar subset of participants who filled out the confidence version of

Surveys 3 and 4. The means associated with these subsets were similar to those of the full sample reported in Figure 1a. (The subset scores ranged from 4.23 to 4.63.) We did not find a difference in confidence ratings of Surveys 2 and 4, $t(49) = 1.21$, $p = .23$, $d = .19$, CI[-.29,.07], and Surveys 3 and 4, $t(38) = 1.9$, $p = .07$, $d = .12$, CI[-.71. .02]. That confidence rating remained high even in the presence of fairly low consistency scores suggests that the confidence ratings participants assigned to their memories might not depend on the consistency of the memories. Focusing on Survey 4, we found that consistency scores and confidence ratings were not significantly correlated ($r = .07$, $p = .50$).

Event Memories. Some of the questions we used to assess event memory (see Table 3) required more than one response to be "accurate," for instance, participants had to state all three crash sites. For each correctly named site, participants received .33 points. For each incorrectly stated site, they lost .16 of a point. Along the same lines, for the airline carrier names (there were two: United and American), participants received .50 points for each correct answer and lost .25 points for each incorrect one. When a score was less than zero, we reassigned it to have the value of zero. As a result, scores were always from 0 to 1. To obtain a score for the temporal ordering of events (Question 11 in Table 3), we calculated the Spearman rank order correlations between the respondent's order and the actual order. We then added one to the result and divided by two, thereby, again, obtaining a value from 0 to 1. The *total accuracy score* was the averaged scores for the five probes. Although, for the flashbulb memories, we compared performance on Surveys 2 – 4 with performance on the baseline Survey 1, limiting our analysis to three testing periods, for event memories, we can and did determine the accuracy at all four testing periods (see Table 4).

Like flashbulb memories for the reception of event information, accuracy scores for event details showed the most dramatic decline within a year and then stablized. We conducted a repeated-measures ANOVA with Time as a within-subject factor and with the total Accuracy score as the dependent measure. There was a main effect for Time, $F(3, 199) = 31.67$, $p < .001$, $n_p^2 = .32$. This main effect could be attributed solely to the drop in average performance from Survey 1 to Survey 2, $t(201) = 8.50$, $p < .001$, $d = .65$, CI[.08,.13].

We draw a distinction between *critical* and *non-critical* probes. We exclude from this discussion the temporal order question (Question 11 in Table 2) because the sequence of events could have been predicted largely on common knowledge alone, e.g., a plane must have hit the World Trade Towers before the buildings collapsed. Of the remaining four probes, two, our *critical* probes (number of planes, crash sites), are more central to the events of 9/11 than the other two, our *non-critical* probes (location of President Bush, names of airlines). People usually do not tell their version of 9/11 without mentioning the number of planes or the crash sites. On the other hand, people often omit the specific names of the airline carriers when talking about 9/11. Whether or not a plane was operated by United does not seem particularly pertinent to the issues likely to be of interest when remembering and/or discussing 9/11. As for the location of Bush at the time of the attack, this detail also seems less central. It is not included, for instance, in the Wikipedia account of the 9/11 attack (http://en.wikipedia.org/wiki/September_11_attacks). We asked 50 Mechanical Turk workers to rate on a scale of 1 to 7 how "critical" each of these facts were to any story one might tell about the 9/11 attack, with 1 being "not critical at all," 7 being "extremely critical." As Table 4 indicates, the crash sites were considered the most critical, followed by the number of planes, $t(49) = 4.24$, $p < .001$, $d = .66$. Both these facts were more critical than the names of airline carriers, $t(49) = 9.43$, $p < .001$, $d = 1.07$, and $t(49) = 7.47$,

$p < .001$, $d = 1.45$, respectively, and the location of Bush, $t(49) = 8.71$, $p < .001$, $d = 1.30$, and $t(49) = 4.50$, $p < .001$, $d = .66$, respectively.

The forgetting observed during the first year can be traced to the non-critical items (again, see Table 4). In a 2 x 4 repeated measures ANOVA, with Element Type (Critical v Non-critical) as one within-subject factor and Time as the other, there was a main effect of Element Type, $F(1, 201) = 133.89$, $p < .001$, $n_p^2 = .40$, for Time, $F(3, 199) = 33.69$, $p < .001$, $n_p^2 = .35$, and an interaction between Element Type and Time, $F(3, 199) = 15.13$, $p < .001$, $n_p^2 = .19$. The main effect of Element Type reflected the excellent retention of the critical items.

In Table 4, one can also observe improvements in accuracy, which appear to arise because of external reminders. First consider memory for the location of President Bush. Michael Moore's film *Fahrenheit 9/11* appeared three months prior to the distribution of Survey 3. In the film, Moore included a long, face-front video of President Bush looking quizzically into a camera after being told *sotto voce* by aide Andrew Card that a plane had crashed into the World Trade Center. This segment of the Moore film was widely discussed. Hirst et al. (2009) reported a marked increase in accuracy from Survey 2 to Survey 3 for the question about Bush's location, with the improvement greater for those who saw the film than for those who did not, a finding replicated in the Four-Survey sample. Did the improvement observed in Survey 3 persist over the next seven years? We conducted a 2 x 2 ANOVA with Time as a within-subject factor (Survey 3 v Survey 4) and Seeing as a between-subject factor (Saw the Moore film v Did not see the film), excluding the six participants who did not answer on Survey 3 the question about seeing the Moore film. The main effect for Seeing was marginally significant, $F(1, 194) = 3.87$, $p = .05$, $n_p^2 = .02$. Neither the main effects for Time nor the interaction were significant, $F(1, 194) = .43$, $p = .51$, $n_p^2 = .00$ and $F(1, 201) = .43$, $p = .51$, $n_p^2 = .00$, respectively. That is, the

Moore film not only refreshed people's memories immediately after seeing the film, but also led to the retention of an accurate memory for at least the next seven years. This finding on Survey 4 was not the result of filling out Survey 3, in that the scores in the New sample did not differ significantly from those reported in Table 4 ($p > .30$). (New proportions: .91 correct, for those who saw the film and .76 for those who did not).

Released on April 29, 2006, five years before the distribution of Survey 4, the film *United 93* also could potentially correct forgotten information, in this case, that United was an involved carrier. Table 4 contains the proportion of participants who correctly listed United as one of the airline carriers involved in the attack as a function of whether or not the participant reported seeing the film. If watching the film mattered, the accuracy scores concerning the name of the carrier should be, and was, greater for those who saw *United 93* than for those who did not, but only for the time period after the film's release (Using Fisher's exact test, for Survey 4, $p = .03$; for Surveys 1 – 3, $p > .30$). For those who saw the film, the difference in performance between Surveys 3 and 4 was marginally significant (McNemar's test, $p = .08$). There was no detectable difference for those who did not see the film. The release of *United 93* may not have improved the memory accuracy of those who failed to see the film, whereas the release of *Fahrenheit 9/11* did, in part because there was not as much discussion about *United 93* after its release as there was for *Fahrenheit 9/11*. According to *Lexus-Nexus*, the number of mentions of *Fahrenheit 9/11* in the *New York Times* one week after its release (27) was 3.38 times greater than the number of mentions of *United 93* in the *Times* for a similar time period (8).

Factors Affecting Performance

As in Hirst et al. (2009), we analyzed the following factors (see Table 3 for details about the surveys): (1) *media attention*, (2) *ensuing conversation*, (3) *emotional intensity of the initial reaction,* (4) *residency* at the time of the attack, and (5) *personal loss and/or inconvenience*.

Flashbulb Memories. *Consistency***.** Similar to Hirst et al. (2009), we found for Survey 4 that consistency did not vary with any of five factors. As Table 5 indicates, emotional intensity, media attention, and ensuing conversation did not correlate significantly with consistency. For *residency* (see Table 6), there was no significant difference between those who lived within the borders of New York City at the time of the attack and those who did not, $t(200) = 1.06$, $p = .29$. Other divisions, e.g., lived in Manhattan, the greater New York area, or below Canal Street, produced similar results. As for personal loss and/or inconvenience, based on responses on Survey 1 to Questions 27 and 28 in Table 3, a participant was coded as experiencing a personal loss and/or inconvenience if they offered at least one concrete example, such as damage to home, loss of business, personal injury to self, friend, or relative, cancellation of school or work, and/or lack of food. We did not include in this list psychological distress (e.g., felt anxious, lost appetite), although good arguments could be made for doing so. Further details can be found in Hirst et al. (2009). Despite the fact that participants seemed to be motivated to continue with the study if they had experienced personal loss and/or inconvenience, we found that this factor also did not affect consistency: average consistency score for those who did experience personal loss and/or inconvenience, $M = .59$, $SD = .21$; those who did not, $M = .61$, $SD = .20$; $t(200) = .79$, $p = .43$.

Bohannon, Gratz, and Cross (2007) argued that flashbulb memory consistency should be greater when the source of the news is a person rather than the media. We divided our sample along these lines. Participants were said to have as the source a person if they indicated that they

learned about the attack through a phone call, by word-of-mouth, or through email or Instant Messages. They were said to have as a source the media if they indicated that they learned about it through TV, radio, or the Internet. We did not consider in this analysis participants who indicated that they heard screams, sounds, sirens, or actually saw something. The consistency of flashbulb memory across the ten-year period did not differ significantly with the source of the news (ps > .20).

Confidence ratings. Many of the factors affecting confidence ratings on Surveys 2 and 3, as reported in Hirst et al. (2009), no longer produced significant results in Survey 4, which we suspect is related to the decline in the factors' importance over time. We still found with the Four-Survey sample that confidence ratings were significantly correlated with (1) media attention and (2) ensuing conversation for Surveys 2 and 3. (Pearson Product-Moment Correlation between total confidence ratings on Survey 2 and media attention in the first year, r = .17; correlation between Survey 3's confidence ratings and media attention in the first three years, r = .21. Similar correlations now between total confidence ratings on Surveys 2 or 3 and ensuing conversation since the attack: S2, r = .16, S3, r = .18, in all instances, p < .01). This result did not extend to the confidence ratings recorded on Survey 4. As indicated in Table 5, by Survey 4, any evidence of a correlation between Survey 4's confidence ratings and a variety of measures of media attention and ensuing conversation disappeared.

As did Hirst et al. (2009), we also find, now with our Four-Survey sample, that confidence ratings varied as a function of personal loss and/or inconvenience in Surveys 2 and 3. By Survey 4, however, just as was the case for media attention and ensuing conversation, personal loss and/or inconvenience no longer had an effect on confidence (For Survey 4, Experienced loss: M = 4.49, SD = .44; Did not: M = 4.58, SD = .38, t(89) = .94, p = .35). As

Hirst et al. also found for Surveys 1 - 3, we did not find a significant correlation between confidence ratings, as measured in Survey 4, and emotional intensity (see Table 5). Finally, although residency was not a factor for Surveys 2 and 3, we did find, for Survey 4, that those living in New York City reported greater confidence than those living outside the City, $t(89) = 3.21$, $p = .002$, $d = .75$, CI[-.45, -.11] (see Table 6).

Event Memory. As did Hirst et al. (2009), we also found, now for Survey 4, that the level of media attention and ensuing conversation was correlated with the accuracy of event memory (see Table 5). The amounts participants talked about the event in the first year following the attack *and* in the last seven years were also significantly correlated with the accuracy of event memory at the 10th year. For media attention, it was only for the last seven years that there was a significant correlation with event accuracy. Even highly significant and emotionally charged public events are not necessarily well retained simply because of how they are initially encoded; they benefit from rehearsal, as instigated through the media in this instance.

The other factors – emotional intensity, residency, and personal loss and/or inconvenience -- did not yield significant results. The correlation between the level of emotional intensity and the accuracy of event memory on Survey 4 was not significant (Table 5). It is noteworthy, that by Survey 4, emotional intensity no longer affected the level of media attention and ensuing conversation (see Table 7). This non-significant correlation may have arisen because the level of emotional intensity declined over time. The averaged emotional rating at the time of Survey 1 was 3.43 out of 5 (SD = .82); at the time of Survey 4, it was 2.54 (SD = .99). The limited opportunities to attend to media concerning 9/11 and the social disapproval that may exist about talking about 9/11 after so many years may also have contributed.

As for residency, unlike both Hirst et al. (2009) and Pezdek (2003), we found no significant effect for residency with our present Four-Survey sample. We compared total accuracy scores for New Yorkers and non- New Yorkers on Surveys 1, 2, 3, and 4 and in each instance, using a t-test, did not find a significant difference (for all comparisons, $p > .50$; see Table 6).

Turning finally to personal loss and/or inconvenience, perhaps because the personal loss and/or inconvenience occurred ten years ago, it did not produce a significant difference in event memory accuracy on Survey 4 (Personal loss and/or inconvenience present*: M = .84,* SD = .16; Absent: $M = .79$, $SD = .21$, $t(200) = 1.56$, $p = .21$). Hirst et al. (2009) suggested that they found an effect of personal loss and/or inconvenience on event memory because it influenced the level of media attention and ensuing conversation. We examined its effect on these two factors, as well as current emotional intensity (see Table 8). Only the ratings associated with the first week after the attack produced significant differences between those with or without personal loss and/or inconvenience on emotional intensity, media attention, and ensuing conversation, $t(200) = 2.21$, $p = .03$, $d = .33$, CI[-.48, -.03]; $t(200) = 2.25$, $p = .03$, $d = .35$, CI[-.51, -.03]; $t(1200) = 2.21$, $p = .03$, $d = .32$, CI[-.43, -.02], respectively. Of course, our definition of personal loss and/or inconvenience is broad. Some participants who fell into this category had a close friend injured or killed, others' personal loss and/or inconvenience was a temporary difficulty in getting to work. Our sample was not large enough to make comparisons across types of personal loss/inconvenience.

It should also be noted, that, contrary to Bohannon et al.'s (2007) suggestion that the source of the news affects event memory, for Survey 4, we found no difference in event memory accuracy when the source was a person and when it was a form of media ($ps > .10$).

Changes in Inconsistent or Inaccurate Memories

What changes occur to inconsistent or inaccurate memories over time? Hirst et al. (2009) offered preliminary data that inconsistent flashbulb memories will be repeated, whereas inaccurate event memories will be corrected. However, as already noted, for that report we had data from only a limited number of testing points – two in the case of flashbulb memories. This ten-year follow-up offers a chance to confirm or dispute the earlier suggestive findings.

Flashbulb Memories. We identified the inconsistent memories on Survey 2 and calculated the proportion that were (1) *corrected* on Survey 3 or (2) *repeated* on Survey 3. We then considered the extent to which these repeated inconsistencies in Survey 3 were, in turn, *corrected* or *repeated* on Survey 4. This latter calculation allowed us to determine whether the inconsistencies in Survey 2 persisted through Survey 3 into Survey 4. By *corrected*, we mean that a previously inconsistent memory now conforms to what was reported on Survey 1. (Of course, participants may respond with neither a correction nor a repetition, but some other alternative, which we treated as *other*. Inasmuch as the proportion of *corrected*, *repeated*, and *other* responses add up to 1.0, we only report here results for *corrected* and *repeated* responses.)

As is clear in Figure 3, inconsistencies on Survey 2 were more likely to be repeated on Survey 3 rather than corrected, a result that replicates with the Four-Survey sample the findings of Hirst et al. (2009), $t(148) = 2.73$, $p = .007$, $d = .23$, CI[.04, .26]. More important, the inconsistent repeated memories of Survey 3 tended to continue to be repeated rather than be corrected on Survey 4, $t(90) = 6.22$, $p < .001$, $d = 1.53$, CI[.33, .65].

These results suggest that not only was a flashbulb memory becoming stable after the first year, as evidenced by the flattening of the forgetting curve after this time period, the emergent stable memory consisted of not just inconsistencies, but, to a remarkable extent,

repeated inconsistencies. For instance, 43.4% of the inconsistencies on Survey 4 were repetitions of an inconsistency on Survey 3. As the data we collected across surveys makes clear, this emergent stable, but not necessarily accurate memory begins to form by the first year.

Table 9 contains some examples of repetitions. Neisser and Harsh (1992) noted that many of the inconsistencies they observed involved temporal confusions (what they called "time slice errors"). That is, the inconsistent memory referred to an event that might have occurred sometime around the time of the flashbulb-memory eliciting event, but did not occur at the time one learned of the event. There may also be temporal errors (that is, confusions about the binding of temporal and other features of an event) in the examples in Table 9. For instance, Participant 4 may have been in the office and on the street on September 11. The inconsistency arises because, when remembering, she may misattribute her location when she learned of the event to her location at some other time on the same day. What is important is that once made, such temporal confusions and other source feature errors (e.g., Johnson et al., 1993) persisted over time.

Event Memories. We calculated the proportion of corrections or repetitions from one survey to the next (Survey N to Survey N+1). For the two items that required a single correct answer (number of planes, location of Bush), for *repetitions*, the numerator was the number of incorrect responses on Survey N that were repeated on Survey N + 1; for *corrections*, it was the number that were corrected. The denominator in both instances was the number of incorrect responses on Survey N. For the two items that required multiple responses to be correct (crash sites, carrier names), for *repetitions*, the numerator was the number of people that repeated an inaccurately remembered crash site or carrier name from Survey N, whereas the denominator was the number of participants who inaccurately remembered the crash site or carrier name on

Survey N. We calculated these proportions for each possible erroneous crash site or carrier name and then summed over them. Furthermore, we calculated the carrier names separately for United, in order to examine the effect of watching *United 93*. For *corrections*, for crash sites, the numerator was the number of participants who failed to specify all three crash sites on Survey N that subsequently remembered at least one of the them on Survey N+1; the denominator, the number of participants who failed to specify all three crash sites on Survey N. For carrier names, the numerator was the number of participants who failed to remember a specified carrier on N, but correctly remembered it on Survey N+1; the denominator was the number of participants who failed to specify an involved airline carrier on Survey N.

We cannot easily compare "repeated" and "corrected" proportions in most instances because their denominators differ. We can examine whether the level of corrections and repetitions changed over time, and whether corrections and/or repetitions differed from chance. If one was above chance, whereas the other was below chance, we might safely assume that they differed. We adopt the simplifying assumption that chance is .50, for instance, that there is an equal chance that an inconsistency will be repeated as opposed to not repeated, or will be corrected as opposed to not corrected. The exact value of chance is not germane to the comparison we plan to make, only whether different measures differed from the prespecified measure of chance. For instance, we can assert that the pattern of corrections changed over time if, for instance, the proportion of corrections went from below chance to chance levels. Different estimates of chance would yield similar patterns of change.

Consider first the critical facts – the number of planes and the crash sites. Given their pervasive presence in most accounts of 9/11, we would expect that, unlike what we found for flashbulb memories, inaccurate recollections on one survey should likely be corrected on the

next survey. As Table 10 indicates, across all four surveys, corrections were indeed more frequent than repetitions for both critical items. In all instances, repetitions were below chance (based on the reasoning in the previous paragraph and using the binominal test), whereas corrections were above chance, $p < .001$.

As for the non-critical facts, first, consider the overall scores associated with the location of Bush. Prior to the release of the film (which happened shortly before Survey 3), repetitions were more likely than corrections, with the latter quite rare. As a result, on Survey 2, for both those who saw the film and those who did not, repetitions were at chance level, $p > .20$, and corrections were below chance, $p < .001$ (relevant data are in the columns labeled "Wrong on Survey 1, Modified on Survey 2"). Immediately after the release of the film, corrections now dominated, with repetitions markedly declining. Thus, for the responses on Survey 3, for those who saw it, corrections were now above chance, $p < .005$, whereas repetitions were below chance level, $p < .001$. For those who did not see the film, corrections were now at chance levels, $p > .20$, whereas repetitions were below chance, $p < 001$. As noted already, the film probably had an effect even for those who did not see it because the notorious segment involving Bush's classroom performance was replayed and discussed in many outlets. As of Survey 4, the impact of the film faded, and repetitions were more common. For those who had not seen the film, repetitions were now more frequent than corrections. For those who saw the film, corrections shifted from above chance on Survey 3 to below chance on Survey 4, $p < .001$, whereas repetitions remained below chance, $p > .01$. For those who did not see the film, on Survey 4, corrections went from chance levels to below chance, $p < .01$, whereas repetitions went from below chance to chance levels, $p > .20$.

Similar effects were observed for the film *United 93*, which came out between Surveys 3 and 4. The overall correction rate remained below chance on Surveys 2 through 4, $p < .01$. However, for those who saw the film, correction rates went from below chance level on Surveys 2 and 3, $p < .01$, to chance levels on Survey 4, $p = .38$. They remained below chance levels across surveys for those who did not see the film, $p < .01$.

General Discussion

The study of flashbulb memories and associated event memories provides an opportunity to study how "real world" events are remembered over an extended period of time. A clear picture emerges in this paper.

Forgetting and Confidence

Although participants clearly formed flashbulb memories of the reception of information about the 9/11 attack, in that they reported an elaborate recollection even after ten years (Brown & Kulik, 1977), they nevertheless experienced forgetting, in the sense that they no longer remembered the original reception event as they initially reported it. Forgetting in this sense occurred mainly in the first year and then leveled off, so that no change in memory performance was detectable between Years 3 and 10. Our results differ from Schmolck et al. (2000), who found that memory continued to decline over time. As we noted in the Introduction, one might argue that we did not find increasing memory distortions for flashbulb memories in our three-year report because we did not have a long enough retention interval. This concern becomes less reasonable when considering the present ten-year follow-up. Our findings for event memories were similar to those we presented for flashbulb memories: rapid forgetting the first year, without any increase in the levels of forgetting thereafter.

Our results, as well as those of others (including Schmolck et al, 2000), also demonstrated what appears to be a characteristic feature of flashbulb memories – the high confidence with which they are held, even with marked levels of forgetting. A delay of ten years did not diminish participants' extraordinarily high level of confidence in their flashbulb memories. We did not follow the procedure of Talarico and Rubin and ask for a memory of an "ordinary" autobiographical event around the time of the 9/11 attack, but their work showed a steady decline in confidence over a 32-week period for "ordinary" autobiographical memories. There is no a priori reason to doubt that this decline would have continued for at least some of the next nine years.

Factors Affecting Consistency, Confidence, and Accuracy

Consistency of Flashbulb Memories. None of our five factors -- media attention, ensuing conversation, residency, personal loss and/or inconvenience, and emotional intensity – had an influence on flashbulb memory consistency. In retrospect, this outcome may not be surprising. There are no doubt multiple roots to consistency. The naturalistic nature of this study does not allow us to isolate one over the others easily. For example, although there is a substantial experimental literature indicating that emotionally evocative stimulus material is often remembered better than emotionally neutral material, there is also evidence that emotion may reduce source memory by disrupting the binding among features of items (e.g., an emotional picture with its location, Mather, Mitchell, Raye et al., 2006; see Kensinger, 2009; Mather & Sutherland, 2011; Phelps, 2004; Reisberg & Heuer, 2004, for reviews and further discussion). How the various affecting the impact of emotion on memory interact in the present instance may depend on how our participants were memorizing and remembering the reception

event. In the laboratory, we could try to control for individual differences in strategy. We cannot do so in the present instance.

Confidence Ratings. At least for the first three years, confidence was influenced by media attention and ensuing conversation. That this effect holds for assessment of media attention and ensuring conversation across this time period suggests that confidence ratings are a function of what occurs subsequent to the time the event is initially encoded, as well as what occurs at the time of the initial encoding. This effect of media attention and ensuing conversation appears to have on confidence ratings may be based on the subjective characteristics of a memory, such as its perceptual or emotional vividness (see Johnson & Multhaup, 1992, for a discussion of the role of subjective characteristics). Presumably, they elicit rehearsals or remindings of the reception event, which, in turn, maintain or increase the vividness of the memory's subjective qualities (e.g., Suengas & Johnson, 1988). Metamemory judgments may also come into play (Johnson & Raye, 2000). For example, people may believe that if they attended to the media or talked about 9/11 a lot, they must have good memories of the event and the circumstances in which they learned about the event. Consequently, as their media attention and ensuing conversations increases, so does their confidence. As to the failure to find an effect after ten years, people simply may not have been attending to the media or talking about 9/11 enough for them to affect either the subjective characteristics of the memory or metamemory judgments, and, in turn, the level of confidence.

Our finding for Survey 4 that people who lived in New York at the time of the attacks gave higher ratings to their confidence in their reception memories than non-New Yorkers is consistent with the idea that metacognitive factors affect confidence about memories for reception details. Echterhoff and Hirst (2006), for instance, argued from a study that

manipulated ease of retrieval that people may assign confidence ratings based on their social expectations concerning what they should or should not remember. New Yorkers may believe that they should, as those most directly affected, remember the circumstances in which they learned of 9/11 and, hence, on the basis of this belief assign a high level of confidence to their memory. A better understanding of the relation between factors that affect confidence and those that affect consistency is needed.

Event Memory. The accuracy of event memory was associated with high levels of media attention and ensuing conversation. These variables, especially media attention, likely increase accuracy because they help correct inaccurate memories and provide occasions for rehearsing accurate memories.

Factors Affecting Recollection

The above discussion focuses on consistency and accuracy, but, as argued, it is also important to consider whether or not someone recollects the reception event, regardless of the consistency of what was recollected. As already noted, in all but one case, participants reported recollections of the six probed-for canonical features. These recollections were accompanied by a high confidence rating. In Survey 2, 94.6% of the total confidence ratings were greater than or equal to 3.5 out of 5; 74.3% of the total confidence rating were greater than or equal to 4 out of 5. For Survey 4, comparable percentages were 96.0% and 86.7%%, respectively. The factors influencing the high levels of confidence could be viewed as the factors establishing whether or not people form and maintain a flashbulb memory, irrespective of its accuracy.

Our results indicate that, for the first three years, media attention and ensuing conversation led to increased confidence and hence a preserved "memory" of the reception event. Finkenhauer et al.'s (1998) emotional-integrative model of flashbulb memories provides a

cue as to why that might be so (for related models, see Conway, 1995). Their model is relevant to Brown and Kulik's interest in recollection per se, as opposed to consistency, because, in their original study, they distributed a survey that examined memories for the death of King Baudouin of Belgium only 7 to 8 months after the death, thereby assessing only claims about remembering, not the consistency of reported memories. Their model suggests that that thinking about the event itself will lead to better "memory" for the reception event. One need only assume that media attention and ensuing conversation led to rehearsal of the event itself to make a connection between our results and Finkenhauer et al.

Changes in Memories

The presence or absence of external influences also accounts for differences in the way flashbulb and event memories change over time. External reminders are present for event memories, and hence, in the case of the media coverage of 9/11 in the United States, corrections prevailed. Of course, external reminders vary as to their prevalence. Certain facts about 9/11, what we referred to as critical facts, are probably included in almost every public account of the events of 9/11. As a result, memory accuracy was high from the beginning and remained high for critical facts. This high level of performance limited the chance for corrections over time, because inaccuracies are rare. However, we found that when errors did occur, there was a marked tendency to correct them in future recollections.

As to non-critical facts, these facts do not surface frequently in accounts of 9/11. This may reflect both political and social influences on the way the media depicts events and the way the public conceptualizes them. Whatever the explanation of why a fact is treated as non-critical, because of their relative infrequency, event memory for non-critical facts declined over time, that is, until an external influence provided a source for possible correction. Why an external

influence suddenly emerges may again be a matter of political and social considerations. Once it does, as with the release of *Fahrenheit 9/11* or *United 93*, people exposed to these external influences, begin to correct their errors. Once a correction was made, the now accurate memory tended to remain accurate.

While corrections might characterize the change occurring in event memories over time, repetition characterizes the change associated with flashbulb memories, and even event memory if there is no external influence. The reason is clear: There is likely a paucity of external reminders when it comes to flashbulb memories. Between the time of the attack and the first year, people appeared to have formed a memory of the circumstances in which they learned of the event and this memory became their story, regardless of any errors it might include. Of course, subsequent discussion with others who might have been with a participant as the time that information about the attack was received could "correct" erroneous elements, but we have no way of assessing the extent to which this factor might come into play in the present study.

The Special Nature of Flashbulb Memories

As we noted in the introduction, psychologists study flashbulb memories in part because they seem relevant not just to long long-term retention, but to two additional topics of general import: memory for trauma and the way in which they allow the personal and public to intersect. How do our present findings bear on these two topics?

Flashbulb Memories and Trauma. The emotionally charged nature of flashbulb memories makes them an ideal vehicle for studying traumatic memories, especially if the flashbulb-memory-inducing event is negatively valenced. Of course, flashbulb memories and their associated event memories have two distinctive characteristics that differentiate them from other traumatic memories. First, almost by definition, they involve learning about an

emotionally charged event rather than directly experiencing the event, as a soldier might when seeing a comrade killed in battle. Second, the trauma associated with flashbulb memories is collective, whereas many of the traumas associated with post-traumatic stress disorder (PTSD) happen to one or at most a few people at any time, as is the case for the soldiers' battle experience. Nevertheless, people can manifest many of the symptoms of PTSD following even indirect exposure to a traumatic event (Galea et al., 2002). Moreover, the collective nature of flashbulb memories provides an advantage for those psychologists inclined to employ a multiple participant methodology, as opposed to case studies.

The present work suggests that the account of a traumatic event, which is autobiographical in nature, may become stable over time, even if it does not reflect what happened. That is, erroneous traumatic memories may become resistant to forgetting over time. To be sure, correction is possible, but, as we noted, the opportunities for correction of autobiographical memories are often limited. This claim, however, is at best suggestive. One reason flashbulb memories may resist forgetting at some point is that they have become highly rehearsed and perhaps highly detailed and structured accounts. Some traumatic memories, however, may not have such features (Brewin & Holmes, 2003, but see Rubin, 2011). Clearly, the present work speaks to the need for more work on the long long-term memory of trauma.

Flashbulb Memories and History. As Neisser (1982) noted, the event associated with a flashbulb memory and the flashbulb memories themselves constitute one of the few times during people's lives that they feel that they are part of history. To a large extent, people's personal lives unfold locally and within a small social sphere. When an event such as the assassination of John F. Kennedy or the attack of 9/11 occurs, the personal timeline of one's life and the historical time line of a nation intersect. Autobiographical memories – in this instance,

memories of the circumstances of learning of the event – and collective memories – in this instance, memories the public holds of the precipitating event – become intricately intertwined. As a result, the study of flashbulb memories bears critically on issues of social identity (Bernsten, 2008; Curci, Luminet, Finkenauer, & Gisle, 2001) and the formation of collective memory (Conway, 1997; Hirst & Meksin, 2008).

Although the present work suggests that autobiographical flashbulb memories and collective event memory both become stable over time, somewhere between the third and tenth year, it also underscores that different factors drive their formation and retention. For instance, autobiographical memories appear to be shaped largely by individual factors affecting both memorizing and remembering (e.g., Johnson, 2006). Moreover, as they stablize, they can be formed around inconsistent as well as consistent recollections. Erroneous, as well as accurate, memories are repeated over time.

Collective event memories also become stabilized over time, but errors may be more likely to be corrected rather than maintained, at least "corrected" along the lines of what an external influence specifies. Whereas the autobiographical memories might reflect to a substantial degree the psychological dynamics of individuals (Conway & Pleydell-Pearce, 2000; Gordon, Franklin, & Beck, 2005; Mather & Carstensen, 2005; but see Nelson & Fivush, 2004; Pillemer, 2009; Wang, 2013) collective event memories rest to a larger degree on the dynamics of social practices and artifacts (Hirst & Meksin, 2008; Olick, Vinitzky-Serouss, & Levy, 2011). People remember events as these social practices and artifacts appear to dictate. Of course, individual psychological dynamics may affect memory for public events, for example, one's expertise, motivations and available schemas may affect how media accounts are processed. Moreover, social practices and artifacts may shape memories for personal autobiographical

events, as the literature on cultural effects on autobiographical memories indicates (Wang, 2013). Similarities and differences in the dynamics underlying the formation and retention of autobiographical flashbulb memories and the more collective memories held of the public event itself are an intriguing domain for future research.

Conclusions

Although participants continued to report elaborate recollections of reception details over the ten years, participants should a considerable level of inconsistency in their long-term retention. There was considerable initial forgetting over the first year. The level of inconsistency remains fairly stable thereafter. Interestingly, the inconsistent flashbulb memories observed in Survey 1 were repeated from Survey 2 to Survey 3 and then in Survey 4. Memories for the events of the 9/11 attack also showed initial forgetting, especially of non-critical details in the first year. Now, however, external influences, such as the films *Fahrenheit 9/11* and *United 93*, led to corrections. Also important was the contribution of media attention and ensuing conversation to event memory accuracy and confidence ratings, at least for the first three years.

Of course, external influences can do more than correct erroneous memories. They can also do the reverse: create erroneous memories (e.g., Johnson, 2007). Hirst and Fineberg (2012), for instance, discussed the role of external influences in the creation of a false memory around the sacrifice of two Flemish brothers during World War I, a false memory that has become a patriotic and widely accepted symbol for the Flemish nationalist movement. In a more experimental setting, Butler, Zaromb, Lyle, & Roediger (2009) showed that historical films can produce memory distortions if they contain historically inaccurate material. In the case of 9/11, however, at least for the facts we explored, external influences corrected memories rather than distorted them.

External influences shape the content of memory through the effects of rehearsal, social contagion, and socially shared retrieval-induced forgetting, among other possibilities (see Hirst & Echterhoff, 2012, for a review). The frequent repetition of facts in the media and in conversations clearly contributed to the leveling off of forgetting. Whether the same "rehearsal" effect could account for the stabilization of flashbulb memories is uncertain, inasmuch as the questions we asked about ensuing conversations were not sensitive enough to address this issue. However, Suengas and Johnson (1988) demonstrated the role of overt and covert rehearsal in maintaining the subjective qualities of memories for laboratory-established "mini-events."

Bahrick (1984) suggested that still-remembered memories are placed into "permastore." Something like this may be happening here as well; however, in developing his theory, Bahrick and colleagues focused on the retention of accurate memories, for example, if participants accurately remembered the name of a street after six years, they remembered it for the long-term. A more straightforward account of the retention of inconsistencies we report here might include the idea that both true and false memories are attributions based on qualitative characteristics of mental experiences and reasoning and metamemory assumptions (Johnson & Raye, 1981; Johnson et al., 1993), as well as Neisser's reframing of the notion of permastore. As Neisser (1982) argued, a sharp resistance to forgetting may develop as a memory consolidates around a highly structured, redundant representation, or in our case, around a story about both learning of 9/11 and the event itself. Inconsistencies that "make sense" then can be folded into the evolving rendering of 9/11 and, as a consequence, remain as part of the representation. This possibility might be particularly powerful for "time slice errors," in that the inconsistency actually occurred and hence may fit well into any evolving story. Also, source feature misattributions among actual events may be even more likely than importing imagined details because of their more

detailed and vivid characteristics (Lyle & Johnson, 2006). Of course, there likely are a variety of factors contributing to how memories become stable. Our point is that if one or more of these are present, stability, whether involving accurate or inaccurate memories, is likely to occur.

The flashbulb and event memories we observed after a 10-year retention interval reflect both the build-up of mnemonic stability and the corrective function of external influences. These dual effects map, at least to some degree, onto two mnemonic mechanisms of interest to cognitive psychologists and neuroscientists: strengthening and construction/reconstruction (sometimes referred to as consolidation and reconsolidation; see Dudai, 2012; Nader & Hardt, 2009). Strengthening/consolidation involves processes by which a memory is transformed into a long-term representation that becomes insensitive to interference (Dudai, 2012). The leveling off of forgetting that we observed may reflect the final end of this strengthening/consolidation process. Construction/reconstruction/reconsolidation refer to phenomena associated with memory malleability. External influences, particularly as reflected in the memory practices of a community (Hirst & Manier, 2009), may lead to the corrections we observed, even after a substantial period of consolidation, because of this feature of memory.

Our research underscores the need to consider both mechanisms. It is only by considering both that one can understand why something as seemingly unforgettable as the 9/11 attack can produce long-held stable memories with marked inconsistencies. We may recollect the events of 9/11 with great confidence over an extremely long period of time. That longevity might speak to the stability of the memory, but it does not ensure its accuracy.

References

Bahrick, H. P. (1983). The cognitive map of a city – 50 years of learning and memory. In G. Bower (Ed.) *The psychology of learning and motivation: Advances in research and theory*. (Vol 17, pp 125-163). New York: Academic Press.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1–29.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional study. *Journal of Experimental Psychology: General, 104*, 54-75.

Baron, R. M. & Kenny. D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Social and Personality Psychology, 51*, 1173-1182.

Baruch, Y. (1999). Response rate in academic studies – A comparative analysis. *Human relations, 52*, 421-439

Bernsten, D. (2008). Flashbulb memories and social identity. In O. Luminet & A. Curci (Eds.) *Flashbulb memories: New issues and new perspectives.* (pp. 197-206). NY: Psychology Press.

Berntsen, D. & Rubin, D. C. (2006). Flashbulb memories and posttraumatic stress reactions across the life-span: Age-related effects of the German occupation of Demark during WWII. *Psychology and Aging, 21*, 127-139.

Berntsen, D. & Thomsen, D. K. (2005). Personal memories for remote historical events: Accuracy and clarity of flashbulb memories related to World War II. *Journal of Experimental Psychology: General, 134*, 242-257.

Bohannon, J. N. III (1988). Flashbulb memories for the space shuttle disaster: A tale of two theories. *Cognition,* 29, 179-196.

Bohannon, J. N. III, Gratz, S., & Cross, V. S. (2007). The effects of affect and input source on flashbulb memories. *Applied Cognitive Psychology, 21*, 1023-1036.

Bohannon, J. N. III & Symons, V. L. (1992). Flashbulb memories: Confidence, consistency, and quantity. In E. Winograd, & U. Neisser (Eds.), *Affect and accuracy in recall* (pp. 65-90). Cambridge: Cambridge University Press.

Bohn, A. & Berntsen, D. (2007). Pleasantness bias in flashbulb memories: Positive and negative flashbulb memories of the fall of the Berlin Wall among East and West Germans. *Memory & Cognition*, *35*, 565-577.

Breslin, C. W. & Safer, M. A. (2011). Effects of event valence on long-term memory for two baseball championship games. *Psychological Science, 22*, 1408-1412.

Brewin, C. R. & Holmes, E. A. (2003). Psychological theories of posttraumatic stress disorder. *Clinical Psychology Review, 23*, 339-376.

Brown, R. & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73-79.

Budson, A. E., Simons, J. S., Waring, J. D., Sullivan, A. L., Hussoin, T., & Schacter, D. L. (2007). Memory for the September 11, 2001, terrorist attacks on year later in patients with

Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Cortex, 43*, 875-888.

Burt, C. D., Kemp, S., & Conway, M. (2001). What happens if you retest autobiographical memory 10 years on?. *Memory & Cognition*, *29*, 127-136.

Butler, A. C., Zaromb, F. M., Lyle, K. B., & Roediger, H. L. (2009). Using popular history films to enhance classroom learning: the good, the bad, and the interesting. *Psychological Science, 20*, 1161-1168.

Catal, L. L. & Fitzgerald, J. M. (2004). Autobiographical memory in two older adults over a twenty-year retention interval. *Memory & Cognition*, *32*, 311-323.

Cohen (1992). A power primer. *Psychological Bulletin, 113*, 155-159.

Coluccia, E., Bianco, C., & Brandimonte, M. A. (2006). Dissociating veridicality, consistency, and confidence in autobiographical and event memories for the Columbia shuttle disaster. *Memory*, *14*, 452-470.

Conway, M.A., (1995). Flashbulb Memories. LEA: Brighton, Sussex.

Conway, M.A. (1997). The Inventory of Experience: Memory and Identity. In D. Jodelet, J. Pennebaker, & D. Paez (Eds.). Political Events and Collective Memories. (pp.21-46). Routledge: London.

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.

Conway, M. A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge acquired through formal education: Twelve years of cognitive psychology. *Journal of Experimental Psychology: General, 120*, 1-22.

Conway, M. A. & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261-288.

Conway, A. R., Skitka, L. J., Hemmerich, J. A., & Kershaw, T. C. (2009). Flashbulb memory for 11 September 2001. *Applied Cognitive Psychology*, *23*(5), 605-623.

Curci, A. & Conway, M. A. (2012). Playing the flashbulb memory game: A comment on Cubelli and Della Sala. Cortex, 49, 352-355.

Curci, A. & Luminet, O. (2006). Follow-up of a cross-national comparison of flashbulb and event memory for the September 11th attacks. *Memory, 14*, 329-344.

Curci, A., Luminet, O., Finkenauer, C., & Gisle, L. (2001). Flashbulb memories in social groups: A comparative test-retest study of the memory of French President Mitterand's death in a French and a Belgian group. *Memory, 9*, 81-101.

Davidson, P. S. R., Cook, S. P. & Glisky, E. L. (2006). Flashbulb memories for September 11th can be preserved in older adults. *Aging, Neuropsychology, and Cognition, 13*, 196-206.

Day, M. V. & Ross, M. (2013). Predicting confidence in flashbulb memories. *Memory*, *22*, 232-242.

Dudai, Y. (2012). The endless engram: Consolidations never end. *Annual Review of Neuroscience, 35*, 227-247.

Echterhoff, G. & Hirst, W. (2006). Thinking about memory for everyday and shocking events: Do people use ease-of-retrieval cues in memory judgments? *Memory & Cognition, 34*, 763-775.

Er, N. (2003). A new flashbulb memory model applied to the Marmara earthquake. *Applied Cognitive Psychology*, *17*, 503-517.

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.

Galea, S., Ahern, J., Resnick, H., Kilpatrick, D., Bucuvalas, M., Gold, J., & Vlahov, D. (2002). Psychological sequelae of the September 11 terrorist attacks in New York City. *New England Journal of Medicine*, *346*, 982-987.

Gordon, R., Franklin, N., & Beck, J. (2005). Wishful thinking and source monitoring. *Memory & Cognition, 33*, 418-429.

Greenberg, D. L. (2004). President Bush's false 'flashbulb' memory of 9/11/01. *Applied Cognitive Psychology, 18*, 363-370.

Hirst, W. & Fineberg, I. A. (2012). Psychological perspectives on collective memory and national identity: The Belgian case. *Memory Studies*, *5*, 86-95.

Hirst, W. & Manier, D. (2008). Towards a psychology of collective memory. *Memory*, *16*, 183-200.

Hirst, W. & Meksin, R. (2008). A social-interactional approach to the retention of collective memories of flashbulb events. In O. Luminet, A. Curci, & M. Conway (Eds.) *Flashbulb memories: New issues and new perspectives.* (pp. 207-225). NY: Psychology Press.

Hirst, W., Phelps, E. A., Buckner, R. L., Budson, A. E., Cuc, A., Gabrieli, J. D., ... & Vaidya, C. J. (2009). Long-term memory for the terrorist attack of September 11: flashbulb memories, event memories, and the factors that influence their retention. *Journal of Experimental Psychology: General*, *138*, 161-176.

Johnson, M. K. (2006). Memory and reality. *American Psychologist, 61*, 760-771.

Johnson, M. K. (2007). Reality monitoring and the media. *Applied Cognitive Psychology, 21*, 981-993.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. Psychological Bulletin, 114, 3-28.

Johnson, M. K., & Multhaup, K. S. (1992). Emotion and MEM. In S.-A. Christianson (Ed.), *The handbook of emotion and memory: Current research and theory* (pp. 33-66). Hillsdale, NJ: Erlbaum.

Johnson, M. K., & Raye, C. L. (l98l). Reality monitoring. *Psychological Review, 88*, 67-85.

Johnson, M. K. & Raye, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 35-86*)*. Cambridge, MA: Harvard University Press.

Julian, M, Bohannon, J. N., & Aue, W. (2009). Measures of flashbulb memories: Are elaborate memories consistently remembered? In O. Luminet & A. Curci (Eds.) *Flashbulb memories: New issues and new perspectives* (pp. 99-122). NY: Psychology Press.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review, 1*, 99-113.

Kopietz, R. & Echterhoff, G. (2014). Remembering the 2006 Football World Cup in Germany: Epistemic and social consequences of perceived memory sharedness. *Memory Studies, 7*, 298-313.

Koppel, J., Brown, A., Stone, C. B., Coman, A., & Hirst, W. (2013). Remembering President Barack Obama's Inauguration and the Landing of US Airways Flight 1529: A comparison of Autobiographical and Event Memory. *Memory*, 21, 798-806.

Koppel, J., Winkel, R. & Hirst, W. (2014). The Role of Implicit Theories in Shaping Memory for Emotions and Other Internal States for September 11, 2001. In preparation.

Kvavilashvili, L., Mirani, J., Schlagman, S., Foley, K., & Kornbrot, D. E. (2009). Consistency of flashbulb memories of September 11 over long delays: Implications for consolidation and wrong time slice hypotheses. *Journal of Memory and Language*, *61*, 556-572.

Kvavilashvili, L., Mirani, J., Schlagman, S., and Kornbrot, D. E. (2003). Comparing flashbulb memories of September 11 and the death of Princess Diana: Effect of time delays and nationality. *Applied Cognitive Psychology, 17*, 1017-1031.

Larsen, S. F. (1992). Potential flashbulbs: Memories of ordinary news as the baseline. In E. Winograd & U. Neisser (Eds). *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp. 32-64). NY: Cambridge University Press.

Lee, P. J. & Brown, N. R. (2003). Delay related changes in personal memories for September 11, 2001. *Applied Cognitive Psychology*, *17*, 1007-1015.

Linton, M. (1986). Ways of searching and the contents of memory. In D. C. Rubin (Ed.), *Autobiographical memory*. (pp. 50-67). NY: Cambridge University Press.

Luminet, O. (2008). Models for the formation of flashbulb memories. In O. Luminet & A. Curci (Eds.) *Flashbulb memories: New issues and new perspectives* (pp. 51-76). NY: Psychology Press.

Luminet, O. & Curci, A. (Eds.) (2008). *Flashbulb memories: New issues and new perspectives.* NY: Psychology Press.

Luminet, O., Curci, A., Marsh, E., Wessel, I., Constantin, T., Gencoz, F., et al. (2004). The cognitive, emotional, and social impacts of the September 11 attacks: Group differences in memory for the reception context and the determinants of flashbulb memory. *The Journal of General Psychology, 17*, 171-181.

Lyle, K.B., & Johnson, M.K. (2006). Importing perceived features into false memories. *Memory, 14*, 197-213.

Mather, M. & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in cognitive sciences, 9*, 496-502.

Mather, M. & Sutherland, M. R. (2009). Disentangling the effects of arousal and valence on memory for intrinsic details. *Emotion Review, 1*, 118-119.

Mather, M. & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*, 114-133.

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

Neisser, U. (1982). Snapshots or benchmarks? In Neisser, U. (Ed.) *Memory Observed: Remembering in Natural Contexts* (pp. 43-48). San Francisco: W. H. Freeman & Company.

Neisser, U. & Harsh, N. (1992). Phantom flashbulbs: False recollections of hearing the news about *Challenger*. In E. Winograd, & U. Neisser (Eds.), *Affect and accuracy in recall* (pp. 9-31). Cambridge: Cambridge University Press.

Neisser, U., Winograd, E., Bergman, E. T., Schreiber, C. A., Palmer, S. E., & Weldon, M. S. (1996). Remembering the earthquake: Direct experience vs. hearing the news. *Memory, 4*, 337-357.

Nelson, K. & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review, 111*, 486-511.

Olick, J. K., Vinitzky-Seroussi, V., & Levy, D. (Eds). (2011). *The collective memory reader*. NY: Oxford University Press.

Paradis, C. M., Florer, F., Solomon, L. Z., & Thompson, T. (2004). Flashbulb memories of personal events of 9/11 and the day after for a sample of New York City residents. *Psychological Reports, 95*, 304-310.

Pillemer, D. B. (2009). *Momentous events, vivid memories*. Harvard University Press.

Pezdek, K. (2003). Event memory and autobiographical memory for events of September 11, 2001. *Applied Cognitive Psychology, 17*, 1033-1045.

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinions in Neurobiology, 14*, 198-202.

Qin, J., Mitchell, K. J., Johnson, M. K., Krystal, J. H., Southwick, S. M., Rasmusson, A. M., & Allen, E. S. (2003). Reactions to and memories for the September 11, 2001 terrorist attacks in adults with posttraumatic stress disorder. *Applied Cognitive Psychology, 17,* 1081-1097.

Reisberg, D. E., & Hertel, P. E. (2004). Memory for emotional events. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 3-41). New York: Oxford University Press.

Rimé, B. (2007a). The social sharing of emotion as an interface between individual and collective processes in the construction of emotional climates. *Journal of Social Issues, 63*, 307-322.

Rimé, B. (2007b), Interpersonal emotion regulation. In J. J. Gross (ed,), *Handbook of emotion regulation*. (pp. 446-485). New York: Guilford Press.

Rimé, B., Finkenauer, C., Luminet, O., Zech, E., & Phillippot, P. (1998). Social sharing of emotion: New evidence and new questions. *European Review of Social Psychology, 9*, 145-189.

Rimmele, U., Davachi, L. & Phelps, E. A. (2012). Memory for time and place contributes to enhanced confidence in memory for emotional events. *Emotion, 12*, 834-846.

Roediger, H. L., Wixted, J. H., & DeSoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. In L. Nadel & W. Sinnott-Armstrong (Eds.) *Memory and the law* (pp. 84-118). Oxford: Oxford University Press.

Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 21-38.

Rubin, D. C. (2005). Autobiographical memory tasks in cognitive research. In A. Wenzel & D. C. Rubin (Eds.), *Cognitive methods and their application to clinical research*. (pp. 219-241). Washington, D.C.: American Psychological Association.

Rubin, D. C. (2011). The coherence of memories for trauma: Evidence from posttraumatic stress disorder. *Consciousness and Cognition, 20*, 857-865.

Rubin, D. C. & Kozin, M. (1984). Vivid memories. *Cognition*, *16*, 81-95.

Schmidt, S. R. (2004). Autobiographical memories of the September 11th attacks: Reconstructive errors and emotional impairment of memory. *Memory & Cognition, 32*, 443-454.

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.

Shapiro, L.R. (2006). Remembering September 11th: The role of retention interval and rehearsal on flashbulb and event memory. *Memory, 14*, 129-147.

Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *Proceedings of the National Academy of Science, 104*, 389-394.

Smith, M. C., Bibi, U., & Sheard, D. E. (2003). Evidence for the differential impact of time and emotion on personal and event memories for September 11, 2001. *Applied Cognitive Psychology, 17*, 1047-1055.

Squire, L. R. & Slater, P. C. (1975). Forgetting in very long-term memory as a assessed by an improved questionnaire technique. *Journal of Experimental Psychology: General*, 1, 50-54.

Stanhope, N., Cohen, G., & Conway, M.A. (1993). Very long-term retention for a novel. *Applied Cognitive Psychology, 7*, 21-33.

Suengas, A. G. & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General, 117,* 377-389.

Talarico, J.M. & Rubin, D.C. (2008). Flashbulb memories result from ordinary memory processes and extraordinary event characteristics. In O. Luminet, A. Curci, & M. Conway (Eds.) *Flashbulb memories: New issues and new perspectives.* (pp. 79-97). NY: Psychology Press.

Talarico, J. M. & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14*, 455-461.

Tekcan, A., Berivan, E., Gülgöz, S., & Er, N. (2003). Autobiographical and event memory for 9/11: Changes across one year. *Applied Cognitive Psychology, 17*, 1057-1066.

Tinti, C., Schmidt, S., Testa, S., & Levine, L. J. (2014). Distinct processes shape flashbulb and event memories. *Memory & Cognition*, *42*, 539-551.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18*, 225-252.

Wang, Q. (2013). *The autobiographical self in time and culture*. NY: Oxford University Press.

Weaver III, C. A. & Krug, K. S. (2004). Consolidation-like effects in flashbulb memories: Evidence from September 11, 2001. *The American Journal of Psychology*, *117*, 517-530.

Wells, G. L. & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, 10, 138-144.

Winningham, R. G., Hyman, I. E., & Dinnel, D. L. (2000). Flashbulb memories? The effects of when the initial memory report was obtained. *Memory*, *8*, 209-216.

Wright, D. B., Gaskell, G. D., & O'Muircheartaigh, C. A. (1998). Flashbulb memory assumptions: Using national surveys to explore cognitive phenomena. *British Journal of Psychology*, *89*, 103-121.

Yarmey, A. D. & Bull III, M. P. (1978). Where were you when President Kennedy was assassinated?. *Bulletin of the Psychonomic Society*, *11*, 133-135.

Table 1

Distribution of Sample for Those Who Completed One or More Surveys, Including Survey 4

Survey	Recruitment location at the time of the first survey								Total
	Boston	New Haven	New York	DC	St Louis	Palo Alto	Santa Cruz	Web-Based	
1, 2, 3, 4	0	22	99	31	35	7	8		202
1, 4	0	4	23	7	3	1	1		39
2, 4	0	7	26	8	0	8	2		51
3, 4	0	18	73	30	2	0	0		123
1, 2, 4	0	7	29	9	2	12	3		62
1, 3, 4	0	0	18	7	1	0	2		28
2, 3, 4	0	17	58	17	3	0	3		98
4 only	0	0	62	41	0	0	0	311	414
No Match									104

Table 2

Proportion of participants in the Four-Survey Sample and the New Sample Fitting Designated Categories[1]

		Four-Survey	**New**
Gender	Female	.72	.60
	Male	.28	.33
Ethnicity/Race	White	.83	.66
	Black	.01	.02
	Asian	.03	.02
	Hispanic	.02	.02
	East/American Indian	.00	.01
	Mixed	.04	.05
Religion	Christian	.45	.40
	Jewish	.18	.06
	Muslim	.01	.01
	Buddhist	.02	.01
	Atheist	.04	.07
	Agnostic	.03	.04
	Other	.07	.05
	None	.10	.14
Politics	Extreme/Moderate Left	.67	.42
	Center	.05	.06
	Extreme/Moderate Right	.12	.15
	Independent	.05	.09
	Other	.01	.01
	None	.03	.04
Student[2]	Yes	.52	.24
	No	.48	.69

Notes.

1 – Figures do not always add up to 1.00 because some participants opted not to respond to some or all the demographic queries.

2 – Figures for Student Status are for at the time of the attack for the Four-Survey Sample, at the time of filling out the survey for the New Sample.

Table 3

Questions on the Surveys Relevant to Current Analyses

FLASHBULB MEMORY
(The six canonical features; Asked on all four surveys)

1) How did you first learn about it (what was the source of the information)?
2) Where were you?
3) What were you doing?
4) How did you feel when you first became aware of the attack?
5) Who was the first person with whom you communicated about the attack?
6) What were you doing immediately before you became aware of the attack?

EVENT MEMORY
(The five facts; Asked on all four surveys)

7) How many airplanes were involved in the attack?
8) What airline or airlines had planes hijacked? How many from each airline?
9) In the vicinity of which cities did the airplanes end up?
10) Where was President Bush when the attack occurred?
11) Many people think that these are the most salient events that occurred in the attack:
a) The Pentagon was hit by a hijacked plane
b) A second World Trade Center Tower was hit by a hijacked plane
c) A World Trade Center Tower collapsed
d) A World Trade Center Tower was hit by a hijacked plane
e) A second World Trade Center Tower collapsed
f) A hijacked plane crashed outside of Pittsburgh
Please indicate the order in which you became aware of each event.
Please indicate the order in which the events actually occurred.

FACTORS AFFECTING PERFORMANCE

Emotional Intensity
(As the questions appeared in Survey 4)

*For the following questions, we'd like you to tell us about your **CURRENT FEELINGS CONCERNING THE ATTACK**. Please indicate your response by marking the appropriate point on the scales provided. Note that you may indicate partial numbers (e.g. 3.5) [Scale was from 1 (low) to 5 (high).]*

12) At this moment, how strongly or intensely do you feel **sad** about the attack?
13) At this moment, how strongly or intensely do you feel **angry** about the attack?
14) At this moment, how strongly or intensely do you feel **fear** about the attack?
15) At this moment, how strongly or intensely do you feel **confusion** about the attack?
16) At this moment, how strongly or intensely do you feel **frustration** about the attack?
17) At this moment, how strongly or intensely do you feel **shock** about the attack?

*For the following questions, we'd like you to tell us about your **FEELINGS CONCERNING 9/11 IN THE TWO WEEKS FOLLOWING THE ATTACK**. Please indicate your response by marking the appropriate point on the scales provided. Note that you may indicate partial numbers (e.g. 3.5) [Scale was from 1 (low) to 5 (high).]*

18) How strongly or intensely did you feel **sad** about the attack?
19) How strongly or intensely did you feel **angry** about the attack?
20) How strongly or intensely did you feel **fear** about the attack?
21) How strongly or intensely did you feel **confusion** about the attack?
22) How strongly or intensely did you feel **frustration** about the attack?
23) How strongly or intensely did you feel **shock** about the attack?

Media coverage
(As the question appeared in Survey 4)

24) In the last seven years, how closely have you followed media coverage about 9/11? [Rate on a scale of 1 (very little) – 5 (very much).]

Ensuing Conversation
(As the question appeared in Survey 4)

25) In the last seven years, how much have you talked about 9/11? [Rate on a scale of 1 (very little) – 5 (very much).]

Residency
(In our analyses, we use the response on Survey 1)

26) Where do you consider home?

Personal Loss and/or Inconvenience
(In our analyses, we use the response on Survey 1)

27) Did you suffer any personal losses in the attack? If so, please specify.
28) Did the attack inconvenience your daily activities in some way? If so, please specify.

Table 4

Criticality and Accuracy Scores for Four Aspects of the Event Memory for 9/11, Compared Across Four Surveys

Event Feature	Criticality Score[4]	Accuracy Score			
		Survey 1	Survey 2	Survey 3[5]	Survey 4[5]
Critical					
Number of Planes	5.59 (1.58)	.95 (.22)	.89 (.32)	.86 (.35)	.88 (.33)
Crash Sites	6.46 (.90)	.95 (.14)	.94 (.17)	.90 (.24)	.89 (.21)
Non-critical					
Location of Bush	4.12 (1.74)				
Saw Moore		.89	.59	.90	.87
Did Not See		.90	.62	.80	.80
Bush Overall		.90	.60	.85	.83
Airline Names	3.69 (1.99)				
United Alone[1]					
Saw *United 93*		.93	.69	.59	.76
Did Not See		.85	.63	.54	.52
United Overall[2]		.86	.64	.54	.55
Airline Overall[3]		.88 (.28)	.75 (.37)	.59 (.42)	.56 (.42)
Temporal Order	----------	.89 (.12)	.83 (.23)	.83 (.23)	.89 (.10)
TOTAL ACCURACY SCORE		.91 (.11)	.80 (.19)	.81 (.21)	.81 (.19)

Notes.

1. Scores based on proportion of participants who correctly recalled United Airline.
2. Includes the five participants who did not report whether they did or did not see *United 93*.
3. Scores based on proportion of participants who correctly recalled both American and United.
4. Using a 1 – 7 scale.
5. *Fahrenheit 9/11* was released 3 months before the distribution of Survey 3; *United 93*, 5 years before the distribution of Survey 4. 105 participants reported watching *Fahrenheit 9/11* on Survey 3; 29 participants reported watching *United 93* on Survey 4.

Standard deviations in parentheses.

Table 5

Correlations between Memory Assessments (FBM Consistency, FBM Confidence, and Event Memory) and Factors Affecting Performance (Emotional Intensity, Media Attention, and Ensuing Conversation). Correlations are between Memory Assessments on Survey 4 and Factor Measures on Surveys 1 through 4.

		Temporal Period in which Factors Were Measured			
Memory Assessment	Factors	Survey 1 (First Week)	Survey 2 (Over the First Year)	Survey 3 (Between Years 1 and 3)	Survey 4 (Between Years 3 and 10)
Flashbulb Memory (Consistency)					
	Emotional Intensity	-.01	.03	-.02	-.03
	Media Attention	.02	.01	.02	-.10
	Ensuing Conversation	-.10	.03	.01	-.10
Flashbulb Memory (Confidence)					
	Emotional Intensity	-.04	-.07	-.02	-.04
	Media Attention	-.09	-.01	-.01	.04
	Ensuing Conversation	.01	.07	-.01	.03
Event Memory (Accuracy)					
	Emotional Intensity	-.04	-.05	-.02	-.06
	Media Attention	.09	.13	.08	.16*
	Ensuing Conversation	.10	.19**	.05	.20**

Notes.
* - $p < .05$
** - $p < .01$

Table 6

Consistency of Flashbulb Memories, Confidence in Flashbulb Memories, and Accuracy of Event Memories as a Function of Residence, Across All Surveys

		Residence	
		New York City	Other
Flashbulb Memory Consistency			
	Survey 2	.60 (.18)	.64 (.19)
	Survey 3	.56 (.24)	.57 (.13)
	Survey 4	.58 (.21)	.61 (.21)
Flashbulb Memory Confidence			
	Survey 2	4.52 (.68)	4.41 (.59)
	Survey 3	4.32 (.99)	4.24 (.88)
	Survey 4	4.63 (.41)	4.31 (.62)
Event Memory Accuracy			
	Survey 1	.91 (.12)	.91 (.11)
	Survey 2	.83 (.19)	.78 (.19)
	Survey 3	.82 (.20)	.80 (.22)
	Survey 4	.84 (.17)	.79 (.21)

Note. Standard deviations in parentheses.

Table 7

Correlations between Averaged Emotional Intensity (EI) and Either Media Attention (MA) or Ensuing Conversation (EC)

	Survey 1 (First Week)		Survey 2 (First Year)		Survey 3 (Between Yrs 1 and 3)		Survey 4 (Last Seven Years)	
EI at time of survey	MA	EC	MA	EC	MA	EC	MA	EC
S1	.14	.15*	.15*	.21**	.16*	.13	-.03	.04
S2	.09	.11	.28**	.32**	.24**	.23**	-.03	-.01
S3	.11	.14	.22**	.27**	.27**	.25**	-.04	-.02
S4	.12	.11	.19**	.18*	.17*	.15*	-.02	-.02

Notes. Emotional intensity was measured at the time the survey was filled out, not when one first learned about the attack. Media Attention and Ensuing Conversation refer to the amount of attention or conversing during the indicated time period (e.g., between years 1 and 3). Highlighted figures are those of most interest. The top row represents the correlations between emotional intensity in the first week and subsequent media attention over the next ten years, indicating the impact of the initial emotional intensity and subsequent media attention and ensuing conversation. Figures on the diagonal examine, for example, the impact of emotional intensity at the end of the year and the amount of media attention and ensuing conversation in the last year.

* - $p < .05$
** - $p < .01$

Table 8

The Effect of Personal Loss and/or Inconvenience on Current Emotional Intensity, Media Attention, and Ensuing Conversation, as rated on Surveys 1 – 4

	Survey 1	Survey 2	Survey3	Survey 4
Emotional Intensity				
With Personal Loss	3.57 (.75)	2.88 (.98)	2.77 (.99)	2.56 (1.00)
Without	3.31 (.83)	2.77 (.86)	2.75 (.79)	2.41 (.95)
Media Attention				
With Personal Loss	4.47 (.64)	3.49 (1.02)	3.14 (1.39)	3.25 (1.06)
Without	4.20 (.90)	3.46 (1.07)	3.33 (1.27)	2.97 (1.15)
Ensuing Conversation				
With Personal Loss	4.52 (.57)	3.72 (1.09)	3.08 (1.32)	2.82 (1.05)
Without	4.30 (.77)	3.49 (1.00)	3.19 (1.11)	2.65 (.90)

Note: Standard deviations in parentheses.

Table 9

Examples of Repeated Inconsistencies Across Four Surveys, for Four Different Participants

Participant 1: Where, What Doing

S1: Kitchen, making breakfast
S2: In dorm room, folding laundry
S3: Ironing in dorm room
S4: In dorm room, ironing

Participant 2: First Communication

S1: Told a fellow biker who I intercepted as she approached the bridge. I told her to turn and go home.
S2: A fellow biker who was watching with me.
S3: I communicated with a guy watching next to me.
S4: The person next to me. I don't know him.

Participant 3: How learned about

S1: Today Show
S2: Girl in dorm rushed in my room and told me
S3: Girl came into my dorm room and told me
S4: A girl came into my room and told me to come quickly

Participant 4: Where

S1: I was sitting in my office
S2: I was standing on the street
S3: Standing near the office on the street
S4: Near the office

Table 10

Proportion of Repetitions and Corrections of Incorrectly Remembered Facts about 9/11

	Wrong on Changed on	Survey 1 Survey 2		Survey 2 Survey 3		Survey 3 Survey 4	
	Type of Change	*Repeated*	*Corrected*	*Repeated*	*Corrected*	*Repeated*	*Corrected*
Critical							
Crash Site		.15	.85	.29	.76	.25	.77
Number of Planes		.20	.60	.00	.88	.26	.92
Non-Critical							
Carrier Names							
AA			.42		.28		.27
United			.36		.26		.35
Saw United 93			.27		.22		.58
Did not			.42		.27		.33
Wrong Carrier		.35		.11		.23	
Bush's Location		.46	.11	.31	.65	.47	.32
Saw Moore Film		.58	.12	.28	.73	.39	.32
Did Not		.40	.10	.35	.54	.57	.31

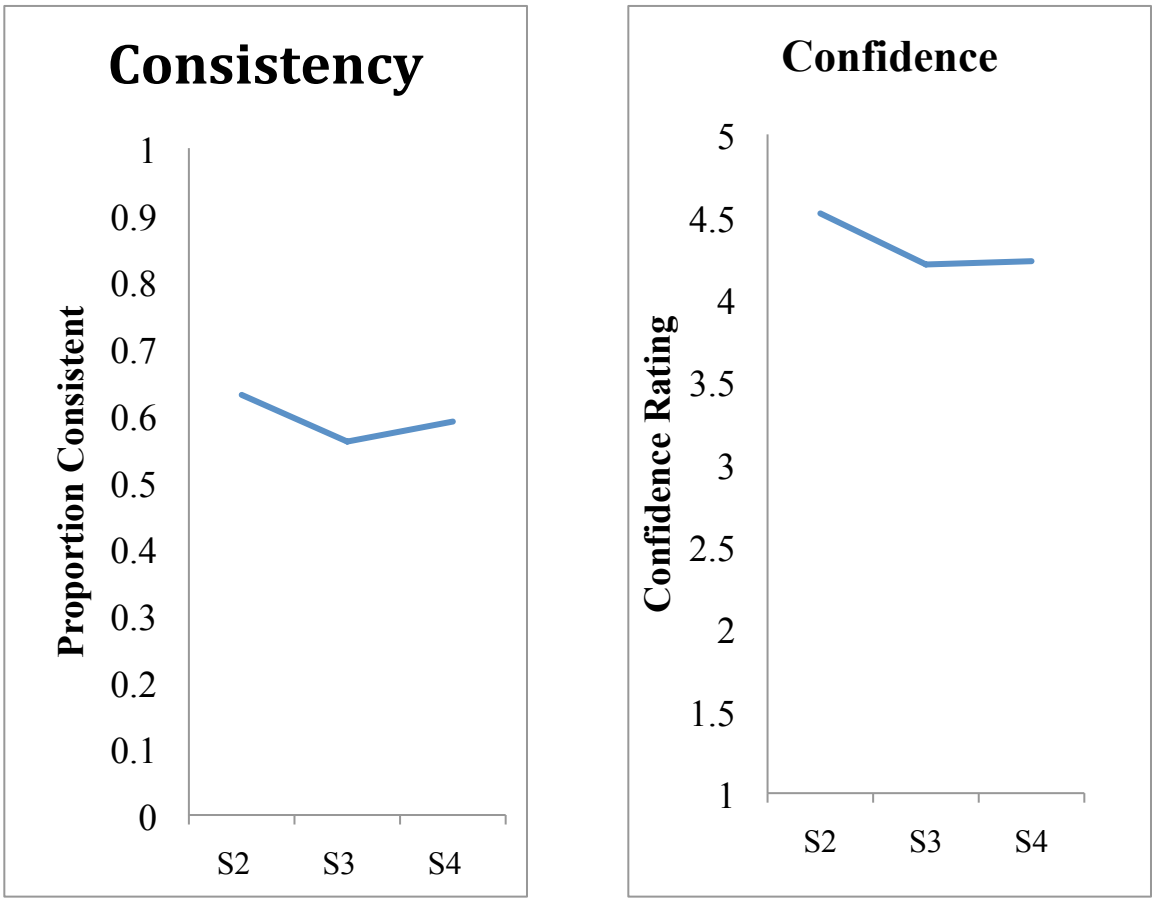

Figure 1. *Consistency and Confidence*. Figure 1a. Consistency score of flashbulb memories, as averaged over six canonical features, for responses on Surveys 2, 3, and 4, with Survey 1 serving as baseline. Figure 1b. Confidence ratings averaged over the six canonical features, for responses on Surveys 2, 3, and 4. Ratings were on a 1 – 5 scale, with 5 as extremely confident.

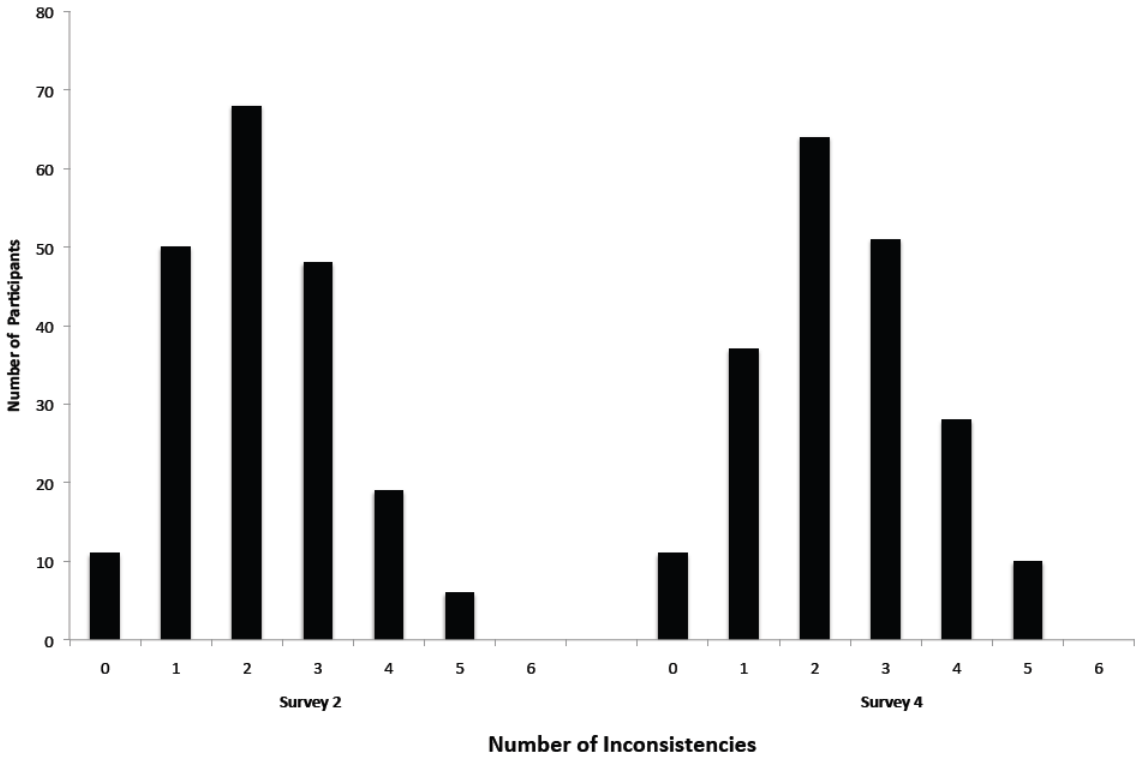

Figure 2. Number of inconsistencies (out of 6) in the flashbulb memories, for the 202 participants in the Four-Survey Sample, for Surveys 2 and 4.

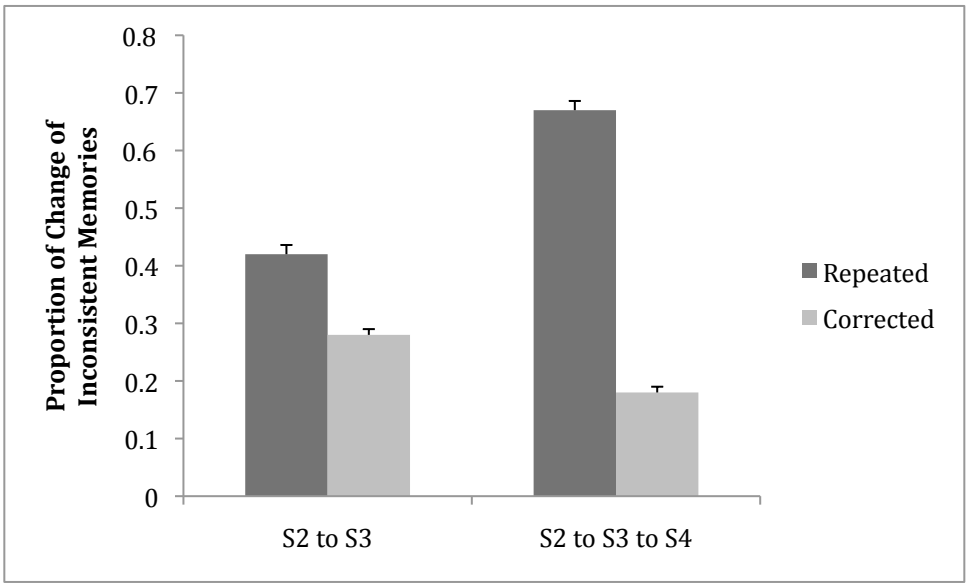

Figure 3. Change of inconsistent memories from one survey to another: S2 to S3: Inconsistent memories from Survey 2 either repeated or corrected on Survey 3; S2 to S3 to S4: Inconsistent memories from S2 repeated on S3 and either further repeated or corrected on S4. Error bars represent standard errors.