THE CONTROVERSY OVER
RECOVERED MEMORIES

Henry L. Roediger, III
Washington University, St. Louis

Erik T. Bergman
Pitney Bowes Technology Center

The authors discuss 4 issues in this commentary: (a) the assumptions and evidence used to support the case for dissociated and recovered memories (noting that the evidence is weak and the assumptions internally inconsistent as well as contradictory to a mass of experimental evidence about human memory); (b) the process by which dissociated memories are said to be recovered (events that were originally very poorly encoded as fragmentary, kinesthetic memories cannot be recovered with accuracy later); (c) 4 bodies of relevant, but neglected, research on human memory (reminiscence and hypermnesia, effectiveness of retrieval cues, priming in implicit memory tests, and intentional forgetting); and (d) the issue of appropriate research strategies to gain evidence on the thorny issues of long-delayed retrieval of information. Current evidence does not support the conclusion that memories of repeated abuse are dissociated and recovered with accuracy years later.

Comments for Psychology, Public Policy, and Law on the Final Report of the American Psychological Association (APA) Working Group on Investigation of Memories of Childhood Abuse will doubtless function as a projective test, telling as much about the commentators’ a priori opinions concerning the controversy over recovered memories as about the report itself. The contentious issues surrounding repressed (or dissociated) memories and their possible later recovery are being thoroughly aired these days. Numerous articles, book chapters, special issues of journals, and entire books have been devoted to these matters. Our reflections here provide new perspectives on the topic, as well as direct readers to some neglected research that seems germane to the issues at hand.

First, we discuss the issue of remembering and forgetting traumatic events and point to inconsistencies in assumptions Alpert, Brown, and Courtois (1998c) make in their analysis of both clinical and experimental studies. Next, we turn to the possible recovery of memories after long periods of time and the little we know about the possibility of such recoveries. Third, we consider other evidence from the memory literature not mentioned in the APA report, which has possible relevance for the issues of the repressed–recovered memory controversy. Fourth, we discuss some issues of external validity and appropriate research strategies that surround these contentious issues.

Forgetting of Traumatic Events

Alpert, Brown, and Courtois (1998c) provided a case for the possibility that memories of great physical and sexual abuse during childhood can be repressed (or
dissociated from other memories) and can be recovered with considerable accuracy many years later in adulthood. The chain of reasoning includes the following assumptions: First, encoding of a single traumatic experience may often be quite thorough (Terr’s [1991, 1994] Type I reaction to trauma) and consequently, memory for single traumas may be quite good. Second, repeated traumas are encoded in a different manner (Terr’s Type II traumas), which frequently renders the events incapable of being consciously recollected. In particular, children experiencing repeated trauma (especially sexual abuse) are said to encode the events by means of kinesthetic and sensory reactions, not as descriptions of what happened on a symbolic level. The memories are said to be fragmentary, and they are not integrated with other aspects of the person’s knowledge about themselves. Alpert et al. (1998c) cited favorably the conclusions from the manuscript of van der Kolk, Fisler, and Vardi (1994) to the effect that traumatic experiences initially are imprinted as sensations or feeling states, and are not translated into personal narratives, in contrast with the way people seem to process ordinary information. This failure of information processing on a symbolic level, in which it is categorized and integrated with other experiences is at the very core of what goes wrong when people are traumatized. (as cited by Alpert et al., 1998, p. 974)

Alpert et al. (1998c) cited Fisler, Vardi, and van der Kolk’s (1994) study showing that when incest survivors described their nontraumatic memories from childhood, even “the nontraumatic memories of incest survivors were more fragmentary and contained more negative affect than the nontraumatic memories of other groups” (p. 971). These memories of abuse are said to be dissociated, or cut off, from the rest of the person’s memories, and in many cases they are not recollected until some event (whether in therapy or outside it) triggers the return of the memories many years later.

The bulk of the Alpert et al. (1998c) report on memory for trauma is devoted to case study and retrospective survey research that bolsters the assumptions summarized above. We find the case they make interesting and provocative, but we agree with the reply by Ornstein, Ceci and Loftus (1998c), that it is far from convincing. One problem is the rather poor quality of the evidence used by Alpert et al., as they themselves often acknowledge (e.g., p. 974). Indeed, Alpert et al. referred to Terr’s (1991, 1994) conceptualization of Type I and Type II traumas, on which so much hinges, as “hypothetical and in need of further research” (p. 968). Actually, this view seems overly optimistic, as discussed below.

The second problem with the Alpert et al. (1998c) report is more difficult to assess: Many of the assumptions made by Alpert et al. and others working in this area contravene generally accepted conclusions derived from psychologists engaged in the experimental study of memory over the past 113 years, since Ebbinghaus’s (1885/1964) original investigations. One of the first questions Ebbinghaus asked (chap. VI) was how repetition affected memory. His research clearly showed that repetition improved memory, and virtually every experiment directed at this issue in the entire history of experimental psychology confirms this result (see Crowder, 1976, chap. 9, for a summary). The one minor qualification is that massed repetition or massed rehearsal (e.g., presenting or saying a word or
other experimental event over and over in rapid succession) sometimes has small or even negligible effects on recall and on certain implicit memory tests; Challis & Sidhu, 1993). However, spaced repetitions (those occurring with other events intervening) virtually always improve retention on any kind of memory measure. To our knowledge, no one has ever reported a strong negative effect, with repeated events remembered much worse than single events. So, given these facts, how can we explain that three psychologists writing an APA report 110 years after Ebbinghaus first showed that repetition improved retention, accept as valid a theory that asserts exactly the opposite, that repeated events are more poorly remembered than single events?

The ostensible answer to this perplexing question is that the normal laws of memory painstakingly worked out by experimental psychologists over the past century are largely irrelevant to clinical psychologists interested in memory for trauma. Alpert et al. (1998c) questioned the relevance of research on “normal” memory in several places. This seems a curious practice by those who describe themselves as following a scientist–practitioner model for clinical psychologists (see Brenneis, 1997). We have examined the evidence on which Terr (1988) based her theory (repeated and single cases of abuse in children), and it does not withstand scrutiny. For example, children who experienced repeated trauma in her study were on average about 2 years, 4 months old, whereas children who experienced single traumas were 3 years, 1 month. Just this difference in age, with memory changing so rapidly in young children, could account for any difference in recollection of the trauma. The experiences seem to differ on other characteristics, too. We see no evidence that memory for repeated events, traumatic or not, is ever worse than memory for single events of the same type. If this point is accepted—and it seems incontestable—it undercuts a central claim in the Alpert et al. (1998c) argument.

Although it is difficult to study memory for trauma by means other than case studies and retrospective reports—both highly fallible techniques—it is possible to study memory for emotional events. The literature is extensive and complex, but many studies (both experimental and naturalistic) show that memory for emotional events is reasonably good, as Alpert, Brown, and Courtois (1998a) noted. Indeed, many people have reported that their memories for high-impact events seem highly accurate, so that the term flashbulb memory was introduced by R. Brown and Kulik (1977) to refer to recollection for such events (such as when one first heard of the President Kennedy’s assassination, or of the explosion of the Challenger). People often feel that they remember with great accuracy even minor details of these emotional events. More difficult to study are emotional

---

1Some studies have used prospective methods by identifying victims of trauma or abuse and then testing their memories for the events over time. Early results from these methods are conflicting; Parker, Bahrick, Merrit, Lundy, and Fivush (1998) reported generally good memories in children for acute trauma (experiencing Hurricane Andrew), whereas L. M. Williams (1994) reported some forgetting in the retrospective accounts of victims with previously documented abuse. However, Pope and Hudson (1995) criticized this latter study on various methodological grounds.

2Neisser and Harsh (1992) showed that even these strong feelings of confidence can sometimes be erroneous. Memories for hearing the news of the space shuttle disaster were only moderately good, despite high levels of confidence and emotion. In some cases, what subjects remembered 3 years after the event was incompatible with what they remembered immediately following the event,
experiences unique to each individual (recollection of one's first sexual experience, the death of a parent, and so on), but to our knowledge no evidence exists that these dramatic moments in the individual's life are less well remembered than public, shared events that form distinctive ("flashbulb") memories. In fact, a recent flashbulb memory study demonstrated that greater personal involvement in an emotionally arousing event (in this case an earthquake) led to exceptionally accurate memory for the event (Neisser et al., 1996). If retention of emotional events is accepted as quite good, then why would some memories be dissociated or repressed?

In answer to this question, Alpert et al. (1998c) made a strong argument that findings from known victims of posttraumatic stress disorder (PTSD) can be generalized to the issue of dissociated–recovered memories of possible victims of sexual abuse. In their reply to the Ornstein, Ceci, and Loftus (1998b) comment, Alpert, Brown, and Courtois (1998b) argued that "because there are no data available to suggest that the trauma response in this population [adults sexually abused as children] varies meaningfully from the trauma response resulting from other kinds of traumatic experience, we believe with some confidence that these results can at least be initially applied to adults sexually abused as children" (p. 1016). However, we find that this generalizability is rather selectively applied, because the responses claimed for victims of abuse differ greatly from the conclusions based on the reports of survivors of PTSD after other types of trauma in several ways. First, the available literature on PTSD from soldiers repeatedly exposed to the horrors of battle shows that, whereas they may forget details and they may merge events together and show other memory errors, they do not forget that they had been in battle or in a war. Their memory for having experienced the core event is intact. However, it is just this complete forgetting of the core event—the alleged forgetting of having suffered sexual abuse even after (or especially after) repeated episodes of abuse—that is at the heart of the whole recovered memory debate. The generalization from PTSD research hardly seems to confirm the analogy being drawn on this important point.

A second frequent feature of PTSD, as Alpert et al. (1998c) noted in several places, is the intrusive nature of flashback memories. The most typical memory difficulty of trauma victims is the frequent and unavoidable recall of their trauma, not its dissociation or repression. What mechanism would plausibly explain both vivid memories of trauma (on some occasions or in some people) and complete forgetting of trauma (on some occasions or in some people)? Except for the distinction between unique (Type I) and repeated (Type II) traumas postulated by Terr (e.g., 1994)—a conception which seems far-fetched at best and almost completely lacking in supportive evidence—we see no reasonable mechanism postulated by Alpert et al. (1998c) or others. At least Terr's idea seems relatively concrete and worked out and almost testable (if it were ethically possible to conduct the relevant experiments). The other theories mentioned as perhaps causing the authors to conclude that these were not simply inaccurate recollections, but false recollections. Similarly, Wright (1993) found memory distortion in recall of hearing the news of the Hillsborough disaster. Merely hearing the news, it seems, is not necessarily sufficient to create a reliable flashbulb memory, and retrospective accounts of even highly emotional events can be fallible.
accounting for such memory failures—psychic numbing, betrayal trauma, and so on—seem even more conjectural.

We hasten to add that the lack of a plausible mechanism to create the possible amnestic effects of trauma on memory does not make the effects untrue or impossible, of course. We agree with Alpert et al. (1998c) that some of the isolated case studies do warrant further investigation (see Schooler, Bendikson, & Ambadar, 1997, for such attempts). As we discuss below, the methods by which more conclusive evidence could be gained are not entirely clear. However, our point here is that the analogy from memories of cases in which PTSD is known to occur seems in several ways to undercut, rather than to support, claims of possible cases of recovered memory of sexual abuse.

Recovered Memories

In the preceding section, we argued that the possible mechanism by which trauma might cause amnesia for the painful events has not been well specified. However, the putative mechanisms for the forgetting of trauma have at least been addressed. Far more mysterious is how painful events, banished to an unconscious state for years through some mechanism of dissociation or repression, could be brought back to consciousness and recollected with great fidelity. Aside from a few case studies noted by Alpert et al. (1998c), no evidence from the voluminous literature on human memory makes us think this is possible. Indeed, the very evidence cited by Alpert et al. (1998c) leads us to conclude that recovered narrative memories even of actual trauma survivors would be highly suspect. This seems to be a logical deduction from the evidence they reported, for three reasons discussed below.

First, traumatic memories are said to be encoded in nonverbal ways—through patterns of sight, sound, kinesthesis—but not in symbolic form. These body memories, as they are often called (van der Kolk, 1994), cannot be recalled with accuracy in words. By definition, the memories are stored in a nonsymbolic form. However, conscious memories of events expressed verbally are reported in symbolic form, as a narrative of events. Therefore, if people have delayed recall of some traumatic event from many years previously, they must translate the kinesthetic body memories into symbolic form. If such a translation is possible at all (which seems unlikely), it should be exactly the sort to be highly influenced by the numerous processes of suggestibility documented by Ornstein et al. (1998a).

Second, Alpert et al.'s (1998c) report is replete with demonstrations of the fallibility of trauma survivors' memories. They cited numerous studies showing that the survivors' memories tend to be poorer than those of comparison groups. In fact, they reported that one study (Fisler, Vardi, & van Der Kolk, 1994) showed that trauma victims' memories for nontraumatic childhood events were more fragmentary and impoverished than were those of control subjects. Why then would they consider recovered memories of other childhood events to be exceedingly accurate?

The two points above, both accepted by Alpert et al. (1998c), would seem to lead logically to the conclusion that the delayed recall or recovered memories of trauma victims may not be accurate. (This conclusion clearly would not apply to memories that had been recalled immediately and recalled repeatedly over time,
although the findings of Neisser and Harsh, 1992, may be relevant here.) To believe differently, as apparently Alpert et al. do, leads to the following antinomy: Trauma victims suffer poor memories for their trauma, memories encoded nonsymbolically and in fragmentary form; however, when these memories are recovered many years later and placed in a verbal narrative, they are accurate and can be trusted. The simultaneous acceptance of both propositions is difficult to justify.

A seed for a third reason to worry about the trustworthiness of recovered memories is also contained in Alpert et al.'s (1998c) report. In their discussion of possible dissociative mechanisms of memory for trauma, they cited evidence showing that psychiatric patients who report childhood sexual abuse also score high on the Dissociative Experiences Scale (DES; Bernstein & Putnam, 1986). However, recent research documents another interesting correlation relevant to this debate. Two recent studies have independently shown that individuals scoring high on the DES are more prone to develop false memories in experimental paradigms (Hyman, Husband, & Billings, 1998; Winograd, Glover, & Peluso, 1998). The Hyman et al. study used a procedure in which suggestions were made to adults that they may have experienced childhood events, using a variant of the "lost in the mall" technique of Loftus (1993). (In fact, parents and siblings had reported that the critical target events had, to their knowledge, not occurred.) Hyman et al. reported that subjects scoring higher on the DES were more likely to develop false memories than those scoring low on the DES.

Winograd et al. (1998) used a list learning paradigm developed by Roediger and McDermott (1995) that creates a robust memory illusion. Roediger and McDermott (1995) had subjects study lists of words (e.g., bed, rest, awake) associatively related to a single (nonpresented) word (e.g., sleep). They showed that subjects were quite likely both to recall the nonpresented word on both immediate and delayed (McDermott, 1996) tests of free recall and were also likely to falsely recognize the missing word. Winograd et al. (1998) gave subjects a variety of individual difference measures before putting them through their paces in this paradigm and found that the probability of false recall and false recognition correlated significantly with the DES. Assuming this pattern holds in further research, the DES seems to predict not only loss of traumatic memory, but also proneness to create false memories. This constitutes a third reason to be skeptical of delayed or recovered memories in people who score high on the DES.

Neglected Memory Research

Ornstein et al. (1998a) concentrated on three areas of memory research highly relevant to issues of the recovered memory debate: developmental aspects of memory, the literature on suggestibility of memory, and the evidence surrounding people's abilities to distinguish between fact and fantasy (reality monitoring) and the ability to discriminate sources of information (source monitoring). These three literatures are clearly germane to the controversy at hand; we consider briefly four other literatures that might inform the debate.

First, we can ask what is known about "recovered memories" in laboratory settings. One hallmark of recovered memories in therapy is that people recall an event after long periods of trying. Is there evidence for such a process in the human
CONTROVERSY OVER RECOVERED MEMORIES

memory literature? That is, in the context of laboratory experiments, if people repeatedly try to recall difficult-to-retrieve events, will they ever recall an event on (say) the third attempt if they have tried hard and failed to recall it on two previous occasions? The answer to this question is yes, and it is surprising that Alpert et al. (1998c) did not cite this voluminous literature on repeated testing. Classic experiments by Ballard (1913) and W. Brown (1923) established the phenomenon of reminiscence: recovery of an event on a later test that could not be recalled on an earlier test. Indeed, virtually all experiments using repeated free recall (tests given with no or minimal overt retrieval cues) reveal recall of events on later tests that could not be recalled on earlier tests. Of course, forgetting between tests also occurs—events recalled on a first occasion may be forgotten during a second occasion. However, perhaps the most remarkable finding from laboratory experiments on repeated testing is that overall recall—the total number of events recalled—often increases over repeated tests. This phenomenon of improved recall over repeated tests in the absence of further study of the material has been labeled hypermnnesia (Erdelyi & Becker, 1974).

There is now a substantial literature on the phenomena of reminiscence and hypermnnesia (for reviews, see Erdelyi, 1996; Payne, 1987; and Roediger & Challis, 1989). Both these phenomena are ones of “delayed recall” or “recovered memories,” so one might expect their discussion to inform the debate on possible recovered memories in therapeutic contexts, but the vast literature surrounding the recovered memory controversy has made little reference to these experimental discoveries (but see Erdelyi’s, 1996, recent book for relevant discussion). Of course, making the connection between these laboratory phenomena and the current controversy presupposes that evidence from experiments using materials such as words, pictures, and stories has bearing on the issues arising outside the lab, an assumption about which Alpert et al. (1998a, 1998c) are skeptical.

Roediger, McDermott, and Goff (1997) considered the implications of the reminiscence–hypermnnesia literature for the issues under current debate. Some points might be briefly summarized here. First, recovered memories in these experimental contexts are clearly of events that were encoded well in the first place (unlike the way repeated trauma is said to be encoded by Alpert et al., 1998c). Second, the experiments on recovered memories in the lab are usually of very recent events, with repeated tests occurring within minutes of the original encoding. In addition, the repeated retrieval attempts usually occur with very little time between tests. Erdelyi and Kleinbard (1978) did report hypermnnesia from study of pictures when subjects took repeated tests (three per day) over a week, but little is known beyond this window of time. If long gaps occur between tests, forgetting rather than hypermnnesia occurs (Wheeler & Roediger, 1992). Third, a critical issue in the repeated testing literature concerns changes in response criteria over repeated tests. Erdelyi and his collaborators have typically used a method of forced recall to study hypermnnesia, in an attempt to control for guessing (e.g., Erdelyi & Becker, 1974). In this procedure, after study of a fixed set of material (say, 60 pictures), subjects try to recall as many as possible over repeated tests, but

---

3 Hypermnnesia in this sense is different from the way Alpert et al. (1998c) used the term. They used it in the older sense, meaning strikingly good memory for a particular event (e.g., Stratton, 1919). Both uses of the term are legitimate but should not be confused.
on each test they are required to write down a fixed number of responses (names of the pictures). The number of required responses is always more items than the experimenter believes subjects can accurately recall. The reason for these precautions is to hold response criterion constant—any increases in recall cannot then be attributed to increased guessing. Increases in recall (hypermnesia) do occur with either forced recall or free recall. However, Roediger, Wheeler, and Rajaram (1993) pointed out that forced recall itself can introduce confusion and errors, which we discuss next.

The experimenter in a forced recall experiment gives the subject credit when he or she successfully produces an item from the studied set. However, subjects are guessing a lot on forced recall tests, so can they reliably tell from trial to trial which memories are correct and which are not? In a preliminary experiment that addressed this question, Roediger et al. (1993) reported considerable confusion: Subjects often judged that erroneous items produced on the test as guesses were actually ones they had studied. Curiously, some of the items that had actually been studied and which subjects produced on the test were deemed to be guesses. These confusions probably represent a form of source monitoring error (Johnson, Hashtroudi, & Lindsay, 1993): Subjects during forced recall may produce many memories; after doing so repeatedly, they have trouble deciding which events actually occurred and which they produced as a result of the demands of the forced recall procedure. Similar processes may operate during repeated retrieval attempts in other contexts, including repeated attempts to retrieve memories of sexual abuse (or any other event) over time, especially when people are encouraged to imagine and to guess. Thoughts and imaginings aroused during one retrieval attempt may carry over to the next attempt and come to mind more readily. With repeated retrieval of putative “events of the past” the rememberer can become more confused about what actually happened, as in forced recall experiments. The rememberer may supply his or her own misinformation. Roediger et al. (1997) discussed these possibilities and attendant evidence in more detail, and Ackil and Zaragoza (1998) reported similar confusions arising in children with this kind of forced recall procedure. In fact, children were more susceptible than young adults to these kinds of errors in their experiment.

A key question in the recovered memory debate is the following: To what extent is it possible for adults to recollect events from their childhoods that have never been previously recollected since childhood? (By events, we refer to any events, traumatic or not). This is a difficult question to answer, because the research must be (in some sense) longitudinal. That is, as minimal conditions, researchers must (a) identify a set of events known to have happened in childhood, and then (b) test for recollection of those events during adulthood, presumably using a set of plausible distractors if the test is one of cued recall or recognition. The second condition is needed to estimate to what extent positive responses to the events that actually happened might be due to guessing or to other inferential processes.

Despite the centrality of this question for the recovered memory debate (or for the study of memory in general), we know of no studies that provide answers. The closest we know (and it is not very close) is M. D. Williams and Hollan’s (1981) study of recollections of high school classmates. They had four people try to recollect as many of their high school classmates as possible after being out of
CONTROVERSY OVER RECOVERED MEMORIES

School for periods varying from 2 to 10 years. The recollections occurred in the form of free recall tests over a period of 10 hours. Cumulative recall curves from two of their subjects appear in Figure 1, showing both true memories (correct recall of classmates) and false memories (recall of people who were not classmates, referred to by M. D. Williams and Hollan as *confabulations*). Accurate recall increased over time, as in studies of reminiscence and hypermnesia.

---

**Figure 1.** Cumulative number of names recalled in M. D. Williams and Hollan's (1981) study of recall of classmates in the senior year in high school. The two functions represent recall of actual classmates (correct recall) and other people who were not classmates (fabrications). The top and bottom graphs represent data from Subject 1 (S1) and Subject 3 (S3).
However, inaccurate recall also increased over time; for all four subjects, the proportion of false responses increased the longer subjects continued trying to remember. That is, at first most recollections were accurate, but the longer they kept at it, the more likely subjects were to produce false recalls. These erroneous recollections were quite possibly of people they actually knew; however, these people were not in their senior class in high school, and therefore would represent source errors.

M. D. Williams and Hollan’s (1981) study is clearly preliminary, but larger scale experiments using similar methods are needed to answer the question of how likely people are to remember childhood events. Obviously, subjects must be tested for memories formed at younger ages, and many more of them should be included in future work. The limiting factor in this type of research is finding a set of events for a child that were known to happen during his or her early years. Still, from the limited evidence at hand, it appears that during protracted periods of recall, false recollections may occur and when they do, they happen late in the period of recollection. Schwartz, Fisher, and Hebert (1998) reached the same conclusion from laboratory experiments on recall: Errors tend to occur late in the recall process.

As we have shown, the repeated testing literature is relevant to the issues in the recovered memory debate, and the general absence of its discussion thus far is surprising. We think three other bodies of work are also relevant and, although they are sometimes mentioned in passing, each deserves further scrutiny in this context. We briefly consider these three in turn.

Tulving and Pearlstone (1966) distinguished between information available in memory (that which is stored) and accessible information (that which can be produced on a particular memory test). They showed that providing retrieval cues relevant to information can lead to much greater recall than when subjects are tested with no overt cues (or under free recall conditions). Their work and that of others spawned a huge literature on the effectiveness of retrieval cues (see Tulving, 1983, chaps. 10 and 11, and Roediger & Guynn, 1996, for partial reviews). Alpert et al. (1998c) sometimes referred to retrieval cues as unlocking seemingly forgotten memories—and they do. However, no one has closely examined the vast literature from experimental psychology on the effectiveness of retrieval cues as it might be applied to the current controversy. Because the issue of cuing is so critical, this omission again seems curious. We note in passing that retrieval cues sometimes have paradoxical effects; what seem to be very good, plausible cues can sometimes inhibit rather than facilitate recall (e.g., Roediger, 1973; Slamecka, 1968). Some cues block recall and so people would actually be better off covering up the cues and testing themselves by free recall (see Nickerson, 1984, and Anderson & Neely, 1996, for reviews of these inhibitory effects of retrieval cues).

Another key issue that deserves further comment is the relevance of implicit memory literature in the recovered memory debate. Alpert et al. (1998c) used the term implicit memory several times but never cited any of the primary (or even secondary) literature on the topic. The term implicit memory was coined by Graf and Schacter (1985) to refer to the indirect testing of memory. Explicit memory tests are those that assess memory directly by instructing subjects to recollect information from a target event. Implicit memory tests assess retention indirectly by giving subjects a task that seems ostensibly unrelated to the target event:
Retention of prior experiences is usually revealed by showing improved performance (priming) on the task from prior events, relative to performance on targets with no prior exposure. Implicit memory tests have revealed that exposure to target events can produce priming in brain-damaged subjects who are otherwise quite poor at recollecting the event explicitly (e.g., Graf, Squire, & Mandler, 1984). In addition, numerous experiments with non–brain-damaged subjects have shown dissociations (interactions) between explicit and implicit memory measures as a function of numerous independent variables (see Schacter, 1987, and Roediger & McDermott, 1993, for reviews).

Alpert et al. (1998c) and others (e.g., van der Kolk, 1994) cited this work as showing the influence of unconscious processes in memory. In a sense, that is true, because implicit memory studies with amnesic patients show that people can demonstrate normal levels of priming on implicit memory tests for information that they cannot consciously recollect (see Moscovitch, Vriezen, & Goshen-Gottstein, 1993). However, most of these studies use what are known as data-driven or perceptual implicit tests (Jacoby, 1983; Roediger & Blaxton, 1987; Tulving & Schacter, 1990). In the criterion test of retention, subjects are usually given degraded forms of words or pictures and asked to guess their identity, with priming reflected in better identification of words or pictures that have been recently studied. These priming effects are thought to reflect potentiation of perceptual processes or systems that occur relatively early in neural processing, before meaning is assigned to the events (e.g., Srinivas, 1993; Tulving & Schacter, 1990).

This early presemantic priming would seem to have little or nothing to do with the type of unconscious mentation that is alleged to be responsible for repression or dissociation of traumatic memories, but nonetheless proponents of the reality of repressed–recovered memories use the concept of implicit memory rather freely (if vaguely). Sometimes they just get it wrong. Alpert et al. (1998c, p. 962), when discussing Hilgard’s (1977) neodissociation theory, wrote that "this normal form of dissociation allows an individual to do two things simultaneously, one requiring more explicit memory, the other more implicit." They seem to be confusing the terms explicit and implicit memory with other concepts from cognitive psychology, probably those of controlled and automatic processing (e.g., Posner & Snyder, 1975; Shiffrin & Schneider, 1977). There may well be some points in the implicit memory literature that could inform the recovered memory debate, but these have not been brought out yet. In addition to perceptual priming tests, another kind of implicit test (conceptual tests) do access meaning of concepts (Blaxton, 1989; Roediger & Blaxton, 1987). However, much less is known about conceptual priming at this point, so the relevance to the recovered memory controversy is uncertain.

A final literature that is worth considering in greater detail is that of intentional or directed forgetting. In directed forgetting experiments, subjects are exposed to material and then later told to forget it. Since the late 1960s (e.g., Bjork, Laberge, & Legrand, 1968; Weiner & Reed, 1969), a voluminous literature has developed on this topic. Because one interpretation of the concept of repression is that it is the act of trying not to think of distressing information (e.g., Freud, 1901/1965, p. 13), the directed forgetting literature takes on heightened interest for the repressed–recovered memory literature. Indeed, Weiner and Reed (1969) interpreted the
finding that recall of to-be-forgotten information is worse than that for to-be-
remembered information as providing support for the concept of repression. 
However, Roediger and Crowder (1972) showed that in the particular experi-
mental paradigm used by Weiner and Reed (1969), a simpler explanation in terms 
of differential rehearsal of the two types of material probably applied. Still, some 
researchers over the years have argued that the concept of retrieval inhibition (akin 
to the idea of repression) is needed to explain directed forgetting results (e.g., 
Bjork, 1989; Geiselman, Bjork, & Fishman, 1983). As with the other literatures 
mentioned above, the situation is not simple. Other researchers using rather similar 
paradigms have shown that instructions not to think of something can have the 
paradoxical effect of making retrieval of the information more likely, as when 
people are told to perform a task but to avoid thinking of white bears; suppression 
of the unwanted thought is quite difficult, and usually the concept becomes more 
(not less) accessible (e.g., Wegner, 1989; Wegner, Schneider, Carter, & White, 
1987). Further research is needed to unravel these paradoxical effects of 
instructions to forget (see Golding & MacLeod, 1998).

The four literatures mentioned here (repeated testing effects, effectiveness of 
retrieval cues, the implicit memory literature, and results from directed forgetting 
paradigms) all deserve closer scrutiny in the recovered memory debate.

Research Strategies in Human Memory

Part of the debate around recovered memories has to do with research 
strategies. Even a casual reader cannot help but be struck that the bulk of the 
evidence cited by Alpert et al. (1998c) on one side of the issue is from case studies 
and from retrospective reports, whereas the bulk of the evidence cited by Ornstein 
et al. (1998b) is based on laboratory experiments or naturalistic studies that are still 
reasonably well-controlled. The limitations of case studies and retrospective 
reports are well-known; indeed, Alpert et al. (1998c) provided an excellent 
summary of the limitations of this sort of evidence and also noted repeatedly the 
preliminary nature of much of the evidence they cited. We appreciate their 
expressions of concern about the nature of the evidence, but the overall conclusion 
we reach is that these methodological limitations render most of their conclusions 
remarkably tentative and speculative, at best. Put succinctly, there is no strong 
evidence that memories of traumatic events can be repressed or dissociated and 
then later recovered intact.

On the other hand, there is a large body of experimental evidence cited by 
Ornstein et al. (1998a) on the suggestibility of memory. A substantial body of 
experimental evidence (buttressed by a long collection of anecdotal cases) shows 
that events that never happened can be vividly remembered, or that events can be 
remembered in ways quite different from the way they occurred (see Roediger, 
1996, and Schacter, 1995, for reviews of the literature). Several studies provide 
evidence for the plausible assumption that distortions occur more easily as the 
retention interval for the events in question increases (e.g., Belli, Windschitl, 
McCarthy, & Winfrey, 1992; Wright, 1993). Not surprisingly, memories for events 
are more accurate immediately after the event has occurred and decline with time 
(Ebbinghaus, 1885/1964). Furthermore, the more distant the memory, the more 
easily its reconstruction can be distorted by suggestibility and misinformation. 
Therefore, when the issue is a childhood event that may not have been encoded
well in the first place, recovery of the events after 20 to 40 years may be safely assumed to have the potential for great error. However, Alpert et al. (1998a) generally rejected this line of reasoning proffered by Ornstein et al. (1998a) and maintained that traumatic memories are different from memories for ordinary events as studied in laboratory settings (see especially p. 1062, although in other places they seem to back off the strong version of this claim).

We note some ironies here. The solid, experimental research is considered nearly irrelevant to the debate, whereas the methodologically questionable case histories and retrospective reports are considered telling evidence (albeit with possible flaws). However, when it serves them, Alpert et al. (1998c) felt free to draw on laboratory evidence. On p. 971 they cited a laboratory study as relevant (Guenther & Frey, 1990), a study in which subjects merely heard a story about sexual abuse read to them. In other cases, when it served their purpose, they drew on lab research on state-dependent retrieval (but see Brenneis, 1997, who criticized their selective use of this evidence) and even research with animals. Some studies of memory distortion have used mildly traumatic events (e.g., getting lost in a mall, plunging one's hand through a window), yet Alpert et al. (1998c) considered this more solid research to be irrelevant. In addition, by favorably citing implicit memory work, the authors endorsed other research whose external validity would be no greater than for other laboratory work (and perhaps less so—the criterial tasks often involve completing fragmented forms of pictures or words).

What qualifies as a case study in psychology has changed over the years, at least in most research areas. Today case studies in, say, cognitive neuropsychology employ compelling designs of single cases, analyzed with repeated measures of the feature of interest and with many controls (both within-subject and between-subject) to rule out possible artifacts (see Caramazza, Berndt, & Basili, 1983, for an example and Caramazza, 1986, for a discussion of this approach). Although considerable debate exists within this field about conclusions that can be securely drawn from case studies of single individuals, these studies are far superior to what passes as a case study in the Alpert et al. (1998c) report. Clinical case studies of memory and trauma seem to have changed little since the late 1890s when Freud and Janet provided their first cases, which are still debated fiercely but are cited in the Alpert et al. report. The case studies cited in support of dissociated and recovered memories have been subjected to withering analysis by Pope and Hudson (1995) and Brenneis (1996), among others. Indeed, Brenneis (1997) noted that two of the “case studies” cited as best support for dissociated and recovered memory were published only as journalistic accounts!

There is considerable debate among memory researchers and experimental psychologists about when experiments need to be designed to maximize external validity (e.g., Banaji & Crowder, 1989; Mook, 1983; Neisser, 1976). There are no easy answers and certainly Alpert et al. (1998c) are correct that not enough cogent research exists on memory for traumatic events. However, as they also noted, because of their nature, traumatic events are very difficult to study experimentally, at least in human populations. Given this valid consideration, we believe there is every reason to be cautious about claims based on case study techniques and retrospective reports. That these represent the primary ways traumas have been studied does not make the techniques any more likely to be valid, especially when
the case studies are often so casual and the collected evidence is so conflicting. After all, as often noted, many people report that they remember traumas only too well and are unable to forget them. Certainly, if we are to give credence to case studies, this more typical reaction to trauma should not be discounted.

**Conclusion**

We have made several interrelated points. First, the evidence that repeated early childhood traumas are repressed or dissociated from other experiences is, at best, highly questionable. Believing such a theory of dissociation from repeated trauma also is inconsistent with long-established findings from the study of memory with numerous experiments showing that repeated events are generally remembered better than single events. In addition, evidence about alleged dissociation of childhood traumas seems different in important ways from evidence of PTSD victims. Nonetheless, the analogy to PTSD to support the case of dissociation of memories of abuse is the centerpiece of the argument.

Second, even assuming memories of trauma were somehow poorly encoded in kinesthetic body memories and dissociated from one's personal narrative of the self, this would seem all the more reason to question the recovery of those memories. Memories poorly encoded cannot be recovered in a more accurate narrative form 20–30 years later. No matter how great the power of retrieval cues, such cues cannot arouse memories that were not encoded well in the first place.

Third, we noted that several bodies of literature from the psychology of memory seem relevant but have not yet been introduced into the debate on recovered memories in an important way. These include phenomena of repeated testing (reminiscence and hypermnesia), the power of retrieval cues in reviving seemingly forgotten memories, the study of priming on implicit memory tests, and the study of intentional or directed forgetting. To this list we might add the numerous experimental attempts to validate the concept of repression, dating from the 1930s (e.g., Rosenzweig & Mason, 1934).

Fourth, we also discussed the thorny issue of the disparate research methods used by proponents and opponents of the concept of dissociated–repressed–recovered memories, although no easy solution to these issues is in sight. However, we agree with Alpert et al. (1998b) that research informed by both clinical and experimental bodies of evidence is likely to pay dividends in the end. We would urge that case study techniques be more rigorous, like those in cognitive neuropsychology and other domains, if genuine progress is to be made.

The APA report on recovered memories has thus far attracted little attention and comment, probably because the report has been hard to obtain, which has been corrected by the publication of the full report in this theme issue. However, Brenneis (1997), a clinical psychologist, examined the report in a *Psychoanalytic Psychology* article, and argued:

The conclusions drawn by the clinicians may be even weaker than assessed by the researchers; conversely, the critique offered by the clinicians of the researchers’ evidence is more impassioned than apt. . . . The inescapable conclusion reached by this reader is that the report and rebuttals by the clinicians reflect an uncritical reading of the clinical literature and an insufficient knowledge of the experimental literature. (p. 533)
In particular, Brenneis noted that clinicians usually underestimate their own considerable influence in affecting events during therapy.

We end by expressing dismay at one theme pervading parts of the Alpert et al. (1998a, 1998b, 1998c) statements, to wit, that this controversy is somehow not a real crisis in the clinical psychology community, but instead is somehow been trumped up by the False Memory Syndrome Foundation. They stated simply that there is no "epidemic" of these recovered memory cases; that many recovered memories are true; that there is no such thing as recovered memory therapy; that no one has ever shown that various therapeutic practices are harmful; and so on (see Alpert et al., 1998b, p. 1020). It is this denial of the obvious that worries us most. Since the APA report was issued, there has been a continuing series of court cases that have generated huge negative publicity for psychology, psychiatry, and related fields in mental health. Even a casual look at the number and enormity of these cases should arouse doubt about the clinicians' complacent attitude. After all, the APA Working Group was convened in response to the large number of these cases and the debate that they aroused. The prevalence of such cases and the tepid response to them by clinical psychologists and organizations such as the American Psychological Association and the American Psychiatric Association hold peril for the profession at large. Clinical psychology as a profession has a tremendous amount to offer on many fronts. To see the gains and successes of the whole profession tarnished by defense of therapeutic practices and techniques that may do more harm than good undermines the whole profession of psychology in general and of clinical psychology in particular.

References


Received February 19, 1997
Accepted October 23, 1998