APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **17**: 251–279 (2003)
Published online in Wiley InterScience 26 February 2003
(www.interscience.wiley.com) DOI: 10.1002/acp.866

Eliciting and Comparing False and Recovered Memories: An Experimental Approach

STEVEN M. SMITH*, DAVID H. GLEAVES, BENTON H. PIERCE, TARA L. WILLIAMS, TODD R. GILLILAND, and DAVID R. GERKENS

Texas A&M University, USA

SUMMARY

We describe an experimental paradigm designed to elicit both recovered and false memories in the laboratory. All participants saw, on a video screen, three critical categorized lists of words mixed in randomly with eighteen filler categorized lists. Those in the blocking condition then had several paper-and-pencil tasks that involved only the 18 non-critical filler lists. An uncued recall test was then given, asking participants to recall all the lists they originally saw on the video screen. Finally, there was a cued recall test that provided category cues for the three critical lists. Substantial memory blocking of critical lists on the uncued test and recovery on the cued recall test was observed in all three experiments. In Experiments 2 and 3, many false and recovered memories were elicited on the cued recall test by including cues for the three critical (forgotten) lists, plus cues for three lists that had never been presented. False memories were distinguishable from truly recovered memories in cued recall by 'know' versus 'remember' judgements, and by confidence ratings; accurately recovered memories were associated with higher confidence. False and recovered memories could not be discriminated based on recall latency. The results repeatedly show powerful effects of memory blocking and recovery. We also show that recovered and false memories can be elicited within a single experimental procedure, and there may be unique characteristics of each. Although we urge caution in generalizing to false and recovered memories of trauma, we suggest that variations of our comparative memory paradigm may be useful for learning about such phenomena.

Scientific debates should have two sides. Most of the laboratory evidence generated by the 'false memory debate', however, has focused on only one side of the debate, examining questions concerning false recall and recognition (e.g. Brewer and Treyens, 1981; Dodson and Johnson, 1993; Johnson and Raye, 1981; Loftus, 1993; Lindsay and Johnson, 1989; Lindsay and Read, 1994; McDermott, 1997; McDermott and Roediger, 1998; Payne *et al.*, 1996; Pezdek, 1994; Read, 1996; Roediger and Goff, 1998; Roediger and McDermott, 1995; Smith *et al.*, 2000; Zaragoza and Koshmider, 1989). Relatively little experimental evidence has examined the other side of this debate, namely, the nature of blocked and recovered memories. The present study is in part an attempt to observe, under controlled laboratory conditions, powerful memory blocking and recovery effects. Such evidence should inform the debate as to whether a memory discovered after a period of apparent

*Correspondence to: Steven M. Smith, Department of Psychology, Texas A&M University, College Station, TX 77843-4235, USA. E-mail: sms@psyc.tamu.edu

amnesia is a false memory, or one that has been blocked from consciousness and later recovered.

In *DSM-IV* (APA, 1994), dissociative amnesia is defined as 'a reversible memory impairment in which memories of personal experience cannot be retrieved in verbal form, (p. 478).' The events that cannot be recalled are 'usually of a traumatic or stressful nature' (p. 478) and the inability to remember is 'too extensive to be explained by normal forgetfulness' (p. 478). Gleaves (1996), Loewenstein (1991), and van der Hart and Nijenhuis (1995) have written more extensive reviews on this subject. Despite some claims that dissociative amnesia is a recently recognized (or invented) phenomenon (e.g. Ofshe and Singer, 1994), it has been recognized by clinicians since the beginning of the nineteenth century (Nemiah, 1979). Dissociative (or psychogenic or hysterical) amnesia was studied and described extensively by Pierre Janet in the 1880s as well as by Freud in some of his early writings. There are also numerous descriptions of dissociative amnesia in the early and recent literature on combat and war trauma (e.g. Bremner *et al.*, 1993; Grinker and Spiegel, 1945; Kolb, 1988; Sargant and Slater, 1941), civilian violence (e.g. Kasniak *et al.*, 1988). Modia (1994) also described total amnesia for childhood in a survivor of the Holocaust. In many of these reports, the authors also described how memory for the traumatic experiences of war could be retrieved through therapy, hypnosis or even narcosynthesis. These authors did sometimes caution that what was retrieved was often a mixture of accurate memory and fantasy (e.g. Kolb, 1988; Sargant and Slater, 1941). Such cautionary notes highlight the need for laboratory research to address these issues.

Reluctance to study recovered memories in the laboratory no doubt arises, at least in part, from the assumption that such memories are initially repressed because of the powerful negative affect associated with the traumatic experiences that are blocked from consciousness. Ethical considerations alone prevent experimenters from using such affectively negative material. Powerful negative affect is assumed to be the cause of repression, the symptoms of which include (but are not limited to) persistent amnesia for an event, followed ultimately by the successful recovery of a memory of the event. The present study challenges the notion that only strong negative affect can cause amnesia, and subsequent recovery. In three experiments, we demonstrate powerful forgetting effects caused by ordinary cognitive mechanisms. These forgetting effects are temporary; the experimentally blocked memories are later recovered when participants are given cues.

Beyond questions of the reality of recovered and false memories are questions about how the two phenomena can be differentiated. To study questions about differences between false and recovered memories, one must use methods that evoke and examine both phenomena. We offer the 'comparative memory' paradigm for the purpose of evoking both false and recovered memories on the same test, allowing direct comparisons of the two phenomena.

In the present study we use a straightforward theoretical framework that can accommodate memory blocking and recovery effects using the simple mechanisms of interference and/or inhibition and retrieval cuing. This approach assumes that a memory can be blocked by the establishment and strengthening of competing memories. Increasing the accessibility of the competing memories should strengthen such a memory blocking effect. Cues associated with the original memory in question, but that do not lead to recall of competing memories, should be capable of reviving the blocked memory.

Several investigators have described how retrieval activities can retroactively affect a memory trace. Retrieving an event can improve subsequent access to the retrieved memory, but it can also decrease the probability of retrieving competing memories. Retrieving some events from a memory set (i.e. a set of items to be remembered) can block access to other events in the set, a phenomenon called 'output interference' (e.g. Roediger, 1974; Rundus, 1973). Output interference has been cited as the cause of part-list and part-set cuing inhibition effects (e.g. Nickerson, 1984). Raaijmakers and Shiffrin (1981) refer to output interference as a biased retrieval set, that is, a tendency to retrieve some items from the target episode at the expense of other event memories.

A particularly powerful effect of output interference was reported by Roediger (1978), who had participants study and recall a set of categorized lists of words. Relative to a free-recall condition, participants who first recalled words from a subset of the studied categories (i.e. the non-critical lists) in cued recall tests were subsequently far less likely to recall words from critical lists. Thus, Roediger showed how the process of recall, itself, was self-limiting; cued recall of some of the target lists restricted subsequent free recall of other (critical) studied lists. Output interference, as described by Rundus (1973), Roediger (1974, 1978), and others, is a phenomenon that occurs as a function of testing, itself. The effects of output interference have been shown to be transitory, dissipating minutes after a recall test has ended (e.g. Smith and Vela. 1991). Nonetheless, output interference effects, such as those of Roediger (1978), may provide a key component of longer-lasting amnesia.

The second component, namely, a shift in output dominance of the responses in a memory set, combined with a powerful output interference effect, could theoretically cause a persistent memory block. Output dominance is a characteristic of category members that describes the 'coming-to-mindedness' or accessibility of members. Operationally defined, when participants are given a category name as a cue, the category members given most frequently are labeled highest in output dominance (e.g. Barsalou, 1985; Ward, 1994). When recalling members of a category, retrieving the more dominant category members, which come to mind first, may suppress recall of less dominant members. Thus, recall of critical items can be decreased if competing items in the memory set are made more dominant (Figure 1). This shift in output dominance should have a memory blocking effect when the entire memory set is accessed because recall of the dominant items in the set should cause output interference, suppressing retrieval of critical items.

Two mechanisms that theoretically could cause a shift in the output dominance of category members are priming effects and retrieval inhibition. Priming effects in category generation tasks show that if certain category members are encountered in an incidental task, those words are more likely to be given as responses in a subsequent task in which participants are asked to list a few members of the category (e.g. Marsh and Landau, 1995; Ward *et al.*, 2002). Thus, priming effects can be described as shifting the output dominance of primed items, making them more dominant. If primed competitors become more dominant, then recall of unprimed critical members of the memory set would be suppressed by output interference during free recall of the entire set.

A different mechanism that could cause such a shift is retrieval inhibition of responses in a memory set (e.g. Anderson, 2000; Anderson and Spellman, 1995; Anderson *et al.*, 1994). Anderson and colleagues have shown that retrieval practice of a subset of items in a memory set inhibits subsequent retrieval of items that are not given extra retrieval practice, a phenomenon referred to as 'retrieval-induced forgetting'. Thus, extra retrieval practice of the non-critical items in a memory set not only primes those items, thereby increasing their output dominance, but it could also inhibit non-practised critical items, decreasing their dominance. Such retrieval inhibition, according to Anderson and Spellman (1995),

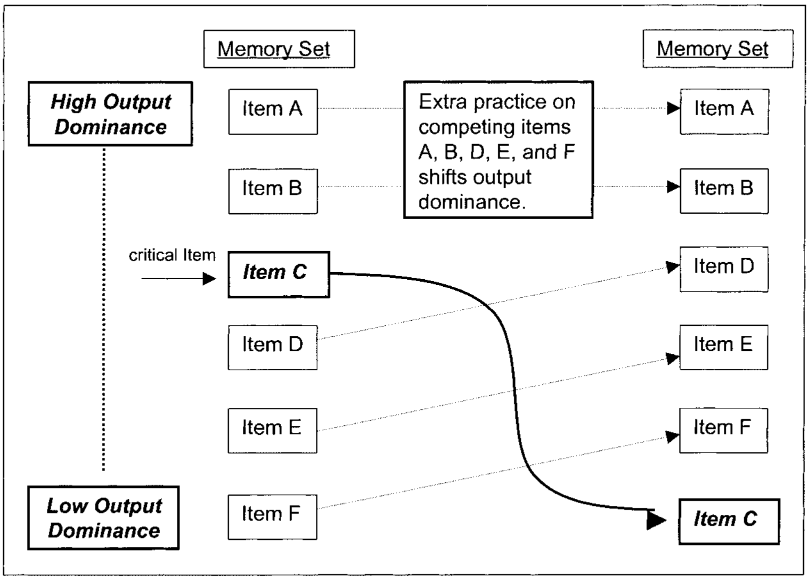

Figure 1. The memory set consists of studied items A through F, shown arranged from highest output dominance (first items retrieved) to lowest output dominance. Extra practice with non-critical items A, B, D, E, and F shift output dominances such that critical Item C becomes lowest in output dominance. During recall of the memory set, the highest dominance items are recalled first, causing output interference that blocks or inhibits retrieval of critical Item C

reduces accessibility of critical non-practised items even if recall is prompted by cues that are independent of cues associated with the original memory set.

Whether extra practice of a subset of items is seen as making those items more dominant via priming, or making non-practised critical items less dominant via inhibition, the result is the same; that is, critical items become low in output dominance (Figure 1). During free recall of the entire set of items, those critical items made systematically low in output dominance become unlikely to be recalled because output interference from retrieved high dominance items will render critical items inaccessible.

According to this approach, persistent forgetting for a critical event can occur if one practices retrieving competing memories within a set of events. Such practice should make the non-critical memories more dominant, causing them to be recalled earlier during attempts to recall the entire set of events, and thereby increasing the likelihood that non-practised critical events will be suppressed. Although we make no definitive statement that such a mechanism is responsible for dissociative amnesia, learning to retrieve memories that compete with memory of a traumatic event (i.e. learning to avoid thinking about the event) could be a coping mechanism used to protect oneself from recalling and re-experiencing painful emotions. This hypothesis is an extension of the theory that emotional trauma can be forgotten because of avoidance of remembering traumatic experiences (Bowers and Farvholden, 1996; Cloitre, 1992). If avoidance is practised and negatively reinforced through elimination or avoidance of pain, then memory of critical events could become habitually blocked (e.g. Bower, 1990; Dollard and Miller, 1950). A similar theoretical account was described by Anderson and Green (2001), identifying retrieval inhibition as a mechanistic substrate for repression, a mechanism that does not rely upon trauma as a necessary factor.

Another well-documented area of memory research has shown that available memories (i.e. events stored in long-term memory) can be inaccessible (Tulving and Pearlstone, 1966), and that appropriate cues can elicit otherwise inaccessible memories (e.g. Smith, 1979; Tulving and Thomson, 1973). This difference between memories that are unavailable versus those that are available but inaccessible roughly parallels Freud's 'unconscious–preconscious' distinction (Erdelyi and Goldberg, 1979). In Tulving and Pearlstone's study, in which participants studied categorized lists of words, more was recalled when category cues were given at the test, in comparison to a free-recall test. Importantly, the difference in recall between the cued and uncued conditions in Tulving and Pearlstone's study was due to the fact that the free-recall group recalled fewer categories than did the cued-recall group. Participants in the free-recall condition failed to access entire categories of words. Giving category cues at recall ostensibly provided access to some otherwise inaccessible categorized lists. Extrapolating from this result, we predicted in the present experiments that categorized lists that are not remembered on a free-recall test can be recalled later if the category names are provided, a form of memory recovery. The categorized lists in question should remain inaccessible until appropriate cues are provided.

Given this theoretical framework, memory blocks for entire categorized lists of words could theoretically be brought about by extra exposure to competing lists, and recovery of the blocked memories should occur if category list names are provided as cues. Figure 1 illustrates this situation, with names of categorized lists as items. If a set of categorized lists is studied, and extra practice occurs for a subset of the originally studied lists, then access to the practised (non-critical) categorized list names should improve (i.e. they take on higher output dominance), and access to an unpractised categorized list (Critical Item C) should become less probable. This situation represents a retrieval block for Categorized List C; the dominance of competing items lead to output interference, preventing recall of the critical item, List C. If participants are given the category name of List C as a memory retrieval cue, however, that cue should provide access to the otherwise blocked memory of Critical List C and the corresponding members of the list.

Although the use of categories of episodes in the present experiments and in the present theoretical discussion is intended to resemble important aspects of prior research paradigms (e.g. Tulving and Pearlstone, 1966), it also resembles certain aspects of clinical interviews with patients. In clinical interviews in which patients are encouraged to recover memories a patient might be asked to recall categories of events, such as cases of physical, verbal, or sexual abuse, or experiences involving particular persons. Although the categories used in the present experiments do not involve clinically relevant materials, they nonetheless provide the same cognitive structures as categories of events that are more clinically realistic.

COMPARING FALSE AND RECOVERED MEMORIES

Despite the limitations of the clinical and experimental data, we suggest that at least under certain circumstances, both false memories and genuine recovered memories may exist. Attention should now be turned to determining the conditions under which each are likely or unlikely to occur, the relative frequency of each, and the degree to which one can discriminate between the two types of phenomena.

With the above empirical and theoretical considerations in mind, we designed the *comparative memory paradigm* to elicit blocked and recovered memories along with false memories within a single experimental treatment. The reason for having both effects within one experimental paradigm is to be able to directly compare the two phenomena in a controlled test. This approach is similar to a signal-detection analysis of recognition memory in the sense that false positives (false memories) and hits (accurate memories) are both produced on the same test. In the signal-detection approach, false and accurate memories are compared to separate the effects of guessing from effects of memory discrimination ability (e.g. Miller and Wolford, 1999). Similarly, the value of the comparative memory paradigm may be that by producing both false and accurately recovered memories, the method might reveal differences between the two phenomena. Another approach to the question of the validity of recovered memories has been taken by Porter *et al.* (1999), who found that memory distortion was likeliest when interviewers were extraverts and participants were introverts, and distortion was greater for participants who scored high on a dissociative scale.

The rationale behind the comparative memory paradigm is this. In a naturally occurring case, a false memory cannot be distinguished from a recovered one, given our current state of knowledge of these phenomena, unless the details of the original episode are clearly and objectively known. We hypothesize, however, that there may be features of the retrieval process that could be used to distinguish recovered memories from created ones. That is, there may be one behavioural 'signature' that is characteristic of recovered memories, and a different signature for created memories. Such hypothetical signatures could be of great value to therapists, for whom it could be very important to distinguish recovered memories from false ones.

For example, metacognitive reports, such as confidence judgements or feelings of knowing, might be useful in distinguishing false memories from recovered ones. That is, people might be more confident in their accurate memories than they are in their false memories. Another type of measure that might distinguish these phenomena could be the time course of retrieval of material relevant to an episode, which might be different for recovered memories than for created memories. Information might be produced rapidly, say, for recovered memories, and slowly for the false memories. Besides possible differences in behavioural signatures, recovered and created memories might also be differentially affected by factors such as suggestion, emotionality, or bizarreness. In short, a paradigm that compares the two phenomena could provide an experimental means to explore these and other questions about recovered and false memories.

Some support for the hypothesis that there are ways of discriminating false and genuine recovered memories can be found in research on reality monitoring (e.g. Johnson and Raye, 1981; Johnson *et al.*, 1988), the ability to discriminate memories of real versus imagined events. Based on their own research and their review of the literature in this area, Johnson and Raye reached the overall conclusion that '[m]emories from external and internal courses [do] appear to differ in class-characteristic ways' (p. 82) and that data support the existence of 'a memory system that preserves information remarkably well; the decision criteria through which this information is filtered, however, allow for some error in attributing memories to sources' (p. 82).

Our initial goals for our research with the comparative memory paradigm were to induce strong memory blocks, to stimulate recovery of the original memories, and to induce false memories, all within a single experimental treatment. In brief, memory blocks were induced using an extension of Roediger's (1978) methods for amplifying output interference effects, recovery was stimulated with category cues, and false memories were produced with a forced-recall procedure in which subjects tried to recall items from categories that were plausible memory targets, but that had not been studied.

In the present experiments we investigated effects of extra retrieval practice of part of a memory set (i.e. filler items) to see if it would block or inhibit retrieval of unpractised (critical) items. Thus, the first empirical question addressed was whether practising competitors can give rise to substantial levels of forgetting. The second question we addressed concerns recovery of memories. In the present experiments we used category cues to evoke memories that had been blocked on an earlier test. Our third intention was to stimulate false memories with the same experimental manipulations used to stimulate recovered memories. This goal was accomplished on a cued-recall test by giving participants category cues that were plausible, but that had not been studied, mixed in with category cues that had been studied and forgotten.

EXPERIMENT 1

The purpose of Experiment 1 was to create and test a laboratory analogue of recovered memories, based on the methods of extra practice of non-critical items and cue-dependent memory. This laboratory analogue needed to meet two major criteria: (1) it must cause a substantial amount of forgetting or blocking of critical memories, and (2) it must be possible to resuscitate the blocked memories so that the recovered material can be remembered nearly as well as non-blocked memories.

Experiment 1 used a method involving extra exposure to a subset of categorized lists to induce a substantial amount of forgetting of critical categorized lists in a free-recall paradigm. After seeing 21 categorized lists of words, participants in the memory blocking condition were given six more tasks, each of which re-exposed participants to 18 of the original lists (filler lists), but not to the three critical lists. Those in the control condition completed mazes and maths problems for an equivalent amount of time. On an uncued recall test (i.e. free recall of all the originally viewed lists), it was expected that participants in the blocking condition would be far less likely than those in the control condition to recall items from the three critical lists.

To revive forgotten list material, we provided category cues on a test subsequent to the initial uncued recall test. It was expected that prompting participants with the critical category cues would elicit memory recovery: i.e. recall of lists that were not recalled on the initial uncued recall test. The strength of the recovered memories was assessed by comparing cued recall levels for critical lists that *were* initially recalled (termed ‘continuous memories’) versus cued recall for critical lists that were not initially accessed (‘recovered memories’).

Method

Participants

The participants were 72 undergraduate students in introductory psychology classes at Texas A&M University who volunteered for partial fulfilment of the course grade. Participants were tested in groups of approximately 10 per session, and each session lasted approximately 2 hours.

Materials
Twenty-one categorized lists (e.g. birds, furniture, fruit) were constructed for the to-be-learned materials. The lists were drawn from norms published by Rosch (1975), and each list contained 15 of the most typical category members.

Design and procedure
Participants were randomly assigned to either the control or blocking conditions, a between-subjects variable. The two groups received the same treatment except that only the blocking group experienced the retrieval practice (blocking) procedure. Three categorized lists were designated as the critical lists, and the remaining 18 lists were designated filler lists.

Presentation of critical lists
All participants were first presented with the 21 categorized lists in an incidental learning procedure. The lists were presented, one at a time, via a television monitor, showing black letters on a white background. Each list of 15 words, accompanied by a category name, remained on the screen for 2 min. During that time participants were to write down the category name and the 15 items, rank-ordering the category members in terms of their judged category typicality from most to least typical. The purpose of this task was to ensure that participants would attend to every category name and to all the list members. The category-ranking task was the only time during the experiment that participants saw the three critical lists, and it was the only task in which stimuli were seen on a television screen.

Intervening tasks
In the control condition participants received two non-verbal tasks that occupied them for the 45 min retention interval. The first task was a set of paper-and-pencil mazes, and the second required participants to compute answers to multiplication problems. These tasks were intended to occupy participants' attention for the same time as the corresponding treatment in the blocking condition.

In the blocking condition, participants were given six tasks, each of which re-exposed participants to the items on the 18 filler categorized lists. These tasks took 45 min to complete. The tasks included a list-recall task, a category name-recall task, a typicality rating task, a recall rating task, a pleasantness rating task, and a size-ranking task. The filler categories, as well as the filler category members, were presented in different orders on each of these tasks. The intent of these tasks was to cause forgetting of the three unpractised critical lists. The three critical lists were never presented on any of the six tasks.

In the list-recall task, participants wrote down all the category members in a list, prompted by the category labels from the 18 filler lists. In the category name-recall task, the 15 members of each list were presented, and participants wrote the appropriate category labels for each list. In the typicality rating task, participants indicated, on a 1–5 rating scale, how typical each member of the list seemed for its category. For the recall rating task participants indicated, on a 1–5 rating scale, how likely it seemed that they would be able to recall the list members. On the pleasantness rating task participants rated how pleasant each item made them feel. Finally, on the size ranking task, participants ranked the items from each category in terms of size.

Initial uncued recall test

The uncued recall test, of which participants had not been forewarned, was the test in which retrieval blocking effects were examined. For the test, participants were asked to recall all the items originally listed in all the categories that were shown on the videotaped list of materials. They were asked to include both category names and all list members, but they were also admonished to write down only those categories and list members that had been shown on the videotaped lists. Participants were given 10 minutes for this test.

Category cued recall test

Following the uncued recall test was the category cued recall test. For this test, the three critical category names were presented on a television screen, one name at a time, and participants were asked to recall all category members that had been presented on that list. Participants were given 2 minutes to recall the items for each of the three critical lists.

Results

Initial uncued recall

A one-way analysis of variance (ANOVA) was computed to assess the effects of blocking (blocking versus control), a between-subjects factor, on the uncued recall test, using the proportion of recall of the three critical lists as the dependent measure. A list of words was counted as recalled if one or more items from the list were given as responses on the uncued recall test. Thus, participants were counted as having recalled items from either 0, 1, 2, or all 3 of the critical lists. As predicted, there was a substantial and significant effect of blocking ($F(1, 70) = 27.40$, $p < 0.0001$, $MSE = 0.073$); fewer critical categorized lists were recalled in the blocking condition than in the control condition (Table 1).

Another one-way ANOVA examined the effect of blocking on the proportion of words recalled from the three critical lists. Again, there was a significant blocking effect ($F(1, 70) = 18.80$, $p < 0.0001$, $MSE = 0.027$); a smaller proportion of the words from critical categorized lists were recalled in the blocking condition than in the control condition (Table 1).

Some participants failed to recall even a single word from any of the three critical categorized lists on the initial uncued recall test. Although only 4% of the control group participants (one of 25 participants) recalled none of the critical lists, 17% of the blocking group (eight of 47 participants) recalled none of the critical items on the uncued recall test.

Table 1. Mean uncued and cued recall as a function of blocking in Experiment 1

	Blocking	Control
Uncued recall		
Proportion of critical lists recalled	0.40	0.75
Proportion of critical words recalled	0.22	0.40
Proportion of recalling no critical lists	0.17	0.04
Cued recall: proportion of words per category		
Continuous	0.72	0.68
Blocked	0.70	0.65

Category cued recall

A categorized list was counted as recovered if it was not recalled on the initial uncued recall test (i.e. if no words from that categorized list were recalled), but it was recalled (i.e. at least one item from the list was recalled) on the subsequent category cued recall test. A continuous memory, in contrast, was a case in which a categorized list had items that were recalled on both the initial uncued recall test and on the subsequent cued recall test. In Experiment 1, all participants recalled at least one item from every critical category on the cued recall tests.

A 2×2 (blocking $\times$ type of memory) ANOVA was computed, using the proportion of words recalled for each list as the dependent measure. Type of memory was recovered versus continuous. The maximum possible for each list was 15 words. There was no effect of blocking ($F(1, 51) = 1.30$, $p = 0.26$, $MSE = 6.02$), nor of type of memory ($F(1, 70) = 1.12$, $p = 0.30$, $MSE = 2.73$). That is, an equivalent number of words were recalled on the cued recall test whether or not a list had been accessed on the initial uncued recall test (Table 1).

Discussion

Substantial and robust memory blocking effects were observed in Experiment 1. Six trials of extra repetitions of non-critical categories resulted in a 35% drop in recall of critical categorized lists on an initial uncued recall test. The magnitude of this memory blocking effect is considerably greater than the forgetting effects reported in many previous laboratory studies. For example, a typical effect of directed forgetting is a 17% drop in recall (e.g. Geiselman *et al.*, 1983), a typical part-list cuing inhibition effect is a 12% decrement in recall (e.g. Roediger and Schmidt, 1980), and a typical effect of retrieval induced forgetting is 10% (e.g. Anderson *et al.*, 1994). The forgetting effect observed in Experiment 1 is equivalent to the output interference effect observed by Roediger (1978) when participants were first required to recall several non-critical lists of words prior to free recall of the critical lists in the memory set. The results of Experiment 1 indicate that procedures that occur before the memory test can produce forgetting effects as great as those found by Roediger (1978). Extrapolating beyond the present results, it does not seem unreasonable that even more profound memory blocking effects could be produced if our experimental treatment were extended beyond the relatively brief manipulations in the present study.

The forgotten categorized lists were recoverable; on the delayed test of cued recall, items could be recalled from all initially non-recalled lists. The number of items recalled per list in cued recall was the same whether or not any items from the list had been recalled earlier on the initial uncued recall test (i.e. whether memories were recovered or continuous). This result is predictable from similar findings by Tulving and Pearlstone (1966), although in that study type of test (free versus cued recall) was manipulated between-subjects. Thus, the present results extend those of Tulving and Pearlstone by showing not only that category cues increased accessibility, but further, that initially inaccessible lists were later recoverable. This result also indicates that the memory blocking and/or inhibition effect occurred for the names of the categorized lists, and not for the individual items in the lists. In this case a list name can be said to represent the 'control element' for each categorized list (e.g. Estes, 1972), defined as a higher-order representation in memory that is associated with each of the items in a list. At an even higher level, the experimental session, itself, serves as a control element for the memory set of categorized lists, as depicted in Figure 1.

EXPERIMENT 2

Experiment 1 documented a procedure for inducing and observing recovered memories in a controlled laboratory setting. The primary goal of Experiment 2 was to observe both recovered and false memories within the same laboratory test. We refer to the methodology used in Experiments 2 and 3 as the 'comparative memory' paradigm because it evokes both recovered and false memories within a single procedure, and allows the two phenomena to be compared at test. The steps involved in this procedure are: (a) incidental learning of 21 categorized lists, (b) interference practice, or practice with non-critical lists, (c) uncued recall test (to assess the retrieval blocking effect), and (d) cued recall test (to evoke and observe both recovered and false memories, which were elicited by including three category cues for lists that subjects did not see previously in the experiment).

Student volunteers participated in a 1-hour study session and a 1-hour test session. During the study session participants were given an incidental learning task in which they saw 21 animal categories (e.g. 'Carnivores'), and ranked the seven items in each category (e.g. badger, cheetah, lynx, raccoon, shark, weasel, jaguar) as to their typicality for their assigned category. This task also required participants to write down every category name and each of the seven category members to ensure that the materials were clearly attended.

Of the original 21 categories, 18 were filler categories and three were critical categories. The three critical categories were not distinguished for the participants in any way from the fillers, and were not seen again by any participants until the final test session. After the incidental learning task subjects in the blocking group were given four extra trials of incidental learning tasks with only the 18 filler categories. The incidental tasks involved rating (as opposed to the prior ranking) the typicality of the members of the 18 filler categories, recalling the 18 category names using the category members as cues for each, judging how well they thought they would be able to recall filler category members, and recalling filler category members from their category names.

There were two memory tests given at the test session, an uncued recall test, and a cued recall test. On the uncued recall test participants wrote down all of the names of animal categories they had seen in Session 1.[1] They were told to write down 24 animal category names on the form, with the instruction that they should recall as many as possible from the original list, and when they could recall no more, they should guess or make up more animal category names until they had 24 on their list. This forced recall procedure, which was similar to a procedure used by Roediger and Payne (1985), was used to minimize the possibility that participants would think of correct category names, but would not write them down because of uncertainty. After they had written 24 names participants were asked to judge their confidence in each response.

On the cued recall test participants were shown six animal category names and asked to recall or guess at the seven animals that had been presented with each category name. A forced recall procedure was again used, in this case requiring seven animal names for each cue. The three critical animal category names and three new category names served as cues. Participants had seen the three critical categories once only, whether they had been in the blocking or control conditions. The three new category names had not been seen in the experiment at all by any participants. Therefore, responses to these three cues were always confabulated. Confidence judgements were obtained for each response after all responses

[1]Although participants recalled all of the categorized list words on the uncued recall test in Experiment 1, they were asked only for the category names in Experiments 2 and 3.

were recorded. If people can reliably distinguish between memory and imagination, then they should be more confident in recovered memories than in confabulated responses.

A secondary goal of Experiment 2 was to examine the forgetting effect after a longer retention interval than the one used in Experiment 1. If the forgetting effect observed in Experiment 1 was merely a function of output interference caused by recent use of non-critical lists, rather than a more persistent shift in output dominances of memory set items, then the effect would not be expected to occur after a long retention interval. In Experiment 2 an immediate test condition used a retention interval similar to that used in Experiment 1, and a delayed test condition used a two-day retention interval prior to the test session.

Method

Participants

In Experiment 2 the participants were 178 students from introductory psychology classes who volunteered for partial fulfilment of the course grade. Approximately 10 participants were tested per session.

Design and procedure

Participants were randomly assigned to one of eight conditions (four blocking groups and four control groups). In the blocking conditions, participants received extra tasks intended to induce retrieval blocking, whereas in the control conditions, the blocking tasks were omitted. Two counterbalancings of the categorized lists (counterbalancings A and B) were used such that three critical categories served as blocked categories and three other critical categories were used as false memory cues in counterbalancing A, whereas the categories were assigned to the opposite conditions in counterbalancing B (see Appendix A). In addition, participants were assigned either to an immediate test condition in which they were tested after a short delay (25 minutes), or to a delayed test condition in which they were tested after a 48-hour delay.

Participants first viewed 21 lists of animal categories with seven members in each list as part of an incidental learning task. The 21 categories and their members were shown on a television screen, one category at a time. Each of the 21 categories was shown for 2 minutes with the category label at the top of the screen and the seven members listed below. Participants were asked to determine how typical each category member was on a scale of 1 to 10, with 10 representing the most typical member of a category and 1 the least typical. Eighteen of the categories were filler categories and three were critical ones. In counterbalancing condition A, *rodents*, *circus animals*, and *dangerous animals* served as the critical categories, whereas in condition B, the critical categories were *reptiles*, *flying animals*, and *carnivores*.

Once participants had rated all 21 categories in the incidental learning task, those in the blocking conditions received extra tasks intended to block recall of the three critical categories. These blocking tasks, which were presented on paper forms rather than on the television screen, consisted of four extra incidental learning tasks with the 18 filler categories. A category ranking task was given in which participants ranked the seven members of each filler category in order of which items best fit the category. Next was a category name recall task in which the seven members of each filler category were shown, and the participant recalled and wrote each category name. Next was a judgement of learning task in which all filler category names and members were listed, and the participant rated each category member as to how likely it was that they would be able to remember the item on a recall test. Finally, participants were given a category cued recall task in which filler category names were listed, and participants recalled and wrote the members of each category. Instead of doing the blocking tasks, participants in the control group did multiplication problems and maze tasks for 25 minutes. This interval was approximately equal to the time taken by the blocking tasks in the blocking condition.

Uncued recall of category names

After the delay subjects were given an uncued recall test involving the 21 categories that had originally been shown on the television screen. Participants were asked to recall as many of the originally shown animal categories as they could remember, and the recall instructions specified that the categories were to be recalled from the material they had viewed on the television screen (none of the extra blocking tasks used television presentations). The test required participants to list 24 animal category names. Although only 21 categories had been originally presented, participants were asked to guess to fill all 24 spaces on their response forms. The purpose of this forced recall procedure was to ensure that participants would not withhold any implicitly retrieved material, fearing that it might not be accurate.

Once they had written down 24 category names, participants were asked go back over the items and give remember/know/guess judgements and confidence ratings for each recalled category name. Participants were given the following instructions:

> Now that you have written down 24 animal category names, you should go back and look at each one you have written, and make two judgments about each response. First, judge whether you remember seeing that category name on the original list, whether you just know that it was just on the list, or whether the name is just a guess. If you specifically remember seeing or having thought about one of the category names you have written, circle 'R' on the line next to that name. If you know it was on the list, but you cannot explicitly remember having seen or thought about it, circle 'K' on the line next to that name. If the name is just a guess, circle 'G'.
>
> Second, judge how confident you are that each name was actually on the original list of animal categories. Make your confidence judgment by writing a number from 1 to 10 on the line next to each category name. Write a '10' if you are completely confident that the name was on the original list, and a smaller number if you are less confident. Write '1' for an animal category if you have no confidence in the response at all; that is, if it is a complete guess.

Cued recall test

After the initial recall test participants received a cued recall test on six categories, including the three critical categories and three animal categories that had not been presented on the original category list. The category names were presented on a television screen with an instruction to the participants to write down only items from that category and to avoid writing down members from other categories. They were told, however, that they were free to guess at category members. Participants were given 2 minutes to recall as many of the seven members of each category as possible. After participants had seen all six categories and recalled as many members as they could for each category, they were then asked to go over their responses and give remember/know/guess judgements and confidence ratings. The instructions for these judgements and ratings were the same as those given for the uncued category name recall task.

Results

At the time of the initial incidental learning task, participants saw only three of the six critical categories. At cued recall they were presented with all six critical category names and asked to recall members of the categories. False memories were operationally defined as responses (after eliminating complete guesses) on the cued recall test to the category names that the participants had not seen in the incidental learning task. The possible range for this variable was 0–21 because there were three false categories with seven possible responses in each category. Recovered memories were operationally defined as list members remembered at cued recall for categories that had not been initially remembered on the uncued recall test. The number of recovered memories also ranged from 0–21 but the actual upper limit was dependent on the number of critical categories *not* remembered at initial recall. Continuous memories were operationally defined as accurately reported (at cued recall) members of each category that the participant did recall on the initial uncued recall test.

Blocking effects

The memory blocking effect was analysed with a $2 \times 2 \times 2$ (blocking $\times$ counter-balancing $\times$ delay) ANOVA, using the proportion of the three studied critical category names that were recalled on the uncued recall test as the dependent measure. Blocking was blocked versus control, a between-subjects factor, and delay was either immediate or delayed, and was manipulated between-subjects. There were no main effects or interactions for counterbalancing, so this term was eliminated from the analyses. There was a blocking $\times$ delay interaction ($F(1, 174) = 6.02$, $p = 0.017$), indicating that the effect of the blocking manipulation on uncued recall of the critical categories was greater in the immediate test condition. As shown in Table 2, participants in the blocking condition only remembered (at immediate recall) an average of 43% of the critical categories, relative to 81% of the same categorized lists in the control condition. For participants in the delayed testing condition those receiving the blocking manipulation recalled 48% of the critical categories, as compared to 63% for those in the control condition.

We also examined the percentages of participants who remembered either none or all of the critical categories. Nearly 20% of the participants in the blocking condition remembered none of the critical categories. In contrast, only 6% of the participants in the control condition recalled none of the critical categories. Conversely, only 7% of participants in the blocking condition remembered all three categories, whereas 44% of those in the control condition remembered all three.

Table 2. Mean uncued recall of category names as a function of blocking and delay in Experiment 2

	Blocking	Control
Immediate	0.43	0.81
Delay	0.47	0.63

To further examine blocking effects, two 2×2 (blocking $\times$ delay) ANOVAs were computed, one using average output position as the dependent measure, and the other using confidence ratings for the critical categories that were successfully recalled on the uncued recall test. For the ANOVA with average output position as the dependent measure, there was no blocking $\times$ delay interaction ($F(1, 152) = 0.67$, $p > 0.05$), but there were main effects for blocking ($F(1, 152) = 34.11$, $p < 0.001$) and delay ($F(1, 152) = 12.67$, $p < 0.001$). Average output position of recalled critical categories was higher for the blocking group (i.e. they remembered the critical categories later in the list; 13.43 versus 9.45), and was higher for the immediate rather than delayed test condition (12.39 versus 9.94).

For the ANOVA that used confidence ratings as the dependent variable there was no blocking $\times$ delay interaction ($F(1, 152) = 0.67, p > 0.05$), but there was a significant main effect for blocking ($F(1, 152) = 23.86$, $p < 0.001$), and for delay ($F(1, 152) = 12.67$, $p < 0.001$). Confidence ratings for recalled critical categories were lower for the blocking group (8.12 versus 9.36 for the controls) and for the delayed recall group (8.33 versus 9.38 for the immediate recall group).

Eliciting false and recovered memories

For both false and recovered memories, we compared the number elicited as a function of blocking and delay. For false memories, we also analysed type of intrusion (a within-subjects factor), comparing the number of filler intrusions (an animal seen on another list) versus new intrusions (an animal not on any of the lists seen by the participant).

Across all conditions, the numbers of false memories (combining filler and new intrusions) ranged from zero (40.4% of sample) to 21 (3.4% of sample), with an average of 6.16 intrusions per participant. There were main effects for blocking ($F(1, 174) = 22.14$, $p < 0.001$)), delay ($F(1, 174) = 8.79$, $p < 0.01$)), and there was a blocking $\times$ delay interaction ($F(1, 174) = 8.50$, $p < 0.01$)). The interaction (Table 3) indicates that the effect of blocking on false memories was greater for delayed ($t(88) = 5.91$, $p < 0.001$) than immediate testing, where the effect was non-significant ($t(86) = 1.17$, $p = 0.25$). The effect of delay only occurred for the control group.

The main effect of type of intrusion, the type of intrusion $\times$ blocking interaction, the type of intrusion $\times$ delay interaction, and the blocking $\times$ delay $\times$ type of intrusion interaction were all non-significant (all Fs < 1.0). The mean numbers of each type of intrusion was approximately 3.1.

For recovered memories, frequencies ranged from zero (38.8% of the sample) to 16, with an overall mean of 2.96. There were main effects for blocking ($F(1, 174) = 17.29$, $p < 0.001$), delay ($F(1, 174) = 6.52$, $p < 0.01$), and the blocking $\times$ delay interaction ($F(1, 174) = 13.21, p < 0.001$). This interaction is shown in Table 4, and indicates that these results mirrored those from the blocking effects. That is, differences between blocking and control groups were greater for the immediate cued recall test ($t(86) = -4{,}91$, $p < 0.001$) than for the delayed recall test where the effect was not significant ($t(88) = -0.44, p = 0.56$).

Table 3. Mean proportions of false memories as a function of blocking and delay in Experiment 2

	Blocking	Control
Immediate	3.7	5.5
Delay	3.0	10.2

Table 4. Mean proportions of recovered memories as a function of blocking and delay in Experiment 2

	Blocking	Control
Immediate	0.06	0.02
Delay	0.02	0.02

We also examined frequency of continuous memories. These ranged from zero (3.4% of sample) to 20. There was a main effect of blocking ($F(1, 174) = 34.08$, $p < 0.001$) and delay ($F(1, 174) = 70.30$, $p < 0.001$), but the blocking $\times$ delay interaction was not significant ($F(1, 174) = 2.94$, $p = 0.09$). The number of continuous memories was greater for the immediate recall condition ($m = 12.31$) relative to the delayed test ($m = 6.88$), and was greater for the control group (10.63) relative to the blocking group ($m = 8.56$).

Comparing false and recovered memories

To compare false, recovered, and continuous memories, we examined confidence ratings for items that were recalled for the critical categories at cued recall (complete guesses, for which confidence was zero, were omitted from this analysis). Across all subjects the mean confidence ratings for recovered, false and continuous memories was 8.74, 6.47, and 9.03 respectively. Using the subset of participants who reported all three types of memories, we computed a $2 \times 2 \times 2$ (blocking $\times$ delay $\times$ type of memory) ANOVA with confidence rating as the dependent measure. Main effects for blocking and delay were not significant ($F(1, 51) = 0.24$, $p = 0.629$, and $F(1, 51) = 0.0$, $p = 0.94$, respectively), nor were any interactions significant. There was a main effect for type of memory ($F(2, 50) = 33.26$, $p < 0.001$). Confidence ratings for false and recovered memories differed significantly ($t(70) = 6.58$, $p < 0.001$), as did ratings for false and continuous memories ($t(89) = 9.69$, $p < 0.001$). Ratings for recovered and continuous memories did not differ significantly ($t(85) = 0.87$, $p = > 0.05$).

Discussion

In Experiment 2, we successfully elicited both false and recovered memories within a single experimental paradigm. We evoked recovered memories by first creating a robust memory blocking effect, and then stimulating blocked memories by presenting appropriate retrieval cues on the cued recall test. This replicated the results of Experiment 1. False memories were also successfully elicited on the cued recall test when non-presented category cues were given.

Numerous indicators demonstrated a powerful memory blocking effect. Participants in the blocking condition remembered 38% fewer critical categories than did participants in the control group, a very sizeable memory blocking effect. Many of the participants in the blocking condition failed to recall any of the critical categories. Relative to the control group, participants in the blocking condition also reported lower confidence when they did recall the critical category names, and they recalled the critical items at later output positions. The finding that the average output position of critical items was later in the blocking condition is consistent with the notion that extra practice with non-critical categories lowers the output dominance of critical categories. This shift in output dominance of critical items was still present, but less pronounced, in the two-day delayed test condition.

False memories were also elicited on the cued recall test. It should be noted that the confabulated responses observed in Experiment 2 are false memories in that they represent events that did not occur. That is, the new category names had never been presented, so remembering that a word had been given in such a category is a false memory. It also should be noted that such false memories are different from those that are elicited with other paradigms, such as the use of associatively structured word lists (e.g. Read, 1996; Roediger and McDermott, 1995), misinformation effects (e.g. Loftus and Hoffman, 1989), or reality monitoring failures (e.g. Johnson and Raye, 1982).

The mere suggestion by the experimenter that non-presented cues had actually been on the original study list appears to have been a strong enough influence to evoke false memories. Three categories of animals on the cued recall test had not been presented during the study session, yet each participant wrote down an average of six responses for the non-presented categories, responses labeled by participants as non-guesses. Smith *et al.* (2001) referred to false memories that result from misattributed events as 'episodic confusion errors', or errors caused by remembering events in the wrong episodic context. About half of the falsely recalled responses were episodic confusion errors, items incorrectly recalled from other lists that had been presented. Half of the falsely recalled items were animals that had not been presented during the experimental sessions.

Furthermore, we found that correctly recovered and false memories differed in terms of confidence ratings. After eliminating total guesses, false memories were associated with lower confidence ratings than were correctly recovered memories. This finding suggests that subjective confidence might provide a means of eventually differentiating correctly recovered memories from certain types of false memories. This laboratory finding parallels others' work examining differences between false and recovered memories of childhood sexual abuse (e.g. Williams *et al.*, 2000).

EXPERIMENT 3

In Experiment 3 we sought to replicate the powerful memory blocking and recovery effects found in Experiments 1 and 2, and we wanted to increase the magnitude of the false memory effect. Demand characteristics from other participants in the group sessions of Experiment 2 could have dampened the false memory effects that were observed. That is, when participants were asked to recall list items for a category they had not seen in the experiment, their scepticism about whether the category had actually been presented could have been bolstered by observing similar subtle responses in others. Therefore, in Experiment 3 participants were given the experimental procedure individually, rather than in groups. A secondary benefit of using individual rather than group administration is that the former more closely matches the therapeutic situation.

As in Experiment 2, we predicted that there would be a memory blocking effect that would be evidenced by the reduced number, later output positions, and lowered confidence of the critical categories remembered on the uncued recall test. We also predicted that category members would be recalled from the blocked critical categories (i.e. critical categories not recalled on the previous uncued recall test). As in Experiment 2, we also predicted that false recall would occur when non-presented category cues were given on the cued recall test.

To compare false and recovered memories evoked on the cued recall test in Experiment 3 we examined confidence ratings, as in Experiment 2, as well as recall latencies. We predicted that confidence levels of false and accurately recovered memories would differ, as they did in Experiment 2. We also hypothesized that recall latency might provide part of a behavioural signature that might be used to distinguish false memories from false and recovered ones, specifically that recovered memories would be recalled faster than false memories.

One further change in the procedure was the addition of a second cued recall test that was given to participants immediately after the first one. The purpose of the second cued recall test was to examine hypermnesia (i.e. a net increase in recall without further study of the target material) that could result from repeated testing, and to observe the persistence or even increase in false memories from the first cued recall test (as found by Hyman *et al.*, 1995 in their study of false childhood memories).

Method

Participants

The 33 participants in Experiment 3 were volunteers from introductory psychology classes. None of these participants were also participants in Experiments 1 or 2.

Materials

The same materials that had been used in Experiment 2 were also used in Experiment 3.

Design and procedure

Although the experimental design and procedure in Experiment 3 were the same as in Experiment 2, for the most part, there were some important changes in Experiment 3. Participants were given the experimental tasks individually, rather than using group sessions, as had been done in Experiment 2. In Experiment 3 all participants were tested after a one-day retention interval. After the recall tests participants gave confidence ratings (1–10) of their responses, as well as guess/know/remember responses. Rather than a single cued recall test, as in Experiment 2, participants completed two identical cued recall tests, the second immediately following the first.

Results

Blocking effects

A one-way ANOVA comparing mean levels of recall of critical categories in the control (70%) versus blocking (40%) conditions showed a significant effect of blocking ($F(1, 31) = 9.18$, $p = 0.005$). Furthermore, whereas all (100%) of the control condition participants remembered at least one of the critical categories, 23% of those in the blocking condition failed to recall even one of the critical categories. Of the participants in the control group, 30% remembered all three critical categories, as compared with only 8% of participants in the blocking condition.

As in Experiment 2, we also examined average output position and confidence ratings for the critical categories recalled on the free recall test. The average output position was somewhat higher (later in recall) for the blocking group ($m = 13.9$) than in the control group ($m = 11.2$), although the difference was not significant ($t(28) = -1.19$, $p = 0.24$). Average confidence ratings for critical categories recalled in the blocking condition ($m = 5.05$) were significantly lower than for the control group ($m = 9.20$), $t(28) = 3.6$, $p = 0.001$.

Eliciting false and recovered memories

Three separate ANOVAs tested the effects of blocking on false, recovered, and continuous memories, using test order (first or second cued recall test) and recognition judgement ('remember' versus 'know') as within-subjects factors. In the analysis of false memories type of intrusion (filler versus new intrusion) was also included as a within-subjects factor, as in Experiment 2.

False memory frequencies ranged from zero (13.6% of sample) to 21, with a mean of 6.92. The main effect of blocking was not significant ($F < 1.0$), nor was there a significant main effect for test order ($F(1, 32) = 2.61$, $p = 0.12$). There was a main effect of type of false memory ($F(1, 32) = 24.47$, $p < 0.001$); more intrusions were fillers rather than new intrusions. There was also a main effect of judgement ($F(1, 32) = 23.13$, $p < 0.001$); more false memories were judged 'know' rather than 'remember'. No interactions involving test order were significant (all Fs < 1.0). All other interactions were significant, including the blocking $\times$ judgement interaction ($F(1, 32) = 17.47$, $p < 0.001$)), the blocking by type of false memory interaction ($F(1, 32) = 5.57$, $p < 0.05$), the type by judgement interaction ($F(1, 32) = 8.96$, $p < 0.01$), and the blocking by type by judgement interaction ($F(1, 32) = 16.76$, $p < 0.001$).

Examination of the three-way interaction (by plotting two-way interactions at different levels of the third variable) indicated that there was a blocking $\times$ type of false memory interaction for memories rated as 'know' ($F(1, 31) = 14.65$, $p < 0.001$), but not for those rated as 'remember', ($F(1, 31) = 0.82$, $p = 0.37$). The two-way interaction (for those with 'know' ratings) suggested that the blocking group had more filler false memories than new false memories whereas the numbers of filler and new were approximately equal for the controls. Thus, the blocking group, which had extra practice on the other lists, remembered more of the filler items, but misremembered the list from which they came. However, this was only for responses subjects rated as 'know'. When participants noted that they truly remembered a response, the blocking group had much fewer filler or new false memories; in fact they reported no new false memories. Also, for false memories rated as 'know', the blocking group reported more than the controls ($F(1, 31) = 4.19$, $p < 0.05$). However, for responses rated as 'remember' the pattern was reverse; controls had more than the blocking group ($F(1, 31) = 14.65$, $p = 0.001$).

With recovered memories, frequencies at either recall test ranged from zero to five with a mean of 1.88 for each participant. There was a main effect for blocking ($F(1, 32) = 4.18$, $p < 0.05$); participants in the blocking condition had more recovered memories than did those in the control condition ($m = 2.55$ versus 1.45). There was also a main effect of test order ($F(1, 31) = 8.23$, $p < 0.01$), with more recovered memories on the second cued recall test. There was also a test order $\times$ judgement interaction ($F(1, 32) = 6.35$, $p < 0.05$). This interaction (Table 5) indicates that recovered memories increased from the first to second cued recall test, but this increase was only for memories rated as 'remembered'. Those rated as 'know' decreased slightly.

Table 5. Mean proportions of recovered memories in Experiment 3 as a function of cued recall test order and Remember/Know judgements

	Know	Remember
Cued Recall Test 1	0.9	0.8
Cued Recall Test 2	0.4	1.4

Table 6. Mean confidence ratings for false, recovered, and continuous memories in Experiments 2 and 3

	Experiment 2	Experiment 3
Recovered	8.7	7.7
False	6.5	6.2

For continuous memories, frequencies at either recall test ranged from zero to 15 with a mean of 2.05. The only significant effect was the main effect for judgement. Across all other conditions, more continuous memories were rated as ‘remember’ ($m = 3.15$) than were rated ‘know’ ($m = 0.95$).

Comparing false and recovered memories

Confidence ratings

Mean confidence ratings for recovered, false, and continuous memories were 7.8, 6.3, and 8.2 respectively. We compared these ratings on a subset of participants who reported all three types ($n = 19$), using a 2×3 (blocking $\times$ memory type; continuous, false, or recovered) ANOVA. There was a significant main effect for blocking ($F(1, 17) = 27.61$, $p < 0.001$); the control group, as compared to the blocking group, gave higher confidence ratings across all three types of memories. There was also a main effect for type of memory ($F(2, 34) = 11.37, p < 0.001$). Confidence ratings were significantly higher for the accurately recovered memories than for false memories ($t(22) = 4.59, p < 0.001$). Ratings for false and continuous memories were also significantly different ($t(25) = 4.00$, $p = 0.001$). Ratings for continuous and recovered memories did not differ significantly ($t(19) = 1.21$, $p = 0.24$). Confidence ratings for false and recovered memories for both Experiments 2 and 3 is shown in Table 6.

Know versus remember judgements

We also compared false and recovered memories in terms of ‘know’ versus ‘remember’ judgements. We initially included all observations to compare percentage of false, recovered, and continuous memories rated as ‘remember’ versus ‘know’. Recovered memories were approximately twice as likely as false memories to be rated as ‘remembered’ (59.7% versus 30.9%). Recovered memories were more similar to continuous memories, 76.8% of which were rated as ‘remembered’.

We calculated for each participant the percentage of false, continuous, and recovered memories that were rated as ‘remember’ versus ‘know’ and compared these figures with pairwise dependent *t*-tests. Only participants that reported all three types of memories ($n = 19$) were included in the pairwise analyses. The average percentage of recovered memories rated as remembered (54%) was significantly higher than the average for false memories (23%) ($t(18) = 5.16$, $p < 0.001$). The average percentage of false memories reported as remembered also differed significantly from the average for continuous memories (59%) ($t(18) = 3.52$, $p = 0.002$). However, the average percentage of recovered memories rated as remembered was not significantly different than for continuous memories ($t(18) = -0.59$, $p = 0.56$).

Recall latency

We compared false and recovered memories in terms of initial response latencies and used two approaches to examine this variable. The first was to compare the amount of time taken to generate an initial response. We also compared false, recovered, and continuous

Table 7. Mean percentages of memories reported across three time intervals in Experiment 3

	Memory type		
	Continuous	False	Recovered
Less than 30 seconds	68	76	65
31 to 60 seconds	20	20	23
Greater than 60 seconds	11	4	12

memories in terms of the distribution of responses across three time intervals: less than 30 seconds, 31 seconds to 1 minute, and greater than 1 minute.

For the first response latency, means for the false and recovered memories were both approximately 9.0 seconds and thus did not differ significantly. Likewise, when comparing average percentages of memory types across the three time intervals, there were no differences (based on pairwise dependent *t*-tests) noted. The majority of responses (whether false, recovered, or continuous memories) occurred during the first 30 seconds of the recall test, and the fewest occurred after 60 seconds (Table 7).

Discussion

The results of Experiment 3 replicated and extended the major findings of the first two experiments. First, although retention was tested after a full day, there was a powerful memory blocking effect caused by the extra practice given to filler categories in the blocking condition. These participants recalled 30% fewer critical category names than did control group participants, roughly similar to the 38% blocking effect found in Experiment 2 and the 35% effect in Experiment 1. As in Experiment 2, both false and recovered memories were elicited on the cued recall tests, demonstrating that both phenomena can be evoked and studied within a single experimental paradigm. In Experiment 3 we also found that false and recovered memories differed in terms of confidence ratings, again replicating the results of Experiment 2. Similarly, a second metacognitive report, Remember/Know judgements, also distinguished false from recovered memories in Experiment 3; that is, a higher percentage of recovered memories than false memories were described as 'remembered'.

Another prediction consistent with the results of Experiment 3 was that repeating the cued recall test resulted in hyperamnesia for correctly recovered items. More items were correctly recalled on the second cued recall test than had been recovered on the first. False memories, however, were not significantly increased as a function of retesting, suggesting that false and recovered memories might be differentially affected by repeating memory tests. In this second experiment but not the first, we found differences in terms of the number of filler versus new intrusions. False memories were most often episodic confusion errors; that is, animals from lists that participants had seen. One prediction was not confirmed, however; recall latency did not reliably distinguish false memories from recovered ones, as had been hypothesized.

GENERAL DISCUSSION

When memories appear to be recovered, do such experiences reflect truly recovered memories that are essentially accurate, or are such events likely to be false memories of events that never happened? Although no immediate resolution of this important question is at hand, in the present paper we acknowledge and experimentally demonstrate the reality of both recovered and false memories. Perhaps more importantly, we have demonstrated that both false and recovered memories can be elicited within a single experimental procedure, allowing for the possibility of directly comparing the two phenomena. The ultimate aim of this comparison is to learn how to discriminate the two in terms of empirical measures, and to discover how the two are influenced by other variables, such as emotionality, retention intervals, and personality factors.

The memory blocking effects observed in the three present experiments were very powerful ones. Although participants in the blocking conditions spent less than an hour giving extra attention to the non-critical filler categories, these brief experiences resulted in a great deal of forgetting of critical categories. Poorer recall of critical items was accompanied by a shift in the average output positions of critical items (when they were recalled), consistent with the notion that extra practice with non-critical filler items produced a shift in the output dominances of items in the memory set (Figure 1). The forgetting effect was persistent; after a one-day retention interval there was still a 30% difference between blocking and control conditions (Experiment 3), although a two-day interval saw a decrease in this difference to 15% (Experiment 2). The persistence of the effect indicates that the shift in output dominances was not merely an effect of output interference from recent practice with filler categories.

The observed forgetting was not likely to have been caused by a reporting bias, because the forced recall procedure in Experiments 2 and 3 encouraged participants to guess when they could recall no additional category names. The degree of forgetting that would occur for critical categories if extra filler practice were substantially increased beyond what has been done in the present experiments is not known, but it is conceivable that even more profound forgetting might result.

Although the reported experiments do not critically evaluate the efficacy of various cognitive mechanisms for the observed forgetting effects, more detailed discussions of such mechanisms have been published (e.g. Anderson and Bjork, 1994; Anderson and Spellman, 1995; Anderson and Neely, 1996; Roediger and Neely, 1982, Nickerson, 1984). An explanation discussed earlier uses the notion of output dominance shifts combined with output interference effects to explain the large forgetting effects we observed. That is, the forgetting manipulation may have decreased the output dominance of critical to-be-remembered items by giving extra exposures to non-critical items, increasing the likelihood that non-critical items would be recalled earlier in the output sequence, consistent with the results. Such a shift would lead to increased output interference for critical items, decreasing recall of those items, as was observed.

A different explanation for the observed forgetting effects is that retrieval inhibition accrued for critical items due to retrieval induced forgetting (e.g. Anderson and Bjork, 1994), a phenomenon based on a procedure quite similar to that used in the present experiments. Such inhibition effects presumably would be observed independently of output dominance shifts. In fact, some studies indicate that shifts in dominance, and consequent output interference effects, cause forgetting only when the lower dominant items are interfering competitors (see Anderson and Neely, 1996). Whether the same independent effects would be observed with the forgetting procedures used in the present experiments is not known.

The direct relevance of this potent forgetting effect to naturally occurring cases of blocked memories, such as the amnesia associated with post-traumatic stress disorder or dissociative disorders, cannot be established from these three experiments. Importantly, the extreme affect associated with emotional trauma may play a unique and important role in causing dissociation (e.g. Metcalfe and Jacobs, 1998). Nonetheless, the use of practised interference to obstruct target memories is consistent with several theoretical accounts of amnesia (e.g. Anderson and Green, 2001; Bower, 1990; Bowers and Farvholden, 1996; Cloitre, 1992; Dollard and Miller, 1950), and deserves greater experimental scrutiny.

Recovery of blocked memories was also observed in all three experiments. We had predicted that using critical category names as cues would elicit correct recall of items even if the categories themselves had not been recalled on the earlier uncued recall test. Consistent with this prediction, substantial numbers of items were correctly recalled for critical category cues that had not been recalled on the previous uncued recall test. Furthermore, recovered memories occurred at about the same rate in cued recall as did continuous memories. This is experimental evidence of recovered memories, that is, memories that were blocked initially, and were recovered when appropriate cues were used. The recovery of material from lists that had been blocked from memory demonstrates the distinction between availability and accessibility (e.g. Tulving and Pearlstone, 1966); forgotten categories of items were inaccessible (as measured by the initial uncued recall test), but they remained available in memory, and could be evoked by appropriate retrieval cues. In the present experiments, the cued recovered memories were as complete as continuous memories, a prediction consistent with Tulving and Pearlstone (1966). The notion that recovered memories were initially inhibited but continuous ones were not is consistent with the observed recovery effects if it is assumed that inhibition occurred at the level of the category label, and that list members were not affected by retrieval inhibition.

The equivalence of continuous and recovered memories in these experiments also indicates that the blocking/inhibition effects occurred at the level of the control element for each list (Estes, 1972), represented by the categorized list name (e.g. ‘vehicles’, ‘house pets’), rather than at the level of the members of each list. Of theoretical significance, this result supports the conclusion that the list members are hierarchically structured, and that retrieval inhibition or interference at the level of the control element does not affect retrieval of items at hierarchically lower levels (e.g. Roediger, 1978). Of practical significance is that it may be possible that entire episodes in one’s life may never be recalled because of interference or retrieval inhibition of control elements of the episode, yet, if the episode were appropriately cued, the details of the episode may be as accessible as details of episodes one has been continuously able to recall.

Although the results support the relative accuracy of recovered memories in the false memory debate, there is also support for the claim that what may seem to be accurately recovered memories may, in fact, be false memories. When given category cues that had never been presented in the experiment, participants often gave responses that were not identified as guesses, essentially claiming to remember details of episodes that did not occur. In Experiment 2, false memories were as likely to be filler intrusions (i.e. items inappropriately recalled from other experimentally presented categories) as new intrusions. In Experiment 3, the majority were filler intrusions, what Smith *et al.* (2001) refer to as ‘episodic confusion errors’. Thus, even though participants falsely remembered to which category animals belonged, they frequently remembered animal names that they had in fact seen as part of the experiment. One could question whether these should be termed false memories, because some of the incorrectly recalled items were actually presented during the experimental procedure. We suggest, however, that these events may reflect a type of false memory that could have some degree of clinical relevance; patients might remember experiences that are based upon accurate memories, but with important details incorrectly inserted, such as when, or where some event occurred, or who was involved. Such inaccurate memories could be due to source monitoring failures (e.g. Johnson *et al.*, 1993). It is also important to note that the cued recall tests in the present experiments elicited a substantial number of intrusions that were animals participants had not seen as part of the experiment. These new intrusions quite clearly fit the definition of 'events that did not happen' (e.g. Roediger and McDermott, 1995). Future research with the comparative memory paradigm might also be adapted to examine recovered and false memories using other similar false memory paradigms, such as the method that uses lists of associates (e.g. Read, 1996; Roediger and McDermott, 1995) or categorized lists (Smith *et al.*, 2000).

There was considerable variation among participants in terms of the frequencies of false and recovered memories. The modal response for both types of memories was zero. For some participants, however, we elicited as many as 21 out of a possible 21 false memories. Future research should investigate which individual differences predict the elicitation of blocked and recovered versus false memories. Hyman and Billings (1998) found scores on the Creative Imagination Scale to predict creation of false memories. Varieties of suggestibility (Brown, 1995) might also predict false memory formation (but see Leavitt, 1997). Other variables such as dissociative experiences might be found to predict both false and recovered memories. Hyman and Billings (1998) found scores on the Dissociative Experiences Scale to predict false memory formation, but they did not study memory blocking and recovery.

We found differences between false and recovered memories on the basis of metacognitive reports, including confidence ratings (Experiments 2 and 3) and Remember/Know judgements (Experiment 3). Two variables measuring response latencies did not prove to be useful in discriminating the two types of memories. One of the major goals of future research on this topic should be experimentally testing other variables that might discriminate false memories from accurately recovered ones.

Caution is advisable in interpreting the observations we have labelled 'false memories' because the criterion we used to operationally define false memories was any response rated above a zero in confidence, with zero indicating a complete guess. Our participants might have used too lenient a criterion for assessing which responses were actually remembered as having occurred as part of the list on the cued recall test. It is also possible that experiment demand characteristics may have been at least partially responsible for some of the observed false memories. It is not clear to what degree other laboratory versions of false memory phenomena are affected by such demand characteristics (although misinformation effects generally involve implications that the misinformation is accurate). Some have expressed concern, however, that false memories may be elicited during psychotherapy because some therapists place heavy social demands on patients to remember certain types of episodes, such as abusive or traumatic events (e.g. Lindsay and Read, 1994). Future research needs to address this gap between laboratory false memories and those that occur in naturalistic settings.

Another important goal of future research in this area should be to determine what factors differentially affect elicitation of recovered and false memories. For example, how are the two phenomena affected by cues, retention intervals, emotionality of events, affective states, demand characteristics, or memory-enhancement techniques? By exploring these naturalistic phenomena in controlled experimental settings we can learn more about the mechanisms that underlie them. The same signatures that occur in experimental paradigms could be investigated in naturally occurring cases.

ACKNOWLEDGEMENTS

The authors are indebted to Michael Anderson, Jonathan Schooler, and Don Read for their useful comments on an earlier version of this manuscript.

REFERENCES

American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders* (4th edn). APA: Washington DC.

Anderson MC, Bjork RA. 1994. Mechanisms of inhibition in long-term memory: a new taxonomy. In *Inhibition in Attention, Memory and Language*, Dagenbach D, Carr T (eds). Academic Press: New York.

Anderson MC, Bjork RA, Bjork EL. 1994. Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition* **20**: 1063–1087.

Anderson MC, Green C. 2001. Suppressing unwanted memories by executive control. *Nature* **410**: 366–369.

Anderson MC, Neely JH. 1996. Interference and inhibition in memory retrieval. In *Memory. Handbook of perception and cognition* (2nd edn), Bjork EL, Bjork RA (eds). Academic Press: San Diego, CA; 237–313.

Anderson MC, Spellman BA. 1995. On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological Review* **102**: 68–100.

Barsalou LW. 1985. Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition* **11**: 629–654.

Bower GH. 1990. Awareness, the unconscious, and repression: an experimental psychologist's perspective. In *Repression and Dissociation: Implications for Personality Theory, Psychopathology, and Health*, Singer JL (ed.). University of Chicago Press: Chicago, IL.

Bowers KS, Farvolden P. 1996. Revisiting a century-old Freudian slip—from suggestion disavowed to the truth repressed. *Psychological Bulletin* **119**: 355–380.

Bremner JD, Steinberg M, Southwick SM, Johnson DR, Charney DS. 1993. Use of the Structured Clinical Interview for DSM-IV-Dissociative Disorders for systematic assessment of dissociative symptoms in posttraumatic stress disorder. *American Journal of Psychiatry* **150**: 1011–1014.

Brewer WF, Treyens JC. 1981. Role of schemata in memory for places. *Cognitive Psychology* **13**(2): 207–230.

Brown D. 1995. Pseudomemories: The Standard of science and the standard of care in trauma treatment. *American Journal of Clinical Hypnosis* **3**: 1–24.

Cloitre M. 1992. Avoidance of emotional processing: a cognitive science perspective. *Cognitive Science and Clinical Disorders*: 19–41.

Dalenberg CJ. 1996. Accuracy, timing and circumstances of disclosure in therapy of recovered and continuous memories of abuse. *Journal of Psychiatry and Law* **24**: 229–276.

Dodson CS, Johnson MK. 1993. Rate of false source attributions depends on how questions are asked. *American Journal of Psychology* **106**: 541–557.

Dollard J, Miller NE. 1950. *Personality and Psychotherapy.* McGraw-Hill: New York.

Erdelyi MH, Goldberg B. 1978. Let's not sweep repression under the rug: towards a cognitive psychology of repression. In *Functional Disorders of Memory*, Kihlstrom JF, Evans FJ (eds). Erlbaum: Hillsdale, NJ; 355–402.

Estes WK. 1972. An associative basis for coding and organization in memory. In *Coding Processes in Human Memory*, Melton AW, Martin E (eds). Winston: Washington, DC; 161–190.

Geiselman RE, Bjork RA, Fishman DL. 1983. Disrupted retrieval in directed forgetting: a link with posthypnotic amnesia. *Journal of Experimental Psychology: General* **112**: 58–72.
Gleaves DH. 1996. The evidence for 'repression': An examination of Holmes (1990) and the implications for the recovered memory controversy. *Journal of Child Sexual Abuse* **5**: 1–19.
Goff LM, Roediger HL, III. 1998. Imagination inflation for action events: repeated imaginings lead to illusory recollections. *Memory & Cognition* **26**: 20–33.
Grinker RR, Spiegel JP. 1945. *War Neuroses*. Blakiston: Philadelphia.
Hyman IE, Billings FJ. 1998. Individual differences and the creation of false childhood memories. *Memory* **6**: 1–20.
Hyman IE, Husband TH, Billings JF. 1995. False memories of childhood experiences. *Applied Cognitive Psychology* **9**: 181–197.
Johnson MK, Foley MA, Suengas AG, Raye CL. 1988. Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General* **117**: 371–376.
Johnson MK, Hashtroudi S, Lindsay DS. 1993. Source monitoring. *Psychological Bulletin* **114**: 3–28.
Johnson MK, Raye CL. 1981. Reality monitoring. *Psychological Review* **88**: 67–85.
Kazniak AW, Nussbaum PD, Berren MR, Santiago J. 1988. Amnesia as a consequence of male rape: a case report. *Journal of Abnormal Psychology* **97**: 100–104.
Kolb LC. 1988. Recovery of memory and repressed fantasy in combat-induced post-traumatic stress disorder of Vietnam veterans. In *Hypnosis and Memory*, Pettinati HM (ed.). Guilford: New York.
Leavitt F. 1997. False attribution of suggestibility to explain recovered memory of childhood sexual abuse following extended amnesia. *Child Abuse & Neglect* **21**: 265–272.
Lindsay DS, Johnson MK. 1989. The eyewitness suggestibility effect and memory for source. *Memory & Cognition* **17**: 349–358.
Lindsay DS, Read JD. 1994. Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology* **8**: 281–338.
Loewenstein RJ. 1991. Psychogenic amnesia and fugue: a comprehensive review. In *Dissociative Disorders: A Clinical Review*, Spiegel D (ed.). Sidran Press: Lutherville, MD; 45–78.
Loftus EF. 1993. The reality of repressed memories. *American Psychologist* **48**: 518–537.
Loftus EF, Hoffman HG. 1989. Misinformation and memory: the creation of new memories. *Journal of Experimental Psychology: General* **118**: 100–104.
Marsh RL, Landau JD. 1995. Item availability in cryptomnesia: assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology: Learning, Memory, & Cognition* **21**: 1568–1582.
McDermott KB. 1997. Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review* **4**: 582–586.
McDermott KB, Roediger HL III. 1998. Attempting to avoid illusory memories: robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory & Language* **39**: 508–520.
Metcalfe J, Jacobs WJ. 1998. Emotional memory: the effects of stress on 'cool' and 'hot' memory systems. *The Psychology of Learning and Motivation* **38**.
Miller MB, Wolford GL. 1999. Theoretical commentary: the role of criterion shift in false memory. *Psychological Review* **106**(2): 398–405.
Modia I. 1994. Forgetting childhood: a defense mechanism against psychosis in a Holocaust survivor. *Clinical Gerontology* **14**(3): 67–71.
Nemiah JC. 1979. Dissociative amnesia: a clinical and theoretical reconsideration. In *Functional Disorders of Memory*, Kihlstrom JF, Evans FJ (eds). Lawrence Erlbaum: Hillsdale, NJ; 303–324.
Nickerson RS. 1984. Retrieval inhibition from part-set cuing: a persisting enigma in memory research. *Memory & Cognition* **12**: 531–552.
Ofshe R, Singer MT. 1994. Recovered-memory therapy and robust repression: influence and psuedomemories. *International Journal of Clinical and Experimental Hypnosis* **42**: 391–410.
Payne DG, Elie CJ, Blackwell JM, Neuschatz JS. 1996. Memory illusions: recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language* **35**(2): 261–285.
Pezdek K. 1994. The illusion of illusory memory. *Applied Cognitive Psychology* **8**: 339–350.

Porter S, Yuille JC, Lehman DR. 1999. The nature of real, implanted, and fabricated memories for emotional childhood events: implications for the recovered memory debate. *Law & Human Behavior* **23**: 517–537.

Raaijmakers JGW, Shiffrin RM. 1981. Search of associative memory. *Psychological Review* **88**: 93–134.

Read JD. 1996. From a passing thought to a false memory in 2 minutes: confusing real and illusory events. *Psychonomic Bulletin & Review* **3**(1): 105–111.

Roediger HL III. 1974. Inhibition in recall from cuing with recall targets. *Journal of Verbal Learning and Verbal Behavior* **12**: 644–657.

Roediger HL III. 1978. Recall as a self-limiting process. *Memory & Cognition* **6**: 54–63.

Roediger HL III, McDermott KB. 1995. Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition* **21**(4): 803–814.

Roediger HL III, Neely JH. 1982. Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology* **36**: 213–242.

Roediger HL, III, Payne D. 1985. Recall criterion does not affect recall level or hypermnesia: a puzzle for generate/recognize theories. *Memory & Cognition* **13**: 1–7.

Roediger HL III, Schmidt SR. 1980. Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning & Memory* **6**: 91–105.

Roediger HL III, Thorpe LA. 1978. The role of recall time in producing hypermnesia. *Memory & Cognition* **6**: 296–305.

Rosch E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* **104**: 192–233.

Rundus D. 1973. Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior* **12**: 43–50.

Sargant W, Slater E. 1941. Amnestic syndromes of war. *Proceeding of the Royal Society of Medicine* **34**: 757–764.

Smith SM. 1979. Remembering in and out of context. *Journal of Experimental Psychology: Human Learning & Memory* **5**: 460–471.

Smith SM, Tindell DR, Pierce BH, Gilliland TR, Gerkens DR. 2001. Source memory failure in episodic confusion errors. *Journal of Experimental Psychology: Learning, Memory & Cognition* **27**: 362–374.

Smith SM, Vela E. 1991. Incubated reminiscence effects. *Memory & Cognition* **19**: 168–176.

Smith SM, Ward TB, Tindell DR, Sifonis CM, Wilkenfeld MJ. 2000. Effects of category structure on created memories. *Memory and Cognition* **27**: 386–395.

Tulving E, Pearlstone Z. 1966. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior* **5**: 381–391.

Tulving E, Thomson DM. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* **80**(5): 359–380.

Van der Hart O, Nijenhuis E. 1995. Amnesia for traumatic experiences. *Hypnos* **22**: 73–86.

Ward TB. 1994. Structured imagination: the role of category structure in exemplar generation. *Cognitive Psychology* **27**: 1–40.

Ward TB, Patterson MJ, Sifonis CM, Dodds RA, Saunders KN. 2002. The role of graded category structure in imaginative thought. *Memory & Cognition* **30**: 199–216.

Williams LM, Siegel JA, Pomeroy JJ. 2000. Validity of women's self-reports of documented child sexual abuse. In *The Science of Self-report: Implications for Research and Practice*, Stone AA, Turkkan JS (eds). Lawrence Erlbaum: Mahwah, NJ; 211–226.

Zaragoza MS, Koshmider JW III. 1989. Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory and Cognition* **15**: 246–255.

APPENDIX A

Filler categories and respective members

Work animals	Amphibians	Primates	Dinosaurs	Large animals
mule	toad	baboon	tyrannosaurus	moose
work horse	bullfrog	gorilla	pterodactyl	tarantula
seeing eye dog	mud puppy	mandrill	stegosaurus	mammoth
bloodhound	newt	ape	brachiosaurus	blue whale
ox	salamander	orangutan	triceratops	giraffe
camel	tree frog	gibbon	brontosaurus	hippopotamus
donkey	tadpole	rhesus	velociraptor	walrus
Animals eaten	**Australian animals**	**Hoofed animals**	**Insects**	**Dogs**
goat	kangaroo	bison	silkworm	wolf
salmon	koala	gnu	aphid	Irish setter
chicken	wallaby	eland	cricket	wolfhound
buffalo	dingo	llama	praying mantis	coyote
pig	Tasmanian devil	reindeer	cockroach	fox
oyster	wombat	gazelle	termite	cocker spaniel
turkey	platypus	elk	grasshopper	rottweiler
Aquatic animals	**Birds of prey**	**Extinct animals**	**Game fish**	**Parasites**
humpback	eagle	dodo	sailfish	leech
dolphin	falcon	mastodon	sturgeon	tapeworm
porpoise	condor	sabre-tooth	marlin	tick
orca	owl	carrier pigeon	barracuda	weevil
sea lion	hawk	duckbill	swordfish	mosquito
beluga	osprey	giant sloth	ray	bloodsucker
narwhal	vulture	moa	grouper	pinworm
Small animals	**House pets**	**Imaginary animals**		
amoeba	Siamese cat	dragon		
flea	hamster	sphinx		
chihuahua	iguana	gryphon		
ant	parakeet	yeti		
gnat	collie	King Kong		
mouse	goldfish	unicorn		
virus	guinea pig	centaur		

APPENDIX B

Critical categories and respective members for counterbalancing A

Rodents	Circus animals	Dangerous animals
beaver	seal	black widow
muskrat	tiger	grizzly
gerbil	chimpanzee	wolverine
woodchuck	elephant	cobra
porcupine	lion	scorpion
gopher	panda	viper
chipmunk	zebra	leopard

Critical categories and respective members for counterbalancing B

Carnivores	Flying animals	Reptiles
badger	bat	lizard
cheetah	monarch butterfly	gecko
lynx	cardinal	rattlesnake
raccoon	moth	turtle
shark	hummingbird	tortoise
weasel	wasp	chameleon
jaguar	sparrow	copperhead