Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 6, 1111–1119

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.6.1111

Assessing the Inhibitory Account of Retrieval-Induced Forgetting With Implicit-Memory Tests

Timothy J. Perfect
University of Plymouth

Christopher J. A. Moulin
University of Leeds

Martin A. Conway
University of Durham

Elizabeth Perry
University of Bristol

Five experiments examined whether retrieval-induced-forgetting effects are observed for implicit tests of memory. In each experiment participants first studied category–exemplar paired associates, then practiced retrieval for a subset of items from a subset of categories before finally completing memory tests for all the studied items. In standard fashion, inhibition was measured as the performance difference of unpracticed items from practiced categories and unpracticed items from unpracticed categories. Across the 5 experiments poorer performance for unpracticed items was seen in conceptual implicit memory (category generation and category matching) but not in perceptual implicit memory (stem completion, perceptual identification). Thus, retrieval-induced-forgetting effects are limited to tests of conceptual memory.

Anderson, Bjork, and Bjork (1994) introduced the retrieval-practice paradigm as a demonstration of the role of inhibition in memory retrieval. This paradigm has three phases: initial learning of a set of category–exemplar paired associates, repeated-retrieval practice for a subset of items from some categories, and a cued-recall test for all items. For example, a participant might study a set of items that includes kinds of fruit (orange, lemon, banana, and pineapple) and kinds of hobbies (gardening, biking, drawing, and ceramics) among other categories. During retrieval practice, half of the items from half of the categories are practiced using a word-stem-completion task (e.g., "Fruit: Or _____"). At the final recall test there are three kinds of items: (a) practiced items from practiced categories (*RP+ items*), (b) unpracticed items from practiced categories (*RP− items*), and (c) unpracticed items from unpracticed categories (*U items*).

The studies showed that repeated practice of a subset of items leads to superior memory for those items compared with items from the unpracticed category. However, this was accompanied by inferior memory for the unpracticed members of the practiced category relative to control items from unpracticed categories. That is, with the above example, *orange* was more likely to be recalled, and *pineapple* less likely to be recalled than the equivalent items for categories not practiced (*ceramics*, *biking*).

In a later development of this paradigm, Anderson and Spellman (1995) showed that retrieval-induced forgetting is not merely found within categories but is instead driven by the relatedness of concepts in memory. Using a procedure similar to that of Anderson et al. (1994), Anderson and Spellman (1995) used categories that included items that could belong in another category (e.g., FOODS: strawberry, crackers; RED: blood, tomato). What they demonstrated was that practicing retrieval of *RED–blood*, not only impaired recall of *RED–tomato* relative to control but also impaired memory for *FOOD–strawberry*, even though a different cue was involved than the one participants practiced. Retrieving *blood* did not impair memory for *crackers*, however. Thus, the retrieval inhibition spreads to concepts related to redness, rather than just within the presented category.

It is this latter finding that Anderson and Spellman (1995) regard as crucial in support of an inhibitory account of the retrieval-induced-forgetting idea. (For a full account of potential noninhibitory accounts of the phenomenon, see Anderson & Bjork, 1994.) They posited the criterion of cue-independent forgetting as the critical demonstration that retrieval-induced forgetting is an inhibitory effect. They reasoned that if the phenomenon was due to either a biasing of the cue associated with the target or an increased associative strength between the practiced cue and target, then the forgetting phenomenon should not be observed with a different test cue. The fact that forgetting was observed indicated that it was the unpracticed target itself that was inhibited, rather than any relation between it and the cue or between the cue and the practiced item. On the basis of this argument, and their supporting evidence, Anderson and Spellman (1995) concluded that "inhibitory processes must be at work" (p. 68).

Timothy J. Perfect, Department of Psychology, University of Plymouth, Drake Circus, Plymouth, United Kingdom; Christopher J. A. Moulin, Department of Psychology, University of Leeds, Leeds, United Kingdom; Martin A. Conway, Department of Psychology, University of Durham, Durham, United Kingdom; Elizabeth Perry, Department of Psychology, University of Bristol, Bristol, United Kingdom.

This work was supported by Economic and Social Research Council Grant R000222681. We thank Dan Burns and Neil Mulligan for helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to Timothy J. Perfect, Department of Psychology, University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom. E-mail: tperfect@plymouth.ac.uk

Terminology Issues

The term *inhibition* has been used loosely in memory research to mean behavioral reductions in output, reductions in accessibility without reduction in availability (cf. Tulving & Pearlstone, 1966), or reduction in availability of the memory representation (see Bjork, Bjork, & Anderson, 1997, for a discussion of this issue). The first of these three uses of the term is the least appropriate, because lower levels of recall can result in many processes that would not be appropriately described as inhibitory.

The second usage has been proposed by Bjork and Bjork (1996; see also Bjork et al., 1997), particularly in connection with the directed-forgetting paradigm. They argued that the directed-forgetting instruction results in participants having reduced access to memory representations that remain unchanged in their activation level. The principal evidence to support this view is that whereas directed forgetting reduces recall of the forgotten list and reduces interference from the forgotten list on the to-be-remembered list, there is equal priming for both lists when tested using indirect methods. Hence Bjork and colleagues (Bjork & Bjork, 1996; Bjork et al., 1997) argued that retrieval is inhibited, but the representation itself is not.

However, Bjork, Bjork, and Anderson (1997) argued for the final interpretation of inhibition in the retrieval-induced-forgetting paradigm, because of the cue-independent nature of the effect reported by Anderson and Spellman (1995). That is, they argued that retrieval practice produces reductions in availability of the memory representation itself. Thus, retrieval practice is thought to produce the most unambiguous form of inhibition, a reduction in the level of activation of an item that is subject to the inhibition. For the remainder of this article we use the term *inhibition* in this final, least ambiguous, sense.

The use of implicit-memory tests is therefore central in deciding which of the two forms of inhibition apply in the retrieval-induced-forgetting paradigm. If retrieval-induced forgetting induces retrieval inhibition, as in the directed-forgetting paradigm, one would expect to see intact implicit memory for items that are not recalled. That is, the items would remain available but are not accessible. On the other hand, if the items are inhibited, then one would expect to see both poorer recall and reduced implicit memory for those items. That is, the items would be less available.

Implicit Memory and Retrieval-Induced Forgetting

Work on retrieval-induced forgetting has largely used recall as the measure of memory on the final test. However, there are several good reasons for using implicit-memory measures. As we have argued above, implicit-memory tests provide an alternative test of the activation level of the unpracticed items. Anderson and Spellman (1995) discussed whether alternative forms of memory testing other than recall are likely to produce the effect. They argued, on the basis of work on other paradigms, such as part-set cuing, that recognition may not be impaired by retrieval practice. However, they presented no evidence for this, and in any case they acknowledged that recognition latency might be slowed. They did not mention any predictions regarding different forms of implicit testing but did argue that "if a representation is truly inhibited, performance impairments arising from that inhibition should generalize to any cue used to test that item" (Anderson & Spellman, 1995, p. 92). This implies that any sufficiently sensitive test of memory that requires activation of the representation should show the effect. Thus there are good a priori reasons for us to believe that at least some forms of implicit memory should show an inhibitory effect following retrieval practice. If, on the other hand, the implicit-memory measures do not show the effect, this would suggest that retrieval-induced forgetting does not induce inhibition in the strong sense, as claimed by Anderson and Spellman (1995).

Using implicit-memory tests also allows memory for items to be tested without reinstating the cue used at study. For example, perceptual identification of the degraded target *orange* does not require presentation of the cue *fruit* that was originally paired with it. The same applies to the likelihood that *orange* will be generated to the stem *or*—. Thus, both these tests provide cue-independent measures of the activation of the target, and thus in Anderson and Spellman's (1995) terms they provide rigorous tests of alternative, noninhibitory accounts of the retrieval-practice effect. If inhibition is observed in these tasks, it cannot be attributed to the cue.

A final advantage is that implicit-memory measures can be applied to materials that were never studied at all. For instance, participants can identify a degraded form of the word *pear* even if *pear* was not one of the fruit words originally studied. Given that repeated retrieval of the word *apple* reduces the recall of the previously studied word *orange*, it is entirely possible that other concepts that share features with "apple" (such as "pear") have also been inhibited. However, one cannot ask people to recall what was never presented. One can, however, ask them to identify the degraded word form of *pear* or to generate a word from the stem *pe*— and to contrast the rate of completions to *pear* with baseline.

Recently, Butler, Williams, Zacks, and Maki (2001) published the first study to examine implicit memory for items in the retrieval-induced-forgetting paradigm. In their study, participants studied category–exemplar pairs in the conventional manner. They also completed a standard retrieval-practice phase. At test, participants were tested in one of five ways: standard cued recall, word-fragment completion, word-fragment cued recall, category-cue-plus-fragment cued recall, and category-plus-word-stem cued recall. Only the first of these conditions showed a retrieval-induced-forgetting effect. Thus, one apparent conclusion from their work is that retrieval-induced forgetting is not observed for implicit-memory tests, suggesting that the retrieval-induced-forgetting effect is similar to retrieval inhibition, as observed in directed forgetting.

However, it is perhaps premature to draw such firm conclusions. In particular, the conditions tested by Butler et al. (2001) do not constitute an exhaustive list of implicit-memory tests. There may be other implicit measures that do show retrieval-induced-forgetting effects. Additionally, as Butler et al. acknowledged, their category-plus-stem cued recall represents the same form of testing as in Anderson et al.'s (1994) study, which did show the effect. Given that they did not replicate the Anderson et al. effect with cued recall, it is hard to draw strong conclusions about their failure to show an effect with implicit memory.

Experiment 1

Our first study had the aim of establishing a retrieval-induced-forgetting effect using a recall measure and contrasting this with the effect observed in an implicit measure. We opted to use a category-member generation task because this required minimal changes to the retrieval cues (cf. Light & Singh, 1987). Under the cued-recall instructions, participants were given the category cues they had seen at study and were asked to recall the exemplars they had seen. Under the category-generation task, participants were given category cues and asked to generate the first four category members they could think of. The retrieval cue was therefore the same in both conditions. We hypothesized that if retrieval-induced-forgetting is a mechanism of retrieval inhibition, as in the work on directed forgetting, we would expect no forgetting effect for the category-generation task, in line with Butler et al.'s (2001) results. On the other hand, if retrieval-induced forgetting represents inhibition, then we would expect to see poorer performance on both the recall and category-generation tasks.

Method

Participants. Twenty participants were recruited, by means of posters distributed around the University of Bristol, Bristol, United Kingdom, to take part in this study. Their mean age was 24.2 ($SD = 3.1$) years.

Materials. There were 36 items in the to-be-learned list, consisting of two exemplars for each of 18 categories (e.g., FRUIT: lemon, banana). The categories were selected from Battig and Montague's (1969) norms, and items within those categories were chosen so that they were common but not one of the two most frequently occurring exemplars. No two items (across all categories) began with the same first two letters.

At the encoding phase, all 36 items were presented with their category cue. At the retrieval-practice phase, participants practiced retrieval for one of the exemplars, for 12 of the categories. Thus there were 12 items that were practiced, 12 that were unpracticed that came from practiced categories, and 12 that were unpracticed and came from unpracticed categories. For each type of item, half were tested using cued recall, and half were tested using category generation.

Procedure. Participants were tested individually in a quiet testing room. They were informed that they would be participating in a test of memory and were shown examples of test items they would see later. Each card contained the category term printed in capitals and an exemplar of that category printed in lower case (e.g., *BUILDING church*). Before the test began participants were told that their task was to try to remember as many of the items as they could. They were then shown the category–exemplar pairs on flash cards at a rate of 5 s per pair, with a 1-s interval between items.

After the items had been studied, participants were then introduced to the retrieval-practice task. As an example they were shown a retrieval cue card containing the category printed in capital letters, together with a two-letter word stem as cue (e.g., *BUILDING ch*). Participants were told that their task was to recall the word they saw earlier that started with the same two letters. They had 10 s in which to generate an answer. If no answer was recalled, an extra letter was added to the word stem provided as a cue (e.g., *BUILDING chu*) and an additional 10 s was allowed. This was sufficient to ensure that the correct item was generated during practice for all participants. Once participants understood the task, they were presented with the test cards. Twelve items (the RP+ items) were practiced in this way, with counterbalancing to control for item effects in the selection of the items to be practiced. The list of 12 items was repeated three times in succession.

After a filled delay of 5 min, participants were given the memory tests. First implicit memory was assessed using three unpracticed categories (with two exemplars) and six practiced categories (with one practiced and one unpracticed item). Participants were told that their task was to say aloud the first four exemplars that they could think of for each category they were given. The experimenter then gave an example unrelated to the test materials to ensure understanding. For each test category, participants were given 20 s in which to generate category members. Then the explicit-memory test was administered for the remaining three unpracticed categories and six practiced categories. Participants were asked to say aloud the two category members they had seen associated with the category in the original list. An example was given to ensure understanding. For each test category participants were given 10 s to recall the test items.

Results and Discussion

The theoretical focus in this article is on inhibitory processes. The extent to which inhibition is present is indicated by the reduction in performance for RP− items compared with U items. All analyses test the prediction of poorer performance in the RP− condition with a one-tailed test with an alpha level of .05. The effects of practice, reflected in performance on the RP+ items are not of interest and are therefore not included in any of the analyses reported here. However, performance on RP+ items is reported in Table 1 for completeness.

On average, participants recalled 53% of the RP− items ($SD = 25\%$) and 71% ($SD = 20\%$) of the U items. This difference was significant, $t(19) = 2.33$, in line with the standard retrieval-induced-forgetting effect. Recall of the RP+ items was near ceiling (96%, $SD = 9\%$). The same analysis was conducted on the proportion of target items generated during the category-member-generation task. Participants generated 44% ($SD = 25\%$) of the RP− items and 61% ($SD = 28\%$) of the U items. Once again, a significant forgetting effect was observed, $t(19) = 2.33$. On average, participants generated 77% ($SD = 26\%$) of the RP+ items.

These data therefore clearly establish that the standard retrieval-induced-forgetting effect is observed for these materials using recall and that it generalizes to category-member generation. The effect of practicing retrieving one member of a category (e.g., FRUIT: orange) meant that it was less likely that a participant would either recall or generate the alternate category member (apple), compared with items in unpracticed categories.

Table 1
Results of Experiments 1–4

Experiment and condition	Item type							
	RP+		RP−		U		N	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1								
Cued recall	.96	.09	.53	.25	.71	.20		
Category generation	.77	.26	.44	.25	.61	.28		
Experiment 2								
Cued recall	.99	.03	.44	.23	.61	.23		
Perceptual ID	8.9	1.8	10.1	1.9	11.1	2.2	12.1	1.4
Experiment 3								
Cued perceptual ID	6.7	1.2	8.4	1.7	8.0	1.9	10.1	1.3
Experiment 4								
Word-stem completion	.72	.23	.37	.14	.36	.12	.28	.08

Note. Cued recall was an explicit-memory test. The following were implicit-memory tests: category generation (items generated to category cue), perceptual ID (number of steps taken to identify degraded word form), cued perceptual ID (perceptual ID in the presence of the category cue), and word-stem completion (number of word stems completed as targets). RP+ = items that were practiced; RP− = items that were not practiced but came from the same category as practiced items; U = unpracticed items from unpracticed categories; N = new items, not previously studied; ID = identification.

Experiment 2

To test the generality of the retrieval-induced-forgetting effect across test format, we conducted a second experiment using a different implicit-memory test. This study replicated the use of cued recall but used perceptual identification of degraded word forms as the implicit measure. There were several reasons for moving away from category-member generation toward this perceptual measure. First, it is less open to the criticism of explicit-memory mediation of the effects. In Experiment 1, the retrieval cues for the explicit-memory task and the implicit-category-member-generation task were identical. Only the instructions differed between conditions. Therefore there is a strong possibility that at least some of the participants were likely to engage in effortful, conscious attempts to retrieve items seen previously. Perceptual identification is less likely to be contaminated by explicit memory (Roediger, Weldon, & Challis, 1989).

Perceptual identification of degraded forms offers generalization in two ways. First, it enables one to test whether inhibition applies to all forms of representation of the nonstudied material. Previous research by Anderson and colleagues (Anderson, Bjork, & Bjork, 1994; Anderson & McCulloch, 1999; Anderson & Spellman, 1995) has focused only on recall-based measures. If retrieval-induced forgetting produces inhibition at the level of item representations in memory—either in semantic memory or in the output lexicon—then one might expect to see inhibitory effects in other forms of memory test. On the other hand, if no inhibitory effects are seen, this would raise questions about what is inhibited in the retrieval-induced-forgetting paradigm.

Perceptual identification of degraded forms also offers a means of testing memory using Anderson and Spellman's (1995) crucial criterion for assessing inhibition, namely cue-independent forgetting. With perceptual-identification tasks, the participant attempts to identify the degraded item without the presence of the original cue that is required by cued-recall and category-member generation. Thus, if participants show slowing in the ability to identify an item, this cannot be attributed to the cue and is therefore, in Anderson and Spellman's terms, clear evidence of inhibitory processes in action.

Method

Participants. Eighteen participants took part in this study, recruited and remunerated as in Experiment 1. Their mean age was 20.4 ($SD = 3.1$) years.

Materials. There were two conditions in this experiment: an explicit condition (cued recall) and an implicit condition (perceptual identification). These two conditions had separate to-be-remembered lists. There were 24 items in each to-be-remembered list, consisting of four exemplars from each of six categories. The categories were selected from Battig and Montague's (1969) norms, as in Experiment 1. Because category generation was not part of this procedure, exemplars that were ranked first or second in the norms were included. At the encoding phase, all 24 items were presented to participants as pairs in category–exemplar form. At the retrieval-practice phase, two exemplars were practiced from four of the presented categories. No items were practiced from the remaining two categories. Thus there were 8 RP+, 8 RP−, and 8 U exemplars.

For the perceptual-identification task, there were an additional 24 items in six categories, which acted as foils. This meant that when presented with a perceptually degraded exemplar, there was a 50% probability that the item had been presented as a to-be-remembered stimuli. The selection of categories and exemplars for this task was based on piloting, which was undertaken to ensure that the all items were identifiable at comparable stages. All exemplars were chosen on the performance of 8 control participants. We examined baseline identification rates for 96 words from 16 categories (i.e., 6 words per category). Words were presented in a random order in the degraded form (without any memory instruction), and we recorded the point at which stimuli were readable (as described in the *Procedure* section). Using these mean number of steps taken to identify words, we rejected 2 words from each category that were identified as being furthest from the grand mean of all 96 words. Then we excluded the two categories that were identified the easiest and an additional two on the basis of high variability between exemplars within that category. From this process, we are confident that all resulting 48 stimuli were of a comparable difficulty at the perceptual-identification stage. The range of mean steps taken until identification across items was 8.75–13.50. The mean identification rate for items was 11.27 steps, with a standard deviation of 1.25 steps. Within each category, counterbalancing ensured that the exemplars were presented equally as RP+, RP−, and U items. Because of our substantial pilot work, the foils were the same exemplars from the same categories for every participant: These were not counterbalanced.

Procedure. There were three phases to both tasks: presentation, practice, and test. The procedure for presentation and practice was identical for the two tasks. Tasks were carried out in a counterbalanced order: Half of each group was tested in the explicit condition first, and half were tested in the implicit condition first. Participants were tested individually in a quiet testing room. They were informed that they would be participating in a test of memory, and they were shown an example of the test category–exemplar pairs. Before presentation began, participants were told to try to remember as many pairs as they could. Word pairs were presented on a computer screen 1 word at a time in a random order for 2 s per pair.

After the 24 to-be-remembered words had been presented, participants were introduced to the practice phase, using the same example presented earlier. The RP+ pairs were presented on the computer screen with only the first letter of the exemplar pair visible. Participants were asked to try to remember the word. If they could not remember it, they were shown an additional letter until they correctly recalled the item. There was no time limit given, but participants were told to ask for the next letter if they felt they could not recall the item. Practicing of the eight items occurred in a random order within one block; the block was presented three times. Therefore, each RP+ item was practiced three times.

Immediately after the practice phase, participants were tested. In the explicit condition, participants were instructed to recall as many of the exemplars as they could remember from each category. They were cued for all six categories using the category name. Recall from the categories was in a counterbalanced order.

In the implicit condition, participants were given two trials of sample degraded stimuli. The perceptual-identification task was arranged so that the to-be-identified word was obscured behind an array of 320 boxes arranged in a grid. At Step 0, none of the word was visible. At Step 1, 15% of the boxes were randomly removed so that some of the word underneath was in view. Then, for each step, 5% of the remaining boxes were randomly removed so that the word became progressively easier to read. The removal of boxes continued at this rate until Step 20 (when about 68% of the word was in view), after which all the word was shown, and the procedure moved on to the next word. Piloting of the procedure (as above) showed that all participants had identified the words well before the 20th step.

Participants were instructed that at each step of the word presentation they should either try to identify the word verbally or say that they could not identify the word. For every incorrect identification attempt or when the participant declared they could not identify the word, another step was revealed. The experimenter recorded when the participant successfully identified the word. Participants were presented all target words in a random order and without cues.

Results and Discussion

The proportion of items recalled was analyzed as before, and a reliable forgetting effect was observed, $t(17) = 2.65$. On average, participants recalled 44% ($SD = 23\%$) of the RP− items and 61% ($SD = 23\%$) of the U items. As in Experiment 1, recall of the RP+ items was close to ceiling ($M = 99\%$, $SD = 3\%$).

Implicit-memory performance was analyzed as the number of steps taken to identify a degraded word form. One should note, however, that a forgetting effect would be observed if more steps were taken to identify RP− items than U items. No analysis was necessary to test the inhibitory account of retrieval-induced forgetting, because no forgetting was seen. It took participants numerically fewer steps to identify RP− items ($M = 10.1$ steps, $SD = 1.9$) than U items ($M = 11.1$ steps, $SD = 2.2$). On average it took participants 8.9 ($SD = 1.8$) steps to identify RP+ items and 12.1 ($SD = 1.4$) steps to identify items not previously studied.

Therefore no reliable inhibitory effect was seen in the implicit measure in this experiment. Is this because the measure was insensitive? Apparently not, because in contrast with the new items, reliable priming was seen for the RP− items, $t(17) = 4.3$, and there was a trend toward priming for the U items, $t(17) = 1.99$, $p < .07$. Similarly, reliably more priming was seen for the RP+ items than both the RP− items, $t(17) = 2.96$, and the U items, $t(17) = 4.24$. Thus the study phase produced reliable priming, and the practice phase produced even more priming for the practiced items. What the practice phase did not do, however, was to induce inhibition of the items related to the practiced items.

This experiment, using a different implicit test, therefore did not replicate the pattern seen with category generation. Before consideration of what aspects of the two tasks may be responsible for the appearance, and nonappearance, of an inhibitory effect, we decided to rule out another potential explanation. In the perceptual-identification task, performance was tested for the item alone without a cue. The standard retrieval-induced-forgetting effect using recall relies on category cues, as did our category-generation task. Would, therefore, our perceptual-identification task show an inhibitory effect if we reinstated the cue at test? We hypothesized that if the presence of the cue is critical in producing the inhibitory effect, then the effect should reappear for the perceptual task when accompanied by the cue. However, if the absence of the effect is due to the perceptual nature of the retention test, then the representation of the cue at test should not alter the pattern of the results.

Experiment 3

Method

Participants. Eighteen participants took part in this experiment. Their mean age was 21.4 ($SD = 3.8$) years. No participants who took part in Experiment 2 were recruited to this study. Remuneration and recruitment was the same as in the previous experiments.

Materials and procedure. This experiment was identical to the perceptual-identification task presented in Experiment 2, with the modification that at the final test stage the category cue for each exemplar was presented centrally above the to-be-identified word. The same materials and counterbalancing as in Experiment 2 were used.

Results and Discussion

There was no measure of cued recall in this study, therefore the only analyses concerned performance on the perceptual-identification task. On average, participants recognized RP+ items in 6.7 ($SD = 1.2$) steps, RP− items in 8.4 ($SD = 1.7$) steps, U items in 8.0 ($SD = 1.9$) steps, and new items in 10.1 ($SD = 1.3$) steps. Apart from a general speeding, attributable to the presence of the category cue, the results are remarkably similar to the previous study. As before, contrast of the RP− and U items revealed no retrieval-induced slowing effect, $t(17) = .67$. This is not because the task lacked sensitivity however, because the RP+ items were solved reliably faster than the RP− and U items, $t(17) = 4.6$ and 3.6, respectively, and the RP− and U items were solved reliably faster than the new items, $t(17) = 5.0$ and 5.9, respectively.

These data therefore replicate Experiment 2 and indicate that the absence of an inhibitory effect in the perceptual-identification data in that study was due to the nature of the test rather than the absence of the cue at the test stage. To examine the generalization of retrieval-induced forgetting to other forms of memory test, we decided to explore performance on word-stem completion in the next study.

Experiment 4

The fourth experiment focused on word-stem completion as a measure of retention. This approach has a number of advantages. First it offers an alternative implicit measure to test the generalizability of the retrieval-induced-forgetting effect. It is also attractive because it mirrors closely the methods used by Anderson, Bjork, and Bjork (1994) in their original demonstrations of the retrieval-induced-forgetting effect. They tested cued recall with word stems, to control for output order in the final test, and showed a reliable forgetting effect. It is a small step to use stems in their more conventional implicit-test format. However, we differed from Anderson et al. in one important way. We did not re-present the cues at the final test. This was because we were keen to retain the cue-independent nature of the final test, to provide a rigorous test of the inhibitory account of retrieval-induced forgetting. If practicing *FRUIT–orange* inhibits the concept "apple," then this should emerge when we test using the stem *ap*—, whether the category cue is present or not. Given that noninhibitory accounts are based on activation from the cue at test, we reasoned that presenting the stem alone at test would preclude noninhibitory accounts of any inhibition effect. We also thought that presentation of the cue at test would prompt explicit retrieval strategies by the participants, and we wished to avoid that.

Method

Participants. Eighteen participants were recruited in the same manner as in the previous experiment. Their mean age was 21.9 ($SD = 4.3$) years. No participants from the previous experiments took part in this study.

Materials. The materials were the same as used in Experiment 2, except that exemplars less than five letters long were replaced. Also all exemplars were rejected that did not have a unique three-letter stem. There were 8 RP+, 8 RP−, and 8 U exemplars, and at test there were 24 foils. In this experiment counterbalancing was full: All exemplars and categories were presented as foils or targets, and as targets all items appeared as RP+, RP−, or U items.

Procedure. The procedure did not differ from Experiment 2, except in the way participants were tested. The presentation, practice phases, and instructions given were identical to Experiment 2. Immediately after practice, participants were presented with a sheet of paper with 48 three-letter stems in a pseudorandom order. These 48 stems were either to-be-remembered target items from the original list or foils. The category labels were not presented. The participant was instructed to complete the three-letter stem with the first word that came to them. No explicit-memory instruction was given. Participants then completed the stems, in writing, in as much time as they needed.

Results and Discussion

In this experiment, there were no explicit-recall data collected, therefore the only analyses concern the implicit stem-completion task. In this task, performance is measured as the proportion of stems completed with target items. Retrieval inhibition, if observed, should therefore be seen as lower stem-completion rates in the RP− condition than in the U condition.

Participants completed 72% (SD = 23%) of the RP+ stems with previously studied items, 37% (SD = 14%) of the RP− items, 36% (SD = 12%) of the U items, and 28% (SD = 28%) of the new, or baseline, items. As in Experiment 2, performance was numerically superior in the RP− condition compared with the U condition, and therefore no analysis was necessary to determine that no retrieval-induced-forgetting effect had been observed. However, as in Experiments 2 and 3, this was not because the implicit test was insensitive to the procedure. Participants completed reliably more items when they were RP+ items than when they were either RP− items, $t(17) = 8.1$; or U items, $t(17) = 7.4$; and reliably more RP− and U items than baseline, $t(17) = 2.5$ and 2.3, respectively.

These data are consistent with the results of Experiments 2 and 3. As in that study, practice had a beneficial effect. There was also reliable priming observed: Stem completion rates were higher for previously studied items than nonstudied items. However, despite both these indications that our test was sensitive to manipulations of retention, we did not observe a retrieval-induced-forgetting effect. Thus the data are in line with those reported by Bjork and Bjork (1996), who showed that recall, but not word-fragment completion, was influenced by a directed-forgetting manipulation.

Our next experiment was designed to further investigate the nature of the inhibitory effect in retrieval-induced forgetting. The pattern so far is that we have shown reliable inhibition using cued recall and category-member generation but no effect at all using perceptual identification or word-stem completion. Why do these tasks differ with regard to the presence or absence of the forgetting effect?

The tasks clearly have very different demands. Cued recall and category-member generation both require retrieval of information and therefore require top-down, conceptual processing. Perceptual identification and word-stem completion, on the other hand, are data-driven tasks. Thus, one potential explanation for the pattern observed so far is that retrieval-induced forgetting may be seen only in tasks requiring top-down processing.

There is an alternative possibility, however. Cued recall and category-member generation are both tests that require access to representations in the conceptual (episodic–semantic) memory system. However, the basis for priming in perceptual identification and word-stem completion may not be in that system. Schacter (1994) argued that perceptual priming is mediated through a preconceptual perceptual representation system. Similarly there is evidence that word-stem-completion priming may be mediated through representations in the output lexicon. Downes et al. (1996) showed an effect called cohort priming when using stem completion as an implicit task. Participants were exposed to materials in which the first phoneme can be pronounced in alternative ways (e.g., the *ba* in *ball* is pronounced differently from the *ba* in *barn*). The normal word-stem priming effect is a bias toward completing a stem with the previously seen word. Cohort priming is a bias toward completing the stem with the same pronunciation of initial phoneme as the target word when the target is not generated. This effect therefore suggests that priming seen in word-stem completion may come from lexical representations, not conceptual ones.

Thus, there are two possibilities regarding the presence or absence of retrieval-induced-forgetting effects in implicit memory. One is that retrieval-induced forgetting is restricted to tasks that involve retrieval, or top-down processing. The other possibility is that retrieval-induced forgetting is restricted to conceptual representations in memory irrespective of the processing demands. What is needed is a task that taps conceptual representations without requiring retrieval. Experiment 5 was designed with exactly this in mind.

Experiment 5

In this experiment we examined inhibition for a task that relies on access to semantic representation of an object but is not retrieval based. The task we selected was category verification. Here, participants are presented with category cue–exemplar pairs and required to verify, as quickly as possible, whether the exemplar belongs in the category. In this paradigm, a retrieval-induced-forgetting effect would be observed as slower verification times for unpracticed items from practiced categories than for unpracticed items from unpracticed categories.

Method

Participants. Thirty-six participants were recruited to take part in this study. Because of an administrative oversight, demographic data were not collected for this experiment. However, because participants were recruited in the same manner as in the previous experiments, there is no reason to believe that the samples differed from previous experiments.

Materials. There were 18 category–exemplar pairs with two exemplars each from nine categories in the to-be-learned list. This meant that there were 6 RP+, 6 RP−, and 6 U items presented at study. At practice, six categories were represented with one item from each. Counterbalancing ensured that each category–item pair appeared equally as an RP+, RP−, and U item. There were 120 pairs presented at test, with 60 invalid category–exemplar pairs (e.g., *time–cushion*) and 60 valid pairs (e.g., *time–minute*). Of the valid pairs, there were 6 each of the target items (RP+, RP−, and U), 6 new targets from practiced categories (RPN), and 6 unpresented targets from unpracticed categories (UN). There were also 30 valid pairs from new categories (Baseline). Of the 60 invalid pairs, 30 items were from presented categories and 30 items were from unpresented categories. The 60 items from unpresented categories came from a further nine categories that were never used at study.

Procedure. The presentation and practice phases were identical to Experiments 1–4, with the exception that there were only 6 RP+, 6 RP−, and 6 U items, and items were presented five times (as opposed to three) at practice. Participants were tested individually in a quiet room. Again, there were three phases to this experiment. All of the procedures used computer presentation, and in the test phase, participants responded to the stimuli using a keyboard. At presentation, participants studied each of the to-be-remembered pairs from the presented categories at a rate of 1 pair every 2 s. They were instructed to try and learn the pairs. Immediately after presentation, the participants practiced recall of the RP+ pairs in the manner described in the foregoing experiments, with five repetitions of each of the items. In the test phase, participants were instructed that they should verify category–exemplar pairs as quickly as possible. They were then presented all the stimuli pairs at a rate of approximately 1 every 5 s. Before each stimuli pair was presented a red fixation spot appeared on the screen. After a random interstimulus interval of between 1 and 2 s, the stimuli pair for verification was presented. If the participant failed to respond, the stimuli disappeared from the screen after 3 s. If the participant did respond, the pair disappeared from the screen immediately. Sixty pairs were presented in each of two blocks, and the participant had an opportunity to rest in the interval.

Results and Discussion

Although Experiment 5 included more categories of items than in previous experiments, we decided to analyze just the RP− and U items initially, for purposes of comparison. Because the dependent measure was response-time correct, we expected inhibition to be manifest as a slowing, rather than as a reduction in accuracy. The response times for RP− and U items were 873 ms ($SD = 206$ ms) and 821 ms ($SD = 179$ ms), respectively, which demonstrates the predicted slowing effect, $t(35) = 2.0$. The response times for the RP+ items was 821 ms ($SD = 173$ ms), which did not differ from the baseline (U items, $t[35] < 1$), suggesting that the additional retrievals did not induce additional priming for the practiced items. Nevertheless, RP+ and U items were reliably faster than new items from the same category (UN items) and new items from new categories (Baseline; all ts > 1.99). These data are shown in Figure 1.

In addition, we also collected response times for correct responses to three other types of items: RPN items were new members of practiced categories that had not appeared as study items. Similarly, UN items were new (unstudied) members of the unpracticed category. Baseline items were members of categories that had not appeared at all during study or practice phases. Comparing RPN and UN items therefore enabled us to see whether the observed retrieval-induced slowing effect generalizes to other members of the practiced category. However, mean response times for RPN (881 ms, $SD = 184$ ms) and UN (876 ms, $SD = 191$ ms) items did not differ. Thus practicing retrieval of *orange* from the cue *fruit* leads to slower verification only of those fruits in the study list, not of all other fruits that may share features with oranges and that may be partially activated by the cue *FRUIT*. We return to this issue in the General Discussion.

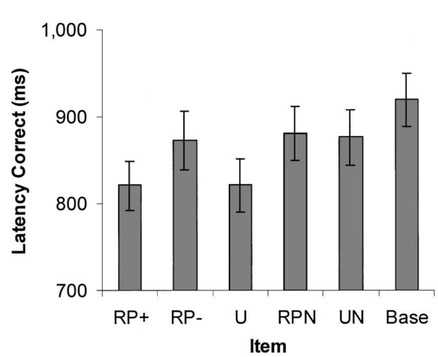

Figure 1. Mean verification times for correct responses to all item types in Experiment 5. Error bars represent the standard error of the mean. RP+ = practiced items from practiced categories; RP− = unpracticed items from practiced categories; U = unpracticed items from unpracticed categories; RPN = new members of practiced categories that had not appeared as study items; UN = new (unstudied) members of the unpracticed category; Base = baseline.

Figure 1 shows the mean response times for all item types and reveals an entirely coherent pattern. Practiced-items and unpracticed-items response times are both quicker than that of either new members of the same categories (RPN and UN) or of the baseline. The same is not true for RP− items. Response times for these items do not differ from response times to the members of the same categories that were not studied. It is as if the RP− items were not on the initial study list.

Thus, the results from this experiment indicate that accessing conceptual representations results in a reliable retrieval-induced-forgetting effect that is specific to previously studied items. Similar items, which share conceptual features with the practiced (and unpracticed) members of the practiced categories, show no evidence of impairment.

It is hard to see how this retrieval-induced slowing effect could be explicitly mediated. First the instructions were clear in requiring the participants to verify category membership. Given that all the necessary information was available at the time of the decision, it is hard to see why participants should adopt a strategy of attempting to recall the previously presented items. Moreover, the mean response times in this task (about 800 ms) seem to preclude a retrieval-based strategy. Thus these data demonstrate a reliable effect of retrieval-induced forgetting on an implicit task that is apparently not mediated by an explicit component.

A further point to note is that the task of category verification patently does not involve retrieval. We have succeeded in demonstrating retrieval-induced forgetting in a task that involves no retrieval. So, although the phenomenon might be induced only by retrieval (Anderson, Bjork, & Bjork, 2000), it is not limited to retrieval in its effects. We return to this point in the General Discussion.

General Discussion

The starting point for the discussion of the retrieval-induced-forgetting effect must be the performance in the cued-recall tests. This most closely matches the conditions used to date to support the idea of retrieval-induced forgetting. In both Experiments 1 and 2, robust retrieval-induced-forgetting effects were observed (18% and 17%, respectively). We therefore have replicated the standard effect using our procedures and materials. The subsequent failures to find retrieval-induced-forgetting effects using implicit tasks cannot therefore be due to some idiosyncrasy of our materials, methods, or participants.

The pattern of performance across the implicit tasks was not as initially expected. We began by contrasting two views of the retrieval-induced-forgetting effect. Either the effect was due to retrieval inhibition, which renders otherwise available items inaccessible to recall, or it was an inhibitory effect that reduces the availability of items in memory. Our initial hypothesis was that if the unpracticed items were truly inhibited, then we would be able to demonstrate poorer performance for those items on a range of implicit-memory tests. This hypothesis has proven inadequate, because we did not find clear retrieval-induced-forgetting effects across all implicit-memory tests, as would be expected from the simplest version of an inhibitory model. However, unlike Butler et al. (2001) we did demonstrate reliable retrieval-induced-forgetting effects on some implicit measures. In the following section we consider the factors that might explain why a retrieval-induced-forgetting effect was seen only with particular tasks.

One potential moderating factor is retrieval. The standard measure of retrieval-induced forgetting that has been repeatedly used in the literature is category-cued recall. If one considers the present set of implicit tasks, then clearly they can be considered to differ in the degree to which they require retrieval. Consistent with this idea, retrieval-induced forgetting was seen in recall and category generation but absent for perceptual identification, which does not require retrieval. However, a retrieval-induced-forgetting effect was absent from the word-stem completion task, which may involve some top-down retrieval processes, but was present for category verification, which was data driven and did not require retrieval. Thus, the final test does not have to involve retrieval to demonstrate a retrieval-induced-forgetting effect.

Turning this conclusion around somewhat, it therefore follows that retrieval is not what is inhibited in the final test. Thus, the pattern of performance across the tests is not consistent with a retrieval-inhibition account of retrieval-induced forgetting. Consequently, the retrieval-induced-forgetting effect is not the same as retrieval inhibition, as observed in the list-based directed-forgetting effect (see Bjork & Bjork, 1996; Macleod, 1997, for a discussion of retrieval inhibition in directed forgetting).

Much the same argument can be made to rule out explicit mediation of the effects. Given that no test is process pure, it might be argued that retrieval-induced-forgetting effects are mediated by the degree to which performance on a putative implicit test is contaminated by explicit memory. However, such an account struggles to explain the pattern of findings across the entire set of tasks. Whereas it is clearly possible that explicit memory could underlie the retrieval-induced-forgetting effect in category generation, the pattern of performance elsewhere is harder to explain in this manner. Perceptual identification failed to show a retrieval-induced-forgetting effect, even when the category cue was used. This reinstatement of the cue certainly induced improved performance. On average, participants were able to identify the target words around 2 steps earlier compared with the no-cue condition. If this improvement were due to explicit memory, then one would have expected a retrieval-induced-forgetting effect to appear, but it did not. Further, the pattern of performance across the two remaining tasks was not consistent with an explicit-mediation account. Word-stem completion did not show the forgetting effect, despite being a top-down task that might encourage explicit memory. Conversely, category verification is a speeded data-driven task with little apparent role for explicit mediation, but this did show the forgetting effect.

A more fruitful avenue of exploration concerns the nature of the implicit-memory tests themselves and the representations that they tap. The different measures may tap into different memory representations to that which is putatively inhibited by retrieval-induced forgetting. Consider first the tasks that do demonstrate a retrieval-induced-forgetting effect, namely cued recall, category generation, and category verification. Clearly these are all tests of conceptual representations. The presence of reliable retrieval-induced-forgetting effects here would dovetail with Anderson and Spellman's (1995) feature-based inhibitory account. The tasks that do not show retrieval-induced forgetting, perceptual identification and word-stem completion, are less obviously conceptually based. As we discussed above, perceptual identification can be considered to tap the perceptual-representation system (Schacter, 1994), whereas word-stem completion, in the absence of a category cue, taps more into the lexical system (Downes et al., 1996). Thus it appears that the present data are consistent with an inhibitory account of retrieval-induced forgetting that is restricted to conceptual level representations.

It is also apparent that the forgetting effect is restricted to the items from the study list. Having studied *apple* and *orange*, then having practiced retrieval of *orange*, only *apple* is inhibited. Access to other fruits, such as *pear*, remains unaffected. This provides quite a theoretical challenge to feature-based accounts of the retrieval-induced-forgetting effect, because presumably pears share many of apples' unwanted features. Our reasoning for this point is as follows: According to Anderson and Spellman's (1995) feature-suppression model, the repeated retrieval of *orange* leads to the competition with *apple*. To prevent retrieval of *apple*, the features of *apple* that are not shared with *orange* are inhibited. Consequently it is harder to recall that the word *apple* was on the study list. However, why should only *apple* be inhibited by this process? Will not some of *apple*'s features themselves be shared with other fruits? For example, pears share many features with apples (e.g., fruit with edible skin, native to Britain, has brown pips). If these features are inhibited by the repeated retrieval of *orange*, then why should *pear* not be inhibited also? What the present data seem to indicate is that retrieval-induced forgetting impairs access to the episodic representation of a trace, rather than impairing access to the semantic features of a concept.

It is interesting to note the absence of any retrieval-induced-forgetting effect does not imply the absence of a retrieval-practice effect for the RP+ item. In all the studies in which retrieval-induced forgetting was absent, items that were retrieved during practice (RP+) later showed more priming on the implicit tasks than did the baseline (U) items. Presumably this reflects the fact that retrieval of an item strengthens its representation at many levels, whereas the competition with the RP− item occurs at only one level. It remains an unexplored issue as to whether different forms of output competition might not produce a different pattern of costs and benefits on implicit-memory tasks.

One final point is worth making. One of the potential advantages of the use of implicit-memory tests was that implicit memory can be tested without using a retrieval cue, thus providing a test of memory that is independent of the study cue. However, in neither of the tasks we used without a retrieval cue did we observe retrieval-induced forgetting. Thus, in the strictest sense our data cannot rule out an associative account of the retrieval-induced-forgetting effects reported here.

References

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory and language* (pp. 265–325). San Diego, CA: Academic Press.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7,* 522–530.

Anderson, M. C., Bjork, R. A., & Bjork E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 608–629.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80,* 1–46.

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition, 5,* 176–196.

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1997). Varieties of goal-directed forgetting. In J. M. Golding & C. M. Macleod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103–137). Mahwah, NJ: Erlbaum.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1314–1319.

Downes, J. J., Davis, E., De Mornay-Davies, P., Perfect, T. J., Wilson, K., Mayes, A. R. M., & Sagar, H. (1996). Stem completion priming in Alzheimer's Disease: The importance of target word articulation. *Neuropsychologia, 34,* 63–75.

Light, L. L., & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 531–541.

Macleod, C. M. (1997). Directed forgetting. In J. M. Golding & C. M. Macleod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing approach. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Schacter, D. L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 233–268). Cambridge, MA: MIT Press.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 10,* 381–391.

Received October 11, 2000
Revision received May 7, 2002
Accepted June 5, 2002 ■