# Episodic generation can cause semantic forgetting: Retrieval-induced forgetting of false memories

JEFFREY J. STARNS and JASON L. HICKS
*Louisiana State University, Baton Rouge, Louisiana*

In two experiments, we tested whether false recognition and false recall were prone to retrieval-induced forgetting, using the retrieval practice paradigm (Anderson, Bjork, & Bjork, 1994). Participants encoded lists of cue–target word pairs associated with a nonpresented, critical theme word and then engaged in retrieval practice for half of the word pairs from half of the lists. As expected, unpracticed targets from practiced lists were recognized (Experiment 1) and recalled (Experiment 2) less well than those from unpracticed lists. In addition, false recognition and false recall of critical items associated with practiced lists was lower than false recognition and false recall of items associated with unpracticed lists. We argue that false memories are prone to inhibitory mechanisms engendered by the retrieval practice paradigm. The results are consistent with the claim that semantically activated critical themes interfere with the episodic retrieval of list words and that inhibition decreases the activation level of these interfering memory representations during retrieval practice.

A common theme among recent studies using the Deese/Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995) is a comparison between the fates of true and false memories. Such comparisons have been made for the subjective experience that accompanies retrieval (e.g., Read, 1996; Roediger & McDermott, 1995), study–test delay (e.g., Toglia, Neuschatz, & Goodwin, 1999), study list repetition (e.g., Benjamin, 2001), level of processing (e.g., Read, 1996; Toglia et al., 1999), and repeated testing (e.g., McDermott, 1996), to name just a few. The focus of the present study concerns whether semantically induced false memories in the DRM paradigm are subject to retrieval-induced forgetting and whether retrieval inhibition affects true and false memories similarly. Questions of this sort have important implications both for the mechanisms underlying retrieval inhibition and for theoretical accounts of false memory in the DRM paradigm.

Several recent studies have adapted paradigms known to induce forgetting of true memories for use in false-memory research (Kimball & Bjork, 2002; Reysen & Nairne, 2002; Seamon, Luo, Shulman, Toner, & Caglar, 2002). Kimball and Bjork compared the fates of true and false memories in a directed-forgetting paradigm (see C. M. MacLeod, 1998, for a review) and in a part-set cuing paradigm (see Nickerson, 1984, for a review). Their directed-forgetting manipulation involved presenting people with two DRM lists, one after the other. People told to "forget" List 1 before studying List 2 were compared with a group told to continue remembering List 1. The instruction to forget decreased recall of List 1 items relative to the instruction to remember, replicating the standard directed-forgetting effect. However, false recall of critical theme words associated with List 1 increased following an instruction to forget. Thus, a directed-forgetting instruction decreased access to true memories but exacerbated false memories. Using a similar manipulation, Seamon et al. found that a directed-forgetting manipulation had no effect on false recall of critical List 1 items.

Kimball and Bjork's (2002) Experiment 2 involved a part-set cuing manipulation. People were given either 4 or 8 of 15 studied DRM items as recall cues in several part-set cuing conditions. The part-set cues consistently produced forgetting of noncued target items during free recall, relative to noncued conditions. Most importantly, the cuing manipulation also decreased false recall of critical items. Reysen and Nairne (2002) reported similar results in two part-set cuing experiments. Thus, across three experiments, part-set cuing interfered with recall of both true and false memories.

Kimball and Bjork (2002) attributed the opposite effects of directed forgetting and part-set cuing on false recall to retrieval impairment of an entire learning episode versus individual elements of an episode, respectively. In directed forgetting, the rise in false recall may occur because impaired episodic access to the context of the study list leaves the semantic retrieval of critical themes unchecked (al-

though Seamon et al., 2002, did not find a similar increase in false recall). In part-set cuing, however, the reduced accessibility of critical items may occur as the result of strength-dependent competition between recalled (or cued) and unrecalled items. Therefore, only in the case of part-set cuing does the impaired access to false memories seem to operate in a manner similar to that for true memories.

We conducted the present study to explore the fates of true and false memories in a different forgetting paradigm, the *retrieval practice* paradigm developed by Anderson and colleagues (e.g., Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995; see Levy & Anderson, 2002, for a brief review). In contrast to directed forgetting and part-set cuing, this paradigm was designed to isolate the cause of forgetting to specific inhibitory mechanisms active in a distinct retrieval practice phase occurring between study and test. In the original incarnation of the paradigm (Anderson et al., 1994), people learned several category–exemplar pairs (e.g., *food–cherry*) from multiple categories (e.g., *food*, *colors*, etc.). In a second phase, retrieval practice in the form of category-plus-stem cued recall occurred for a subset of items, from a subset of the categories (e.g., *food-ch_____*). For ease of exposition, we will label the subset of items from practiced categories that are individually practiced as P–P (practiced item, practiced category) and the remaining category members that are not individually practiced as U–P (unpracticed item, practiced category). Items from the entire categories that never appear in the practice phase are labeled U–U (unpracticed item, unpracticed category).[1] In a final phase, category label cues or category-plus-stem cues are provided to prompt retrieval of all studied items, whether previously practiced or not.

The standard finding in this paradigm is that P–P items are better recalled and U–P items are worse recalled in the final test phase, as compared with items from unpracticed categories (U–U items). The suppression hypothesis (Anderson & Neely, 1996) explains the impaired recall of U–P items as being caused by their inhibition during the intermediate retrieval practice phase. As P–P items receive retrieval practice, the activation of competing U–P items is suppressed to reduce interference with target-item retrieval, causing reduced U–P item accessibility on subsequent tests of memory (Anderson et al., 1994; Anderson & Spellman, 1995). Retrieval-induced forgetting has been found in a variety of contexts, including self-performed actions (Koutstaal, Schacter, Johnson, & Galluccio, 1999), eyewitness memory (Shaw, Bjork, & Handal, 1995), personality traits (M. D. MacLeod & Macrae, 2001), and geometric shapes (Ciranni & Shimamura, 1999).

We combined the retrieval practice and DRM methodologies in our study because, according to the activation/monitoring theory of false memory (Roediger, Balota, & Watson, 2001; Roediger, Watson, McDermott, & Gallo, 2001), the high levels of false recall and recognition in the DRM paradigm reflect spreading semantic activation (Collins & Loftus, 1975) from the studied associates (e.g., *blouse*, *sleeves*, *button*) to the critical, nonpresented item (e.g., *shirt*). This view implies that both critical items and studied items are activated during the study episode, which creates the possibility that retrieval practice will inhibit false remembering of critical themes. Although the term *inhibition* can and has been used loosely to describe a variety of phenomena (C. M. MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003), we will follow Anderson's suppression hypothesis by defining inhibition as a "process that deactivates the representation of an item in memory" (Anderson & Spellman, 1995, p. 70). As such, we sought to determine whether retrieval practice decreases the activation level of both U–P list words and critical themes that are conceptually activated by the lists. The DRM paradigm provides a novel method to test whether the source of a competitor's activation, directly perceived versus internally activated, differentially affects the item's potential inhibition.

On the basis of the suppression hypothesis, any activated items that interfere with target retrieval in the practice phase should have their activations inhibited (Anderson & Neely, 1996). Thus, critical themes activated in the study phase will be suppressed in the retrieval practice phase if they compete for retrieval with the items being practiced. Critical theme words will be associated with the conceptual cues given to guide target retrieval in the practice phase, and, because critical themes will be activated during the study phase, they will be associated with the same general context cues as will the list words (Hicks & Hancock, 2002). For these reasons, critical theme words should compete for retrieval during the practice phase, and the activation level of critical theme representations should be reduced by inhibitory mechanisms. This, in turn, should decrease false recall and false recognition because the extent to which critical themes are activated is a large influence on false memory rates (Roediger, Watson, et al., 2001).

The prediction that critical theme activation will be reduced during the practice phase rests on a key assumption regarding the unique inhibitory processes involved in retrieval. Solely on the basis of the activation/monitoring theory, one might predict that extra exposure to DRM list associates in the practice phase would provide more opportunities for activation to spread from list words to the critical theme words, resulting in *higher* activation levels. This logic assumes that *retrieving* a target item during retrieval practice activates associatively related concepts in the same way as does *encoding* a target item in a study list. However, studies using the retrieval practice paradigm have shown that worse recall of U–P items only occurs when retrieval is required during the practice phase; merely re-presenting a subset of items for further study is not sufficient to create forgetting of items not re-presented for study (e.g., Anderson, Bjork, & Bjork, 2000). Thus, retrieval practice uniquely involves processes that suppress interfering information and that focus activation on the target item. Therefore, we would expect DRM list associates to activate critical items when they are encoded in the study phase, but to inhibit the activation of critical items when they are retrieved in the practice phase.

Finally, the fuzzy-trace theory of false memory (Brainerd, Wright, Reyna, & Mojardin, 2001; Reyna & Brainerd, 1995) apparently predicts no effect of retrieval practice on

false memories. According to this theory, false memories in the DRM paradigm result from the substantial overlap in meaning between the critical item and a gist trace created during encoding of list associates. In other words, fuzzy-trace theory does not specifically predict that the critical item itself is activated during study of its associates; rather, the critical item is viewed on tests of recall or recognition as consistent with the gist created during the encoding of its associates. Even if the critical item is activated during study and inhibited during retrieval practice, those mechanisms should not affect the perceived overlap between the critical item and the gist trace when considered during retrieval. A recognition test (Experiment 1) provides the best test of this hypothesis, because specific target items are made explicitly available for a consideration of overlap with a gist trace.

## EXPERIMENT 1

To test these predictions, we had participants study multiple DRM lists in the form of paired associates and then undergo retrieval practice for half of the items from half of the lists. We used the strongest prime of each critical item as the cue word during the study and practice phases. For example, the strongest prime of the word *shirt* is *blouse*, and people therefore studied *blouse–sleeves*, *blouse–button*, *blouse–pants*, and so forth. When a DRM list received retrieval practice, the strongest associate and target fragments were used to cue retrieval of some of the studied targets (e.g., *blouse–s_e_v_s*). The critical item (e.g., *shirt*), of course, was neither studied nor practiced.

We used item recognition to test participants' responses to DRM list items, critical items, and other lures. Although few researchers have used recognition as the final test of memory in the retrieval practice paradigm (Anderson, De Kok, & Child, 1997; Hicks & Starns, 2004), we thought it appropriate here for important reasons. A recognition test allows control over the stimuli tested, and in the present case this meant that participants were forced to entertain a variety of lure items on the test, including critical items, weakly related lures, and unrelated lures. A recognition test also offers a "best-case" situation for participants to retrieve items from memory, given that a copy cue is presented for judgment (Hicks & Starns, 2004). Furthermore, a recognition test of individual target items in the absence of category cues can be considered one form of "independent" cue test (cf. Anderson & Spellman, 1995). This last point implies that the activation of target items is considered directly, without influence of competition arising from the use of a category or word cue. Thus, a recognition test served as the most conservative method to test our hypotheses.

### Method

**Participants**. Fifty-six Louisiana State University undergraduate students participated in exchange for extra credit in their psychology courses.

**Materials**. We constructed 16 associative lists using the University of South Florida word norms (Nelson, McEvoy, & Schreiber, 1998) by selecting the 12 words that were most likely to prime each critical item as a free associate. The strongest associate for each list theme served as the cue word during the encoding phase and subsequent practice phases. Eight other associates were the target words for each list, so that each list consisted of eight cue–target pairs (e.g., *blouse–sleeves* for the list theme *shirt*). The 4th, 7th, and 10th associates were not studied but were used as related lures on the recognition test. We computed the mean associative strength of the target words to the critical item from each list, ordered the lists in terms of this score, and divided them into two counterbalancing sets of eight lists each. There was a wide range of mean associative strength values, from the lowest score of .10 for the list theme *chemistry* to the highest score of .42 for the list theme *old*.

For the practice phase, we prepared word fragments by replacing some of the letters in the studied target words with underscores (e.g., *blouse–s_e_v_s* for *blouse–sleeves*). No fragments were used that could be correctly completed with other studied words. We constructed a single random order of word fragments to ensure that no more than two fragments from a given list theme appeared in succession. Four counterbalancing conditions varied which four of the eight studied lists received practice and which four of the eight target words in each of those lists received practice.

The recognition test included the 64 target words from the study phase, 24 weak lures (3 from each studied list theme), the critical item from each list, and 32 unrelated lures from the nonstudied lists. The test words were randomized in eight blocks of 16 items each. Each block contained a word from each of the studied lists, weak lures from each of the three lists, a critical item from one of the lists, and four unrelated lures. We arranged the blocks in such a way that if a practiced item from a given list was tested in one block, a unpracticed item from the same list was tested in the next block, and vice versa. Similarly, critical items and related lures from unpracticed and practiced lists were alternated between blocks.

**Procedure**. After the participants read and signed the consent form, they were told that they would see pairs of words and that they should learn these words by relating the target word to the cue word. The participants studied the 64 cue–target pairs from a given set in random order for 3 sec each and then solved multiplication problems for 5 min. Next, the participants completed a memory test in which fragments of the target words that they had studied were presented with underscores to indicate missing letters. They were also told that each fragment would appear with the appropriate cue word, and that they should solve the fragments with a target word that they remembered being associated with that cue in the study phase. A beep sounded as each fragment appeared, and each cue–fragment pair remained on the screen for 10 sec. The participants wrote their solutions to the fragments on separate pages in an answer booklet. When the participants could not retrieve a solution for a word fragment, they were told to move on to the next fragment that appeared on the screen. Three rounds of retrieval practice were given, and a unique random order of the practiced fragments was used for each round. After the third round, the participants solved multiplication problems for another 5 min.

For the final test phase, a label with an "O" written on it was placed on the "F" key, and an "N" label was placed on the "J" key. The participants were told that they would be presented with individual words and should press the "O" label if the word had appeared in *any* previous phase of the experiment or to press the "N" label if the word had not appeared previously in the experiment. The participants were instructed to make this decision with both speed and accuracy and were warned not to guess. After the participants made decisions on all 128 test words, they were debriefed and thanked for their participation.

### Results

In what follows, we present cued recall performance for the retrieval practice phase, the hit and false alarm rates for each item type and each condition of interest (P–P,

U–U, U–P), and recognition latencies for target items. For all inferential analyses, the probability of a Type I error was set at .05 and a Bonferroni correction was applied to all pairwise comparisons.

**Retrieval practice**. The retrieval practice phase produced high levels of cued recall. The proportions correct were .86 ($SE = .02$) for the first practice round, .87 ($SE = .01$) for the second round, and .88 ($SE = .01$) for the third round.

**Recognition accuracy**. The proportions of positive recognition responses to targets, critical item lures, and related (weak) lures are shown in Table 1 for each retrieval practice condition (P–P, U–U, U–P). These data demonstrate a consistent pattern of lower proportions to U–P items than to U–U items. A 2 (practice condition: U–U vs. U–P) $\times$ 3 (item type: target, critical theme lure, weak lure) repeated measures ANOVA revealed a significant effect of item type [$F(2,110) = 193.43$, $MS_e = .04$]. Pairwise comparisons demonstrated that targets ($M = .65$) and critical lures ($M = .70$) were recognized more often than weak lures ($M = .21$), with the former two item types not statistically different. There was also a significant effect of practice condition [$F(1,55) = 20.45$, $MS_e = .02$] and no significant interaction of this factor with item type [$F(2,110) = 1.20$]. The main effect of practice condition reflects higher recognition for U–U items ($M = .55$) than for U–P items ($M = .44$), and the lack of an interaction demonstrates that this relationship held for list words [$t(55) = 3.94$, $p < .001$], critical themes [$t(55) = 2.90$, $p < .005$], and related lures [$t(55) = 2.19$, $p < .05$]. P–P targets were clearly recognized better than either U–U or U–P targets, and false alarms to unrelated lures were quite rare ($M = .03$; $SE = .05$).

**Recognition latencies**. We also examined latencies to positive recognition responses for target items. Each participant's latency distribution, for each condition, was trimmed at two standard deviations, resulting in an overall trimming of 6% of the data. P–P items were recognized fastest ($M = 874$ msec; $SE = 21$), U–U items next fastest ($M = 1,097$ msec; $SE = 34$), and U–P items the slowest [($M = 1,193$ msec; $SE = 43$), $F(2,110) = 58.29$, $MS_e = 25,674.69$]. Pairwise comparisons declared all three practice conditions significantly different from one another.

## Discussion

We found significant U–P impairment in item recognition of list items, conceptually replicating previous reports (Anderson et al., 1997; Hicks & Starns, 2004). More importantly, we found a reduction in the level of false memories for list themes that received retrieval practice. Recognition probabilities were lower, and recognition latencies were slower, for all U–P item types related to the list theme, suggesting that these items were suppressed during retrieval practice.

However, an alternative account of our recognition results must be considered. It is also possible that participants used a more conservative recognition criterion (cf. Hicks & Marsh, 1998) for words related to practiced lists, and that this is why fewer unpracticed items consistent with prac-

ticed themes were recognized. Specifically, participants may have deliberately considered whether test items were associated with practiced lists before rendering a recognition decision, most likely when no other obvious evidence of prior experience (e.g., recollection) was available. Therefore, test items that seemed subjectively weak and that participants did not remember practicing might have been called new more often (e.g., "I don't remember seeing it in the practice phase, but I remember practicing other items with a particular, related cue word, so it must be new.").

However, such a criterion-shift interpretation seems unlikely upon careful consideration. First, the stimuli tested in this experiment were associative in nature and not necessarily categorical in nature. Many of the target words in the list themes are not easily identified as being associated with a particular cue word when not seen in the context of that cue word on the recognition test (e.g., the target *pocket* paired with the cue *blouse*). Thus, it would be difficult for a person to identify some test words, especially lures never seen at all, with a unique cue label. Second, we argue that this deliberate reasoning process, as applied selectively to items from practiced lists, should have taken more time than the average 90 msec beyond that required for U–U items.

Third, a criterion-shift interpretation implies that people assume *all* of the studied members of a practiced theme were practiced, whereas only half of the studied items from practiced lists were presented in the practice phase. Therefore, if participants truly required stronger evidence in memory for every test word consistent with practiced lists and were perfectly able to associate individual test words with a unique practiced list cue, they still must have ignored the fact that many of these items were studied but not encountered in the practice phase. In other words, participants must have ignored a critical source of studied items (U–P items) when using the more stringent criterion. In direct contrast to this possibility, Hicks and Starns (2004) showed that retrieval-induced forgetting occurred even when participants were given a source test in which "studied, not practiced" was one of the response options. Thus, worse recognition occurred for U–P items even when a lower criterion response was available. In light of these data, although a criterion-shift account cannot be em-

**Table 1**
**Experiment 1: Probability of Positive Recognition Responses to Targets, Critical Theme Lures, and Weak Lures as a Function of Retrieval Practice Condition**

| Condition | Targets | | Critical Lures | | Weak Lures | |
|---|---|---|---|---|---|---|
| | M | SE | M | SE | M | SE |
| P–P | .93 | .01 | – | – | – | – |
| U–U | .68 | .02 | .74 | .03 | .23 | .02 |
| U–P | .62 | .02 | .65 | .04 | .19 | .02 |

Note—P–P refers to items from practiced lists that were individually practiced; U–U refers to items from, or related to, unpracticed lists; U–P refers to items that were not individually practiced but were from, or related to, practiced lists.

pirically ruled out by the present experiment, we consider it unlikely that the pattern of data resulted from a selective criterion shift for practiced categories.

## EXPERIMENT 2

To further address the concern that the recognition results in Experiment 1 might have been influenced by a criterion shift for practiced lists, we conceptually replicated Experiment 1 using a final cued-recall test, which is used often in retrieval-induced forgetting research (e.g., Anderson et al., 1994; Smith & Hunt, 2000). If the consequence of retrieval practice is to actively suppress the activation of interfering U–P list words and critical themes, we would expect to find impairment for true and false memories in cued recall. If the consequence of retrieval practice was a more conservative recognition criterion in Experiment 1, rather than a reduction in memory strength, then we would expect no differences in the true and false recall of U–P items relative to U–U items in cued recall.

### Method

**Participants**. Fifty-eight Louisiana State University undergraduate students participated in exchange for extra credit in their psychology courses.

**Materials**. We used the study and practice materials from Experiment 1. For the final cued-recall test, we prepared test booklets with the cue word (e.g., *blouse*) for each studied list at the top of a different page. The cue words were presented in a random order, and we created a counterbalancing order by switching the last four and the first four cues.

**Procedure**. The procedure was exactly the same as in Experiment 1 until the final test phase began. At this point, the participants were informed that they would see the eight cue words that they had encountered in the first phase of the experiment, and to remember as many of the targets that were paired with each cue as possible. They were told to turn over the test booklet and begin recalling words for the first list cue when they heard the computer beep, and to turn to the next page in the booklet upon each subsequent beep. The computer was set to beep every 40 sec, which, in practice, was ample time for recall of the lists.

### Results

**Retrieval practice**. The proportions of correct recall were .86 ($SE = .02$) for the first practice round, .90 ($SE = .02$) for the second round, and .92 ($SE = .02$) for the third round.

**Cued recall**. Table 2 displays the probabilities of recall for target items and critical theme items. P–P target items clearly benefited from the practice phase; their recall was greater than the recall of U–U items. Of most interest, correct recall of U–P target items and false recall of U–P critical items were lower than the comparable recall levels for U–U items. A 2 (practice condition) × 2 (item type) ANOVA confirmed this pattern by revealing a significant main effect of practice condition [$F(1,57) = 30.05$, $MS_e = .02$]. There was no significant effect of item type [$F(1,57) = 2.45$], nor was there an interaction [$F(1,57) < 1.0$]. The lack of an interaction indicates that retrieval-induced forgetting affected both list targets [$t(57) = 7.75$, $p < .001$] and critical theme words [$t(57) = 3.02$, $p <$

### Table 2
**Experiment 2: Recall Probability of Targets and Critical Theme Items as a Function of Retrieval Practice Condition**

| Condition | Targets | | Critical Items | |
|---|---|---|---|---|
| | M | SE | M | SE |
| P–P | .65 | .02 | – | – |
| U–U | .28 | .02 | .33 | .03 |
| U–P | .17 | .01 | .23 | .03 |

Note—P–P refers to items from practiced lists that were individually practiced; U–U refers to items from, or related to, unpracticed lists; U–P refers to items that were not individually practiced but were from, or related to, practiced lists.

.005]. None of the related items (i.e., the 4th, 7th, and 10th lure associates from Experiment 1) were falsely recalled.

### Discussion

Consistent with the results of Experiment 1, retrieval-induced forgetting was found for both true and false memories in the form of cued recall. The effect on true memories is consistent with the growing literature on forgetting demonstrated with final cued recall (e.g., Anderson, Bjork, & Bjork, 1994, 2000). Importantly, the cued-recall results replicated our recognition findings in a test that was not sensitive to changes in response criterion, suggesting that, when viewed together, the results reflect a suppression mechanism.

As with Experiment 1, however, we must entertain an important alternative account for the cued-recall results. In this case, the alternative is output interference during recall. When cued to recall a practiced list, participants may have first produced the strengthened P–P items, which may have interfered with later attempts to access the remaining U–P items. Items from unpracticed lists would not suffer from this interference, because none of their category members were strengthened during the practice phase. To determine whether output interference might be a tenable explanation for our results, we performed an analysis comparing U–P impairment (U–U performance minus U–P performance) for participants who recalled P–P items relatively early or late in their recall sequences (M. D. MacLeod & Macrae, 2001). Following M. D. MacLeod and Macrae, we calculated the mean output position of P–P and U–P items for each participant, subtracted these two scores, and performed a median split on these values to divide the participants into an early P–P group and a late P–P group.[2] If output interference was responsible for the impaired U–P recall, then the U–P decrement should have been most pronounced for participants who recalled P–P items early in the recall sequence. For studied words, impairment for the early P–P group ($M = .07$) and the late P–P group ($M = .11$) did not differ [$t(50) = 1.60$, n.s.]. Critical theme impairment also did not differ for the early P–P group ($M = .12$) and the late P–P group [$M = .10$, $t(50) = .27$, n.s.]. Furthermore, we calculated U–P impairment scores only for the 11 participants who, on average, recalled U–P list words before P–P list words, and even these participants showed impairment for both

list words ($M = .12$) and critical themes ($M = .14$). On the basis of these analyses, output interference appears to have had little, if any, impact on the results, and it certainly cannot completely account for the observed retrieval-induced forgetting.

## GENERAL DISCUSSION

The selective retrieval of some items from DRM lists reduced recognition and recall of unpracticed items from those lists, replicating the retrieval-induced forgetting found with a variety of other stimulus materials (Levy & Anderson, 2002). More importantly, retrieval practice also reduced false recognition and false recall of critical, non-presented items. These results are consistent with the theoretical position that individual competitors to target items were inhibited during the retrieval practice phase in order to facilitate target recall (e.g., Anderson et al., 1994). Our findings also suggest that the source of a competitor's activation need not be episodic in nature; that is, semantically activated items interfere with cued recall and require suppression as well (see also Blaxton & Neely, 1983). The suppression of critical items following retrieval practice suggests a strong role for associative activation mechanisms in the creation, retrieval, and, in the present case, forgetting of false memories in the DRM paradigm (e.g., Roediger, Balota, & Watson, 2001).

The present results have implications for the effect of interitem similarity on retrieval-induced forgetting, although this issue was not a main focus of our study. Encouraging participants to integrate a set of items sharing a retrieval cue reduces or removes retrieval-induced forgetting (Anderson & McCulloch, 1999), and inhibition is similarly reduced when practiced and unpracticed items are members of the same subcategory (Bäuml & Hartinger, 2002). Bäuml and Kuhbander (2003) have argued that critical theme words from associative lists should not show retrieval inhibition, because these items bear strong associative relationships with the words on their lists. This prediction seems reasonable when one applies the feature-pattern suppression account proposed by Anderson and Spellman (1995) to associative-list stimuli. In the study phase, list words activate the semantic features of critical theme words, and these features define critical theme representations that might be "remembered" later in the experiment. In the practice phase, critical theme feature patterns compete for retrieval as participants attempt to recall P–P items, and all features unique to critical theme words must be inhibited. However, all of the features that critical themes share with the practiced items will actually be strengthened, and, if there are many overlapping features, this may offset the effect of inhibition and lead to little or no retrieval impairment for critical theme words. If one assumes that associated words share many features, there should be no retrieval-induced forgetting of critical theme words.

Our results demonstrate that retrieval-induced forgetting can occur for words that are highly associated to the practiced items (see also Bäuml & Kuhbander, 2003, Experiment 2). Although these results seem to contradict the feature-based account just presented, they are potentially consistent with the two-factor theory of similarity proposed by Anderson, Green, and McCulloch (2000). This theory identifies two types of similarity: target–competitor similarity and competitor–competitor similarity. Although target–competitor similarity reduces retrieval-induced forgetting for reasons mentioned previously, competitor–competitor similarity enhances forgetting. This occurs because the features shared among competitors are repeatedly inhibited as part of multiple interfering patterns, leading to more overall inhibition. Our critical theme stimuli were comparably associated to both the practiced and the unpracticed halves of their lists, and their association with the unpracticed items (i.e., competitor–competitor similarity) may explain the retrieval-induced forgetting we observed. One way to explore this possibility is to manipulate the associative strength of practiced and unpracticed list items to critical theme words. Perhaps little or no retrieval-induced forgetting of false memories will be observed when a critical theme's strongest associates receive retrieval practice (a condition favoring target–competitor similarity), whereas larger false memory inhibition effects will result when a critical theme's weakest associates are practiced (a condition favoring competitor–competitor similarity).

Regarding other retrieval inhibition paradigms, our results are consistent with studies of part-set cuing in the DRM paradigm (Kimball & Bjork, 2002; Reysen & Nairne, 2002) by demonstrating an effect of retrieval impairment on both true and false memories. Both the part-list cuing procedure and the retrieval practice procedure are thought to induce forgetting at the item level (Anderson & Neely, 1996), as opposed to the list-context forgetting induced in the directed-forgetting procedure. However, despite the general similarity between forgetting caused by part-set cuing and that caused by retrieval practice, the theoretical mechanisms are different.

One common mechanism implicated in part-set cuing, and relevant to the present study, is strength-dependent competition (e.g., Raaijmakers & Shiffrin, 1981; Rundus, 1973). In brief, according to this theory, the successful recall (or presentation as cues) of some items increases their strength. The probability of an item's coming to mind during recall will therefore be greater for strengthened items, as opposed to nonstrengthened items. And because people sample with replacement during recall, as strengthened items are repeatedly retrieved, fewer other items are retrieved, and people reach their stopping criterion before recalling many nonstrengthened items. Thus, it is not the case that part-set cues actually decrease the activation level of non-cued items. Instead, part-set cues simply decrease the likelihood that noncued items can compete (i.e., come to mind) with strengthened cued items during a fixed recall period.

Moreover, the strength-dependent mechanism ostensibly plays a role in recall but not in recognition (e.g., Ratcliff, Clark, & Shiffrin, 1990). Selectively strengthening some studied items does enhance their later recognition, but it

does not impair the recognition of nonstrengthened items (Ratcliff et al., 1990). This null strength effect occurs in recognition because nonstrengthened items that otherwise would not come to mind during recall are made directly available during a recognition test. Only a theory that posits the active suppression of U–P items (Anderson & Spellman, 1995) should show consistent results across recall and recognition tests (see Hicks & Starns, 2004, for a similar argument). Therefore, the present results are more consistent with a view that posits the inhibition (i.e., reduced activation) of unpracticed items from practiced lists.

Assuming that false memories in the DRM paradigm are generally caused by strong semantic activation of list themes (e.g., Roediger, Balota, & Watson, 2001), our results demonstrate that retrieval practice can have consequences for information that is associatively related to studied items but that was not itself studied. Others have shown that retrieval practice of studied items impairs episodic retrieval of U–P items (e.g., Anderson et al., 1994; Ciranni & Shimamura, 1999), and even that generation of nonstudied, semantically related items impairs episodic retrieval of studied U–P items (Bäuml, 2002). Our results extend such impairment to the false retrieval of semantically related, nonstudied items (see also Bäuml & Kuhbander, 2003).

In conclusion, our results demonstrate that false memories are subject to the inhibitory forgetting mechanisms that operate on true memories, at least as measured in the retrieval practice paradigm. This forgetting was evident in tests of item recognition and cued recall, attesting to the generality of the effect. Moreover, the forgetting of critical items associated with target recall during retrieval practice is evidence that factors outside of one's awareness contribute to the forgetting of true and false memories.

## REFERENCES

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, & Cognition*, **20**, 1063-1087.

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

ANDERSON, M. C., DE KOK, D., & CHILD, C. (1997, November). *Retrieval-induced forgetting on a test of recognition memory*. Poster presented at the annual meeting of the Psychonomic Society, Philadelphia.

ANDERSON, M. C., GREEN, C., & MCCULLOCH, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1141-1159.

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

ANDERSON, M. C., & NEELY, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 265-325). San Diego: Academic Press.

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.

BÄUML, K.-H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science*, **13**, 356-360.

BÄUML, K.-H., & HARTINGER, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory*, **10**, 215-224.

BÄUML, K.-H., & KUHBANDER, C. (2003). Retrieval-induced forgetting and part-list cuing in associatively structured lists. *Memory & Cognition*, **31**, 1188-1197.

BENJAMIN, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 941-947.

BLAXTON, T. A., & NEELY, J. H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory & Cognition*, **11**, 500-510.

BRAINERD, C. J., WRIGHT, R., REYNA, V. F., & MOJARDIN, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 307-327.

CIRANNI, M. A., & SHIMAMURA, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1403-1414.

COLLINS, A. M., & LOFTUS, E. F. (1975). A spreading–activation theory of semantic memory. *Psychological Review*, **82**, 407-428.

DEESE, J. (1959). On the prediction of occurrence of particular verb intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22.

HICKS, J. L., & HANCOCK, T. W. (2002). Backward associative strength determines source attributions given to false memories. *Psychonomic Bulletin & Review*, **9**, 807-815.

HICKS, J. L., & MARSH, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1105-1120.

HICKS, J. L., & STARNS, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review*, **11**, 125-130.

KIMBALL, D. R., & BJORK, R. A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General*, **131**, 116-130.

KOUTSTAAL, W., SCHACTER, D. L., JOHNSON, M. K., & GALLUCCIO, L. (1999). Facilitation and impairment of event memory produced by photograph review. *Memory & Cognition*, **27**, 478-493.

LEVY, B. J., & ANDERSON, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences*, **6**, 299-305.

MACLEOD, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1-57). Mahwah, NJ: Erlbaum.

MACLEOD, C. M., DODD, M. D., SHEARD, E. D., WILSON, D. E., & BIBI, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 163-214). San Diego: Academic Press.

MACLEOD, M. D., & MACRAE, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, **12**, 148-152.

MCDERMOTT, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, **35**, 212-230.

NELSON, D. L., MCEVOY, C. L., & SCHREIBER, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Available at http://www.usf.edu/FreeAssociation/.

NICKERSON, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition*, **12**, 531-552.

RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93-134.

RATCLIFF, R., CLARK, S. E., & SHIFFRIN, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 163-178.

READ, J. D. (1996). From passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, **3**, 105-111.

REYNA, V. F., & BRAINERD, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning & Individual Differences*, **7**, 1-75.

REYSEN, M. B., & NAIRNE, J. S. (2002). Part-set cuing of false memories. *Psychonomic Bulletin & Review*, **9**, 389-393.

ROEDIGER, H. L., III, BALOTA, D. A., & WATSON, J. M. (2001). Spreading activation and the arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). Washington, DC: American Psychological Association.

ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.

ROEDIGER, H. L., III, WATSON, J. M., MCDERMOTT, K. B., & GALLO, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, **8**, 385-407.

RUNDUS, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, **12**, 43-50.

SEAMON, J. G., LUO, C. R., SHULMAN, E. P., TONER, S. K., & CAGLAR, S. (2002). False memories are hard to inhibit: Differential effects of directed forgetting on accurate and false recall in the DRM procedure. *Memory*, **10**, 225-237.

SHAW, J. S., BJORK, R. A., & HANDAL, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review*, **2**, 249-253.

SMITH, R. E., & HUNT, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, **28**, 503-508.

TOGLIA, M. P., NEUSCHATZ, J. S., & GOODWIN, K. A. (1999). Recall accuracy and false memories: When more is less. *Memory*, **7**, 233-256.

## NOTES

1. We have used the labels P–P, U–P, and U–U instead of the more traditional labels for item-based conditions in retrieval-induced forgetting work (Rp+ for practiced items, Rp− for unpracticed items from practiced categories, and Nrp for categories not practiced, respectively) for an important reason. The labels used herein specify in a more neutral manner how the conditions are literally defined by the experimental procedure, whereas the more traditional labels could be taken to imply a particular pattern of performance (i.e., greater than baseline for practiced items, less than baseline for unpracticed items from practiced categories).

2. The 5 participants who failed to recall any U–P items were removed from this analysis, and, of the remaining participants, only data from the 26 participants with the lowest and the 26 participants with the highest difference between P–P and U–P output positions were analyzed.