

Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition

JOHN G. SEAMON, CHUN R. LUO, JONATHAN J. KOPECKY, CATHERINE A. PRICE,
LEEATT ROTHSCHILD, NICHOLAS S. FUNG, and MICHAEL A. SCHWARTZ
Wesleyan University, Middletown, Connecticut

What is the effect of retention interval on accurate and false recollection in the Deese, Roediger, and McDermott (DRM) procedure? Previous research has suggested that false recall is more persistent than accurate recall, but the recognition results have been inconsistent. In two parametric studies, we tested recall and recognition for the same DRM lists, over retention intervals that ranged from no delay to a 2-month delay. We found that accurate and false memory were diminished by increases in retention interval, false memory persistence was present for recall and recognition, greater persistence for false memory than for accurate memory was more readily observed for recall than recognition, and the high-threshold (P_r), signal detection (d'), and nonparametric (A') recognition measures differed in their sensitivity for detecting change. The effect of retention interval on accurate and false memory is consistent with expectations from fuzzy trace theory. In the DRM procedure, truth is not more memorable than fiction.

It is ironic that Deese's (1959) study of intrusion errors in free recall has generated more empirical research some 4 decades after it was published than when it first appeared. In noting the influence of the *Zeitgeist* in science, Bruce and Winograd (1998) suggested that Deese's paper became important to memory researchers only when they stopped focusing on correct recollection and actively investigated false memory. Bruce and Winograd are undoubtedly correct, but we also suspect that part of the reason for the resurgence of interest in this particular paper is the fact that Deese's procedure so readily produces false memories.

Specifically, Deese (1959) presented participants with lists of semantic associates that converged on nonpresented words. For example, for the list consisting of *thread, pin, eye, sewing, sharp, point, pricked, thimble, haystack, pain, hurt, and injection*, the converging associate was *needle*. When he presented participants with lists of this type for free recall, Deese found that the participants often erroneously recalled the nonstudied converging associates. Roediger and McDermott (1995) modified and extended this procedure. In one study, these researchers presented participants with lists of semantic associates that produced a high frequency of false recall, followed by a recognition test composed of studied words, related non-

studied words (the converging associates, hereafter referred to as *critical words*), and unrelated nonstudied words. For any recognized word, the participants had to make a remember/know judgment by stating whether they specifically remembered the word's occurrence at study or merely knew that the word had been presented in the absence of any specific recollection. Not only did Roediger and McDermott find that the critical words were falsely recognized as frequently as the list words, they reported that these false recognitions were often characterized as remembered in the remember/know judgment task. Thus, the participants recognized words that were never presented, and they indicated that they specifically remembered some aspect of their prior occurrence at study.

Since the publication of Roediger and McDermott's (1995) paper, a large number of studies involving the Deese, Roediger, and McDermott (DRM) procedure have been conducted (e.g., Hicks & Marsh, 1999; Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996; Seamon, Luo, & Gallo, 1998; Tussing & Greene, 1999). Roediger, McDermott, and Robinson (1998) organized much of this research in a chapter on the role of associative processes in false memory. Included in that chapter was a review of various study and test factors that influence false recollection in the DRM procedure. One of those factors is retention interval. Memory researchers have manipulated this variable since Ebbinghaus (1885/1913) first found that nonsense syllables were forgotten rapidly immediately after study and then showed a more gradual decline over time. Obviously, there are numerous differences between Ebbinghaus's procedure and the DRM procedure,

This research was presented as a poster at the 41st Annual Meeting of the Psychonomic Society, November 2000 in New Orleans. Appreciation is expressed to Charles Brainerd, Kathleen McDermott, Suparna Rajaram, and Mieke Verfaellie for helpful comments. Correspondence should be addressed to J. G. Seamon, Department of Psychology, Wesleyan University, Middletown, CT 06459-0408 (e-mail: jseamon@wesleyan.edu).

including the use of lists of unrelated nonsense syllables by Ebbinghaus and lists of semantic associates in the DRM procedure. But for our present understanding of false memory, it is important to know whether the effect of retention interval is the same on accurate and false recollection.

Brainerd, Reyna, and Brandse (1995) were the first to examine the issue of accurate and false memory persistence, defined as memory stability over time, and they noted its theoretical and practical importance. Theoretically, it is important to determine which characteristics are shared by accurate and false memory. Practically, it is important in forensic interviews, where decisions on how to interpret consistent or inconsistent recollections over time can depend on jurors' beliefs about whether accurate memory is more stable than false memory. To date, relatively few studies have addressed this issue of the stability of accurate and false memory over time, and their results have not always been consistent. A review of this research suggests that greater consistency has been observed when memory was tested by recall than when it was tested by recognition.

Retention Interval and Accurate and False Recall

Three DRM studies have consistently demonstrated that false memory can be more persistent than accurate memory over various retention intervals. First, McDermott (1996, Experiment 1) presented participants with 24 auditory lists of semantic associates. Each list was followed by one of three different activities: an immediate free recall test followed by 30 sec of math problems, 30 sec of math problems followed by an immediate recall test, or 30 sec of math problems with no immediate recall test. Two days later, all the participants were given a final free recall task in which they were asked to recall all of the words that had been presented at study. On the immediate test of each list, McDermott found that the proportion of accurate recall exceeded false recall; on the 30-sec delayed test of each list, the proportions of accurate and false recall did not differ, since the short delay reduced accurate recall but had no effect on false recall. Results from the final free recall test showed that the proportions of accurate and false recall were lower than those from the initial test, yet false recall exceeded accurate recall, both for lists tested previously and for those tested for the first time. McDermott cautioned against making direct comparisons between the immediate and the delayed test sessions, because the recall tests, she said, differed not only in delay, but also in the fact that the initial tests involved single lists, whereas the final test involved all lists.

Second, Toglia, Neuschatz, and Goodwin (1999, Experiment 2) conducted a parametric study of the effect of retention interval on accurate and false recall in the DRM procedure. In this study, participants were presented with five auditory lists of associatively related word lists in either a blocked or a random format. Following study, the participants were given a single free recall test immediately, 1 week, or 3 weeks later. Toglia et al. found that block-

ing associatively related words at study led to greater recall of list words and critical words at all retention intervals than did the random format. More important, however, were the findings that recall probability was greater for critical words than for list words at each retention interval and that list word recall decreased over retention intervals, whereas critical word recall did not. In fact, the false recall of critical words was as high after a delay of 3 weeks as it was on the immediate test.

Finally, Thapar and McDermott (2001, Experiment 1) manipulated retention interval and level of processing in an incidental learning study employing DRM lists. Participants made pleasantness, color, or vowel decisions about visually presented study words from 24 lists, followed by a surprise free recall test, either immediately after study or after delays of 2 or 7 days. These researchers observed an interaction of word type and retention interval, since list words were recalled better than critical words on the immediate test but poorer than critical words after 1 week. Thapar and McDermott concluded that whereas list words and critical words both demonstrated forgetting, false recall was less affected by retention interval than was accurate recall. These researchers reached the same conclusion when they statistically equated the initial probability of recall for all conditions to account for potential differences in initial learning of list words and critical words. Although the findings of Toglia et al. and Thapar and McDermott differed with respect to the effect of retention interval on false recall, the three DRM recall studies were alike in demonstrating that retention interval could differentially affect accurate and false recall of DRM lists. False memories can be more persistent than accurate memories, as was shown by a greater decline in accurate recall than in false recall between immediate and delayed tests (McDermott, 1996; Thapar & McDermott, 2001; Toglia et al., 1999).

Retention Interval and Accurate and False Recognition

Less consistent are the findings from research on retention intervals and accurate and false recognition. Some recognition studies have mimicked the recall results, whereas others have not. For example, Payne et al. (1996, Experiment 1) presented participants with 16 auditory DRM lists, followed by a recall test or math problems to solve immediately after each list was presented. After all of the study lists were presented, the participants received a recognition memory test, consisting of list words, critical words, and unrelated words, either immediately or after a 1-day delay. Payne et al. found that false recognition of critical words was higher than accurate recognition of list words over all conditions and that a 1-day delay decreased accurate recognition more than false recognition. These results are consistent with the recall findings that showed greater persistence of false memory than of accurate memory.

Thapar and McDermott (2001, Experiment 2) observed an interaction of retention interval and word type, in their

previously described study of levels of processing with DRM lists, when they tested memory by recognition. Recognition scores decreased faster over retention interval for list words than for critical words, since list words, especially in the deep encoding condition, were recognized better than critical words on the immediate test but differed little from critical words on the 2-day and 7-day tests. Thus, both Payne et al. (1996) and Thapar and McDermott demonstrated a greater decline in accurate recognition than in false recognition over retention intervals. In addition, Payne et al. demonstrated that false memory exceeded accurate memory on a delayed test; Thapar and McDermott did not.

Still other researchers have not observed greater false memory than accurate memory persistence in the DRM procedure. For example, following the intentional learning of auditorily presented DRM lists, both Lampinen and Schwartz (2000) and Neuschatz, Payne, Lampinen, and Toglia (2001) gave participants a recognition memory test either immediately after study or 2 days later. In both studies, recognition scores decreased over retention intervals for list words and critical words in a similar manner, and false memory did not exceed accurate memory after a 2-day delay. Brainerd, Wright, Reyna, and Mojardin (2001, Experiment 2) also presented participants with auditory DRM lists, followed by a recall test or math problems to solve immediately after each list was presented. After study was completed, these participants received two recognition memory tests, one immediately after study and another after a 1-week delay. The second test included previously tested list words and critical words, as well as list words and critical words tested for the first time. In the conditions most comparable to the preceding studies, where the participants did not recall each list previously and words were tested only once by recognition, either immediately or after a 1-week delay, Brainerd et al. (2001) found more false recognition of critical words than accurate recognition of list words, but recognition for both decreased at a comparable rate over the delay.

Thus, unlike the DRM recall studies, in which greater persistence of false memory than of accurate memory has been consistently observed (McDermott, 1996; Thapar & McDermott, 2001; Toglia et al., 1999), DRM recognition studies have yielded different outcomes. Some studies found greater declines in accurate recognition than in false recognition over retention intervals (Payne et al., 1996; Thapar & McDermott, 2001), whereas others reported no difference (Brainerd et al., 2001; Lampinen & Schwartz, 2000; Neuschatz et al., 2001).

Several factors, however, urge caution about this summary of the recognition results. First, the inconsistent outcomes are based on a relatively small number of studies that manipulated retention interval in the DRM procedure, and many of those studies employed short delays of 1 or 2 days. Second, several additional studies, employing different stimulus materials and procedures, have indicated that a different effect of retention interval on false recog-

niton is possible. For example, Brainerd et al. (1995) presented children with an auditory list of words, followed by an immediate auditory recognition test and a repetition of that test 1 week later. In this within-subjects design, persistence could be observed following repeated testing. Brainerd et al. (1995) found that children showed persistence in their hits and false alarms from the first test to the second and that false alarms were sometimes more persistent than were hits. Brainerd and Mojardin (1998) reported similar results with short narratives instead of words. Seamon, Luo, Schlegel, Greene, and Goldenberg (2000), using a between-subjects design, presented children and adults with categorized pictures to study, followed by a recognition test, either immediately after study or after a 3-day delay. Seamon et al. (2000) found that accurate recognition decreased over the delay, whereas false recognition for category-related but nonstudied pictures actually increased for both age groups. An increase in false recognition after a delay has also been observed in the retention of prose material over 1 week (Sulin & Dooling, 1974) and of autobiographical events over 1–3 months (Barclay & Wellman, 1986). Thus, whereas only two studies have found greater persistence for false recognition than for accurate recognition in the DRM procedure, evidence for greater persistence of false memory, and even an increase in false memory over time, is available from other recognition tasks.

Current Research Goal

It is presently difficult to provide a cogent summary of the effect of retention interval on false memory, even when limiting the summary to only those studies that used DRM lists, because those various studies have differed in numerous ways. Those differences include, but are not limited to, the treatment of retention interval as a between- or a within-subjects variable, the number and length of retention intervals employed, the number and length of DRM lists studied, the list presentation modality at study, whether the participants studied the lists intentionally or incidentally, whether the initial levels of learning for list words and critical words were equated, and for studies involving recognition, whether the recognition scores were expressed as a measure of sensitivity and, if so, which measure was employed. Thus, a great many factors could potentially contribute to the observed results when retention interval is manipulated in the DRM procedure. To tease apart the possible effect of each of these variables is not possible in a single study. But, we can take advantage of the existing research and provide a fresh look at the effect of retention interval on recall and recognition in this procedure. From Toglia et al. (1999), retention interval can be manipulated parametrically as a between-subjects variable and can include lengthy delays. From Thapar and McDermott (2001), the initial levels of learning of list words and critical words can be equated to observe any changes in retention over delay. And because recognition sensitivity can vary with the measure employed (Snod-

grass & Corwin, 1988), different recognition measures can be compared, to determine which one provides the most sensitive measure of retention.

Accordingly, our goal was to provide a comprehensive examination of accurate and false memory in the DRM procedure by (1) measuring recall and recognition with the same materials, (2) parametrically varying retention intervals over a wide range in a between-subjects design, to avoid possible contamination effects of prior testing, (3) equating the initial learning of list words and critical words, and (4) examining recognition scores with a variety of sensitivity measures.

To provide a theoretical framework for this research, we employed Brainerd and Reyna's fuzzy trace theory (e.g., Brainerd et al., 1995; Brainerd, Reyna, & Poole, 2000; Reyna & Lloyd, 1997). According to this view, memory judgments can be based on verbatim or gist traces. Verbatim traces represent the surface details of physical stimuli, whereas gist traces represent the meaning or theme of the stimuli. In the DRM procedure, correct recall or recognition of list words is hypothesized to be based largely on verbatim traces, whereas false recall or recognition of critical words is assumed to be based mostly on gist traces. A gist representation corresponds to the theme of a list represented by the converging associate (e.g., *needle* for the list consisting of *thread, pin, eye, sewing*, and so forth).

Fuzzy trace theory holds that verbatim and gist traces can differ in durability, with gist traces more durable than verbatim traces, perhaps owing to the greater vulnerability of verbatim than of gist traces to interference effects over time (Brainerd et al., 1995; Reyna & Lloyd, 1997). This difference in durability leads to predictions about the effect of retention interval on accurate and false memory in the present experiments. Specifically, when memory is tested over a broad range of retention intervals, false memory can be more persistent than accurate memory, if false memory is based on gist representations that are more durable than verbatim representations. Thus, false memory should show less of an effect of retention interval than does accurate memory for the same stimulus materials. In fact, because of hypothesized differences in trace durability, differences in accurate and false memory may be more pronounced after longer delays than after shorter delays.¹ However, durability for both verbatim and gist traces should eventually diminish, so that a very long retention interval will yield little or no accurate or false memory, regardless of how memory is tested.

EXPERIMENT 1

In this experiment, we examined accurate and false recall in the DRM procedure over retention intervals involving no delay, a 2-week delay, or a 2-month delay. If false recall is more persistent than accurate recall, as is suggested by both fuzzy trace theory and the results of previous research (McDermott, 1996; Thapar & McDermott, 2001; Toglia et al., 1999), we should find a greater decrease in the recall of list words than of critical words over these retention intervals. In fact, following Toglia et al.,

who reported no effect of retention interval on critical word recall for delays of 1 or 3 weeks, list word recall should decrease over each retention interval, whereas critical word recall may be unchanged, at least for the 2-week delay. We included a lengthy 2-month delay condition in order to observe whether false recall would diminish over an extended delay, whereas we omitted a short delay condition of 1 or 2 days because Toglia et al. found no effect on false recall for delays of up to 3 weeks.

Method

Participants. The participants were 60 Wesleyan University undergraduates who received introductory psychology credit or served voluntarily as part of a course experiment. None had participated in any related memory research.

Materials. We used the same 16 word lists that Gallo, Roberts, and Seamon (1997) used. These lists were selected from those provided by Roediger and McDermott (1995) in their appendix. Each list was composed of 15 words that were all converging associates of a nonstudied critical word. The strongest associates to the critical word normally occurred first in each list, and the order of the words was constant. To help ensure that our results were not dependent on the particular lists sampled, the 16 lists were divided into two sets of 8 lists, labeled A and B. Within each delay condition, half of the participants had previously studied Set A, and half had studied Set B.

Procedure. During study, the participants listened to an auditory tape containing eight lists of 15 words presented in blocked fashion. The words were spoken by a male voice at a rate of 1.5 sec per word, with a tone separating each list. The participants, who were informed that a recall test would follow the study words, were tested in small groups under intentional learning conditions. No deception was employed, since all the participants were instructed to try to remember as many words as possible for the subsequent recall test. In addition, they were told that the memory test would follow the study words either immediately or after a delay that could involve weeks, but they were not informed about the associative basis of the word lists or the false memory effect. As will be seen in the recall results, the participants in the delayed conditions showed no evidence of rehearsing the study stimuli during the delays, but this possibility could not be eliminated in an intentional learning design. Except for the delays between study and test, this procedure was generally similar to that used by other DRM researchers (e.g., Gallo et al., 1997, uninformed condition; Payne et al., 1996, Experiment 1; Roediger & McDermott, 1995, Experiment 2).

After the study words were presented, the participants were informed about their specific time of testing. One group of 20 participants was tested immediately, a second group of 20 participants was tested after a 2-week delay, and a third group of 20 participants was tested after a 2-month delay. At the time of test, the participants were given a recall sheet and asked to free recall as many words as they could remember from the previously heard auditory tape. No time limit was placed on the free recall. After the test was completed, the participants were debriefed and were asked to refrain from discussing the experiment with other students until the project was completed.

Results

The results were analyzed for correct recall of list words, false recall of critical words, and false recall of noncritical words (intrusions). Stimulus study set (List Sets A vs. B) did not systematically affect performance and will not be considered.

Mean recall proportions. The top two rows of Table 1 show the mean recall proportions for list words and critical words at each retention interval. Two principal obser-

Table 1
Experiment 1: Unadjusted and Adjusted Mean Recall
Proportions of List Words and Critical Words

Item Type	Retention Interval		
	Immediate	2 Weeks	2 Months
Unadjusted Means			
List words	.17	.07	.04
Critical words	.28	.27	.12
Adjusted Means			
List words	1.00	.39	.21
Critical words	1.00	.96	.42

Note—The adjusted means for the 2-week and 2-month intervals are calculated as a proportion of the immediate recall test for list words and critical words.

vations can be made. First, the mean recall proportions were greater for critical words than for list words at each retention interval. Second, retention interval influenced accurate and false recall differently. The free recall of list words decreased dramatically from the immediate to the 2-week delay test, followed by an additional, but smaller, decrease for the 2-month test, as performance approached the floor. The free recall of critical words, however, showed little effect of retention interval from the immediate to the 2-week delay test, since false recall showed a decline only for the 2-month delay condition.

The results from analyses of variance that included retention interval (immediate vs. 2-week vs. 2-month delay) and stimulus list set (A vs. B) as between-subjects variables and word type (list words vs. critical words) as a within-subjects variable confirmed these observations.² Analyses showed that the mean proportion recalled was greater for critical words (.22) than for list words [.09; $F(1,54) = 54.81, MS_e = 0.22, p < .0001$] and that this difference was reliable at each retention interval (all $ps < .005$). Moreover, there was an effect of retention interval [$F(2,54) = 11.23, MS_e = 0.51, p < .0001$] and an interaction of retention interval and word type [$F(2,54) = 4.41, MS_e = 0.04, p < .02$]. Although retention interval influenced both list words [$F(2,54) = 53.75, MS_e = 0.10, p < .0001$] and critical words [$F(2,54) = 5.84, MS_e = 0.16, p < .01$], the pattern of results was not the same. For list words, Newman-Keuls comparisons indicated that accurate recall systematically decreased from the immediate to the 2-week test and from the 2-week to the 2-month test (both $ps < .05$). For critical words, the same comparisons showed no decrease in false recall from the immediate to the 2-week test ($p > .05$) but less false recall for the 2-month test than either the immediate or 2-week test (both $ps < .05$). Thus, Table 1 shows greater persistence for false recall than for accurate recall over the retention intervals.

One potential problem with the present results is that the initial levels of recall for list words and critical words were not the same. As was noted by Thapar and McDermott (2001), some researchers, such as Loftus (1985), have suggested that a linear scale for the probability of recall cannot be assumed if two functions (in this case, accurate recall and false recall) begin at different points on

the scale (i.e., the immediate test condition). According to Thapar and McDermott, a linear scale assumes that reductions of equivalent probabilities represent reductions of equivalent magnitude, whereas reductions of different probabilities represent reductions of different magnitudes. They addressed this issue by statistically equating the initial probability of recall by setting it to 1 for each immediate recall condition and calculating the probabilities of recall for list words and critical words for each delayed test as a proportion of recall from the corresponding immediate test. We followed the same procedure on an individual participant basis and present the adjusted mean recall proportions for list words and critical words in the bottom two rows of Table 1.

Adjusted recall proportions. Analyses of the adjusted scores of all the participants, excluding the immediate test scores that were the same for all the participants, revealed the same pattern of results as that for the unadjusted scores. As is shown in Table 1, a 2-week delay produced a major drop in accurate recall but had virtually no effect on false recall, whereas a 2-month delay produced only a modest additional decrement in accurate recall but a substantial drop on false recall. These results were supported by an analysis of variance that indicated that the adjusted recall scores were greater for critical words (.69) than for list words [.30; $F(1,36) = 40.69, MS_e = 3.14, p < .0001$] and for the 2-week (.67) than for the 2-month delay [.32; $F(1,36) = 9.62, MS_e = 2.56, p < .005$]. Most important, there was an interaction of retention interval and word type [$F(1,36) = 8.22, MS_e = 0.64, p < .01$]. Critical word recall exceeded list word recall at both delays (both $ps < .05$), but only critical word recall was reliably reduced by extending the retention interval from 2 weeks to 2 months ($p < .001$).

Free recall intrusions. Finally, an analysis of the mean frequency of noncritical word intrusions in the recall for each condition revealed a marginal effect of retention interval on intrusion frequency [$F(2,54) = 2.92, MS_e = 131.15, p < .07$]. The mean frequency of noncritical intrusions tended to be higher in the 2-month test ($M = 7.00$) than in the immediate test ($M = 2.75$) or the 2-week test ($M = 2.40$) conditions. If the rate of noncritical intrusions is collapsed over the immediate and 2-week conditions to compare the present rate with Toggia et al.'s (1999) rate over comparable study and test conditions (their immediate, 1-week, and 3-week delay conditions following blocked list presentation), the rates for the present experiment ($M = 2.58$) and Toggia et al.'s ($M = 2.04$) are similar. Our finding that noncritical intrusions tended to increase after a 2-month delay suggests that the participants in this condition tried hard to recall the previously studied lists but that their memory was severely impaired.

Because this experiment was procedurally similar to that of Toggia et al. (1999), it is appropriate to compare our findings with theirs. Both experiments presented participants with auditory DRM lists, followed by free recall at different retention intervals. However, in the condition most similar to the present experiment, Toggia et al. presented participants with five lists of 12 words (60 studied

words) in a blocked format, whereas this experiment presented participants with eight lists of 15 words (120 studied words) in the same format. Although Toggia et al.'s mean recall proportions for comparable retention intervals were approximately double those of the present experiment (see Toggia et al., Table 1, p. 245), Toggia et al.'s participants studied half the number of words that the participants in our experiment studied. If the absolute levels of accurate and false recall are calculated from the mean proportions of list words and critical words recalled (their Table 1 and our Table 1), these levels are highly similar across experiments. Thus, whether the comparison is by mean recall proportions or absolute levels of recall, the pattern of results across both experiments is the same.

In summary, all DRM recall studies have consistently shown that false recall is more persistent than accurate recall over retention intervals (McDermott, 1996; Thapar & McDermott, 2001; Toggia et al., 1999; the present experiment). Moreover, both Toggia et al. and the present experiment showed that accurate recall decreased, whereas false recall did not, when the retention interval was measured in weeks. The present research extended that finding by showing that false recall was reduced by a lengthy 2-month delay. Together, these results show that whereas false recall is more persistent than accurate recall, both forms of memory are ultimately subject to the same effects of interference and/or decay that occur over time.

EXPERIMENT 2

In this experiment, we presented participants with the same word lists used previously, but we tested memory by recognition instead of recall, over retention intervals involving no delay, a 2-day delay, a 2-week delay, or a 2-month delay. We added a 2-day delay condition to this experiment because many previous DRM recognition studies have employed a short delay of 1 or 2 days (Lampinen & Schwartz, 2000; Neuschatz et al., 2001; Payne et al., 1996; Thapar & McDermott, 2001). We examined the correct and incorrect recognition proportions and measures of memory sensitivity for list words and critical words at each retention interval. On the basis of predictions from fuzzy trace theory, where gist representations are hypothesized to be more durable than verbatim representations over time, and previous research that has demonstrated greater persistence for false recognition than for accurate recognition (Payne et al., 1996; Thapar & McDermott, 2001), we should observe a greater decrease in the recognition of list words than in that of critical words over the retention intervals. However, because this effect has not been consistently observed (Brainerd et al., 2001; Lampinen & Schwartz, 2000; Neuschatz et al., 2001), the hypothesized difference in persistence for accurate and false recognition might be a more modest or subtle effect than that observed with recall.

Method

Participants. The participants were 120 Wesleyan University undergraduates who received introductory psychology credit or served

voluntarily as part of a course experiment. None had participated in any related memory research.

Procedure. The word lists and study conditions were identical to those from the first experiment, except that the participants in this experiment were instructed to remember as many words as possible for a subsequent recognition test that followed the study words, either immediately or after a delay that involved days or weeks.

After all of the 120 study words were heard, the participants were informed about their specific time of testing. One group of 30 participants was tested immediately after the study words were presented, a second group of 30 participants was tested after a 2-day delay, a third group of 30 participants was tested after a 2-week delay, and a fourth group of 30 participants was tested after a 2-month delay. As compared with the first experiment, we increased the sample size for each delay condition by 50% in this experiment, to provide additional statistical power for detecting any potential main effects or interactions. Half of the participants in each retention interval condition studied List Set A; the other half studied List Set B. The stimulus set that was not used during study provided distractors for the recognition test.

The visual recognition memory test consisted of 64 words. There were 3 words from each studied list (serial positions 1, 8, and 10), the nonstudied critical word from each studied list, 3 words from each nonstudied list (serial positions 1, 8, and 10), and the critical word from each nonstudied list, all presented sequentially on a test sheet in a random order. Each word was accompanied by the words *yes* and *no* for a yes/no recognition decision and by the letters *R* and *K* for a remember/know judgment. The participants were instructed to examine the words in sequential order and to make a recognition decision for each word on the list. They were told to circle *yes* for any word that they recognized from the study tape or *no* for any word they failed to recognize. In addition, the participants were instructed to make a *remember* or a *know* judgment for each word they recognized by circling either *R* or *K* on the answer sheet. Similar to previous research (e.g., Rajaram, 1993; Roediger & McDermott, 1995), the participants were instructed to circle *R* if they consciously recalled the word from the study lists, such as the way the word was spoken or what they were thinking about at the time. Alternatively, they were told to circle *K* if they were sure that the word was presented but could not remember its specific occurrence or any related details from the study portion of the experiment. After the test was completed, the participants were debriefed and were asked not to discuss the experiment until it was completed.

Results

The results were analyzed for accurate and false recognition with three different memory sensitivity measures. The mean recognition proportions, along with the remember/know judgments, will be described before comparing the different sensitivity measures. As in the first experiment, stimulus study set (List Set A vs. B) did not systematically affect performance and will not be considered in any analysis.

Mean recognition proportions. Table 2 presents the mean recognition proportions for list words and critical words and remember/know judgments for each retention interval. It shows that the mean recognition proportions were greater for studied than for nonstudied words and for related than for unrelated critical words at all retention intervals. These results demonstrate accurate memory for list words and false memory for related critical words. Table 2 also shows the effect of retention interval on recognition. Retention interval strongly influenced both accurate and false recognition as the hit rate for studied

Table 2
Experiment 2: Mean Recognition Proportions for Studied and Nonstudied Words
and Related and Unrelated Critical Words

Item Type	Retention Interval							
	Immediate		2 Days		2 Weeks		2 Months	
	Mean	R/K	Mean	R/K	Mean	R/K	Mean	R/K
List words								
Studied	.67	.39/.28	.59	.27/.32	.52	.23/.29	.40	.21/.18
Nonstudied	.17	.04/.14	.22	.03/.20	.24	.05/.19	.31	.15/.16
Critical lures								
Related	.76	.45/.32	.68	.33/.35	.64	.28/.36	.44	.19/.25
Unrelated	.18	.05/.14	.27	.03/.24	.23	.03/.19	.34	.14/.20

Note—R, remember judgment; K, know judgment. Instances in which remember and know proportions do not sum to the overall proportion reflect rounding to two decimal places.

words and the false alarm rate for related critical words decreased over retention interval, whereas the false alarm rates for nonstudied words and unrelated critical words generally increased with greater delays. Although the baseline false alarm rates for nonstudied words and unrelated critical words were similar at each retention interval, these false alarm rates varied over retention interval, thereby rendering any direct statistical test of the retention interval variable problematic. Therefore, all the statistical tests of the recognition data were based on scores that were derived from different recognition sensitivity measures.

Remember/know judgments. Table 2 also shows the mean proportions of remember and know judgments for studied and nonstudied list words and related and unrelated critical words. Consistent with previous research (e.g., Gallo et al., 1997; Gallo, Roediger, & McDermott, 2001; Payne et al., 1996; Roediger & McDermott, 1995), remember judgments exceeded know judgments for studied words and related critical words for the immediate test, whereas know judgments were more frequent than remember judgments for nonstudied words and unrelated critical words. For each of the longer retention intervals, remember and know judgments were generally comparable for both studied words and related critical words, but know judgments typically exceeded remember judgments for nonstudied words and unrelated critical words. Both Payne et al. and Gardiner and Java (1991) observed a similar shift in remember/know judgments after a delay. These results suggest that the remember/know judgment pattern that is typically observed for an immediate test largely disappears after a delay.

Three different sensitivity measures. Because DRM researchers have used different measures of recognition sensitivity, we compared three different measures: a high-threshold measure, Pr , a signal detection measure, d' , and a nonparametric measure, A' . According to Snodgrass and Corwin (1988), the high-threshold measure is more sensitive to changes in discrimination than is either the signal detection or the nonparametric measure. For each procedure, two different sensitivity measures were computed to provide separate estimates of item-specific memory and false memory susceptibility. For item-specific memory, the mean estimate of sensitivity was computed from hits and false alarms to studied and nonstudied list words from

individual participants' responses. For false memory susceptibility, the estimate was computed from false alarms to related and unrelated critical words. This procedure, used previously by Koutstaal and Schacter (1997) and Seamon, Luo, et al. (2002; Seamon, Lee, et al., 2002), treated false recognitions of related critical words as "hits" in order to provide a measure of the extent to which the participants were fooled by these words. In terms of fuzzy trace theory, the Pr , d' , and A' measures of false memory susceptibility can be viewed as measures of "gist" memory that differ from item-specific or "verbatim" memory.

Table 3 shows the mean recognition sensitivity scores for each measure for list words and critical words over the retention intervals. Unlike Experiment 1, where the recall proportions were adjusted to equate list word and critical word recall on the immediate test, no adjustments were made to these recognition scores. As Thapar and McDermott (2001) noted, measures such as d' theoretically provide a linear transformation to the data that allows any interaction of retention interval and word type to be interpretable, regardless of the list word and critical word results on the immediate test. Moreover, the nonparametric A' scores for list words and critical words were virtually identical on the immediate test. Thus, we will examine the unadjusted Pr , d' , and A' measures in turn.

For the high-threshold measure, Pr , we subtracted false alarms for nonstudied words from the recognition of studied words, and we subtracted false alarms for unrelated critical words from related critical words, both on an individual participant basis. Pr can range from 0 to 1.00, with higher scores indicating greater sensitivity than do lower scores and 0 indicating chance performance. Table 3 shows that the Pr measures for list words and critical words generally decreased over the retention intervals, with sensitivity somewhat higher for critical words than for list words at all but the longest delay. After 2 months, Pr scores approached chance for both word types. These results were supported by analyses of variance that showed an effect of retention interval [$F(3,112) = 28.25$, $MS_e = 1.68$, $p < .0001$], and word type [$F(1,112) = 13.69$, $MS_e = 0.38$, $p < .001$], but no interaction of these variables ($F < 1$). Sensitivity decreased from the immediate to the 2-day test [$F(1,56) = 8.15$, $MS_e = 0.63$, $p < .01$], was unchanged from the 2-day to the 2-week test ($F < 1$),

and decreased from the 2-week to the 2-month test [$F(1,56) = 31.76, MS_e = 1.32, p < .001$].

Two additional analyses were conducted on these data. First, an analysis that excluded the 2-month delay condition, owing to possible floor effects, revealed the same pattern of results: Memory sensitivity decreased over delay [$F(2,84) = 8.14, MS_e = 0.64, p < .001$], sensitivity was greater for critical words than for list words [$F(1,84) = 9.55, MS_e = 0.32, p < .005$], and there was no interaction of these variables ($F < 1$). Second, an analysis that included only the immediate and 2-day test results was conducted to compare the present results with those of prior DRM studies that employed short retention intervals of 1 or 2 days (Lampinen & Schwartz, 2000; Neuschatz et al., 2001; Payne et al., 1996; Thapar & McDermott, 2001). Once again, there were main effects of retention interval [$F(1,56) = 8.15, MS_e = 0.63, p < .01$] and word type [$F(1,56) = 5.77, MS_e = 0.15, p < .05$] but no interaction of these variables ($F < 1$).

To summarize the high-threshold recognition results, the *Pr* sensitivity scores showed that lengthening the retention interval impaired list word and critical word recognition and that false recognition of critical words was greater than accurate recognition of list words over all but the longest delay, where sensitivity for both word types approached chance.

Table 3 also shows the results according to the signal detection measure, *d'*, which can range from 0 to approximately 4.00, with scores between 1 and 2 usually indicating good recognition performance and 0 indicating chance (Neath, 1998). The observed pattern of results mimicked that of the high-threshold measure. Once again, sensitivity for list words and critical words typically decreased over the retention intervals, and sensitivity was higher for critical words than for list words. Analyses of variance showed an effect of retention interval [$F(3,112) = 20.22, MS_e = 17.99, p < .0001$] and word type [$F(1,112) = 12.30, MS_e = 6.16, p < .001$] but no interaction of these variables [$F(3,112) = 1.69, MS_e = 0.85, p > .15$]. Sensitivity decreased from the immediate to the 2-day test [$F(1,56) = 8.15, MS_e = 0.63, p < .01$], was unchanged from the 2-day to the 2-week test ($F < 1$), and decreased from the 2-week to the 2-month test [$F(1,56) = 31.76, MS_e = 1.32, p < .001$].

The first additional analysis that excluded the 2-month delay condition, owing to possible floor effects, revealed the same pattern of results: Memory sensitivity decreased over delay [$F(2,84) = 5.30, MS_e = 5.69, p < .01$], sensitivity was greater for critical words than for list words [$F(1,84) = 9.85, MS_e = 5.81, p < .01$], and there was no interaction [$F(2,84) = 1.92, MS_e = 1.13, p > .15$]. However, the second additional analysis that included only the immediate and 2-day delay tests revealed a different pattern. In this analysis, there was an effect of retention interval [$F(1,56) = 4.89, MS_e = 5.85, p < .05$], but there was no reliable effect of word type [$F(1,56) = 2.15, MS_e = 1.27, p > .10$] and no interaction of these variables ($F < 1$).

In summary, the *d'* scores showed that list word recognition and critical word recognition were impaired by lengthening the retention interval and that false recognition was greater than accurate recognition when analyses included the 2-week interval, but not when an analysis excluded the 2-week and 2-month intervals. This measure approached chance for both list words and critical words after a 2-month delay.

Finally, Table 3 presents the results for the nonparametric estimate of sensitivity, *A'*, which can range from 0 to 1.00, with higher scores indicating greater sensitivity than lower scores and .50 indicating chance performance. Table 3 shows that the *A'* measures for list words and critical words were highly similar and that both decreased at a comparable rate over the retention intervals. After a 2-month delay, performance was close to chance for both measures. Unlike the previous analyses involving *Pr* and *d'*, no effect of word type was present. These results were supported by analyses of variance that showed that retention interval affected list words and critical words in the same manner. There was an effect of retention interval [$F(3,112) = 27.92, MS_e = 0.88, p < .0001$] but no effect of word type and no interaction of these variables (both $F_s < 1$). Sensitivity decreased from the immediate to the 2-day test [$F(1,56) = 8.39, MS_e = 0.26, p < .01$], was unchanged from the 2-day to the 2-week test ($F < 1$), and decreased from the 2-week to the 2-month test [$F(1,56) = 28.83, MS_e = 0.93, p < .0001$].

The first additional analysis that excluded the 2-month delay condition, owing to floor effects, revealed the same

Table 3
Experiment 2: Recognition Sensitivity Scores for List Words and Critical Words by Three Different Measures

Recognition Measure and Word Type	Retention Interval			
	Immediate	2 Days	2 Weeks	2 Months
High threshold: <i>Pr</i>				
List words	0.50	0.37	0.28	0.10
Critical words	0.58	0.41	0.41	0.10
Signal detection: <i>d'</i>				
List words	1.70	1.20	0.84	0.35
Critical words	1.84	1.47	1.51	0.55
Nonparametric: <i>A'</i>				
List words	0.84	0.77	0.72	0.56
Critical words	0.86	0.75	0.75	0.55

results: Memory sensitivity decreased over delay [$F(2,84) = 7.85, MS_e = 0.22, p < .001$], but A' scores for list words and critical words did not differ, and there was no interaction of retention interval and word type (both $F_s < 1$). The same outcomes were observed by the analysis that included only the immediate and 2-day delay tests. There was an effect of retention interval [$F(1,56) = 8.39, MS_e = 0.26, p < .01$], but there was no reliable effect of word type ($F < 1$) and no interaction of these variables [$F(1,56) = 1.05, MS_e = 0.02, p > .30$].

In summary, the results from the A' scores were similar to those from the Pr and d' measures in showing that list word recognition and critical word recognition were impaired by lengthening the retention intervals and that A' for both list words and critical words was close to chance after a 2-month delay. However, the A' measure differed from the other measures in failing to find any difference between the level of accurate and false recognition for list words and critical words in any analysis.

A comparison of sensitivity measures. The three measures, Pr , d' , and A' , were alike in showing that true memory sensitivity and false memory susceptibility were diminished over retention intervals. Each measure showed a decrease from the immediate test to a 2-day test, no change from the 2-day to the 2-week test, and an additional drop close to chance from the 2-week to the 2-month test. Clearly, accurate recognition and false recognition were influenced by retention interval. This finding is in agreement with previous DRM recognition studies in showing that a short delay of 1 or 2 days can reduce accurate and false recognition, regardless of recognition measure (Lampinen & Schwartz, 2000; Neuschatz et al., 2001; Payne et al., 1996; Thapar & McDermott, 2001). We extended that research by describing the change in recognition performance over a broad range of retention intervals.

More important, however, were our findings on the effect of retention interval on accurate and false recognition. Unlike the recall results from the first experiment, the recognition results from this experiment showed no reliable interaction of these variables for any recognition measure. The Pr and d' measures in Table 3, however, do suggest an interaction of retention interval and word type, especially for the immediate and 2-week tests. Yet, post hoc analyses of the three measures for these test conditions revealed a marginal interaction for only one measure, d' [$F(1,56) = 3.24, MS_e = 2.10, p < .08$]. This trend toward greater persistence for false recognition than for accurate recognition is consistent with the direction of the recall results, but, as was suggested in the introduction to this experiment, the effect is clearly modest and subtle. A strong, unambiguous demonstration of greater persistence of false recognition than of accurate recognition was not demonstrated, and this outcome mirrors the inconsistent results that have been observed in prior recognition research (e.g., Brainerd et al., 2001; Lampinen & Schwartz, 2000; Neuschatz et al., 2001; Payne et al., 1996; Thapar & McDermott, 2001). Those studies and the present study are consistent in showing that accurate memory is not more persistent than false memory, but greater persistence

for false memory has not been consistently observed. We will suggest a possible reason for this result in the General Discussion section.

We did find that the three recognition measures differed with respect to the effect of word type. Both the high-threshold and the signal detection measures revealed an effect of word type on performance that the nonparametric measure failed to observe. Specifically, for both Pr and d' , susceptibility to critical words was greater than sensitivity to list words for analyses that included all retention intervals, as well as those that excluded the 2-month interval. This finding demonstrates persistence of false recognition over lengthy retention intervals. Strong false memory over time was also demonstrated by the fact that both of these measures showed more false recognition than accurate recognition for the 2-week retention interval alone (both $ps < .01$). Yet, for analyses that included only the immediate and 2-day tests, the high-threshold measure was alone in showing this effect of word type; the signal detection and nonparametric measures failed to demonstrate this effect.

Which measure or measures should we consider most appropriate? Among the three measures, Snodgrass and Corwin (1988) found the high-threshold measure to be the most sensitive for discriminating change, and our results confirm the sensitivity of this measure. Whereas the high-threshold and signal detection measures were more sensitive than the nonparametric measure in detecting differences in word type over all delays, the high-threshold measure was most sensitive over short delays. But even with the high-threshold measure, the magnitude of the word type effect over short delays was small. Thus, our results for the immediate and 2-day delay tests and those from other DRM recognition studies that employed a short delay (Lampinen & Schwartz, 2000; Neuschatz et al., 2001; Payne et al., 1996; Thapar & McDermott, 2001) are not necessarily divergent. If the high-threshold measure is used, our results support Payne et al.; if the other measures are employed, our results are consistent with the remaining studies. The important point, we think, is that false recognition may or may not be greater than accurate recognition over short delays, depending, to varying degrees, on recognition measure, number and length of lists, presentation rate, and so forth. This fact may help explain some of the inconsistent results in the literature found with short retention intervals.

Several conclusions can be drawn from this experiment. First, researchers should consider using either the high-threshold or the signal detection measure as a preferred recognition measure in the DRM procedure. The high-threshold measure was most sensitive to detecting a difference in word type over short retention intervals, whereas the signal detection measure was most sensitive to detecting a possible interaction of retention interval and word type. Second, false memory persistence can be demonstrated over lengthy delays, and false memory can exceed accurate memory. When the retention interval was measured in days, false recognition was greater than accurate recognition for one measure (Pr). When the retention in-

terval was measured in weeks, false recognition was greater than accurate recognition for two measures (Pr and d'). When the retention interval was measured in months, both accurate and false recognition approximated chance for all the measures (Pr , d' , and A'). Finally, false memory could be more persistent than accurate memory, when the retention interval was measured in weeks and recognition is measured by d' , but this effect was modest and subtle. Thus, the relative persistence of accurate and false memory can be influenced by the sensitivity of the recognition measure and the range of retention intervals employed.

GENERAL DISCUSSION

In two experiments, this research examined accurate and false memory in the DRM procedure over retention intervals that ranged from no delay to a 2-month delay. Together, they demonstrated several noteworthy points. First, when memory was tested by free recall, accurate memory for list words decreased faster than did false memory for critical words. In fact, false recall after a 2-week delay was as strong as it was on an immediate test. Second, when memory was tested by recognition, accurate and false memory decreased over retention interval, but false memory was only marginally more persistent than accurate memory, depending on recognition measure and retention interval. Finally, when memory was tested by recall or recognition, accurate and false memory approached floor or chance levels following a delay of 2 months. Thus, the results of both experiments, which used the same word lists and study conditions, are in general agreement in showing that whereas both accurate and false memory are eventually lost over time, false memory can be more persistent than accurate memory. In light of these findings, it is important to reconsider prior research with the DRM procedure regarding false memory persistence.

Retention Interval and Accurate and False Memory

Similar to previous DRM recall studies (McDermott, 1996; Thapar & McDermott, 2001; Toglia et al., 1999), the present research showed greater persistence for false recall than for accurate recall. The variety of procedures and retention intervals employed in these four different studies attests to the robustness of the conclusion that false recall is more persistent than accurate recall, whether time is measured in days or weeks. In addition, the results from the present research are in general agreement with previous DRM recognition studies that indicate that false recognition can be more persistent than accurate recognition. Rather than accept the null hypothesis of no difference in relative persistence, there are now three studies with different procedures (Payne et al., 1996; Thapar & McDermott, 2001; the present experiments) that have provided at least some measure of evidence that false recognition can be more persistent than accurate recognition. These results, coupled with the previously cited findings comparing accurate and false recognition over time in other memory paradigms (e.g., Barclay & Wellman, 1986;

Brainerd & Mojardin, 1998; Brainerd et al., 1995; Seamon et al., 2000; Sulin & Dooling, 1974), provide converging evidence that false recognition can be more persistent than accurate recognition, again whether time is measured in days or weeks.

Theoretical and Practical Aspects

Why were differences in persistence for accurate and false memory more readily observed for recall than for recognition in the present experiments? Following fuzzy trace theory, one possibility is that the effect of retention interval on accurate and false memory is dependent on the availability and accessibility of verbatim and gist traces for previously studied DRM lists. As was previously suggested, the strength of both verbatim and gist traces is diminished over time, so that both accurate and false memory are reduced over retention intervals. However, because gist traces are hypothesized to be more durable than verbatim traces (Brainerd et al., 1995; Reyna & Lloyd, 1997), false memory can be more persistent than accurate memory. When memory is tested by recall, both verbatim and gist traces decrease in strength and, consequently, availability over time. But gist traces, owing to their greater durability, may be more accessible than specific verbatim traces, leading to greater persistence for false recall than for accurate recall. On the other hand, when memory is tested by recognition, both verbatim and gist traces again decrease in availability over time, but the presence of list words and critical words on the recognition test can serve, at least to some degree, to offset the prior differences in accessibility for verbatim and gist traces, thereby making a difference in the relative persistence of accurate and false recognition more difficult to observe.

Summary

The present findings support several hypotheses about retention in the DRM procedure, whether tested by recall or recognition. First, accurate memory is not more persistent than false memory. Second, false memory can be more persistent than accurate memory. Third, false memory does not increase over time. False memory may increase over a short delay for word lists previously recalled during study (McDermott, 1996) or for words or pictures that are categorically related to previously studied stimuli (Brainerd et al., 1995; Seamon et al., 2000), but there is no support for a general increase in false memory in the DRM procedure over a broad range of retention intervals (Toglia et al., 1999; the present experiments). Finally, both accurate and false memory are effectively eliminated by a lengthy retention interval measured in months.

In closing, we noted in the introduction that, for practical as well as theoretical reasons, it is important to understand whether the effect of retention interval is the same for accurate and false recollection. Recently, Brainerd et al. (2000) used fuzzy trace theory to question a number of legal assumptions pertaining to memory. Included in those assumptions are the beliefs that truth is more memorable than fiction and recollections of true events are more consistent than false events. Jurors' beliefs, for ex-

ample, can be influenced by the assumption that accurate memory is more persistent than false memory over time. The present research provides no support for this assumption. From parametric measures of recall and recognition in the DRM procedure, the opposite appears to be true.

REFERENCES

- BARCLAY, C. R., & WELLMAN, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory & Language*, *25*, 93-103.
- BRAINERD, C. J., & MOJARDIN, A. H. (1998). Children's false memories for sentences: Long-term persistence and mere testing effects. *Child Development*, *69*, 1361-1377.
- BRAINERD, C. J., REYNA, V. F., & BRANDSE, E. (1995). Are children's false memories more persistent than their true memories? *Psychological Science*, *6*, 359-364.
- BRAINERD, C. J., REYNA, V. F., & POOLE, D. A. (2000). Fuzzy-trace theory and false memory: Memory theory in the courtroom. In D. F. Bjorklund (Ed.), *False-memory creation in children and adults: Theory, research, and implications* (pp. 93-127). Mahwah, NJ: Erlbaum.
- 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.
- BRUCE, D., & WINOGRAD, E. (1998). Remembering Deese's 1959 articles: The Zeitgeist, the sociology of science, and false memories. *Psychonomic Bulletin & Review*, *5*, 615-624.
- DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.
- EBBINGHAUS, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Columbia University, Teacher's College. (Original work published 1885; reprint edition, New York: Dover, 1964)
- GALLO, D. A., ROBERTS, M. J., & SEAMON, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271-276.
- GALLO, D. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*, 579-586.
- GARDINER, J. M., & JAVA, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, *19*, 617-623.
- HICKS, J. L., & MARSH, R. L. (1999). Attempts to reduce the incidence of false memory with source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1195-1209.
- KOUTSTAAL, W., & SCHACTER, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory & Language*, *37*, 555-583.
- LAMPINEN, J. M., & SCHWARTZ, R. M. (2000). The imperistence of false memory persistence. *Memory*, *8*, 393-400.
- LOFTUS, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 397-406.
- McDERMOTT, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, *35*, 212-230.
- NEATH, I. (1998). *Human memory*. Pacific Grove, CA: Brooks/Cole.
- NEUSCHATZ, J. S., PAYNE, D. G., LAMPINEN, J. M., & TOGLIA, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, *9*, 53-71.
- PAYNE, D. G., ELIE, C. J., BLACKWELL, J. M., & NEUSCHATZ, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language*, *35*, 261-285.
- RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89-102.
- REYNA, V. F., & LLOYD, F. (1997). Theories of false memory in children and adults. *Learning & Individual Differences*, *9*, 95-123.
- ROBINSON, K. J., & ROEDIGER, H. L., III (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*, 231-237.
- 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, McDERMOTT, K. B., & ROBINSON, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, and C. Cornoldi (Eds.), *Theories of memory* (Vol. 2, pp. 187-245). Hove, U.K.: Psychology Press.
- SCHACTER, D. L., VERFAELLIE, M., & PRADERE, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory & Language*, *35*, 319-334.
- SEAMON, J. G., LEE, I. A., TONER, S. K., WHEELER, R. H., GOODKIND, M. S., & BIRCH, A. D. (2002). Thinking of critical words during study is unnecessary for false memory in the Deese, Roediger, and McDermott procedure. *Psychological Science*, *13*, 527-532.
- SEAMON, J. G., LUO, C. R., & GALLO, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, *9*, 20-26.
- SEAMON, J. G., LUO, C. R., SCHLEGEL, S. E., GREENE, S. E., & GOLDENBERG, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory & Language*, *42*, 120-146.
- SEAMON, J. G., LUO, C. R., SCHWARTZ, M. A., JONES, K. J., LEE, D. M., & JONES, S. J. (2002). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory & Language*, *46*, 323-340.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50.
- SULIN, R. A., & DOOLING, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, *103*, 255-262.
- THAPAR, A., & McDERMOTT, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*, 424-432.
- TOGLIA, M. P., NEUSCHATZ, J. S., & GOODWIN, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*, 233-256.
- TUSSING, A. A., & GREENE, R. L. (1999). Differential effects of repetition on true and false recognition. *Journal of Memory & Language*, *40*, 520-533.

NOTES

1. Persistence can be demonstrated in different ways. First, many studies, including the present research, have employed a between-subjects design wherein participants are tested either immediately or after a delay. Greater false memory than accurate memory persistence is demonstrated by a greater decline in accurate memory than in false memory over the retention intervals. Other studies have used a within-subjects design (e.g., Brainerd et al., 1995) in which the same participants are tested at all intervals. Greater false memory than accurate memory persistence is defined as greater consistency in false memory than in accurate memory over repeated tests. In addition, a hybrid design can combine aspects of both designs if participants are given a second recognition test after a delay that includes items previously tested, as well as those tested for the first time (e.g., Brainerd et al., 2001). Accurate or false memory persistence per se, however, can be demonstrated simply by above-chance levels of performance on a delayed test, and the relative level of accurate and false memory performance at any retention interval is not the same as relative stability over time. We thank Charles Brainerd for making this point.

2. Strictly speaking, word type (list words vs. critical words) is not manipulated as an independent variable in recall studies, although it is an independent variable in recognition studies. However, a number of researchers have treated word type as a within-subjects variable in order to compare overall levels of accurate and false recall (e.g., Robinson & Roediger, 1997; Roediger & McDermott, 1995; Thapar & McDermott, 2001). We followed this practice as well.