False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing

ANJALI THAPAR
Bryn Mawr College, Bryn Mawr, Pennsylvania

and

KATHLEEN B. McDERMOTT
Washington University, St. Louis, Missouri

The effects of retention interval and level of processing on false recall and false recognition of associates were examined. False recall and false recognition were induced by presenting subjects with words closely associated with a nonstudied word. Both level of processing and retention interval affected false recall (Experiment 1) and false recognition (Experiment 2) in the same direction with which they affected accurate recall and accurate recognition. That is, semantically processed lists exhibited higher levels of later false recall and false recognition than did superficially processed lists. Furthermore, a decline in false recall and false recognition occurred across retention intervals of 0, 2, and 7 days. However, the decline in false recall and false recognition was less pronounced than the decline in accurate recall and accurate recognition. Results are consistent with source monitoring and fuzzy trace explanations of false recall and false recognition.

Roediger and McDermott (1995) revived and developed a paradigm previously used by Deese (1959), which provides a straightforward way to study false recall and false recognition. Subjects study a list of 15 associated words (e.g., *bed*, *rest*, *awake*, *tired*, *dream*) and are subsequently given a recall or recognition test; interest is in the probability with which subjects will erroneously recall or recognize a specific word that is highly associated with the presented word but was not itself presented (e.g., *sleep*). False memories arising from this paradigm are thought to represent experimental analogs of inferences people draw routinely in activities, such as conversation, reading, and observing events (Roediger & McDermott, 2000).

In the past several years, researchers have used this paradigm as a means of determining how fundamental variables, known to have predictable effects on accurate memory, affect false recall and false recognition (see Roediger, McDermott, & Robinson, 1998, for a review of the emerging literature). The goal of the present experiments was to investigate the role of two important, but little-understood variables with respect to false recall and false recognition in this paradigm: retention interval and level of processing during the study phase.

Retention Interval

The primary question addressed by these experiments was how the probability of false recall and false recognition changes across time. Thus far, most of the evidence with respect to retention interval has been indirect, in that retention interval has been covaried with repeated testing (McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996). This feature was by design, in that the primary question of interest in these studies was how prior testing affected later false recall. However, a between-subjects manipulation of retention interval is necessary to fully understand the effect of retention interval on false memory.

The principal question addressed by the present experiments was whether (and in what direction) false recall and false recognition change across time. Theories that have been recently applied to false memories would lead one to predict an increase in the *ratio* of false recall to accurate recall (and in the ratio of false recognition to accurate recognition). That is, as studied items are forgotten, there will be a greater reliance on the processes that tend to produce false memories. For example, both *fuzzy trace theory* (Reyna & Brainerd, 1995) and the *source monitoring framework* (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981) predict that as item-specific, perceptual details fade across time, subjects will place increasing emphasis on gist (fuzzy trace theory) or memory for cognitive/emotional processing (source monitor-

ing framework) at the expense of detailed, perceptual memory. An important point is that these relative changes speak only to the relation between critical and studied items, holding that the ratio of critical items to studied items recalled or recognized should increase across time. However, the increase in the ratio of critical words to studied words could occur with many different patterns: absolute decreases, no change, or absolute increases in false memories across retention interval. That is, all the aforementioned possibilities can occur under the predicted conditions of an interaction, such that there is forgetting for studied items, and there is some, but less, forgetting for critical items, no change for critical items, or an increase in recall or recognition of critical items across time. In contrast, activation theories of false recall and false recognition, at least those following from the semantic priming tradition (Neely, 1991), would predict a rapid decay of false memories (see Roediger, Balota, & Watson, 2001).

The single experiment to use a between-subjects design to examine the effect of retention interval in this paradigm was reported by Toglia, Neuschatz, and Goodwin (1999), who found no decrease in false recall across delays of 1 week and 3 weeks. That is, following study of five 12-word associative lists, subjects recalled .45, .42, and .46 of the critical nonpresented items when tested immediately, following a 1-week delay, or following a 3-week delay, respectively. These results are very surprising because they suggest that false memories in the Deese–Roediger–McDermott (DRM) paradigm are not vulnerable to forgetting.

The primary question addressed by these experiments was whether retention interval exerts an effect on false recall and false recognition. A secondary question concerned the relation between studied and critical items. For example, if forgetting was observed for both studied and critical items, would there be an interaction such that the decline was more dramatic for one than for the other?

Level of Processing

Level of processing of the studied lists was also manipulated in both experiments. Predictions with respect to level-of-processing effects on false memories can be made on several grounds. For example, in many (but not all) experiments, the critical nonpresented items behave almost as if they were presented in the list. People recall them with a probability similar to that of recalling the studied items (e.g., McDermott, 1996; Roediger & McDermott, 1995; Schacter, Verfaellie, & Pradere, 1996); in addition, people claim to be able to recollect the instance of presentation of the nonpresented associates, as manifested in Tulving's (1985) remember/know procedure (Norman & Schacter, 1997; Payne et al., 1996; Roediger & McDermott, 1995). The nonpresented associates show testing effects similar to those found for studied items (McDermott, 1996; Roediger & McDermott, 1995), and the nonpresented associates even manifest perceptual priming (McDermott, 1997). Therefore, it might reason-

ably be expected that the nonpresented associates show level-of-processing effects just as do studied words: Lists encoded with respect to their meaning are better remembered than are lists processed at a superficial level (Craik & Lockhart, 1972; Craik & Tulving, 1975) and therefore might be expected to show a higher probability of false recall and false recognition.

Further support for this prediction can be found in studies that have manipulated the presentation order of studied words. When several of these associative sets are presented together in one long list, the probability of false recall and false recognition is greater when the sets are blocked together (e.g., 15 *sleep*-related words, followed by 15 *needle*-related words and 15 *window*-related words) than when the words from the three lists are randomly intermixed (Mather, Henkel, & Johnson, 1997; McDermott, 1996; Toglia et al., 1999). One could make the claim that blocked presentation leads to richer, more meaningful encoding of the studied words in a way similar to the level-of-processing manipulation.

However, there are also reasons to make the opposite prediction regarding the effect of level of processing on false memories (i.e., reasons to predict that deep processing should reduce the probability of later false recall and false recognition). Specifically, the probability of false recall tends to decrease with slower presentation rates, at least within the intervals of about 0.5 to 10 sec (McDermott & Watson, in press), whereas veridical recall is (of course) enhanced by slower rates. It could be argued that slower presentation rates would allow for deeper, more meaningful processing; therefore, we could reasonably predict that, relative to shallow processing, deep processing would diminish false recall and false recognition. On logical grounds, this pattern makes sense: The better subjects are at remembering what was in the list, the fewer errors of commission should be made.

We consider now the empirical results directly addressing whether/how level of processing affects false recall and false recognition in the DRM procedure. There are four reported attempts to address this question: Two of the papers (Read, 1996; Tussing & Greene, 1997) reported null effects, and two (Rhodes & Anastasi, 2000; Toglia et al., 1999) reported that semantically processed lists led to greater false recall than did superficially processed lists. Our prediction was that we would observe a level-of-processing effect because the two published experiments obtaining null effects also had features that made the results difficult to interpret. Specifically, Tussing and Greene (1997) showed that deep processing and shallow processing led to near-equivalent probabilities of false recognition (.82 and .81 for deep and shallow processing, respectively), but they also failed to find a level-of-processing effect on the hit rate. Therefore, level of processing may not have been manipulated effectively. Similarly, Read found no effect of level of processing on false recall (.73 and .76 for deep and shallow conditions, respectively), but he only used a single list in this experiment, and, with only one observation per subject, his data

may not have been reliable enough to detect an effect. Thus, although it is possible that level of processing has no effect on false memories in this paradigm, we believed this issue deserves a more careful examination. The prediction was that we would obtain level-of-processing effects similar to those seen by studied items (Rhodes & Anastasi, 2000; Toglia et al., 1999), which would allow us to observe performance across retention interval from different starting points on the scale.

The Present Experiments

We report two experiments in which we examined the effects of level of processing and retention interval on false recall (Experiment 1) and false recognition (Experiment 2) of the critical nonpresented words in the DRM procedure. We tentatively predicted that level of processing would affect studied items and critical items in a qualitatively similar manner. Retention interval was a much more open question; theoretical accounts led us to predict an interaction between item type (studied/critical) and retention interval, but the direction of change (if any) for critical items was unspecified in this prediction.

EXPERIMENT 1

The effect of level of processing and retention interval on false recall of the critical nonpresented words was investigated in Experiment 1. Level of processing was manipulated by having the subjects orient either to the color of individual words within a list, to count the number of vowels in each word within a list, or to rate the pleasantness of each word within a list. (It was initially hypothesized that color might lead to more superficial processing than vowel counting, such that we would observe recall under three levels of encoding, as in Craik & Tulving, 1975; however, as can be seen by the results, these two levels produced nearly identical performance in all conditions.) Retention interval was varied between subjects in that the final free recall test was given either immediately after the presentation of all lists or after 2- or 7-day delays.

Method

Subjects. Ninety-nine undergraduate students from an introductory psychology class at Williams College participated for extra credit. The subjects were tested in groups of up to 4 students. Groups were randomly assigned to one of three retention intervals (immediate test, 2-day delayed test, 7-day delayed test). The resulting groups consisted of 33 participants in each of the three retention intervals.

Design and Materials. A 3×3 design was used in which encoding task (color, vowels, and pleasantness) was manipulated within subjects, and retention interval (immediate, 2-day, or 7-day) was manipulated between subjects. The materials used were the first 12 items of the 24 lists reported in the Appendix of Roediger and McDermott's (1995) article, with the following exception: The three lists that corresponded to the critical words *king*, *man*, and *girl* were replaced by three lists corresponding to the critical words *trash*, *smoke*, and *smell* (see Stadler, Roediger, & McDermott, 1999, for the items in these three lists). The 24 lists were randomly divided into three sets of 8 lists each. The three list sets were then assigned to the three encoding conditions, such that each set occurred in each encoding condition an equal number of times across sub-

jects. As in Roediger and McDermott's (1995) study, the ordering of words within lists was from strongest to weakest associate (i.e., words most strongly associated with the critical nonpresented item occurred at the beginning of the list). The subjects were told prior to the presentation of each list which encoding orientation they were to apply to the upcoming list; the order of occurrence of these encoding orientations was random.

Procedure. The subjects were told that they would be presented with lists of words on flash cards and that their memory for these words would be tested later. The subjects were then given instructions for the three encoding conditions: pleasantness rating (rate the pleasantness of the meaning of the word on a 1–5 scale), vowel counting (count the number vowels in the word), and color of word instructions (write the color in which the word was written; blue, green, orange, purple, red, and black were the colors used). Regardless of processing instructions, the subjects recorded their responses on rating sheets provided by the experimenter. Each word was presented for 4 sec.

The subjects were given response sheets and told that for each word presented in a list, they were to write the number corresponding to their pleasantness rating of the word, the number corresponding to the number of vowels in the word, or the name of the color in which the word appeared (depending on the encoding instruction assigned to the list). The orienting condition of each list was given immediately prior to list presentation.

On completion of the last study list, the subjects in the immediate retention interval condition were presented with a blank sheet of paper and asked to write down as many of the studied words as they could remember (regardless of the prior judgment task) without guessing. The subjects were given unlimited time to complete the free recall task. The subjects in the delayed test conditions departed after presentation of the final study list and returned at the specified time (2 or 7 days later) for their final free recall test.

Results and Discussion

A summary of the results can be found in Figure 1, in which the mean proportions of studied words and critical nonpresented words recalled are presented as a function of retention interval and level of processing.[1] In general, there was a decline in recall across retention interval for studied items processed deeply (.36, .12, and .09 for the pleasantness condition) and shallowly (.09, .03, and .03 for the color condition; .09, .03, and .02 for the vowel condition). Similarly, a decline occurred for critical nonpresented items (.24, .16, and .14 for the pleasantness condition; .06, .04, and .05 for the color condition; .06, .04, and .02 for the vowel condition).

A mixed-factor analysis of variance (ANOVA) was performed to analyze the effects of level of processing (color of word, vowel counting, and pleasantness rating), item type (studied, critical nonpresented), and retention interval (immediate, 2 day, 7 day) on free recall performance. The ANOVA yielded reliable main effects of level of processing [$F(2,192) = 97.31, MS_e = 1.29$] and retention interval [$F(2,96) = 29.26, MS_e = 0.48$] but no reliable effect of item type ($F < 1$). Deep processing led to greater probabilities of recall than did shallow processing, and declines in recall probabilities occurred across the retention interval.

Two of the two-way interactions reached significance. Specifically, item type interacted with retention interval [$F(2,96) = 13.73, MS_e = 0.10$], such that studied items exhibited greater declines across time than did critical
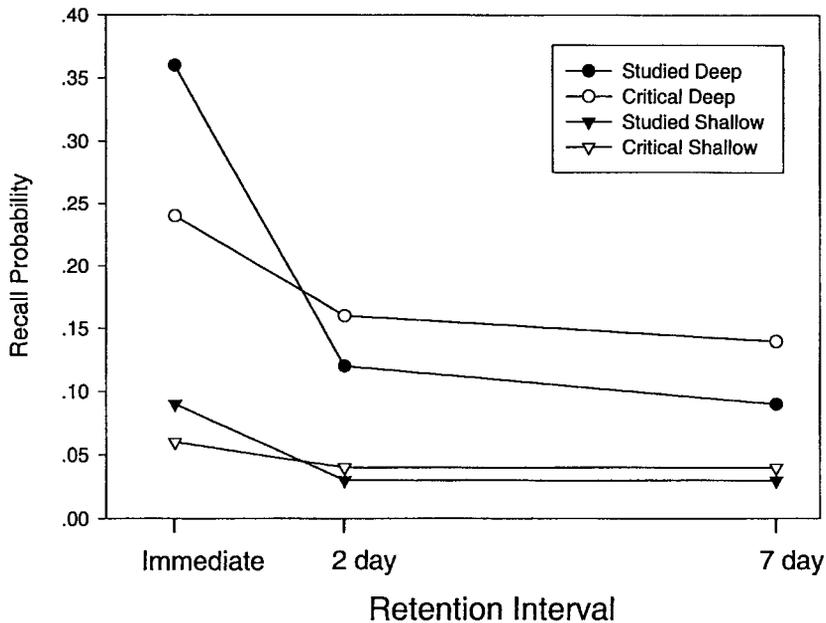
Figure 1. Probabilities of accurate and false recall as a function of level of processing and retention interval in Experiment 1.

items. Both item types showed declines across retention interval, however. Collapsing across level of processing, both studied and critical items demonstrated forgetting [$F(2,96) = 84.19$, $MS_e = 0.17$; and $F(2,96) = 4.03$, $MS_e = 0.02$, respectively]. Breaking these findings down further, we can see that the studied items exhibited a significant drop between the immediate and 2-day delayed tests [$t(64) = 9.13$, $SEM = 0.013$] and a marginally significant drop from 2- to 7-day delays [$t(64) = 1.93$, $SEM = 0.008$, $p < .06$]. Similarly, critical items manifested a drop from the immediate test to 2-day delayed test [$t(64) = 2.04$, $SEM = 0.021$], but no further drop thereafter ($t < 1$). Interestingly, although there was a higher probability of recalling studied items than critical items on the immediate test, this pattern was reversed after 1 week. That is, the overall probabilities of recalling studied and critical items were .18 and .12, respectively, on the initial test [$t(32) = 3.19$, $SEM = 0.017$]; after 1 week, these probabilities were .05 and .07 for studied and critical items, respectively [$t(32) = 3.25$, $SEM = 0.008$].

Retention interval interacted also with level of processing [$F(4,192) = 10.04$, $MS_e = 0.13$], such that deep processing led to greater declines across retention interval than did shallow processing. Both deeply and shallowly processed lists manifested forgetting, however [$F(2,96) = 22.60$, $MS_e = 0.670$; and $F(2,96) = 10.11$, $MS_e = 0.036$, respectively]. This claim of an interaction needs to be qualified, however, due to floor effects for the shallow condition. Another way of interpreting this interaction is that level of processing had a greater effect on the immediate test than on the delayed tests. Even after 1 week, though, there was a reliable level-of-processing

effect [$F(2,64) = 20.06$, $MS_e = 0.177$], and this was true for both studied and critical items [$t(32) = 5.97$, $SEM = 0.012$; and $t(32) = 3.98$, $SEM = 0.028$, respectively] (where the deep condition was contrasted with a combination of the two shallow conditions).

The third two-way interaction (item type $\times$ level of processing) did not reach significance ($F < 1$). However, the three-way interaction was reliable [$F(4,192) = 5.04$, $MS_e = 0.03$]. Retention interval affected studied items more prominently than critical items, and this was especially true for the deep level-of-processing condition, as can be seen in Figure 1. Again, though, the relatively small effect of retention interval on shallowly processed items must be viewed cautiously due to floor effects in the shallow processing condition.

Note that the analysis strategy adopted here assumes a linear scale for the probability of recall. That is, although studied and critical items (and deeply and shallowly processed items) began at different points on the scale, we assume that drops of equivalent probabilities represent drops of equivalent magnitudes and drops of different probabilities represent drops of different magnitudes. This claim has been contested (Loftus, 1985; although see Slamecka, 1985). One way of addressing this issue (aside from replication and extension, which was the objective of Experiment 2) is to artificially equate the initial probability of recall by setting it equal to 1 for all conditions. That is, each data point after the immediate test is calculated as a proportion of recall on the immediate test. This method, too, leads to the conclusions that studied and critical items both manifested forgetting, but forgetting was more pronounced for studied items. That is, recall of studied items dropped

from 1.00 to .35 to .26 across the intervals, whereas recall of critical items dropped from 1.00 to .65 to .60. We also applied this same approach to the deep condition only, in which no floor effects are present; this approach also leads to the same conclusion. Correct recall dropped from 1.00 to .33 to .25 across retention interval, whereas false recall dropped from 1.00 to .67 to .63 across the same interval. In sum, this approach yields conclusions that converge with those from the analysis of the raw data.

The primary conclusions to be drawn from this experiment are as follows. First, declines were observed across retention interval for both studied and critical items. That is, forgetting was observed for both item types (although the term *forgetting* is a bit unusual when there was no initial presentation to forget, as in the case of critical items). Second, the degree of forgetting was greater for studied items than for critical items. This can best be seen in the deep processing condition, in which no floor effects cloud interpretation of the findings. That is, the interaction predicted by the source monitoring framework and fuzzy trace theory was obtained; moreover, the interaction neither was in the form of different directions of effects (e.g., forgetting for studied items and increases across time for critical items) nor was it caused by stability across retention interval for critical items. Rather, retention interval affected studied and critical items in a qualitatively similar fashion, but the effect was more pronounced for studied items. Third, robust level-of-processing effects were seen for critical items, and these were qualitatively similar to those for studied items (replicating Rhodes & Anastasi, 2000; Toglia et al., 1999). That is, semantic processing elicited greater probabilities of false recall than did superficial processing; this pattern was observed at all retention intervals for both studied and critical items.

Experiment 2 was designed to confirm and extend these findings to a recognition memory test. The primary results to be replicated were (1) the decline in false memories across retention interval and (2) the retention interval × item type interaction, which shows greater forgetting for studied items than for critical items. Moreover, we sought to replicate the level-of-processing effect and to determine whether the three-way interaction in Experiment 1 disappeared under conditions in which no floor effects were exhibited for the shallow processing condition. The study phase of Experiment 2 was identical to that of Experiment 1; only the test phase differed.

EXPERIMENT 2

The purpose of Experiment 2 was to replicate the basic pattern of results for Experiment 1 and to extend the findings to the domain of recognition memory.

Method

Subjects. Seventy-two undergraduate students from an introductory psychology class at Williams College participated for extra credit. The subjects were tested in small groups (4 students or fewer); 24 subjects were tested after each of the three retention intervals.

Design and Materials. The design and materials used in Experiment 2 were similar to those used in Experiment 1, with the exception that a yes/no recognition memory test replaced the free recall test used in Experiment 1. The recognition test consisted of 144 words. These items consisted of 24 words that had actually been presented during the study phase (one from each study list), 24 critical nonpresented targets that corresponded to the 24 study lists, 24 weakly related nonpresented words (one from each list taken from Positions 13–15 of Roediger & McDermott's, 1995, lists), and 72 unrelated distractor items.

Procedure. The procedure was identical to that of Experiment 1, with the exception that following presentation of the last study list, the subjects in the immediate retention condition were presented with the yes/no recognition test. The remaining subjects returned either 2 days or 7 days later for the recognition test, which was self-paced.

Results and Discussion

A summary of the recognition data is reported in Table 1 (raw data) and Figure 2 (signal detection sensitivity measures).[2] Table 1 reports the mean probabilities of recognition for each condition, irrespective of the baseline false-alarm rates, which increased dramatically across retention interval (from .04 on the immediate test to .28 and .31 on the 2- and 7-day delayed tests). Because the baseline false-alarm rates increased across retention interval, we performed statistical analyses on d'. This transformation has the added feature of theoretically being a linear scale, such that any interaction of item type and retention interval should be readily interpretable, regardless of the level of performance on the immediate test.

An examination of the data in Figure 2 suggests that the pattern obtained for recognition is much like that for recall: Level-of-processing effects were exhibited for both studied and critical items at all three retention intervals; moreover, recognition performance declined across time, and this decline was more pronounced for studied items than for critical items. Specifically, accurate recognition (d' for studied items) declined across retention interval from 3.47 to 1.80 to 1.56 in the pleasantness condition, from 1.62 to 0.84 to 0.87 in the color condition, and from 1.90 to 0.97 to 0.62 in the vowel condition. Similarly, the d' value corresponding to false recognition declined from 2.71 to 1.58 to 1.64 for critical items in the pleasantness condition, from 1.59 to 0.95 to 0.88 in the color condition, and from 1.52 to 1.16 to 0.79 in the vowel condition.

A three-factor ANOVA on the d' measures confirmed these observations. As in Experiment 1, we observed main

Table 1
Probability of Accurate and False Recognition in Experiment 2 as a Function of Retention Interval and Level of Processing

| Level of Processing | Retention Interval | | | | | |
| | Immediate | | 2-Day Delay | | 7-Day Delay | |
	Studied	Critical	Studied	Critical	Studied	Critical
Pleasantness	.90	.67	.79	.76	.75	.77
Color	.42	.34	.52	.56	.57	.58
Vowel	.34	.32	.50	.53	.50	.55

Note—Baseline false-alarm rates for unrelated lures were .04, .28, and .31 for immediate, 2-day delayed, and 7-day delayed conditions, respectively.
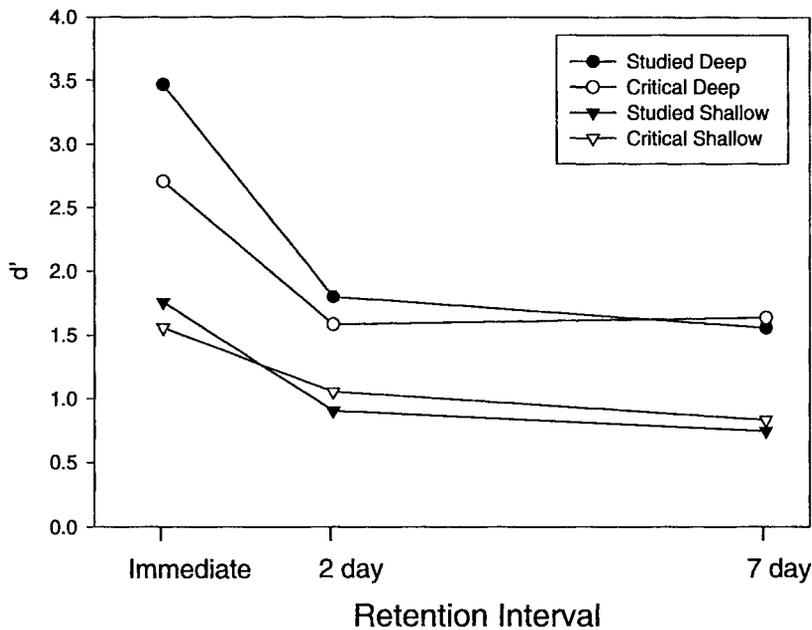
Figure 2. Values of *d′* for accurate and false recognition as a function of level of processing and retention interval in Experiment 2.

effects of retention interval [$F(2,69) = 22.96$, $MS_e = 48.41$] and level of processing [$F(2,138) = 56.55$, $MS_e = 46.34$] but no main effect of item type [$F(1,69) = 2.35$, $MS_e = 0.91$].

In addition, both the retention interval × item type interaction and the retention interval × level of processing interaction were reliable [$F(2,69) = 6.16$, $MS_e = 2.38$; and $F(4,138) = 3.47$, $MS_e = 2.85$], respectively; the interaction between level of processing and item type was marginally significant [$F(2,138) = 2.80$, $MS_e = 1.18$, $p < .06$]. The three-way interaction was not reliable [$F(4,138) = 1.30$, $MS_e = 1.55$].

The retention interval × item type interaction demonstrates that (as in Experiment 1) the effect of retention interval was more pronounced for studied items than for critical items. Follow-up tests confirmed that forgetting was indeed manifested for both studied items [$F(2,69) = 28.30$, $MS_e = 36.12$] and critical items [$F(2,69) = 12.04$, $MS_e = 14.67$]. For both deeply and shallowly processed studied and critical items, there was reliable forgetting from the immediate test to the 2-day test [smallest $t(46) = 2.06$], but not from the 2-day test to the 7-day test [largest $t(46) = 1.03$]. Furthermore, as in recall, recognition performance for studied items exceeded that for critical items on the immediate test [$t(23) = 3.48$, $SEM = 0.111$]. However, after 1 week, the two probabilities (for studied and critical items) did not differ ($t < 1$).

The retention interval × level of processing interaction indicates that (as in Experiment 1), deep processing led to a greater decline across retention interval than did shallow processing. Both deeply and shallowly processed lists manifested forgetting [$F(2,69) = 25.18$, $MS_e = $

33.43; and $F(2,69) = 12.34$, $MS_e = 19.87$, respectively]. Unlike in the free recall data, however, this interaction does not appear to have been caused by a floor effect. As noted in the discussion of Experiment 1, another way of interpreting this interaction is that level of processing exerted a greater effect on the immediate test than after a delay. However, even after 1 week, there was a reliable level of processing effect [$F(2,46) = 13.61$, $MS_e = 10.71$], and this was true for both studied and critical items [$t(23) = 3.99$, $SEM = 0.203$; and $t(23) = 4.04$, $SEM = 0.199$, respectively] (where the deep condition was contrasted with a combination of the two shallow conditions).

The marginally significant item type × level of processing interaction demonstrates that level of processing exerted a greater effect on studied items than on critical items. The level-of-processing effect was, however, reliable for both item types [$F(2,138) = 44.85$, $MS_e = 31.12$; and $F(2,138) = 30.05$, $MS_e = 16.40$, respectively].

In short, the pattern of results observed in recognition mirrors closely the pattern seen in free recall (with the exception of the three-way interaction in free recall, which appears to have resulted largely from the floor effect in the superficial processing conditions). The primary conclusions of Experiment 1 were replicated with a recognition test.

GENERAL DISCUSSION

The results of the two experiments reported here reveal fundamental patterns in false memory induced by associative lists. First, the probabilities of false recall and false recognition declined across time. Second, the

declines in false recollection across time were less pronounced than the declines seen in the probabilities of accurate recall and recognition. That is, although false memories were "forgotten," this process occurred at a slower rate than forgetting of accurate memories. Third, when lists were processed with respect to their meaning, both false recall and false recognition were enhanced, relative to superficial processing. We consider below how these data relate to the literature and then how they contribute to theoretical accounts of false memories.

Retention Interval

Toglia et al. (1999) found equivalent levels of false recall (approximating 50%) across delays as long as 3 weeks. Perhaps a key difference between Toglia et al.'s experiment and the present experiments is the number of study lists (6 in Toglia et al. and 24 in the present experiments); furthermore, Toglia et al. chose 6 lists most effective at eliciting false recall (see Stadler et al., 1999). In light of these factors, it is probably the difference between our study phase and Toglia et al.'s study phase that produces the different effects; the difference could lie in power differences (from the differing number of lists) or from true differences in the nature of the recall phase that are produced by retrieving differing numbers of lists.

The results reported here are consistent with the predictions of both the source monitoring framework and fuzzy trace theory. Both theories predict that, as the perceptual details are forgotten (or given less emphasis during retrieval), there will be a relative increase in the emphasis placed on more semantic/cognitive/emotional/gist features (see Barclay & Wellman, 1986, Leicht, 1968, Reyna & Kiernan, 1994, Sachs, 1967, Spiro, 1980, and Sulin & Dooling, 1974, for empirical evidence of this process in other paradigms). Our data go beyond the prediction of an interaction, however, in that they show that (at least in this paradigm), false recall and false recognition decline across time. However, this decline is less dramatic than the forgetting demonstrated with studied items.

It is important to note that although the sensitivity measures declined across retention interval, the absolute probabilities of false recognition increased across retention interval (see Table 1). That is, the absolute probabilities increased, but the relative probabilities decreased. In real-world situations, in which "unrelated baseline" false-alarm rates are not specified and therefore cannot be used as a correction factor, it appears to be the case that the likelihood of false recognition arising from inferences formed during encoding increases across retention interval. This absolute increase was observed for recognition only; recall declined across retention interval, by any analysis strategy.

Another pattern of interest emerging from these data is that on the immediate test, there was slightly greater recall of studied items than of critical items (see Figures 1 and 2). By 1 week later, this pattern had disappeared and actually was numerically reversed (with a very slight numerical advantage for critical items). This difference was very small (and insignificant in recognition) but nevertheless appeared in all four relevant comparisons (see, too, McDermott, 1996, Experiment 1).

When reviewing the existing literature in associative recall and recognition, Roediger et al. (1998, p. 225) concluded that "false recall is extremely robust, either remaining constant or increasing over a time interval, during which memory for the studied events is diminishing . . . in either case, there is an increase in false memories relative to veridical memories." It now appears that this conclusion was only partially correct. The increase in false memories relative to veridical memories was on target; however, the claim of steady or increasing false recall was premature. This conclusion was largely based on experiments in which retention interval was manipulated within subjects and was covaried with repeated testing; it may still hold under these conditions.

For example, McDermott (1996, Experiment 2) showed an increase in false recall over a delay of 1 day, but this increase was seen under what might be considered unusual circumstances. That is, subjects were given multiple study–test trials, and over the trials, they were able to decrease their false recall probabilities considerably (e.g., from .57 to .32 in one condition). They then returned 1 day later to receive a final test, and a slight but reliable increase in false recall was observed from the last trial of the 1st day to the final trial (1 day later). However, it seems reasonable to posit that the probability of false recall on the final trial of the 1st day had been somewhat artificially decreased; subjects had edited out false recalls over trials. Thus, one can think of the enhanced probability across the day as a reintroduction of errors that had been committed previously but had eventually been edited out when subjects were given enough practice. When a delay was introduced in the absence of intervening practice, there was a rebound effect, such that subjects fell back into their prior recall patterns, a kind of regression to the mean. In other words, although we do not have reason to question the replicability of McDermott's (1996) results, we are suggesting that their generalizability may not be wide. Indeed, in a separate experiment in this same paper (Experiment 1), McDermott found very low probabilities of false recall after a 2-day delay (and a greater probability of false recall [.20] than of accurate recall [.12]). However, no comparable final free recall test was obtained immediately because the question of interest was in determining the effects of immediate free recall tests on a later final free recall test.

Turning to recognition, Payne et al. (1996) manipulated retention interval between 2 min and 24 h. In one of their conditions, they observed relatively stable false recognition of critical items across retention interval (.59 to .57 when unrelated false alarms are subtracted from the critical false-alarm rates). In another condition, however,

there was a hint of a decline (from .59 to .48) across retention interval, although these conditions were not compared statistically.

In sum, it appears that a between-subjects manipulation of retention interval leads to declines in the level of both false recall and false recognition (at least when controls are introduced for the differing unrelated nonstudied base rates).

Level of Processing

As mentioned in the introduction, two published studies have reported failures to find level-of-processing effects in false recognition in this paradigm. However, due to the limitations of these studies (described in the introduction) and the recent reports by Toglia et al. (1999) and Rhodes and Anastasi (2000) that showed level-of-processing effects similar to those reported here, we conclude that robust effects of level of processing are indeed present in both false recall and false recognition in this paradigm. The effects in the present experiments were robust and replicated across two measures of performance and three retention intervals for each measure.

These data demonstrate a pattern opposite to that which might be predicted by the presentation rate findings (with longer processing times leading to diminished false recall) but consistent with the pattern that would be predicted from the blocked/random effect seen in false recall and false recognition.

Conclusion

In summary, we have shown that the likelihood of false recall and false recognition declines in a fashion similar to (although less pronounced than) the decline observed for studied items. In addition, we have observed robust level-of-processing effects in false recall and false recognition of nonpresented words that are associated to studied words; these effects were similar to those exhibited by studied items.

REFERENCES

BARCLAY, C. R., & WELLMAN, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory & Language*, **25**, 93-103.

CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **4**, 671-684.

CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 671-684.

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

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.

KOUTSTAAL, W., SCHACTER, D. L., GALLUCCIO, L., & STOFER, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology & Aging*, **14**, 220-237.

LEICHT, K. L. (1968). Recall and judged frequency of implicitly occurring words. *Journal of Verbal Learning & Verbal Behavior*, **7**, 918-923.

LOFTUS, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 397-406.

MATHER, M., HENKEL, L. A., & JOHNSON, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, **25**, 826-837.

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

MCDERMOTT, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review*, **4**, 582-586.

MCDERMOTT, K. B., & WATSON, J. M. (in press). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory & Language*.

NEELY, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processing in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

NORMAN, K. A., & SCHACTER, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, **25**, 838-848.

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.

READ, J. D. (1996). From a 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.

REYNA, V. F., & KIERNAN, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, **30**, 178-191.

RHODES, M. G., & ANASTASI, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review*, **7**, 158-162.

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. Surprenant (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, & MCDERMOTT, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, **9**, 123-127.

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, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187-245). Hove, U.K.: Psychological Press.

SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, **2**, 437-442.

SCHACTER, D. L., ISRAEL, L., & RACINE, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language*, **40**, 1-24.

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.

SLAMECKA, N. J. (1985). On comparing rates of forgetting: Comment on Loftus (1985). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 812-816.

SPIRO, R. J. (1980). Accomodative reconstruction in prose recall. *Journal of Verbal Learning & Verbal Behavior*, **19**, 84-95.

STADLER, M. A., ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, **27**, 494-500.

SULIN, R. A., & DOOLING, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, **103**, 255-262.

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

TULVING, E. (1985). Memory and consciousness. *Canadian Psychologist*, **26**, 1-12.

TUSSING, A. A., & GREENE, R. L. (1997). False recognition of associates: How robust is the effect? *Psychonomic Bulletin & Review*, **4**, 572-576.

NOTE

1. For Figure 1 and for Figure 2, we have collapsed across the two shallow levels (vowel and color) because they produced equivalent results across all conditions in both Experiment 1 and Experiment 2. (We had initially hypothesized that the two superficial levels might lead to different levels of accurate and false recall; however, they did not differ and will subsequently be viewed as replications of each other.)

2. The d' value was calculated in the usual manner for studied items. For critical nonpresented items, the false-alarm rate was compared with the unrelated false-alarm rate (as if the critical nonpresented false-alarm rate were actually a hit rate). Thus, this analysis was a bit unconventional: d' was computed to compare two different classes of false alarms (see Koutstaal, Schacter, Galluccio, & Stofer, 1999, and Schacter, Israel, & Racine, 1999, for similar transformations, using A' to compare recognition of studied and critical items). Values of 1 and 0 lead to indeterminate calculations of d' and were therefore replaced with .99 and .01.