Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall

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

The *testing effect* is the finding that taking an initial test enhances the likelihood of later recall. The present report examines the extent to which this benefit of testing comes with a cost: an enhanced likelihood of erroneously recalling incorrect information. Subjects were given short lists of semantic associates (e.g., *hill, valley, climb*); each list converged upon a related nonpresented word (e.g., *mountain*). After presentation of some lists, the subjects received no initial test; after others, one initial free recall test; and after others, three successive free recall tests. The probabilities of final free recall (and the probability of reporting vivid recollection of the moment of encoding) of both studied and related, nonstudied words (e.g., *mountain*) were highest when three initial tests had been taken, intermediate following one initial test, and lowest when no initial test had occurred. The beneficial effects of testing carry the cost of increases in erroneous memory for related information.

Experiencing a momentous or noteworthy event can lead one to recount the story repeatedly. Each subsequent retelling can be considered a recall test, in which recollection of the experienced event is blended with the occasional inference (in the service of telling a coherent story) or embellishment (with the goal of spicing up the story to hold the attention of the listener; see Tversky & Marsh, 2000). How do such retellings influence one's later memory for the original event? Do embellishments become incorporated into memory? Of interest in the present report is how repeated retrieval attempts may alter one's memory for related events that did not actually occur. Specifically, memory for encoded information and for associatively related (but not presented) information is examined as a function of whether (and how often) the event was recalled in the intervening retention interval.

Prior Testing Enhances the Amount of Accurate Information Later Remembered

The idea that an initial memory test can help to immunize against forgetting has a long history. This phenomenon is known as the *testing effect* (Glover, 1989; Spitzer, 1939). To the extent that a retrieval episode is an important determinant of later memory performance, the number of retrieval tests (not simply the presence or absence of an initial retrieval) should exert an influence on later memory. This, too, has been investigated. For example, Glover

found that the greater the number of initial tests taken, the greater the enhancement to a final criterial test of free recall, at least when the initial tests were spaced apart. Wheeler and Roediger (1992) showed that spacing out the initial tests is not necessary, although they investigated forced recall (not free recall). That is, after studying 60 pictures, subjects were required to produce 60 responses (corresponding to names of the pictures) during the initial and the final criterial tests. The likelihood of later producing a name corresponding to a studied picture was greatest in a condition in which three initial tests had been taken, intermediate in a one-test condition, and lowest in a no-test condition. The use of forced recall in the final criterial test, however, makes interpretation of these results difficult. Were the subjects on the final forced recall test simply *producing* more studied items after having taken more tests (due to increased guessing or other non–memory-based processes), or were these responses based on an increase in recollection of having seen those items?

Prior Testing May Enhance Later Memory for Erroneous Yet Plausible Information

Less clear than the influence of testing on later accurate memory is the role that initial tests can play in fostering later memory for nonencountered events, facts, or aspects of a studied situation (i.e., what is sometimes termed *false memory*). Consider first the effect of taking a single initial test on later memory performance (relative to no initial test). Schooler, Foster, and Loftus (1988) showed that if an initial forced choice recognition test contains only distractors (i.e., no studied item), the process of choosing one of the distractors can enhance the likelihood of choosing that same distractor later, in a situation in which the studied item is indeed present (cf. Roediger & Marsh, 2005). Similarly, Roediger, Jacoby, and McDermott (1996) found

that when misleading postevent information was followed by an initial test, the later acceptance of that misinformation as having been experienced was enhanced (relative to a condition in which no initial test occurred). Similar findings have been reported in the semantic memory domain (Bacon, 1979; Brown, Schilling, & Hockensmith, 1999).

Roediger and McDermott (1995) showed that the testing effect could occur for false recognition in the absence of any overt misleading information. After studying 15 semantic associates (e.g., *hill*, *valley*, *climb*), people either took an initial free recall test (see Deese, 1959) or performed a distractor task. On a subsequent final recognition test covering multiple lists, the false alarm rate for related, nonstudied associates (e.g., *mountain*) and the hit rate were enhanced for the condition in which a prior test had occurred. Furthermore, the likelihood of claiming to be able to remember the specific instance of presentation of the nonpresented word (in the remember/know paradigm; Tulving, 1985) was enhanced by prior testing. One cautionary note, though, is that the replicability of this finding has been variable (e.g., Norman & Schacter, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger, McDermott, Pisoni, & Gallo, 2004; Schacter, Verfaellie, & Pradere, 1996; Spitzer, 1939), despite the fact that multiple other findings from this article have been replicated dozens of times (see McDermott & Watson, 2001; Roediger, McDermott, & Goff, 1997; Roediger, McDermott, & Robinson, 1998). The effect of an initial free recall test on later false recognition of semantic associates appears intermittently.

Using a variant of the procedure described above (sometimes referred to as the *DRM procedure*), McDermott (1996) showed that on a final free recall test given 2 days after initial encoding, the likelihood of recalling a nonstudied associate (e.g., *mountain*) was twice as great when a test had occurred after encoding of each list than when it had not. The effect here was robust, which may be attributable to the use of free recall as the final criterial measure (instead of recognition; Darley & Murdock, 1971; Richardson, 1985; Runquist, 1983).

If the effects of a single test on later false memory have been variable, what about manipulations of the number of initial tests? This question is not well studied and is the primary focus on the present experiment. Henkel (2004) showed that when people were asked to imagine having perceived pictures and were then given successive recall tests, the likelihood of mistakenly classifying previously imagined items as having been perceived increased (see, also, Goff & Roediger, 1998; also, Hyman, Husband, & Billings, 1995). Hence, Henkel concluded that "repeated attempts at remembering . . . may create a fertile environment for source misattributions" (p. 27), which could include remembering what was only thought, inferred, or imagined as having overtly occurred (Johnson, Hashtroudi, & Lindsay, 1993).

Changes in Recall Also Occur Across Successive Initial Tests

Although initial testing has been discussed thus far as an independent variable—with the number of prior tests being manipulated and examined on a final criterial test—one can also examine how recall patterns change across those initial recall tests in the conditions in which multiple tests occur. The probability of accurate recall sometimes increases across initial tests, a phenomenon known as *hypermnesia* (Erdelyi & Becker, 1974; Payne, 1987; Roediger & Thorpe, 1978). Preliminary evidence suggests that a similar finding can occur for related, nonpresented information as well (Payne et al., 1996).

One concern with studies investigating hypermnesia, or the increase in recall across successive, identical tests, is that subjects may loosen their criteria across tests. That is, by the third test, what is reported may be some blend of what is recalled and what seems like a reasonable guess. This concern is not so much relevant to the issue of looking at the effects of initial recall on later recall but is of critical importance in trying to interpret effects across the initial tests.

One potential solution is to give forced recall tests (Erdelyi & Becker, 1974), on which people are required to produce a fixed number of items on the recall test, even if that process entails making educated guesses. Although this procedure makes sense from a methodological perspective—to control recall criteria—it points to an interesting question: What is the effect of knowingly embellishing and guessing on later recall? Does generation of guesses lead to later source-monitoring difficulties, with the number of times one guesses enhancing the likelihood of later false memory?

The Goals of the Present Experiment

The experiment reported herein was designed to examine the effect of repeated retrieval attempts on the likelihood of later memory for nonpresented items. Specifically, the number of initial tests given to the subjects after each list (zero, one, or three) was manipulated, followed by a final free recall test. The primary question of interest was whether recall of the nonpresented associates would vary with prior testing history.

Also of interest was whether requiring subjects to guess (i.e., to produce 15 words) on the initial (one or three) tests would exacerbate the effects of testing on later false recall (relative to the case in which the initial tests were free recall tests, with warnings against guessing). Prior work in other paradigms has suggested that guessing initially can increase the likelihood of later false recall (Ackil & Zaragoza, 1998; Henkel, 2004; Roediger et al., 1996; Roediger, Wheeler, & Rajaram, 1993).

A third goal was to explore whether prior testing would affect peoples' likelihood of claiming to be able to vividly recollect the specific moment of presentation of studied and of related yet nonstudied words. A variant of the remember/know procedure was used (Tulving, 1985; see Gardiner & Richardson-Klavehn, 2000, for a review). Unlike most experiments in which the remember/know procedure has been used, we used this procedure with free recall, instead of recognition (see, also, Hamilton & Rajaram, 2003; Tulving, 1985). Recently, Hamilton and Rajaram have demonstrated that not all recalled words on a free recall test

are given *remember* responses and have argued that this procedure could be used in an effort to purify measures of recollection (by simply looking at responses claimed to be remembered).

METHOD

Subjects

Sixty-four undergraduate students participated for credit toward a course research requirement.

Design and Materials

A 3 (number of initial tests: zero, one, or three; within-subjects) × 2 (recall instructions: free or forced; between-subjects) mixed design was used.

Twenty-four lists of semantic associates from Roediger and Mc-Dermott's (1995) article served as materials. Lists were divided into four groups of six for counterbalancing purposes; each set of six lists was rotated through the four conditions (not presented at study, presented once but not tested, presented once and tested once, and presented once and tested thrice) an equal number of times across subjects. Lists were recorded digitally in a female voice in such a way that the average presentation time for each list of 15 words was 15.6 sec (ranging from 14.5 to 17.5 sec). The lists were presented via stereo speakers.

Procedure

Within each experimental session, 5 subjects or fewer were tested.

Initial tests. The subjects were told that they would hear 18 short lists of words presented via computer speakers and were asked to pay close attention to the lists, because their memory for the lists would sometimes be tested (from one to three times). A 30-sec distractor task followed each list (to eliminate the recency effect; Glanzer & Cunitz, 1966); during this 30-sec task, the subjects were required to count the number of even digits in an auditory sequence. Following this task, the subjects heard either the next list (in the no-test condition) or the word "recall," signaling them to recall the last list that they had heard. The subjects did not know whether a given list would be tested until after the distractor task.

In the free recall condition, the experimenter stressed to the subjects that they should recall as many words as possible without guessing. Ninety seconds into the recall period, they heard a tone, which signaled that they had only 30 sec remaining and that they should draw a line under the last word produced thus far and continue to try to recall more words.

The subjects in the forced recall group were instructed to spend the first 1.5 min recalling only words presented in the immediately preceding list. The tone occurring after 90 sec was their signal to draw a line and begin guessing, producing words that could plausibly have been present in the list. They were asked to generate as many guesses as necessary to fill in all 15 slots. Note that both the free and the forced recall groups performed identical tasks for the first 90 sec of the recall period for each list; the groups differed only in the aim of the final 30 sec of each recall period.

The subjects were further told that sometimes, upon completing a recall test, they would be asked to recall the list again for another 2 min and that they would sometimes recall a list as many as three successive times without any further presentations of the list or any feedback. It was stressed that the entire allotted time should be used for recall. On every test, a tone sounded when 30 sec remained, at which time the subjects drew a line under the last word recalled. The one-test and three-test conditions differed only in that following Test 1, people were sometimes asked to recall the list again and sometimes moved on to the next study list. It was never stated that a third initial test always followed the second, but it was indeed the case that no two-test condition was used.

Final free recall. After presentation of the 18 lists and the accompanying recall sessions, the subjects received instructions for the final free recall test and an explanation of the *remember/know* distinction.

All the subjects were given 15 min in which they were to recall all the words they could retrieve from the study lists *but were not to guess*. In addition, as the subjects recalled the words, they were asked also to classify each word with respect to whether they *remembered* its original occurrence in the lists or they simply *knew* it had occurred (Rajaram, 1993; Tulving, 1985). Detailed instructions and examples were given to ensure that the subjects understood the distinction. In summary, the subjects were told that remember judgments should be reserved for words for which they held a specific, vivid memory of the occurrence of the word on the speakers. For example, they were told that if they recollected something distinctive in the speaker's voice when she said the word or if they recalled the word that had preceded or followed it, they should place an R (for *remember*) by the word. If, however, they were certain that a word had occurred on the list but lacked the feeling of being able to mentally reexperience the exact moment of presentation, they should place a K (for *know*) beside the word. It was emphasized that the remember/know distinction should be made with respect to the original presentation episode (i.e., the question was not whether the subjects remembered or knew they had previously recalled the word). The subjects were encouraged to ask questions to clarify the remember/know distinction.

The entire experimental session took approximately 1 h 45 min.

RESULTS

The primary data of interest are the final free recall probabilities for the various test conditions. Data from initial recall tests, although primarily an independent variable, will also be reported in brief. The alpha level was set at $p < .05$, and partial eta squared (*pes*) indicates effect size.

Initial Tests

First, the data from the free recall portion (the data from the first 90 sec) for the conditions in which three successive initial tests were taken will be considered.[1] The free/forced variable exerted no significant effects (and entered into no interactions) for the first 90 sec, and therefore, the data in Table 1 collapse across this variable. Studied words were recalled more often than nonstudied associates [$p = .57$ and $.48$, respectively; $F(1,62) = 6.84$, *pes* $= .10$]. More important, an interaction appeared such that the likelihood of recalling studied words decreased slightly across Tests 1–3, whereas the opposite pattern occurred for critical nonstudied associates [$F(2,124) = 6.93$, *pes* $= .10$].

Next, the *increment* in recall for the various conditions in the final 30 sec of the recall periods (not shown in Table 1) will be considered. The expected finding was obtained: The subjects in the forced recall condition produced the critical nonstudied associates with a high probability (.26) during the last 30 sec of the recall period, when they were asked to guess. They did not, however, generate many more studied items in this period (.06). Furthermore, the subjects in the free recall condition did not produce many items during these last 30 sec (less than .01 proportion of studied and critical items).

Table 1
Recall Probabilities on the Initial Tests as a Function of
Test Number for the Condition in Which Three Successive
Tests Were Taken (With Standard Errors)

| | Item Type | | | |
| | Studied Words | | Nonstudied Associates | |
Test Number	M	SE	M	SE
1	.578	.013	.461	.030
2	.563	.013	.482	.033
3	.558	.015	.497	.031

Final Free Recall

The effect of number of tests on final recall was of primary interest. As can be seen in Figure 1, testing history did affect the likelihood of recall of both studied words and critical nonstudied associates. The probability of recalling a studied word from a list not previously tested, tested once, or tested three successive times was .21, .33, and .44, respectively. The probability of recalling a critical nonstudied associate from the zero, one, or three prior test conditions was .27, .35, and .37. Thus, testing had a profound effect on the likelihood of later recall. This pattern appeared both for the subjects who had previously been asked to generate guesses on the initial recall tests (the forced recall condition) and for the subjects who had been cautioned against guessing (the free recall condition); for this reason, Figure 1 collapses across this variable.

A 3 (number of prior tests) $\times$ 2 (item type) $\times$ 2 (instructions for prior test) ANOVA was in line with the conclusions above. Specifically, two significant effects emerged: a main effect of the number of prior tests [$F(2,124) = 33.55$, $pes = .35$] and an interaction between item type
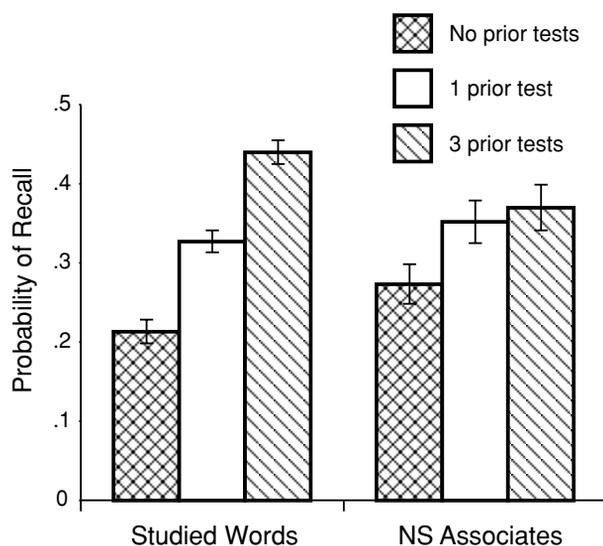

Figure 1. The proportion of studied words and nonstudied (NS) semantic associates recalled on the final free recall test as a function of number of prior tests. Error bars display the standard error of the mean.

and the number of prior tests [$F(2,124) = 8.86$, $pes = .13$]. The free/forced variable did not exert a main effect, nor did it interact with any variables (all Fs < 1.4).

Collapsing across other variables, the likelihood of recall for the no prior tests, one prior test, and three prior tests conditions was .24, .34, and .41, respectively. The item type $\times$ number of prior test interaction demonstrates that this testing effect was more pronounced for studied items than for critical items. Follow-up ANOVAs on the studied items and the nonstudied associates separately demonstrated that each of them exhibited reliable simple main effects of number of prior tests [$F(2,124) = 119.2$, $pes = .66$, for studied words, and $F(2,124) = 4.84$, $pes = .07$, for critical nonstudied words].

The remember responses will now be considered. Recall that the subjects were asked to assign a remember or know tag to each recalled word. Remember responses were to be reserved for situations in which the person could vividly recollect the presentation of the word during the study phase (i.e., not whether they remembered having recalled it on a prior test). To the extent that remember judgments tap the same recollective processes as free recall, we would expect to obtain almost exclusively remember responses; furthermore, we would expect the independent variable to exert qualitatively similar effects on the remembered words and on the overall free recall data. Due to difficulties with independence assumptions surrounding remember/know data, the know data are not analyzed; instead, *remembering* is examined as a purified form of recollection.

Figure 2 shows the proportion of words both recalled and given remember judgments as a function of number of prior tests and item type. The probability of a remember judgment increased with prior testing (.14, .20, .28 for studied items and .13, .13, .22 for nonstudied associates for the zero, one, and three prior testing conditions, respectively). The lack of a difference in remember judgment for nonstudied associates as a function of whether no or one initial test had occurred is puzzling, and replication is probably warranted before too much is made of this null effect.

A 3 $\times$ 2 $\times$ 2 ANOVA comparable to that done on the overall recall data was performed on the remember data. This analysis manifested a reliable main effect of number of prior tests [$F(2,124) = 25.84$, $pes = .29$], a reliable main effect of item type [$F(1,62) = 6.80$, $pes = .10$], and an interaction between item type and number of prior tests [$F(2,124) = 4.15$, $pes = .06$]. The greater the number of tests intervening between the encoding and the final retrieval episodes, the higher the probability that people would claim to be able to remember the *initial study* episode (probabilities of .14, .17, and .25 for zero, one, and three prior tests, respectively). Furthermore, the subjects were more likely to claim to remember the moment of presentation of studied words (.21) than of nonstudied associates (.16). The interaction of the two variables emerged because the effect of prior testing was greater for studied words than for nonstudied associates. An examination of simple effects showed that the number of prior tests exerted a reliable effect on both studied words [$F(2,124) = $
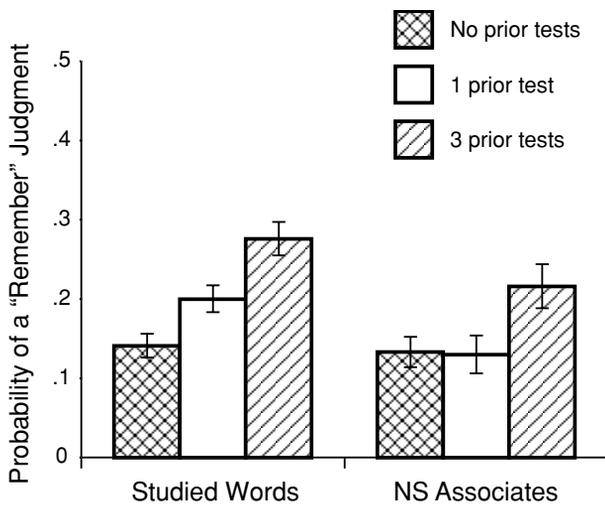
Figure 2. The proportion of studied words and nonstudied (NS) semantic associates assigned remember judgments on the final free recall test as a function of number of prior tests. Error bars display the standard error of the mean.

49.89, *pes* = .45] and nonstudied associates [$F(2,124)$ = 8.11, *pes* = .12].

Contingency analyses reveal much the same picture. For example, consider the words recalled on the final test after three initial free recall tests had been taken. Often, they had been produced on all three initial tests (p = .92 and .80 for studied words and nonstudied associates, respectively); rarely did words appear on the final recall test without first being recalled on at least one of the three initial tests (p = .01 and .08 for studied words and nonstudied associates, respectively). These data, too, suggest that the link between initial recall and final recall, although similar for studied and critical nonstudied associates, was tighter for studied words. That is, final free recall of nonstudied associates was less tied to whether the individual word had been initially recalled than was final free recall of studied words. Similar conclusions hold when remember and know contingency data are examined.

DISCUSSION

Five key findings emerged from this data set. First, the number of initial recall tests taken influenced the likelihood that the subjects would later recall words that had been studied and, importantly, words semantically associated to studied words but that were not themselves studied. Second, an examination of the proportion of words that the subjects claimed to be able to recollect from the initial study episode (i.e., those receiving remember judgments) showed qualitatively the same pattern: Prior testing enhanced the likelihood of reports of remembering, for both studied and related, nonstudied words. Third, the effect of prior tests was more pronounced for studied words than for nonstudied associates (when measured both by recall probability and by remember judgments). Fourth, the remember/know data for free recall revealed that the

subjects did not claim to be able to recollect the presentation of every word recalled; instead, they gave the alternative, know judgments, with a moderate frequency. Fifth, in the condition in which the subjects took three initial tests, forgetting across the three tests was observed for studied words, whereas the critical nonstudied associates were recalled with increasing frequency across tests. The relation of these findings to those in the literature and the broader implications of this work will now be considered.

The finding that taking a test enhances the likelihood of later recognition endorsement has been variable and inconsistent with respect to both accurate recognition (Darley & Murdock, 1971) and false recognition of related lures (Roediger et al., 1998). The finding that testing affects later recall, however, has been more readily exhibited for both accurate recall (Darley & Murdock, 1971) and false recall (McDermott, 1996). Roediger et al. (1998) have suggested that effects of testing are likely present in both cases but are much more easily observed on free recall tests, which are often more sensitive measures than are recognition tests. The present results extend prior work by showing that it is not mere presence/absence of testing that matters but that the number of tests taken mediates the effect.

An unexpected finding was that forcing guesses on initial tests did not influence later probabilities of recall. It had been anticipated that encouraging the subjects to guess (in the forced recall condition) would enhance the likelihood that they would later recall their guesses as having been studied (Ackil & Zaragoza, 1998; Henkel, 2004; Payne et al., 1996; Roediger et al., 1993) and may have exacerbated the effects of prior testing. In hindsight, the specific version of forced recall used may have been crucial. Instead of the typical forced recall procedure, in which subjects are given a certain amount of time to recall a given amount of information (even if guessing is required), a slightly different procedure was used in the present experiment. That is, the subjects were asked to perform free recall for the first 90 sec and then to guess, as needed, for the final 30 sec in order to fill in a total of 15 words. The process of withholding guesses until the final 30 sec and demarcating those guesses may have helped the subjects tag guesses as such, so that those guesses were later less confusable with studied words than they would have been if the guessing had been intermixed with recollection throughout the recall period (see Henkel, 2004, for a similar idea).

One might wonder why this specific forced recall procedure was utilized. The idea was that a comparison of the tests for the three-test condition might have produced interesting data whereby the subjects would (for example) have guessed *mountain* on Test 1 (in the last 30 sec of the recall period) and then transferred that response to the beginning (free recall) portion of their recall on the second and third tests. This pattern of greater increases in false recall on Tests 2 and 3 for the forced recall group than for the free recall group was not observed, however.

A finding that at first glance may appear surprising is that no hypermnesia was found (for studied items) across

the initial three tests; instead, forgetting was observed. Hypermnesia is known to be a reliable yet somewhat fragile phenomenon, appearing more often with pictures and imaginable stimuli than with abstract words, with longer lists (rather than with shorter lists), and with lengthy recall periods (Payne, 1987; Payne et al., 1996). In the present case, short lists of words (few of which were concrete) were used. Against this backdrop, it is not terribly surprising that the present results were in line with the very similar results in Payne et al. in showing slight levels of forgetting across the three initial tests (although see Brainerd, Payne, Wright, & Reyna, 2003, for a different pattern). The finding of decreases in accurate recall across initial tests makes the increase in false recall across those tests all the more impressive, however.

The remember judgments given on the final free recall test are also noteworthy, especially in light of the recent discussion by Hamilton and Rajaram (2003), who observed that

the free recall task, considered to be the quintessential measure of conscious recollection, also elicited Know responses that represent a failure to evoke a specific and vivid recreation of the study episode . . . Thus, contrary to the widespread assumption, it appears that a "pure" measure of conscious recollection may be more accurately provided by Remember responses and not by other measures of explicit memory including free recall. (p. 53)

Note that their point of view is that remember responses represent purified recollection and that free recall, which is often also taken as such, is not made up solely of remember experiences. Only 66% of the recalled words in their experiment were given remember judgments by subjects (cf. Tulving, 1985, whose subjects remembered 88% of all recalled words). In the present study, approximately 64% of all the *recalled* studied words were assigned remember responses, whereas the percent of *recalled* nonpresented associates given a remember judgment was about 48%. Not all words produced on a free recall test are accompanied by vivid recollection.

The present data, along with those in Hamilton and Rajaram (2003), add to a recent, growing literature suggesting that dual-process models of retrieval might be well suited to free recall (Brainerd et al., 2003; Brainerd, Wright, Reyna, & Payne, 2002). To summarize, dual-process theories of recognition have long held that the process of determining whether a recognition probe was previously studied or not can be driven by a fast-acting heuristic process of familiarity assessment (Atkinson & Juola, 1973; Jacoby & Dallas, 1981; Mandler, 1980). When this process returns very high or very low feelings of familiarity, a quick decision is made. When intermediate levels of memory strength (or familiarity) are returned, though, a second, slower, more controlled process kicks in, often called *recollection*. This more deliberate decision process evaluates whether the evidence strength is sufficient to classify an item as having been previously studied.

In comparison with dual-process theories of recognition, the recent dual-process theories of free recall have some similarities and some important differences (Brain-

erd et al., 2003; Brainerd et al., 2002). The conceptualization by Brainerd, Reyna, and colleagues under the rubric of fuzzy trace theory posits that the initial, fast operation in recall is the "direct access" operation, which "predominates at the start of output and provides direct access to verbatim traces of target presentations" (Brainerd et al., 2002, p. 121). The second operation, "which is slower and less accurate, waxes as recall proceeds. This operation *reconstructs* targets by processing gist traces of the meaning content of studied lists" (Brainerd et al., 2002, p. 121).

As was noted by Brainerd et al. (2002), this new conceptualization offers a marked departure from classic approaches to understanding free recall. Consider the classification of the second, slower process in recognition (referred to above as *recollection*); it has been also referred to as being *recall* or *recall-like* (Horton, Pavlick, & Moulin-Julian, 1993; Mandler, 1980). The implication of this conceptualization is that *recall* is a single process and that this process represents both one component of recognition and the sole retrieval process used in free recall.

The data in the present experiment do not provide a rigid test of the dual-process theory of free recall (nor were they designed to); they are, nonetheless, fully consistent with the idea in several respects. First, the finding of high probabilities of free recall of nonpresented words (McDermott, 1996; Roediger & McDermott, 1995) was one of the impetuses for the dual-process theory of recall put forth by Brainerd, Reyna, and colleagues. More novel is the finding that a large proportion of words inaccurately recalled on the final free recall test were labeled *remembered*. That is, the subjects claimed to remember the specific moment of presentation of nonpresented words that they recalled. This finding is reminiscent of the remember judgments previously seen on recognition tests (e.g., Roediger & McDermott, 1995). Finally, the finding that not all studied words recalled were remembered (but instead, some were simply *known* to have been presented) is consistent with the idea that recall has two bases: one underlying vivid conscious recollection and the other a more general familiarity.

In conclusion, it appears that recall of a list of words (and by extension, repeated retellings of a story) indeed affects one's later memory. The effect is facilitative in the sense that a greater amount of *correct* information is produced, but this increase comes at the cost of a greater amount of erroneous information. As expressed by Henkel (2004),

Repeated attempts to remember as much information as possible involve many different cognitive processes, such as retrieving information that comes to mind readily, searching for additional information, and evaluating that information, any of which may involve imagination, embellishment, and reconstructive processes. The repetition that occurs across multiple retrieval efforts may serve to strengthen the confidence and vividness with which one both veridically and falsely recollects information. (p. 42)

REFERENCES

ACKIL, J. K., & ZARAGOZA, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology*, **34**, 1358-1372.

ATKINSON, R. C., & JUOLA, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583-612). New York: Academic Press.

BACON, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 241-252.

BRAINERD, C. J., PAYNE, D. G., WRIGHT, R., & REYNA, V. F. (2003). Phantom recall. *Journal of Memory & Language*, **48**, 445-467.

BRAINERD, C. J., WRIGHT, R., REYNA, V. F., & PAYNE, D. G. (2002). Dual-retrieval processes in free and associative recall. *Journal of Memory & Language*, **46**, 120-152.

BROWN, A. S., SCHILLING, H. E. H., & HOCKENSMITH, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology*, **91**, 756-764.

DARLEY, C. F., & MURDOCK, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, **91**, 66-73.

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

ERDELYI, M. H., & BECKER, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology*, **6**, 159-171.

GARDINER, J. M., & RICHARDSON-KLAVEHN, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 229-244). Oxford: Oxford University Press.

GLANZER, M., & CUNITZ, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, **5**, 351-360.

GLOVER, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, **81**, 329-399.

GOFF, L. M., & ROEDIGER, H. L., III (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition*, **26**, 20-33.

HAMILTON, M., & RAJARAM, S. (2003). States of awareness across multiple memory tasks: Obtaining a "pure" measure of conscious recollection. *Acta Psychologica*, **112**, 43-69.

HENKEL, L. A. (2004). Erroneous memories arising from repeated attempts to remember. *Journal of Memory & Language*, **50**, 26-46.

HORTON, D. L., PAVLICK, T. J., & MOULIN-JULIAN, M. W. (1993). Retrieval-based and familiarity-based recognition and the quality of information in episodic memory. *Journal of Memory & Language*, **32**, 39-55.

HYMAN, I. E., HUSBAND, T. H., & BILLINGS, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology*, **9**, 181-197.

JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.

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

MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.

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

MCDERMOTT, K. B., & WATSON, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory & Language*, **45**, 160-176.

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. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, **101**, 5-27.

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.

RICHARDSON, J. T. E. (1985). The effects of retention tests upon human learning and memory: An historical review and an experimental analysis. *Educational Psychology*, **5**, 85-114.

ROEDIGER, H. L., III, JACOBY, J. D., & MCDERMOTT, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory & Language*, **35**, 300-318.

ROEDIGER, H. L., III, & MARSH, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 1155-1159.

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., & GOFF, L. M. (1997). Recovery of true and false memories: Paradoxical effects of repeated testing. In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 118-149). Oxford: Oxford University Press.

ROEDIGER, H. L., III, MCDERMOTT, K. B., PISONI, D. B., & GALLO, D. A. (2004). Illusory recollection of voices. *Memory*, **12**, 586-602.

ROEDIGER, H. L., III, MCDERMOTT, K. B., & ROBINSON, K. J. (1998). The role of associative processes in producing false remembering. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187-245). Hove, U.K.: Psychology Press.

ROEDIGER, H. L., III, & THORPE, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition*, **6**, 296-305.

ROEDIGER, H. L., III, WHEELER, M. E., & RAJARAM, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 97-134). New York: Academic Press.

RUNQUIST, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, **11**, 641-650.

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.

SCHOOLER, J. W., FOSTER, R. A., & LOFTUS, E. F. (1988). Some deleterious consequences of the act of recollection. *Memory & Cognition*, **16**, 243-251.

SPITZER, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, **30**, 641-656.

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

TVERSKY, B., & MARSH, E. J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, **40**, 1-38.

WHEELER, M. A., & ROEDIGER, H. L., III (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, **3**, 240-245.

NOTE

1. The initial recall results when only one test was taken do not differ statistically from the results of Test 1 in the three-test condition and so will not be considered further. There should have been no difference, because the only difference between the conditions was whether other tests followed the initial recall for a list, and the subjects did not know whether additional tests would occur until after the first test had been completed.