MEMORY, 2004, *12* (6), 748–761

The role of rehearsal and generation in false memory creation

Elizabeth J. Marsh
Washington University in St. Louis, MO, USA

Gordon H. Bower
Stanford University, CA, USA

The current research investigated one possible mechanism underlying false memories in the Deese-Roediger-McDermott (DRM) paradigm. In the DRM paradigm, participants who study lists of related words (e.g., "table, sitting, bench . . .") frequently report detailed memories for the centrally related but non-presented critical lure (e.g., "chair"). One possibility is that participants covertly call to mind the critical non-presented lure during the study phase, and later misattribute memory for this internally generated event to its external presentation. To investigate this, the DRM paradigm was modified to allow collection of on-line thoughts during the study phase. False recognition increased following generation during study. False recognition also increased following study of longer lists; this effect was partially explained by the fact that longer lists were more likely to elicit generations of the critical lure during study. Generation of the lure during study contributes to later false recognition, although it does not explain the entire effect.

Much enthusiasm has been generated for a straightforward word-learning paradigm that yields unusually high levels of false memories (see Roediger, McDermott, & Robinson, 1998, for a review). Roediger and McDermott (1995) resurrected and extended a paradigm first introduced by Deese (1959) in which participants study lists of related words such as *sour*, *candy*, *sugar*, *bitter*, *good*, *taste*, *tooth*, *nice*, *honey*, *soda*, *chocolate*, *heart*, *cake*, *tart*, *pie*. Deese (1959) found that participants later recalled "sweet" at high levels even though it was not actually presented. Roediger and McDermott replicated this result and also showed that participants falsely recognised critical lures such as "sweet" (84%) as frequently as they correctly recognised old list items (86%). Perhaps most intriguing, however, was the finding that participants were extremely confident that the critical lure was presented, and claimed that they actually "remembered" its presentation rather than simply "knew" that it was presented (e.g., Gardiner, 1988). Further research has demonstrated that participants are quite willing to describe their false memories in detail. They believe that the critical lure was presented early in the list (Read, 1996) and they are willing to say which of two speakers spoke the word (Payne, Elie, Blackwell, & Neuschatz, 1996). It is as if they are remembering an actual event.

Correspondence should be addressed to Elizabeth J. Marsh, Psychological & Brain Sciences, Duke University, 9 Flowers Drive, Box 90086, Durham, NC 27708-0086, USA.

We are grateful to Roddy Roediger and Jessica Lang Nowinski for reading and commenting upon an earlier draft of this paper. Keith Payne provided both comments on the paper and assistance with the path analysis. John Seamon and an anonymous reviewer wrote thoughtful reviews that were extremely helpful in the revision of this paper. We thank John Gabrieli, Kenneth Kurtz, Ewart Thomas, and Barbara Tversky for useful comments concerning this research. Nicole Dudukovic, Cheryl Galvin, Katie Hill, Nnanna Ogbu, and Onuoma Omordia helped to collect and code some of the data reported in this paper. The first author was supported by an NRSA postdoctoral fellowship from the National Institute of Mental Health (NIMH), #1F32MH12567.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210344000170

A possible explanation for a memory illusion that seems phenomenologically real is that participants *are* actually remembering a real event. However, this event was generated internally (e.g., a thought) rather than presented externally. Thus, the memory error in the Deese-Roediger-McDermott (DRM) paradigm may actually be a source error: people misattribute internal events to external perception (Johnson & Raye, 1981). Underwood (1965) first advanced the idea of a stimulus-driven internal event, or Implicit Associative Response (IAR), to explain why participants falsely recognise associates of presented words at a higher rate than unrelated control words. As described by Underwood, IARs occur during the study phase in response to the to-be-remembered words. While studying a list of words, associates come to mind and these thoughts may later be misattributed to perception. This idea is consistent with the finding that words that come to mind easily in free association tasks also elicit more incorrect ''old'' responses on recognition memory tests (Rubin, 1983).

Several findings support the idea that participants generate associates during the DRM study phase. First, participants do report thinking of the critical lures during the study phase. When asked to estimate the amount of covert rehearsal of words during study, participants report rehearsing the critical lures as much as studied words. They report lower rates of rehearsal for non-critical lures (Mather, Henkel, & Johnson, 1997).

Second, the memories for these hypothesised generations of the critical lures behave similarly to internal memories studied in other paradigms. For example, similar levels of priming on implicit memory tests are observed for critical lures and internal memories created in ''read-generate'' paradigms. In ''read–generate'' paradigms, participants create internal memories by covertly generating words in response to cues (e.g., Jacoby, 1983; Slamecka & Graf, 1978). For example, in response to the cue hot–c__, subjects would either read or generate the opposite ''cold''. On tests requiring completion of word-stems and word-fragments with the first words that come to mind, previously read words are produced at a higher rate than baseline non-studied words. Previously generated words are produced at a rate intermediate between read and non-studied words (Blaxton, 1989; Masson & MacLeod, 1992). When DRM lists are used, studied list words yield standard perceptual priming effects, non-studied words yield no priming, and critical lures yield intermediate levels of perceptual priming (McDermott, 1997). The finding of intermediate levels of priming for critical lures suggests that they were generated during study, since never-studied (and never-generated) events do not lead to priming.

Similarly, the phenomenal characteristics of memories for the critical lures are more like memories of imagined events than external events. Real events are remembered as having more perceptual, spatio-temporal, and emotional content than are imagined events, whereas memories of imagined events are more associated with the processing involved in their generation (Hashtroudi, Johnson, & Chrosniak, 1990; Johnson, Foley, Suengas, & Raye, 1998b; Schooler, Gerhard, & Loftus, 1986; Suengas & Johnson, 1988). Thus, it is not so much that imagined memories contain less information but rather that they are characterised by different kinds of information. Accordingly, Mather et al. (1997) found that memories of critical lures were less associated with sound and feelings than were memories for real events; in fact, memory errors were reduced when the memory test made this qualitative difference salient to participants. According to the source monitoring view, the more different internal events are from perception, the less likely they are to be misattributed to perception (e.g., Johnson, Foley, & Leach, 1988a; Lindsay, Johnson, & Kwon, 1991). Spoken words are more similar to thoughts than are written words. Accordingly, auditory presentation sometimes increases the occurrence of false memories, at least on visual tests (Gallo, McDermott, Percer, & Roediger, 2001a; Kellogg, 2001; Smith & Hunt, 1998; Vogt & Kimble, 1978). Similarly, false alarms should decrease when external sources are less similar to internal events. Thinking of a word is not similar to looking at a picture; accordingly, studying pictures of list items decreases false alarm rates (Israel & Schacter, 1997).

The Source Monitoring Framework (Johnson & Raye, 1981; Johnson, Hashtroudi, & Lindsay, 1993) easily accounts for the unusually high levels of false alarms in the DRM paradigm. In general, internally generated memories are better remembered than are externally presented memories (Slamecka & Graf, 1978), so it is not surprising that if lures are internally generated during encoding they would be well remembered in the DRM paradigm. In addition, DRM lists contain many associates of the critical lure, providing multiple opportunities for the generation of a given internal associate (event). Repeated generation of the internal event should lead to an even stronger memory trace (e.g., Johnson, Raye, & Durso, 1980), and may also lead to a boost in the item's familiarity (Jacoby, 1999). This strong memory trace may be misattributed to perception, especially if participants are not explicitly oriented towards assessing the source of their memories (e.g., R. Marsh, Landau, & Hicks, 1997). Recognition memory tests typically encourage participants to label items as "old" without emphasising the possible sources (e.g., internal or external) of their memories. Correctly retrieving an internal event but failing to attribute it to internal sources boosts levels of false alarms. Consistent with this idea, Goff and Roediger (1998) found that the more frequently participants imagined doing some action, the more likely they were later to claim that they had actually performed the actions.

How can one study whether false memories in the DRM paradigm involve the misattribution of internal events to perception? In general, because the hypothesised internal sources are by definition covert, the first step is to devise a way to observe, measure, and manipulate those internal events (as did Rundus & Atkinson, 1970, in their attempt to measure hypothesised covert rehearsals). Recently, two papers were published that used modified DRM procedures to measure thoughts of the critical lures during study. Goodwin, Meissner, and Ericsson (2001) told subjects to say "everything that comes to mind as you work on the tasks". Verbal reports were collected both during the study phase and also during the written recall test. Using path analyses, Goodwin et al. reached the conclusion that lures verbalised during study were likely to be falsely recalled (r = .319), supporting the idea that explicit generation of the lure during study can be linked to later false recall.

Seamon and colleagues reached a somewhat different conclusion about the role of rehearsal processes in false memory creation, titling their paper "Thinking of critical words during study is unnecessary for false memory in the DRM procedure" (Seamon, Lee, Toner, Wheeler, Goodkind, & Birch, 2002, p. 526). Their subjects were told to "think out loud by saying any word or thought that came to mind, even if it was unrelated to the word lists". Rehearsal of the critical lures increased significantly when presentation was slowed from 2 to 5 seconds. When the critical lure had been rehearsed, false recall increased at both presentation rates. However, false recall still occurred for never-rehearsed critical lures, and false recognition was unaffected by prior rehearsal.

Our own research addresses similar questions, with slightly different procedures. Our research was conducted prior to the publication of the Goodwin et al. and Seamon et al. studies, and so our studies were not directly motivated by their results. However, our results fit nicely with theirs; we return to the integration of the three studies in the discussion.

In order to observe, record, and manipulate the hypothesised internal events in the DRM paradigm, participants were instructed to make overt their thoughts during the study phase. In Experiments 1a and 1b, participants were directed to report the first two words that came to mind during study of each target word; of interest was whether subjects reported thinking of the critical non-presented words during the study phase. In Experiment 2, participants rehearsed all to-be-remembered words aloud; of interest were intrusions of non-presented words into the rehearsal set. Participants later took a recognition memory test that required them to remember whether each target word was externally presented. Of interest is (1) whether participants do indeed call to mind the critical lures during the study phase, and whether they do so at a higher rate than for non-critical lures, and (2) whether generation during study is related to later false recognition.

EXPERIMENT 1a LIST LENGTH AND FREQUENCY OF GENERATION

Various experimental manipulations may affect false alarm rates by changing the frequency with which the critical lure is called to mind. We examined the effect of list length on generations of the lures during study, and the consequent effect on false recognition. Robinson and Roediger (1997) found that studying a greater number of related words increased the likelihood that participants would falsely recognise the critical lure. In the current experiment, we tested the hypothesis that longer lists of related words would lead to more frequent generations of the critical lure, thus leading to stronger internal memories of the lure, which might be misattributed to perception. Participants studied lists of 3, 6, 9, and 12 words; they then took a recognition memory test. The control group studied the to-be-remembered words without explicitly recording any thoughts, as in a typical DRM experiment. The experimental group recorded their first two thoughts in response to each of the studied words, thus allowing us to examine the relationship between generation and false recognition.

Method

Participants. A total of 60 Stanford University students participated in the experiment, 40 in the Overt Thought condition and 20 in the Control condition. Each was paid $10 for his/her time, and was recruited via campus-wide fliers. Participants were tested either individually or in pairs.

Materials. A total of 20 lists from Roediger and McDermott (1995) were used: *anger*, *black*, *bread*, *cold*, *doctor*, *foot*, *fruit*, *girl*, *high*, *king*, *man*, *mountain*, *music*, *river*, *rough*, *slow*, *soft*, *spider*, *thief*, *window*. Each list consisted of associates of a critical non-presented lure. These lists were randomly assigned to five sets of four lists. Across subjects, these sets were rotated through five list lengths: 0, 3, 6, 9, or 12 words. A 0-word list was never presented. A 3-word list contained the top three associates from the list; the 6-word list contained the top six associates from the list, and so on. Each list was studied in blocked format such that all list items were studied sequentially, beginning with the highest associate. The same order of lists was used for all participants. Thus, five study booklets were constructed to counterbalance list length. Study booklets for the Overt Thought condition contained two lines following each target word, in order for subjects to record their thoughts. Study booklets for the Control condition simply contained the target words.

Recognition tests were blocked in the same order as presentation (following procedures of Roediger & McDermott, 1995). Following Roediger and McDermott (1995), the recognition test contained 140 items, of which 40 were list items, 20 were critical lures, 40 were related lures, and 40 were unrelated lures. The list items were high associates, chosen from position six or higher of the list. The related lures were from the Russell and Jenkins (1954) association norms, and were low, non-presented associates of the critical lure. The unrelated lures were not listed as associates of the critical lure in the Russell and Jenkins association norms, but were matched to the critical lure in word length and word frequency as defined by the Kucera-Francis frequency norms.

Procedure. Participants were told they would be participating in a memory experiment. They were instructed that they would be presented with a series of words, one at a time. They were told that each word would be printed on a booklet page. For subjects in the Control group, each page contained only a single printed word. Subjects were told simply to study the words for the upcoming test. For subjects in the Overt Thought condition, each page contained the printed-to-be-remembered word plus two lines for recording responses. They were instructed to write down the first two words they thought of in response to the target word. They were told that it was fine if multiple target words made them think of the same responses, and that they should report their natural first two responses rather than trying to be creative. All subjects were told not to turn the page to the next word until they heard a tone. Following the instructions, the experimenter started a computer program that beeped every 6 seconds. In this fashion, subjects completed all 120 pages in the booklet.

After the study phase, participants did unrelated visual tasks for 15 minutes. For example, they were asked to draw the stimulus pattern that would come next in a sequence of abstract, spatially related stimuli.

Following the delay, all participants took the recognition test. They were cautioned against guessing. Participants were instructed that they should only call a word ‘‘old’’ if it had been visually presented earlier. Subjects in the Overt Thought condition were explicitly told that words they had written in the booklet did not count as ‘‘old’’ words unless they had also been pre-printed in the booklet.

Results

Recognition. The recognition data are shown in Table 1. The false alarm rate to critical items from studied lists (M = .56, collapsing across the 3, 6, 9, and 12 word-list conditions) was lower than the hit rate (M = .89), $t(59)$ = 11.94, SEM = 03. However, it was substantially higher than the false alarm rate to critical items from non-studied lists (the zero list condition; M = .14), related lures from studied lists (M = .14), and unrelated lures

TABLE 1
Proportion of items called "old" on the recognition memory test, Experiment 1a

		List length				
		0	*3*	*6*	*9*	*12*
List items	Overt Thought	.05	.91	.90	.91	.89
	Control	.12	.87	.86	.88	.92
FA critical	Overt Thought	.10	.40	.61	.73	.67
	Control	.21	.39	.49	.53	.53
FA Related	Overt Thought	.02	.05	.04	.07	.26
	Control	.09	.15	.18	.14	.36
FA Unrelated	Overt Thought	.02	.03	.02	.03	.00
	Control	.12	.11	.09	.08	.08

from studied lists (M = .04) (all ps < .001). Thus there was substantial false recognition of critical lures in both study conditions.

In order to isolate the effect of having studied related words, the base-rate performance in the zero list condition was subtracted from performance in the studied list conditions. This corrects for base-rate differences in false alarm rates across lure type (e.g., the base-rate for critical lures was higher than for unrelated lures) and study condition (e.g., the base-rate was higher in the Control condition). Such correction procedures have been used in numerous DRM studies (e.g., Gallo et al., 2001a; Seamon, Luo, & Gallo, 1998). This correction was done on all dependent variables: hits, false alarms to critical lures, false alarms to related non-critical lures, and false alarms to unrelated non-critical lures.

A 2 (condition) × 4 (list length) ANOVA was computed on corrected recognition scores. There was no main effect of list length; participants were equally good at recognising studied words belonging to lists of 3, 6, 9, or 12 words (F < 1). Participants in the Overt Thought condition showed higher levels of corrected recognition than did Control participants, $F(1,58) = 3.69$, MSe = .11, p < .07. However, study condition did not interact with list length, $F(3,174) = 1.32$, MSe = .01.

We next considered false recognition of non-critical lures. A 2 (condition) × 4 (list length) ANOVA was computed on corrected false recognition of related lures. There was no main effect of condition (F < 1). There was a main effect of list length, $F(3,174) = 16.70$, MSe = .03. Subjects falsely recognised more related non-critical lures when they had studied longer lists. This effect of list length did not interact with condition (F < 1). False alarms to unrelated lures were close to floor, especially once the base-rates were subtracted. Given that, the main effect of list length should not be over-interpreted. The data suggest, however, that reading additional list items reduced false alarms of unrelated words, $F(3,174) = 3.51$, MSe = .01.

Of greatest interest are the data on false recognition of the critical lures. A 2 (condition) × 4 (list length) ANOVA was computed on corrected false recognition of critical lures. Again there was a main effect of list length, $F(3,174) =$ 12.16, MSe = .05. Replicating Robinson and Roediger (1997), subjects falsely recognised more critical lures when they had studied longer lists of related words. There was also a main effect of condition, $F(1,58) = 12.91$, MSe = .22. Overall, subjects in the Overt Thought condition falsely recognised more critical lures than did Control participants. However, the interaction between list length and study condition was not significant,[1] $F(3,174) = 1.77$, MSe = .05.

The relationship between study generation and recognition. An important goal of the experiment was to examine the effect of list length on generation of the critical lure in the Overt Thought condition. As shown in Table 2, longer lists elicited more generations of the lure during the study phase, $F(4,156) = 113.4$, MSe = .50.

Table 2 also shows proportion false recognition for each of the list lengths, broken down by whether or not the critical lure was generated during the study phase. Both list length (r = .41) and number of prior generations (r = .43) were correlated with false recognition. Longer lists yielded more false memories. Similarly, previously generating the lure increased its probability of later being falsely recognised. A path analysis examined whether generation of the critical lure mediated the effect of list length on false recognition. That is, does list length have an effect

[1] The reader may notice that although the interaction between list length and condition was not significant, the data suggest a possible interaction. Given that list length affected both conditions, we do not think it would have been particularly problematic if list length had had more of an impact in the Overt Thought condition. However, we note that if anything, it is the results from the Control condition that are surprising; false recognition scores were much lower than they were in Robinson and Roediger (1997). Given that the data in the experimental condition closely replicate previous work, we believe the data from the Overt Thought condition are appropriate for modelling the illusion.

TABLE 2
Mean number of generations of critical lures, and conditional probabilities of false recognition given 0 or 1+ generations of critical lures for each list length

	List length				
	0	*3*	*6*	*9*	*12*
Mean number generations	0.13	0.94	1.88	2.57	3.07
p (false recognition/ generated)	.40 (15)	.52 (101)	.68 (123)	.74 (142)	.72 (140)
p (false recognition/not generated)	.07 (145)	.20 (59)	.38 (37)	.67 (18)	.21 (19)

For each conditional probability, the number of observations is shown in parentheses.

on false recognition in part because longer lists increase generations of the critical lure? Figure 1 depicts the relevant path analysis, which was based on linear regressions. List length predicted both generation (ß = .58) and false recognition (ß = .41). When generation was included as a covariate, however, the effect of list length on false recognition was reduced (ß = .24). This difference was significant via a sobel test (z = 14.98, $p <$ .0001). List length was correlated with generation, and part of list length's effect on false recognition was due to its effect on generations of the critical lure (see Baron & Kenny, 1986, for a description of mediational analyses). These analyses defined the "generation" construct as number of previous generations, the continuous variable. The analyses were repeated treating "generation" as binary (i.e., whether or not the lure had been generated, regardless of frequency). The same conclusion was reached; generation partially mediated the effect of list length (significant via a sobel test, z = 13.50, $p < .001$).

Thus, both list length and prior generation mattered. That is, there was an effect of list length over and above generation. The effect of list length occurred for never-generated lures, and was still significant (although reduced) for generated lures even after analyses covaried out the effect of generation. However, generation also clearly mattered. There was an even higher false alarm rate, and a comparatively smaller effect of list length, for lures that had previously been generated. That is, when generation was taken into account, the effect of list length was significantly reduced. Part of the way that list length affects false recognition is via its effect on generation of the critical lure during study.

Discussion of Experiment 1a

The data support the modification of procedures to allow for the collection of on-line thoughts. A robust false memory effect was obtained using a modified paradigm in which participants explicitly wrote down their thoughts in response to the list items. This study procedure might have been expected to *decrease* the illusion, as the modified procedure slowed the presentation rate (to 6 seconds per word, in both conditions) to allow subjects to record their thoughts (at least for false recall, slower rates actually decrease false memories; McDermott & Watson, 2001). In addition, explicit recording of thoughts might have been expected to help subjects to later attribute those memories to thought rather than presentation. Nevertheless, as in previous studies, the false alarm rate to critical lures was very high. False recognition of critical lures was almost as high as correct recognition of studied items, and was well above the false alarm rate to non-critical lures.

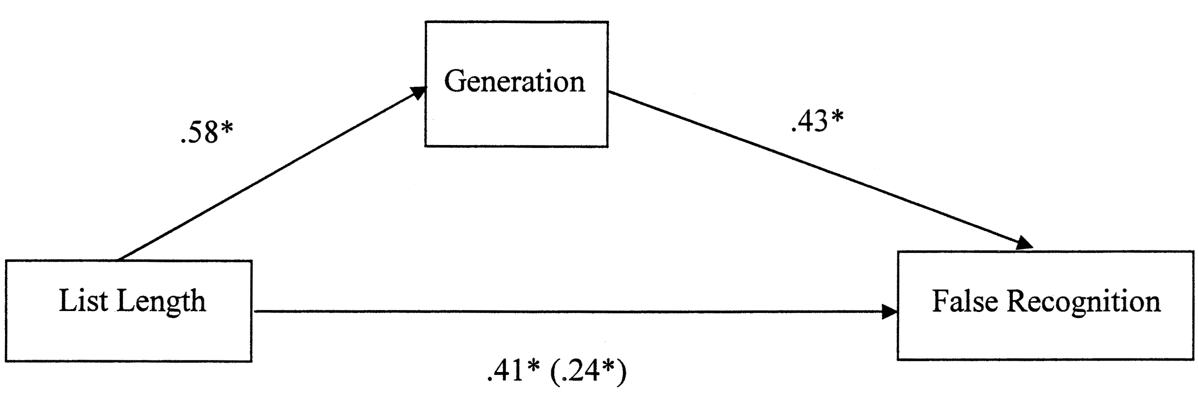

Figure 1. Path analysis (based on linear regressions) showing that generation of critical lures partially mediates the effect of list length on false recognition.

That is not to say, however, that making thoughts overt changes nothing about the subjects' performance. Making thoughts overt, and recording them, may be a ''deeper'' encoding task (Craik & Lockhart, 1972), as suggested by the higher corrected recognition performance in the Overt Thought condition. The memory illusion was also larger in the Overt Thought condition than the Control condition, but this difference was not unexpected. Drawing the participant's attention to his or her thoughts quite possibly could increase their frequency (although we would argue this would be a main effect of thought frequency, and not one that would interact with any variables of interest). In addition, making thoughts overt yields internal events that are more similar to perception (that is, sources are more similar to the typed source in the Overt Thought condition than in the Control condition), thus also leading to an overall increase in false recognition. However, in both study conditions, false recognition increased with list length, and thus we believe it is appropriate to use the data from the Overt Thought condition to model the illusion.

Importantly, examination of the generation data suggests the effect of list length on false recognition occurred in part because longer lists yielded more generations of the lure. The effect of list length was partially, but not totally, mediated by prior generation. Clearly, although generating the lure during study increased the false recognition effect, it was not mandatory for the illusion.

Two final comments on the data are warranted. First, subjects were very unlikely to falsely recognise lures that had never been generated during study of 12 related words. We note simply that this data point is based on few observations; the critical lures were almost always generated when participants had studied 12 words. Second, we find it interesting that the level of false recognition in the control condition is intermediate between the Overt Thought never-generated and Overt Thought generated groups. One possibility is that control subjects are sometimes generating the lures, albeit not as frequently as subjects in the Overt Thought group—leading to intermediate levels of false recognition.

EXPERIMENT 1b

We include here a short report of a secondary experiment, to make two main points. First, given that Seamon et al. found no relationship between rehearsal and false recognition, we wished to replicate the correlation between generation at study and later false recognition. Second, we used longer lists in order to examine generations of non-critical lures. Of interest is whether a similar relationship between generation and false recognition exists for these lures. Otherwise, Experiment 1b is a replication of the Overt Thought condition of Experiment 1a.

Method

Participants. A total of 24 Stanford University students participated in the experiment. They were tested in small groups of four or fewer participants. They were recruited via campus-wide fliers and received monetary compensation for their participation.

Materials. Each participant studied ten lists of 15 words taken from Roediger and McDermott (1995). The ten lists used were associates of the lures *black*, *cold*, *chair*, *high*, *mountain*, *music*, *slow*, *sweet*, *rough*, and *window* lists. All lists were presented in blocked format, beginning with the strongest associate and progressing to the weakest associate. Each participant was given a booklet containing 150 8.5 × 5.5 inch pages. Each page contained the to-be-remembered word plus two blank lines. Participants studied the lists in one of four different presentation orders.

The recognition test was created in the same manner as for Experiment 1a, except that it contained 70 test items, of which 20 were studied words, 10 were critical lures, 20 were related lures, and 20 were unrelated lures.

Procedure. The procedure was the same as in the Overt Thought condition of Experiment 1a, except that subjects recorded their first two thoughts to each of 150 target words.

Results

Recognition. Participants were good at recognising old words ($M = 91\%$). The false alarm rate was higher for related ($M = 10\%$) than for unrelated lures ($M = 2\%$), $t(23) = 8.67$, $SEM = .01$. A significant memory illusion occurred; participants falsely recognised many more critical lures (81%) than related or unrelated lures ($ps < .001$). Although this false alarm rate approached the hit rate, the hit rate was still significantly higher, $t(23) = 2.77$, $SEM = .04$.

Relationship between study generation and recognition. As in Experiment 1a, subjects generated the critical lures during the study phase; on average, each critical lure was generated four times during the entire study session. The number of critical lure generations was a significant predictor of proportion of lures incorrectly identified as "old", $F(1, 134) = 9.21$, $MSe = .11$. That is, as in the previous study, number of generations and proportion false recognition were positively correlated ($r = .254$).

Of particular interest are the non-critical lures. These were generated at a much lower rate than were critical lures. Related lures ($M = .17$) were generated at a higher rate than unrelated lures ($M = .04$), $t(23) = 5.22$, $SEM = .03$, but in both cases the average number of generations was less than one. A regression analysis confirmed that the number of study generations predicted false recognition (collapsing over the subject factor), $F(1, 70) = 7.59$, $MSe = .10$. A similar correlation ($r = .313$) between number of study generations and proportion false recognition was obtained as with critical lures ($r = .254$). Similarly, a within-subject ANOVA examined the effect of number of lure generations (0, 1, or 2+) on proportion false alarms to non-critical lures. Only 11 subjects were included in the analysis as only the generated non-critical lures at all three levels. False recognition increased from 3% at the zero-level, to 25% after one generation, to 41% after two generations, $F(2, 20) = 4.12$, $MSe = .10$.

Discussion of Experiment 1b

Both of the experiment's main aims were achieved. First, we replicated the results of the prior experiment. Both studies were consistent with the original hypothesis: when participants generated a lure during study, they may have been unsure later as to whether it had been presented or not. The more often they thought of and wrote down a word, the more likely they were to fall prey to this kind of memory discrimination error. These results suggest that one reason why the Roediger and McDermott lists lead to unusually high false alarm rates is because the lists are constructed to make participants think about particular non-presented words a number of times. Lists that do not have this characteristic, that do not make subjects think of the critical lure, do not yield high levels of false memories (e.g., Gallo & Roediger, 2002; Roediger, Watson, McDermott, & Gallo, 2001b). This mechanism is not specific to the DRM paradigm and as such also predicts false recognition of non-critical lures. That is, subjects falsely recognised more non-critical lures when they had also generated them during the study phase.

However, it is noteworthy that the false alarm rate was much lower for the non-critical lures, even once the number of generations was taken into account. Generating a non-critical lure did not increase false recognition to the levels observed when a critical lure had been generated. For example, critical lures were incorrectly recognised 50% of the time even when never generated; this increased to 77% with one generation and 83% with two generations. Non-critical lures were only incorrectly recognised 4% of the time if never generated; this increased to 20% with one generation and 36% with two generations (note that these values differ slightly from those reported in the previous results section; these values are based on all 24 subjects rather than only the subset that generated at all three levels for both kinds of lures). While generation increases false recognition of all lures, generation cannot be the only factor involved in false recognition. If it were, the false alarm rate would be much closer to zero for never-generated critical lures, and higher than 20% for generated non-critical lures. Generation increases the memory illusion, but it is not the only cause. In the general discussion, we will re-visit the issue of how much of a role is played by generation, and what other factors might be involved in the illusion.

EXPERIMENT 2 AN ALTERNATIVE SOURCE OF INTERNAL EVENTS

In the first two experiments, we explored the idea that participants covertly call to mind the critical lure during the study phase, in response to studying other related words. The idea that such internal events are created during the study phase is critical to the current research programme. Without the creation of internal events, misattribution of thoughts to perception would not be possible.

The paradigm used in Experiments 1a and 1b was by necessity a modified version of the DRM paradigm. To be able to study the hypothesised covert responses, they must be made overt. However, changing the paradigm may also change the phenomenon under study. Thus, in Experiment 2, we used a different paradigm to observe the occurrence of internal events.

Internal events can be created through multiple means. Events may be generated in response to a cue, such as when the experimenter asks a participant to generate opposites or semantic associates (e.g., Slamecka & Graf, 1978) or to imagine a line drawing corresponding to a word label (e.g., Raye, Johnson, & Taylor, 1980). Internal events may be created as part of another task, such as when intermediate sums are computed while solving an arithmetic problem in one's head (e.g., Dosher & Russo, 1976). Internal events may also be created naturally, such as when one dreams (e.g., Johnson, Kahan, & Raye, 1984). Similarly, there may be more than one way in which internal events are created in the DRM paradigm. In Experiment 2, we used a different methodology to examine another source of internal events, namely intrusions of the critical lure into the rehearsal set. Participants were instructed to practise the list items aloud as much as possible. Of interest was whether they would naturally intrude the critical lures into their rehearsal sets, and whether those intrusions would be related to later false recognition. Note that because participants were instructed ''to rehearse only list items'', rehearsals of the critical lures would reflect memory errors during the encoding phase.

We also note that the method of Experiment 2 is more similar to that used by Seamon et al. and Goodwin et al. than was the method in Experiments 1a and 1b. In Experiments 1a and 1b, we found critical lure generation was related to false recognition whereas Seamon et al. did not find a connection between rehearsal and false recognition. Thus of interest is the replication of a relationship between rehearsal and false recognition in a verbal rehearsal paradigm (as opposed to a written generation paradigm).

Method

Participants. A total of 14 Stanford University students participated in exchange for monetary compensation. Each participant was tested individually.

Materials. Each participant studied ten lists of 15 words taken from Roediger and McDermott (1995); the *chair*, *mountain*, *cold*, *rough*, *sweet*, *black*, *music*, *slow*, *high*, and *window* lists were used in the current experiment.

Each participant was given a booklet containing 160 8.5 × 5.5 inch pages. Presentation of each list was composed of 16 pages; the first 15 pages involved presenting each list item on a separate page, and the 16th page informed the participant that the end of the list had been reached. The lists were always presented in the same order. The same test as in Experiment 1b was used.

Procedure. The experiment paralleled Experiment 1b with one critical difference: Participants were told that they should rehearse the to-be-remembered words aloud. Subjects rehearsed continuously during the presentation of the list items. The lists were presented in blocked format; the participants received a visual cue at the end of each list so that they would stop rehearsing that list and begin rehearsing the new list. All sessions were recorded so that rehearsal could later be examined for intrusions of the critical lures.

Results

Recognition. Overall, participants correctly recognised 88% of the studied words. They falsely recognised more non-critical related (8%) than unrelated lures (3%), $t(13) = 3.47$, $SEM = .02$. The false alarm rate was much higher for the critical lures (51%) than for the non-critical lures, $t(1,13) = 6.05$, $SEM = .08$, but was still lower than the hit rate, $t(13) = 4.19$, $SEM = .09$.

Relationship between study generation and recognition. On average, participants generated 24% of the critical lures at least once; generated lures were produced 2.8 times on average. None of the non-critical lures was generated during the study phase. Thus, although the rate of critical lure generation was much lower when participants were rehearsing study lists than when reporting any items that came to mind (Experiment 1a), it was still substantially above the rate of non-critical lure generation. It should be noted that these generations are themselves memory errors; subjects were instructed to rehearse only list items.

Of the critical lures that were generated during the study phase, 62% were later accepted as ''old''. Of the critical lures that were *not* generated during the study phase, only 47% were accepted as ''old''. This difference was significant when the data were analysed with a within-subject analysis. For each participant who had generated at least one critical lure, the proportion of lures called ''old'' was computed. We then compared mean proportion called ''old'' for lures that had never been generated versus those that had been generated at least once. Given that participants generated at least one lure during the study phase, they judged as ''old'' a greater proportion of generated lures (M = .69) than never-generated lures (M = .44), $t(10)$ = 4.89, SEM = .05.

Discussion of Experiment 2

Participants who rehearsed the critical lure during study later judged a greater proportion of generated lures to be ''old'' than they did for lures that were never rehearsed. These results parallel the finding from Experiments 1a and 1b: items that were generated during study were later more likely to be called ''old''. The result holds even though participants in Experiment 2 were not instructed to generate thoughts. If anything, the rehearsal instructions should act as a kind of dual task, making it difficult for participants to think of anything other than the presented list words during the study phase. Even so, participants still sometimes intruded the critical lure, albeit at a lower rate (24%). It is not surprising that the rate of critical lure intrusion was lower in this study, as subjects were explicitly told to rehearse only list items. Rehearsals of the critical lures were themselves memory errors—only lures that had already been misattributed to the lists would have been rehearsed. In contrast, in Experiments 1a and 1b, lures might have been generated during study for a variety of reasons; generations were not limited to intrusions in a rehearsal set but also likely occurred as responses to a studied word (without subjects necessarily thinking those thoughts had been part of the lists). Similarly, in both the Goodwin et al. and Seamon et al. studies, subjects reported everything that came to mind, and generations were not limited to lures misidentified as list items. Thus, Experiment 2 is the first demonstration that limits generations to rehearsal set intrusions, allowing us to pinpoint specifically one way the critical lure can be generated during study, with consequences for memory.

GENERAL DISCUSSION

Using two different methods to observe critical lure generations, we found that subjects did generate the critical lures during the study phase. More importantly, in all three studies, study phase generation was related to later false recognition. Memory errors were more frequent for words that had been repeatedly generated, as multiple generations likely led to stronger traces. This effect held for both critical lures and non-critical lures. Related non-critical lures were more likely to be generated during study than were unrelated lures; accordingly, false alarm rates were higher for related lures. Longer lists yielded more generations of the critical lure, and accordingly elicited higher false alarm rates. However, not all observed effects can be explained by prior generation. For example, false recognition was always higher for critical lures, even when non-critical lures had been generated as frequently. Similarly, the effect of list length was only partially mediated by prior generation, and false recognition of never-generated critical lures was still quite high.

Both our data and those of Goodwin et al. and Seamon et al. converge on several points. In three different laboratories, using slightly different methodologies to observe and record critical lure generation, it was found that subjects generated the critical lures during the study phase. In all studies, generation of the critical lures during study was related to later memory performance. However, in none of the studies did generation of the critical lures explain *all* of the variance in later memory performance. Thus, drawing on the results from the three laboratories, we can say with confidence that (a) subjects think of the critical lure during the study phase, (b) there is a relationship between thinking of the critical lure and later false memory, but (c) this only partially explains the illusion. Across studies, the data support the hypothesis that *sometimes* the DRM memory illusion arises from the misattribution of thoughts to perception.

Our data diverged from Seamon et al. on one point: Seamon et al. found no relationship between critical lure generation and later false recognition, whereas we did. As described already, Seamon et al. did find a relationship between generation of the critical lures and later memory performance—but only for false recall, not false recognition. There were numerous procedural differences between our studies and their study. We describe in turn the three that seem most likely to have affected the results: (a) the presentation rate, (b) the instructions given during the study phase, and (c) the composition of the recognition tests. First, our studies used a slower presentation rate (6 s) than did Seamon et al. They manipulated presentation rate (2 s vs 5 s), and found that rehearsal increased with longer presentation rates. thus, with our 6 s presentation rate, levels of rehearsal were likely higher in our studies, increasing our likelihood of finding an effect of rehearsal on recognition. Second, the studies also differed in their methods for collecting on-line thoughts. Our rehearse-aloud study differed in that subjects were instructed to only rehearse list items, whereas Seamon et al. encouraged verbalisation of everything that came to mind. Two of our studies collected generations via written associations in booklets instead of recording continuous rehearsals. Instructions or other study variables may affect the frequency with which the critical lures come to mind during study. Indeed, the Goodwin et al. (2001) paper suggested that subjects themselves differ in which strategies they use to encode the words, and these strategies likely affect whether the lures come to mind. Finally, we point to a difference in the composition of the recognition tests. Ours contained list items, critical lures, related non-critical lures, and unrelated non-critical lures (in Experiment 1, for both studied and unstudied lists). The only distractors in the Seamon et al. study were list items and critical lures from *unstudied* lists. Our recognition test, with more semantically related lures, might have been more difficult for subjects. On an easier test, subjects might sometimes have made decisions based on semantics (e.g., ''I know I didn't read a list of sleep words, so this must be new''). On a more difficult recognition test, subjects might be more likely to think back to the study phase, to try and remember specific details to help them make their old–new decisions. Thus, a difficult test might be more likely to capture the consequences of study variables such as number of generations.

Studies from three laboratories found a relationship between critical lure generation and later false memory, but all studies also found that this relationship only partially explained the illusion. Only *sometimes* does the DRM memory illusion arise from the misattribution of thoughts to perception. The question becomes: how large is the role of source errors in the DRM illusion? In part this depends on how ''source monitoring'' is conceptualised, and the nature of the to-be-monitored trace. We think it means something different to monitor activation (possibly summed over events) for its source, as compared to monitoring a memory of a specific thought. Although we believe both types of monitoring occur, we discuss here only the idea of generating the lure during study, and then later incorrectly monitoring its source (see Roediger, Balota, & Watson, 2001a, for a discussion of activation monitoring). Clearly there are situations where this does not occur, and the illusion persists. As described already, false memories still occur at high rates, even if people fail to generate the lure during study. This result held in the present studies, and in previous talk-aloud studies (Goodwin et al., 2001; Seamon et al., 2002). Similarly, generating non-critical lures does not boost false recognition to the levels observed with critical lures. Generation and source misattributions cannot fully explain the illusion. There are situations that preclude the generation of the lure during study, and still yield false memories. For example, false memories persist following such rapid study presentation (20 ms per word) that subjects do not remember the list items. Subjects were unlikely to consciously generate the lures during study; consistent with this, these false memories were associated with ''know'' rather than ''remember'' responses (Seamon et al., 1998).

If the illusion were driven primarily by source misattributions of the sort investigated here, we would expect variables known to reduce source misattributions in other false memory paradigms to also eliminate the DRM illusion. Explicit instructions to monitor source have reduced other types of false memories, supporting the role of source misattributions in unconscious plagiarism (R. Marsh et al., 1997), eyewitness suggestibility (Lindsay & Johnson, 1989), and false fame paradigms (Multhaup, 1995). However, source instructions have yielded mixed results with the DRM paradigm. Warnings against the illusion, which may implicitly direct people towards the sources of their memories, only slightly reduce the illusion, especially if only given just before the test (e.g., Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001b; McDermott & Roediger, 1998). Subjects are able to take advantage of source-specifying characteristics in free recall (Hicks & R. Marsh, 1999) but not recognition. In fact, a source test orienting subjects to different external sources actually *increased* false recognition (Hicks & R. Marsh, 2001)! While the aforementioned studies directed subjects towards source information in general, they did not specifically orient subjects to attributing a critical lure to their own thoughts. Directing subjects towards the memory characteristics that tend to be associated with thoughts reduced the memory illusion in some circumstances (Mather et al., 1997). Similarly, and most critically, the memory illusion was reduced when (at test) subjects received the option "I did not hear this word but generated it on my own" (Multhaup & Conner, 2002). In both of these studies, however, although the illusion was significantly reduced it was *not* eliminated.

Thus, while we believe source monitoring to be involved, like Seamon et al. (2002), we do not wish to argue that all DRM false memories can be explained by means of misattributions of thoughts. Neither our results nor those from studies manipulating test instructions suggest that only source misattributions are involved. Rather, both types of studies support the idea that *sometimes* participants generate the critical lures during study, and later misattribute those generations to presentation.

If generation and source monitoring only partially explain the illusion, then what else is involved? Our data cannot directly answer that question, we can only speculate. For example, consider the list length manipulation. False alarm rates remained higher for longer lists even after controlling for frequency of generation. One possibility is that the longer lists affect activation levels in a way other than explicit generations of the critical lure. Another possibility is that participants are more likely to extract and remember the theme or gist of a longer list, increasing the likelihood of falsely recognising prototypical exemplars (e.g., Brainerd & Reyna, 1990; Israel & Schacter, 1997). It is likely that it is some combination of mechanisms, such as misinterpretations of activation, gist extraction, and source misattributions, that makes the memory illusion so much stronger in the DRM paradigm than in other laboratory analogues of false memories.

Manuscript received 18 October 2002
Manuscript accepted 20 May 2003
PrEview proof published online 18 November 2003

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 657–668.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*, 3–47.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684.

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

Dosher, B. A., & Russo, J. E. (1976). Memory for internally generated stimuli. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 633–640.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. III (2001a). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 339–353.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin and Review*, *4*, 271–276.

Gallo, D. A., & Roediger, H. L. III (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory & Language*, *47*, 469–497.

Gallo, D. A., Roediger, H. L. III, & McDermott, K. B. (2001b). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*, 579–586.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.

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

Goodwin, K. A., Meissner, C. A., & Ericsson, K. A. (2001). Toward a model of false recall: Experimental manipulation of encoding context and the collection of verbal reports. *Memory & Cognition*, *29*, 806–819.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging*, *5*, 119–126.

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1195–1209.

Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old–new recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 375–383.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577–581.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior*, *22*, 485–508.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 3–22.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, *16*, 337–342.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*, 371–376.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Johnson, M. K., Kahan, T. L., & Raye, C. L. (1984). Dreams and reality monitoring. *Journal of Experimental Psychology: General*, *113*, 329–344.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

Johnson, M. K., Raye, C. L., & Durso, F. T. (1980). Reality monitoring: Second perceptions and thoughts. *Bulletin of the Psychonomic Society*, *15*, 402–404.

Kellogg, R. T. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 913–919.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, *52*, 297–318.

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 886–897.

Masson, M. E. J., & MacLeod, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General*, *121*, 145–176.

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. (1997). Priming on perceptual implicit memory tasks can be achieved through presentation of associates. *Psychonomic Bulletin & Review*, *4*, 582–586.

McDermott, K. B., & Roediger, H. L. III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory & Language*, *39*, 508–520.

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.

Multhaup, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology & Aging*, *10*, 492–497.

Multhaup, K. S., & Conner, C. A. (2002). The effects of considering nonlist sources on the Deese-Roediger-McDermott memory illusion. *Journal of Memory & Language*, *47*, 214–228.

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 and Language*, *35*, 261–285.

Raye, C. L., Johnson, M. K., & Taylor, T. H. (1980). Is there something special about memory for internally generated information? *Memory & Cognition*, *8*, 141–148.

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.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*, 231–237.

Roediger, H. L. III, Balota, D. A., & Watson, J. M. (2001a). 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, and Cognition*, *21*, 803–814.

Roediger, H. L. III, McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in producing false memories. In M. Conway, S. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, UK: Psychology Press.

Roediger, H. L. III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407.

Rubin, D. C. (1983). Associative asymmetry, availability, and retrieval. *Memory & Cognition*, *11*, 83–92.

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, *9*, 99–105.

Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test*. (Technical report no. 11). Office of Naval Research, University of Minnesota.

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 in unnecessary for false memory in the DRM procedure. *Psychological Science*, *13*, 526–531.

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for non-conscious processes. *Psychological Science*, *9*, 20–26.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 171–181.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin and Review*, *5*, 710–715.

Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, *117*, 377–389.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122–129.

Vogt, J., & Kimble, G. A. (1978). False recognition as a function of associative proximity. *Journal of Experimental Psychology*, *99*, 143–145.