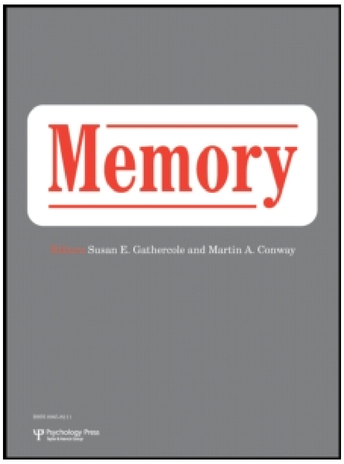

Memory
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683358

True and false memories in the DRM paradigm on a forced choice test

Yana Weinstein [a]; Kathleen B. McDermott [a]; Jason C. K. Chan [b]
[a] Washington University in St. Louis, MO, USA [b] Iowa State University, Ames, IA, USA

First published on: 19 April 2010

To cite this Article Weinstein, Yana , McDermott, Kathleen B. andChan, Jason C. K.(2010) 'True and false memories in the DRM paradigm on a forced choice test', Memory, 18: 4, 375 — 384, First published on: 19 April 2010 (iFirst)
To link to this Article: DOI: 10.1080/09658211003685533
URL: http://dx.doi.org/10.1080/09658211003685533

MEMORY, 2010, 18 (4), 375–384

Psychology Press
Taylor & Francis Group

True and false memories in the DRM paradigm on a forced choice test

Yana Weinstein and Kathleen B. McDermott
Washington University in St. Louis, MO, USA

Jason C. K. Chan
Iowa State University, Ames, IA, USA

Participants studied lists of semantic associates that converged on a non-presented critical word (e.g., *sleep*; Deese, 1959; Roediger & McDermott, 1995) and took a two-alternative forced choice test. At test, each critical non-presented word was paired with a studied word from the same list. The test was administered either immediately or 7 days after the study phase. Accuracy in distinguishing between the non-presented critical word and the studied list word was above chance at immediate testing. After a 7-day retention interval, however, accuracy did not differ from chance performance: participants were as likely to choose the non-presented critical word as the studied list word.

Keywords: False memory; Forced choice tests.

Since it was first developed in the mid-1990s, the Deese (1959)/Roediger and McDermott (1995; DRM) paradigm has become a popular method for studying false memories. In this paradigm, participants are presented with semantic associates (e.g., bed, rest, awake) of a critical word (in this example, sleep) that is never actually studied but is very frequently mistaken for a studied word at test. The original authors and researchers from countless other labs have replicated this effect on production (i.e., free recall) and identification (i.e., recognition) tests of memory (see Gallo, 2006, for review). These non-presented critical words tend to have the same characteristics as studied words: they are often classified as *old* on recognition tests, spontaneously produced on recall tests, and their encoding is often remembered with vivid detail (Roediger & McDermott, 1995). Indeed, in some cases the probability of falsely recalling a critical word can exceed that of correctly recalling a studied word (McDermott, 1996). Furthermore, critical words can be accompanied by vivid details similar to those of true recollection (Gallo, McDermott, Percer, & Roediger, 2001). Thapar and McDermott (2001) showed that after a delay between study and test, false recall and recognition increased and exceeded true recall and recognition, which conversely dropped with delay. An intriguing question that arises from these data is whether participants would be able to correctly pick a studied word if it were paired with the non-presented critical word on a forced choice test, either immediately or after a long retention interval. Despite the considerable literature on DRM false memories, this question has never been addressed directly.

In a standard yes/no recognition task participants classify test items on the basis of memory strength and their criterion. For instance, participants could apply a liberal criterion to words they

Address correspondence to: Yana Weinstein, Department of Psychology, Washington University in St. Louis, One Brookings Drive, St Louis, MO 63105, USA. E-mail: y.weinstein@wustl.edu

http://www.psypress.com/memory DOI:10.1080/09658211003685533

know to be highly related to studied words. This would result in a dual strategy whereby *old* responses to critical words would be made on different bases from *old* responses to studied words (Miller & Wolford, 1999). However, there is much evidence against this theory, including the high level of *remember* false alarms to critical words (Roediger & McDermott, 1995). Forced choice tests avoid issues of criterion because the criterion is effectively set at zero, and these tests may also be easier for participants to perform accurately (Macmillan & Creelman, 2005). It has also been argued that yes/no and forced choice tasks recruit different memory processes. For instance, Bastin and Van Der Linden (2003) showed a reversal in performance on the two tasks between younger and older adults, with younger adults performing better on a yes/no recognition test and older adults performing better on a forced choice test. One explanation for this effect is that forced choice recognition relies more on familiarity than does yes/no recognition, which recruits more recollective processes (e.g., Cook, Marsh, & Hicks, 2005; Gardiner, Java, & Richardson-Klavehn, 1996). Given these differences between the two tasks, we thought it important to determine how well participants would be able to distinguish between true and false memories on a forced choice recognition test after studying DRM lists.

It is somewhat surprising that despite its popularity in other false memory investigations (e.g., McCloskey & Zaragoza, 1985; McDermott & Chan, 2006), forced choice recognition is rarely employed in the DRM paradigm. Two other papers have done so, although in neither was a non-presented critical word paired with a presented list word from the same list. Westerberg and Marsolek (2003, Exp. 2) presented participants with standard DRM lists and DRM lists in which the critical word replaced one of the list words.[1] At test, the omitted critical word from the former list was paired with the presented critical word from the latter, and the task was to indicate which word had been studied. Similarly, the presented list word from the former was paired with the omitted list word from the latter. Westerberg and Marsolek found that participants correctly chose the presented critical word on only 0.54 of trials, while they were able to correctly choose the presented list word on 0.63 of trials. Participants thus found it harder to distinguish between presented and omitted critical words than presented and omitted list words. While this result suggests that non-presented critical words are activated by the presentation of their associates and thus are relatively difficult to tell apart from critical words that have actually been presented, it does not tell us whether this false memory is as strong as the true memory of a studied associate. Gallo and Seamon (2004) also used a forced choice task, but in their version of the test participants had to choose between critical words from studied and non-studied lists, so their data do not speak to the question of the relative strength of true and false memories. The present paper is thus the first to directly pit critical words against studied words in a two-alternative forced choice test to determine the relative strength of true and false memories in the DRM paradigm.

Our experimental design included three conditions that paired the critical word with a list word from the middle of that same list. In one condition the list had not been studied at all, so both words were new. In another condition the list had been studied and both words actually appeared in the list, so the two words were both old. In the third and most important condition the list had been studied but the critical word had not been presented in the list. In this condition the list word was old and the critical word was new, but the critical word would have been activated by the study of its associates. With this condition our goal was to determine whether participants are able to distinguish between the presented list word and non-presented critical word. According to an activation account (Roediger, Balota, & Watson, 2001; Underwood, 1965), chance performance could occur on this task if activation of the critical word from studying its associates were as strong as activation of the list word from study. A preference for the non-presented critical word over a presented list word would point to the existence of a very strong false memory of the critical word arising from such activation.

An idiosyncrasy of the DRM paradigm is that list and critical words cannot be counterbalanced. Due to the way the lists are constructed, list and critical words are not matched on lexical and orthographic characteristics. More specifically, the critical words tend to have higher word frequencies than the list words. Consequently, predictions for performance in the two conditions

[1] Participants in this experiment also studied and were tested on lists of unrelated words, but these are not relevant to the discussion.

described above can be made from the literature on word frequency effects in forced choice recognition. Glanzer and Bowles (1976) showed that when both words on a forced choice test are new, the more-frequent word is chosen more often. When both words are old, on the other hand, the less-frequent words tend to be chosen. For the case in which both words are new, we predict that the critical word will be chosen more often. When both words are old, the list word should be chosen more often according to Glanzer and Bowles's findings, but the situation is complicated by the fact that the studied critical words also receive activation from study of associates, which may eliminate or even reverse the expected preference for the less-frequent list words. In yes/no recognition, McDermott and Roediger (1998) and Miller and Wolford (1999) showed that the critical word, when studied, had a very high probability of being correctly recognised. This makes sense because of the nature of the lists: each list consists of the 15 strongest associates of the critical word. Hence, when the critical word is presented in the list it is activated not only because it is actually studied, but also through study of the associates, and this increased activation may mean that participants are more likely to choose the critical word than the list word when both have been studied, even though the word frequency literature would predict the opposite. In addition to looking at the probabilities of choosing the critical and list words when neither and when both were studied, we also performed regressions to establish whether the difference in frequency between words in each pair affected recognition decisions.

Another goal of the current paper was to examine the behaviour of false memories over time. Gallo's (2006, pp. 64–65) meta-analysis of 14 experiments in six different papers investigating yes/no recognition of DRM lists at long retention intervals revealed a somewhat inconsistent pattern of results: only 9 of the 14 studies showed a larger decrease in hits for list words relative to false alarms for critical words, when both were corrected for false alarms to unrelated words. There is some evidence that sleep may be responsible for selectively promoting recall of non-studied critical words relative to the studied words over a delay (Payne et al., 2009). However, none of these studies employed a forced choice test to compare true and false memories directly. We included a condition with a 7-day retention interval in our study to fill this gap in the literature. Of course, it is a ubiquitous finding in memory that people forget information over time (Ebbinghaus, 1885/1964). Thus, as the delay between study and test increases, studied items become less easy to distinguish from new items on virtually any recognition test. To address this issue, on some trials we paired list words from studied lists with list words from non-studied lists. These trials served to ensure that participants in our experiment still had some memory of the lists after the long retention interval. As the distractors in these pairs were not semantically related to studied items, they were not expected to have any memory strength above baseline familiarity. Participants were thus expected to perform above chance on these pairs, even after the 7-day retention interval, and any decrease in performance could be attributed to a general decrease in discriminability between studied and new items. However, performance on pairs in which a non-presented critical word was paired with a studied associate from the same list was also expected to suffer to the extent that false memories of critical words are more persistent than true memories of list words. If "forgetting" of false memories occurs more slowly than forgetting of true memories, participants could perform below chance when asked to distinguish between a studied list word and non-presented critical word after a long retention interval (i.e., they could be more likely to pick the non-presented critical word than the studied list word).

In summary, our primary goal was to determine whether participants would be able to distinguish between a non-presented critical word and a studied list word (i.e., to correctly choose the presented word from a studied list, not the associated critical non-presented word). In pursuing this goal, efforts were made to address the effects of word frequency on recognition judgements. Our secondary goal was to see how performance on the two-alternative forced choice test was affected by a 7-day retention interval.

METHOD

Participants

A total of 96 Washington University undergraduates participated in the experiment and were either given course credit or financial reimbursement for their time.

Design

The experiment employed a 2 (list type: critical word in list/critical word not in list) × 2 (retention interval: immediate/7-day) between-participants design. The two between-participants variables (list type and retention interval) were manipulated orthogonally with 24 participants in each of the four cells. That is, half the participants completed the test immediately after the study phase and the other half after a week delay; half the participants studied lists that contained the critical word and the other half studied lists that did not contain the critical word.

Materials

A total of 36 lists of 15 semantically related words each were selected from the Roediger, Watson, McDermott, and Gallo (2001) norms. These lists were divided arbitrarily into two 18-list sets. Half of the participants in each condition studied List Set A, and the other half studied List Set B. In addition, alternate versions were created for each list to enable presentation of the critical word. Specifically, the fifth associate was removed and replaced with the critical word. Because at test the critical words were paired with the eighth associate, it was necessary to counterbalance where these words appeared in the lists. Thus, for half of the lists the critical word was presented in the fifth position in the list while the eighth associate remained in the eighth position. For the other half of the lists the critical word was presented in the eighth position, with the eighth associate taking the fifth position. Aside from this modification, words within each list were arranged in order from strongest to weakest associative strength to the critical word.

Procedure

All participants studied 18 lists presented in a random order. Words appeared sequentially in 18-point font for 2000 ms, separated by a 500-ms interstimulus interval. Participants advanced through the study phase by pressing a key to move on to the next list, and the encoding session lasted about 12 minutes. At the end of the study phase participants were either dismissed and returned a week later to complete the test, or took the test immediately.

The test consisted of the same 54 word pairs for all participants; the prior history of some of the words differed across conditions. There were two types of pairs: list pairs and critical pairs. The 18 list pairs were made up of the twelfth associate from a list in List Set A paired with the twelfth from a list in List Set B chosen arbitrarily but kept constant across all participants (see Appendix A for the resulting pairs). Thus, for all participants, one of the words in each list pair was old and the other was new.

The 36 critical pairs were made up of the critical word of each list paired with the eighth associate from that same list. There was one such pair for each of the 36 lists. Since participants had studied only half of these lists (either List Set A or List Set B), 18 of the pairs were made up of two *new* items. These trials were included to gauge baseline preference for choosing the critical word relative to the eighth associate in each list when neither had been studied. The other 18 pairs differed depending on the list type. For participants who studied lists that did not contain the critical words, one word in the pair (the eighth associate) was old, while the other (the critical word) was new. For participants who studied lists that contained the critical word, both of those items were old.

The 54 pairs described above were presented in a different random order for each participant, but all item pairings remained the same across participants. On each trial, two words were presented side by side. The relative position of the two words was selected randomly by the program. Participants were instructed to choose the word that they thought had appeared in the study phase. No mention was made of trials where two words had been studied, or neither word had been studied; participants were simply told to make their best guess on each pair. Following their selection, participants rated how confident they were in their decision on a 6-point scale (from 1 = *complete guess* to 6 = *very confident*).

RESULTS

List pairs

List pairs were those in which the twelfth associate from a studied list (*old* word) was paired with the twelfth associate of a non-studied list (*new* word).

Figure 1 presents the probability of correctly selecting the studied word from the 18 list pairs in all conditions. A 2×2 analysis of variance (ANOVA) with retention interval and list type as the between-participants variables revealed a main effect of delay, $F(1, 92) = 49.9$; $MSE = .02$; partial $\eta^2 = .35$; $p < .001$, such that participants tested immediately correctly identified the studied list words more often than those tested at a 7-day delay ($M = .80$ and .62 respectively). Importantly, participants did perform above chance in the delay condition: a one-sample *t*-test comparing performance to the chance level of .50 in the 7-day delay condition yielded a significant difference, $t(47) = 6.92$; $d = .99$, $p < .001$. Performance on these pairs did not differ between list type conditions, $F(1, 92) < 1$, and there was no interaction between list type and delay, $F(1, 92) < 1$.

Table 1 presents mean confidence ratings by accuracy in the immediate and 7-day delay conditions, collapsed across list type conditions. Confidence ratings were subjected to a 2×2 mixed-design ANOVA with accuracy as the within-participants variable and retention interval as the between-participants variable. Three participants did not produce any incorrect responses in the immediate condition, so they were excluded from this analysis. Participants were overall more confident immediately ($M = 3.90$) on the 6-point confidence scale across hits and false alarms) than when tested after a 7-day delay, ($M = 2.89$); $F(1, 91) = 29.33$; $MSE = 1.62$; partial $\eta^2 = .24$; $p < .001$. Participants were also more confident on hits ($M = 3.96$) than on false alarms ($M = 2.81$); $F(1, 91) = 115.11$; $MSE = .55$; partial $\eta^2 = .56$; $p < .001$. Finally, there was an interaction between accuracy and retention interval such that confidence on hits decreased over time more markedly than did confidence on false alarms, $F(1, 91) = 27.60$; $MSE = .55$; partial $\eta^2 = .23$; $p < .001$.

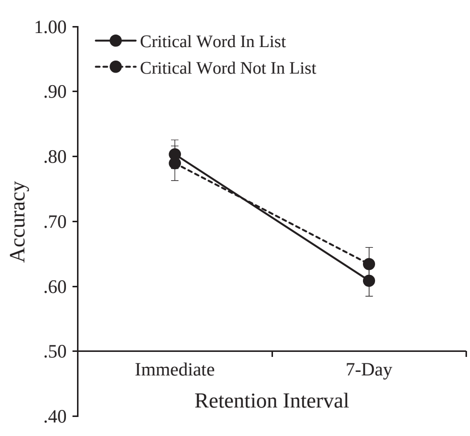

Figure 1. Accuracy (probability of choosing the list word from the studied list) as a function of list type and retention interval for list words from studied lists paired with list words from non-studied lists. Error bars represent *SE*.

TABLE 1
Mean confidence ratings by accuracy and retention interval on list pairs

Retention interval	*Hits*	*False alarms*
Immediate	4.77 (0.11)	3.03 (0.19)
7-day delay	3.19 (0.16)	2.59 (0.14)

Mean confidence ratings (on a scale from 1 = *complete guess* to 6 = *very confident*) by accuracy and retention interval on list pairs (*SE* in brackets). List pairs are those in which the twelfth associate from a studied list was paired with the twelfth associate from a non-studied list. Accuracy data for these pairs are presented in Figure 1. Confidence ratings were collapsed across list type conditions, as neither accuracy nor confidence differed across these two conditions. Three participants' data were excluded from these analyses because they did not produce any false alarms on the immediate test.

Critical pairs

Critical pairs were those in which the critical word of each list was paired with the eighth associate of that same list. The data for these pairs are presented in Figure 2 in terms of the probability of choosing the critical word. For all participants, 18 of these pairs came from lists that had not been studied, and thus the probability of choosing the critical word in these pairs

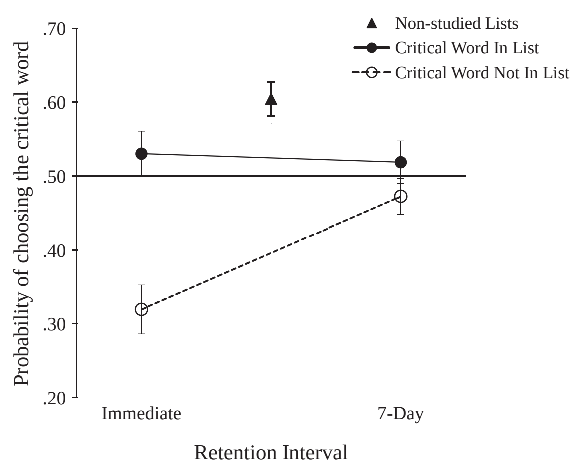

Figure 2. Probability of choosing the critical word when neither word in the pair had been studied (triangle), when the list word only had been studied (dashed line), and when both words in the pair had been studied (solid line), immediately and after a 7-day delay. Error bars represent *SE*.

($M = .60$) reflects the baseline preference for the critical word over the list word. The data for these pairs are presented in Figure 2 collapsed across conditions, because no differences between conditions could be expected as neither of the words in these pairs had been studied in any condition. A one-sample *t*-test comparing the probability of choosing the critical word to the chance level of .50 revealed that participants were significantly more likely to choose the critical word than the eighth associate when neither had been studied, $t(95) = 7.21$, $d = .74$, $p < .001$.

Table 2 presents mean confidence ratings for each of the critical pair comparisons, split by which word was chosen. Confidence ratings for the non-studied list comparison are presented in the top row of the table. Not surprisingly, these judgements were made with low confidence ($M =$ 2.91; note that these confidence ratings were averaged across the two retention intervals), and there was no difference in confidence ratings by recognition decision ($p = .20$; i.e., list words and critical words were chosen with equal confidence).

For the other 18 pairs, item status differed depending on list type condition. The probabilities of choosing the critical word on these 18 pairs in each condition were subjected to a 2×2 ANOVA with list type and retention interval as the between-participants variables. This analysis revealed a main effect of retention interval, $F(1, 92) = 5.75$; $MSE = .02$; partial $\eta^2 = .06$; $p =$ .02; a main effect of list type, $F(1, 92) = 19.05$; $MSE = .02$; partial $\eta^2 = .17$; $p < .001$; and an interaction between delay and list type, $F(1, 92) = 7.79$; $MSE = .02$; partial $\eta^2 = .08$; $p <$ .001. In order to explore these effects, we considered each list type condition separately.

For participants who had studied lists that contained the critical word (represented by the solid line in Figure 2), both of the items in the pair were *old*. The probability of choosing the critical word in this condition ($M = .52$) was not significantly different from the chance level of .50 for these participants across the two delay conditions, $t(47) = 1.17$, $p = .25$. Moreover, the retention interval did not have an effect on participants' responses to these pairs ($p = .78$).

Confidence ratings for these pairs are shown in the middle two rows of Table 2. These confidence ratings were subjected to a 2×2 mixed-design ANOVA with recognition decision (critical word/list word) as the within-participants variable and retention interval as the between-participants variable. Participants were overall more confident immediately ($M = 5.14$ across both critical word and list word responses) than when tested after the delay ($M = 3.18$); $F(1, 46) = 67.60$; $MSE =$ 1.45; partial $\eta^2 = .56$; $p < .001$. Participants also reported numerically higher confidence ratings when they picked the list word ($M = 4.27$) than when they picked the critical word ($M = 4.11$); however, this difference was only marginally significant, $F(1, 46) = 3.45$; $MSE = .18$; partial $\eta^2 = .07$; $p = .07$. There was no interaction between recognition decision and retention interval.

For participants who had studied lists that did not contain the critical word (represented by the dashed line in Figure 2), these pairs consisted of one *old* word (the eighth associate) and one *new* word (the critical word). In this condition, retention interval had a significant effect on responses, $F(1, 46) = 13.71$; $MSE = .02$; partial $\eta^2 = .23$; $p <$ .001. When these participants were tested immediately, they did not tend to choose the critical word ($M = .32$) over the studied word: a *t*-test comparing the probability of choosing the critical word to the chance level of .50 showed a significant difference, $t(23) = 5.45$, $d = 1.11$, $p < .001$. The probability of choosing the critical word on an immediate test when it had not been presented in the list was also significantly lower than when it had been presented in the list, $F(1, 46) = 21.90$; $MSE = .02$; partial $\eta^2 = .32$; $p < .001$. At a 7-day delay, on the other hand, performance dropped to chance and participants were no longer able to correctly select the studied list word ($p = .27$ in the one-sample *t*-test

TABLE 2
Mean confidence ratings by recognition decision, list type, and retention interval on critical pairs

List type/retention interval	*List word*	*Critical word*
Non-studied lists[a]	2.95 (0.11)	2.88 (0.10)
Critical word in list		
Immediate	5.33 (0.11)	5.06 (0.16)
7-day	3.21 (0.24)	3.15 (0.24)
Critical word not in list [b]		
Immediate	5.17 (0.13)	4.17 (0.21)
7-day	3.43 (0.21)	3.25 (0.21)

Mean confidence ratings by recognition decision, list type, and retention interval on critical pairs (*SE* in brackets). Critical pairs are those in which a critical word from a list was paired with the eighth associate of that same list. Recognition decision data for these pairs are presented in Figure 2.
[a]Confidence ratings for this comparison were collapsed across list type and retention interval conditions as for recognition data.
[b]One participant correctly picked the list word on every trial, so they were excluded from this analysis.

comparison to .50). Furthermore, at a 7-day delay there was no longer a difference in the probability of selecting the critical word between participants who had actually studied them and those who had not ($p = .23$).

Confidence ratings for these pairs are shown in the bottom two rows of Table 2. These confidence ratings were subjected to a 2×2 mixed-design ANOVA with recognition decision (critical word/list word) as the within-participants variable and retention interval as the between-participants variable. One participant correctly picked the list word on every trial, so this participant was excluded from this analysis. Participants were overall more confident immediately ($M = 4.67$ across both critical word and list word responses) than when tested after the delay ($M = 3.34$); $F(1, 45) = 28.09$; $MSE = 1.47$; partial $\eta^2 = .38$; $p < .001$. Contrary to all the other critical pairs, participants were also more confident when they picked the list word ($M = 4.28$) than when they picked the critical word ($M = 3.70$); $F(1, 45) = 26.07$; $MSE = .31$; partial $\eta^2 = .37$; $p < .001$. Finally, there was an interaction between recognition decision and retention interval such that confidence when picking the list word decreased with time more markedly than did confidence when picking the critical word, $F(1, 45) = 4.04$; $MSE = .31$; partial $\eta^2 = .22$; $p = .001$.

Word frequency

In order to examine whether differences in word frequency between the critical and list words in each critical pair affected recognition decisions, regressions were performed for the pairs where both the critical and list words were new, and the pairs where both were old. Word frequency data were obtained from the English Lexicon Project (Balota et al., 2007) for the critical words and the list words they were paired with for 34 of the 36 lists (word frequencies were unavailable for two of the lists). Log word frequency of critical words ($M = 9.95$) was significantly higher than log word frequency of list words ($M = 8.79$); $t(33) = 3.82$, $p = .001$. The difference in word frequency (log) between the critical and list words was entered into a regression with the probability of choosing the critical word as the dependent variable. For pairs in which neither word was studied, word frequency (log) was a significant predictor of the probability of choosing the critical word (b = .39, $p = .021$), accounting for 16% of the variance; such that the more frequent the critical word was relative to the list word, the more often it was chosen when neither had been studied. The same analysis was also carried out for critical pairs in which both words had been studied. Word frequency (log) accounted for 27% of the variance in responses (b = .52, $p = .002$), but in this case, the more frequent the critical word was relative to the list word it was paired with, the *less* often it was chosen when neither had been studied.

DISCUSSION

The main goal of this paper was to investigate performance on a forced choice test in which one word had been studied and the other had been only activated by the study of associates. In summary, although participants were quite accurate in distinguishing between the studied list words and non-presented critical words when tested immediately following study, they were not able to do so after a 7-day retention interval. At the outset three questions were identified: First, how well can participants distinguish between the non-presented critical word and the studied list word? Second, how will performance be affected by a longer retention interval? Third, will participants prefer the critical word or the list word when neither have been studied and when both have been studied? Below we address each of these questions in turn.

Is activation of critical words that arises from the study of associates (Roediger et al., 2001; Underwood, 1965) equivalent to that which arises from actually studying the list words? The present results suggest that the answer is no in the absence of a long retention interval. That is, participants performed well above chance on the immediate test. One mechanism that could explain this success is participants' awareness of subjective characteristics that could help distinguish between true and false memories. For instance, Norman and Schacter (1998) and Mather, Henkel, and Johnson (1997) found that participants recalled more details about studied words than non-presented critical words. The activation account of false memory in the DRM paradigm also includes a monitoring component (McDermott & Watson, 2001). If this recollective process is engaged, it could enable participants to select the list word even if the critical word seemed equally familiar.

Because not all DRM lists are equal in their ability to induce a false memory for the critical word, one might wonder whether participants were more likely to select the critical word than the list word on a subset of the lists. Stadler, Roediger, and McDermott (1999) collected recognition norms for false recognition of the critical word and true recognition of list words from various serial positions. As the present study used 31 of the 36 lists normed by Stadler et al., we can directly compare the two datasets. They found a large scope in the effectiveness of DRM lists in producing false recognition, with false recognition of the critical word ranging from 0.53 to 0.84 for those 31 lists. Subtracting these figures from true recognition of the list word in serial position eight, the majority of the lists (21 of 31) produced greater levels of false recognition of the critical word compared with true recognition of the eighth associate. In our data, on the other hand, participants only chose the critical word more often than the eighth associate on 2 of those 31 lists (*sweet* and *soft*). Not surprisingly, however, the lists that produced the highest true recognition relative to false recognition in Stadler et al.'s study were the same ones that produced the most accurate recognition on our forced choice task.

A secondary goal of the present paper was to determine participants' ability to distinguish between true and false memories after a longer retention interval. Previous attempts to address this issue have produced mixed results: while some studies report an increase in false memory and a decrease in true memory over time (e.g., Thapar & McDermott, 2001), other studies have reported the opposite pattern (e.g., Colbert & McBride, 2007), although the evidence seems to be weighted slightly in the former direction (Gallo, 2006, pp. 64–65). However, data from free recall tests more firmly support the conclusion that false memories supersede true memories over time (Gallo, 2006, p. 64).

The forced choice test data presented in the current paper contribute to this debate. With time, as forgetting of studied words occurs, performance on a forced choice test will tend towards chance, or the base rate at which each word would be chosen if neither had been studied. In our particular case this would lead to the critical word being chosen more often than the list word, as seen in the critical pairs from non-studied lists. After a 7-day retention interval, participants were no longer able to accurately distinguish between the non-presented critical word and the studied associate from the middle of the same list, but they did not choose the critical word more often than the list word. Importantly, this observation implies that participants were making responses to these pairs on the basis of memory rather than trying to choose between two words that both seemed new. In fact, performance in this condition was identical to that in which both the list and critical words were studied.

In order to control for word frequency effects, recognition decisions were also examined when both the critical word and the list word were new (i.e., that particular list had not been studied) and when both words were old (i.e., the critical word was presented in the list instead of another associate). When neither the critical nor the list words were studied, participants were more likely to choose the critical word. In addition, the critical words were on the whole more frequent than the list words, and the tendency to choose the critical word when neither had been studied increased as the gap in frequency between the two widened. This result is in line with Glanzer and Bowles's (1976) finding that in a situation where both words are new, the more frequent word will be chosen more often. However, Glanzer and Bowles also demonstrated the opposite pattern for situations where both items are old: the less-frequent word should be chosen more often. While we did find that the bigger the gap in word frequency between the two, the more likely the list word was to be chosen, participants were overall no more likely to choose the less-frequent list word than the critical word. This suggests that studying associates produced extra activation of the critical word (Roediger et al., 2001; Underwood, 1965), which had an effect on recognition judgements such that participants were more likely to pick the critical word than would be expected in the absence of such activation. However, this activation was not strong enough to cause participants to choose the critical word more often than the list word.

Comparing our data to the extant free choice recognition literature, it may at first glance appear that participants are showing a greatly reduced susceptibility to false memories in the results reported here. For instance, the original Roediger and McDermott (1995) article demonstrated that on an immediate test, critical lures were falsely recognised with the same probability as list words were correctly recognised, accounting for the relevant baselines (their Experiment 2). However, two important differences must be noted between these studies and our own. First, yes/no recognition studies that have shown greater levels of false recognition than true recognition have tended to be those in which presentation of the words was auditory (and the test was visual). In fact, Gallo's (2006, Table 4.1, p. 78) meta-analysis of 16 experiments that used visual presentation shows that presented list words were called old 10% more often than were non-presented critical words. Second, there is evidence that when targets and lures are similar (as was the case in our critical and list word pairs, as both were taken from the same list), it is easier to perform accurately on a forced choice than a yes/no recognition test; this has been demonstrated both with healthy participants (Hintzman, 1988) and brain-damaged patients (Migo, Montaldi, Norman, Quamme, & Mayes, 2009) whose performance on the forced choice task was preserved relative to performance on the yes/no recognition task. This effect is particular to related targets and lures because the strength of distributions of the two are highly correlated, and thus small differences in activation can be tracked reliably. This strategy would be ineffective on a yes/no recognition test, where items are presented individually and thus cannot be compared in this way. The results we report show that participants are able to use this strategy effectively on an immediate forced choice test, but not after a 7-day delay.

In sum, two important findings emerged from this experiment. First, participants can accurately distinguish between non-presented critical words (false memories) and presented list words (true memories) when tested immediately. Second, participants are no longer able to do so after a 7-day delay: following this retention interval, the critical non-presented word is as likely to be chosen as a presented list word.

Manuscript received 10 September 2009
Manuscript accepted 5 February 2010
First published online 19 April 2010

REFERENCES

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Bastin, C., & Van Der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and age. *Neuropsychology*, *17*, 14–24.

Colbert, J. M., & McBride, D. M. (2007). Comparing decay rates for accurate and false memories in the DRM paradigm. *Memory & Cognition*, *35*, 1600–1609.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720–725.

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

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology*. Oxford, UK: Dover.

Gallo, D. A. (2006). *Associative illusions of memory*. New York: Psychology Press.

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

Gallo, D. A., & Seamon, J. G. (2004). Are nonconscious processes sufficient to produce false memories? *Consciousness & Cognition*, *13*, 158–168.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology*, *50*, 114–122.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory*, *2*(1), 21–31.

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York: Cambridge University Press.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgements and memory characteristics questionnaire compared. *Memory & Cognition*, *25*, 826–837.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*, 1–16.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language*, *35*, 212–230.

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

McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition*, *34*, 1273–1284.

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 and Language*, *39*, 508–520.

McDermott, K. B., & Watson, J. M. (2001). The rise and false of false recall: The impact of presentation duration. *Journal of Memory & Language*, *45*, 160–176.

Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1198–1215.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, *106*, 398–405.

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, J. D., Schacter, D. L., Propper, R. E., Huang, L., Wamsley, E. J., Tucker, M. A., et al. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory*, *92*, 327–334.

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

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

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

Stadler, M. A., Roediger, H. L. III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*, 494–500.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*, 424–432.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *102*, 917–939.

Westerberg, C. E., & Marsolek, C. J. (2003). Sensitivity reductions in false recognition: A measure of false memories with stronger theoretical implications. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 747–759.

APPENDIX

List Set A			List Set B		
Critical word	Eighth associate	Twelfth associate	Critical word	Eighth associate	Twelfth associate
sleep	blanket	nap	wish	bone	get
mountain	plain	climber	flag	wave	emblem
spider	tarantula	animal	pen	scribble	marker
shirt	polo	jersey	army	Marines	war
slow	traffic	quick	man	strong	handsome
trash	rubbish	dump	rubber	foam	glue
health	strong	vigor	chair	sofa	stool
window	open	breeze	music	horn	jazz
high	building	over	anger	wrath	mean
river	flow	brook	soft	touch	downy
foot	walk	inch	bread	flour	slice
justice	jury	supreme	smell	whiff	fragrance
lamp	desk	on	stove	wood	gas
cup	coffee	stein	cold	chilly	air
doctor	physician	stethoscope	smoke	chimney	pipe
rough	coarse	sand	sweet	nice	heart
girl	niece	date	needle	thimble	injection
city	New York	Chicago	thief	rob	crime