APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **20**: 97–114 (2006)
Published online 1 November 2005 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1170

Do Instructional Warnings Reduce False Recognition?

CARMEN E. WESTERBERG* and CHAD J. MARSOLEK
University of Minnesota, USA

SUMMARY

After previously encoding lists of related words (e.g. *bed*, *rest*, *awake*, etc.) associated with one critical word (e.g. *sleep*), participants frequently falsely recognize critical words as having been previously presented. Past research indicates that warning participants of this memory illusion can reduce false recognition of critical words. However, the memory processes responsible for this reduction are not known. We investigated whether the increase in critical-word memory performance reflects changes that are specific to the processing of critical words, or alternatively, changes that are applied generally to processing in all conditions. Different participants were warned (in two different ways) or not warned before encoding, and recognition sensitivity for critical words, related words and unrelated words was tested. The warnings increased memory performance equally across all conditions, not just for critical words. These results help to more effectively conceptualize false recognition and reductions in false recognition in this paradigm. Copyright © 2005 John Wiley & Sons, Ltd.

A wealth of past research indicates that human memory is prone to distortions (e.g. Bartlett, 1932; Deese, 1959; Kelley & Jacoby, 1990; Loftus & Palmer, 1974; Neisser, 1967). Because false or distorted memories can be problematic in everyday life, an important contemporary question is whether false memories can be reduced or prevented. In this article, we report an investigation of the impact of instructional warnings on false recognition in the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995), as revealed by signal detection measures of memory performance. Our main question is whether warnings specifically affect the processing of critical words in particular. Alternatively, warnings may have an effect that influences the processing of all types of words. This question is important because it addresses the distinctiveness of false memory effects exhibited in the DRM paradigm. The results from the present investigation will help to more effectively conceptualize false memories and reductions in false memories observed in this paradigm, and this information may be useful for developing successful strategies for reducing false memories in real-world situations.

*Correspondence to: C. E. Westerberg, Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, USA. E-mail: c-westerberg@northwestern.edu

Contract/grant sponsor: Center for Cognitive Sciences (National Institute of Health and Human Development); contract/grant number: HD07151.
Contract/grant sponsor: Office of Vice President for Research and Dean of the Graduate School at the University of Minnesota.

MEASUREMENT OF FALSE RECOGNITION

The DRM paradigm is a common and useful technique for studying false memories in a controlled laboratory setting. In the original version of this paradigm, participants are presented with lists of 'related' words (e.g. *bed, rest, awake, dream,* etc.) that are associated with one 'critical' word (e.g. *sleep*) that is not actually presented. In the recognition version of this paradigm, participants are subsequently given an old/new recognition test, which includes a subset of the presented related words and the unpresented critical words. Typically, memory performance is measured by calculating the hit rate for the related words and the false-alarm rate for the critical words. Because the false-alarm rate for critical words often matches or exceeds the hit rate for related words, false alarms to critical words are thought to reflect false memories for these words.

Application of signal detection theory (SDT; Green & Swets, 1966) and the inclusion of additional conditions provides another, more theoretically constraining measure of false recognition in the DRM paradigm. In the present study, we used a slightly modified version of the DRM paradigm. These modifications enabled us to apply SDT to measure recognition performance for multiple types of words. Our modifications are as follows.

In typical DRM experiments, critical words are not presented during encoding, and memory for the unpresented critical words and some of the presented related words is subsequently tested. We will refer to test words that were presented at encoding as old words, and to test words that were not presented at encoding as new words. Thus in typical DRM experiments, the recognition test includes new critical words and old related words. In our modified procedure, some of the critical words are actually presented along with their associated lists during encoding, replacing one of the related list words. Therefore, in addition to new critical words and old related words, our recognition test also includes old critical words and new related words. Furthermore, we also present unrelated-word lists during encoding in addition to the standard related-word lists, and both old and new unrelated words are included in the recognition test. An increasing number of DRM experiments have adopted similar modifications to the typical procedure (e.g. Miller & Wolford, 1999; Westerberg & Marsolek, 2003), and the results from these studies indicate that these modifications do not affect the key finding observed in standard DRM experiments. When slightly modified versions of the DRM paradigm are employed, the false-alarm rate for new critical words is greater than or equivalent to the hit rate for old related words.

The assumptions of SDT as applied to recognition memory are as follows. To make a recognition decision, participants assess the memory strength (the level at which a test item reactivates a recently encoded representation of that item or very similar items) of each test item. The mean memory strength will generally be larger for old items than for new items, resulting in two overlapping probability distributions along the memory strength dimension. In addition, a criterion is set along the memory strength dimension, such that items located to the right of the criterion are given 'old' responses, and items located to the left of the criterion are given 'new' responses (see Figure 1). Under these assumptions, it is possible to calculate independent measures of recognition sensitivity and bias from the observed hit and false-alarm rates. Sensitivity is the ability to discriminate between previously presented and previously unpresented items on the recognition test, and it is measured as the distance between the means of the old- and new-item distributions. Bias is the tendency to produce predominantly 'old' or 'new' responses

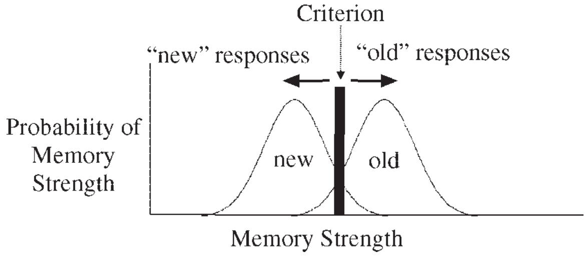

Figure 1. According to SDT, new and old test items are normally distributed along a memory strength dimension. Items with a memory strength greater than the criterion are given 'old' responses

on the recognition test, and it is measured as the distance between the intersection of the old- and new-item distributions and the criterion.

Applying SDT analyses to the modified DRM paradigm, Westerberg and Marsolek (2003) observed lower sensitivity for critical words than for both related words and unrelated words, and a larger bias to respond 'old' to critical words than to both related words and unrelated words. Calculating signal-detection measures in this paradigm is beneficial for at least two reasons. First, the effect of lower sensitivity for critical words than for related and unrelated words is more informative than the simpler observation of high false-alarm rates for critical words. High false-alarm rates for critical words are ambiguous, because they do not indicate whether participants can discriminate between presented and unpresented critical words. High critical-word false-alarm rates could be accompanied by substantially higher critical-word hit rates, indicating an effective ability to recognize and discriminate old versus new critical words. Alternatively, they could be combined with equal magnitude hit rates for critical words, indicating no ability to recognize and discriminate between old versus new critical words (see Miller & Wolford, 1999). Second, with appropriate indices, recognition sensitivity and bias are independent measures (a change in sensitivity does not change the range of values that can be attained in the bias measure; Snodgrass & Corwin, 1988). The effect of lower sensitivity for critical words than for related and unrelated words more strongly constrains theory than the effect of a larger bias to respond 'old' to critical words than to related and unrelated words or the effect of high false-alarm rates for critical words. Fewer theories are able to account for the sensitivity effect than the bias effect or critical-word false alarm effect. Both storage-based and decision-based theories can account for the latter effects equally effectively (see Wickens & Hirshman, 2000; Wixted & Stretch, 2000), but only storage-based theories can account for the sensitivity effect and associated results from analyses of receiver-operating-characteristics (see Westerberg & Marsolek, 2003).

For these reasons, the most useful and constraining measure of false recognition may be significantly lower recognition sensitivity for critical words that for the other two word types. In the present experiments, we will further test whether this difference in sensitivity is crucial for a compelling conceptualization of false recognition and reductions in false recognition.

INSTRUCTIONAL WARNINGS AND REDUCTIONS IN FALSE MEMORY

The ease with which false memories can be generated highlights the need to understand how they can be prevented. Previous research has addressed whether warning participants about the nature of the DRM paradigm can reduce false memory expression in that paradigm (e.g. Anastasi, Rhodes, & Burns, 2000; Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McCabe & Smith, 2002; McDermott & Roediger, 1998; Neuschatz, Benoit, & Payne, 2003; Neuschatz, Payne, Lampinen, & Toglia, 2001). In these experiments, participants were shown sample lists of related words and their associated critical words, and they were told that people frequently falsely remember the critical words when given memory tests (see Table 1 for a summary of previous warning experiments).

When warnings are administered after the encoding phase but before the test phase, they do not appear to reduce false recognition. Equally high false-alarm rates for critical words are observed under warned and unwarned conditions (Anastasi et al., 2000; Neuschatz et al., 2001; but also see Gallo et al., 2001, and McCabe & Smith, 2002, for exceptions). Although measures from SDT were not calculated, these results suggest that warnings must influence encoding processes to be effective in reducing false memories.

In contrast, when warnings are administered before the encoding phase, they do affect false recognition. This was demonstrated in a couple of studies through a reduction in false-alarm rates for critical words in the warned condition (Gallo et al., 1997; McCabe & Smith, 2002), but it also has been demonstrated in other studies that involved measurements of both hit and false-alarm rates for critical words. Although SDT measures of recognition sensitivity per se were not reported, Gallo et al. (2001), McDermott and Roediger (1998), and Neuschatz et al. (2003) all found hit and false-alarm rates that indicated increases in the ability to discriminate old and new critical words when warnings were given before the encoding phase compared with when warnings were not given.

It should be noted that Gallo and colleagues (2001) did calculate three alternative measures of false recognition aside from measuring hit and false-alarm rates (a high-threshold correction procedure, d', and A'). In their study, the recognition test included both critical words and related words from lists that were not presented during encoding, and false alarms for the alternative analyses were defined as 'old' responses to items whose associated lists were not presented at encoding. Thus, results from these alternative analyses reflect the tendency to falsely recognize an item as a function of whether or not its associated list was previously presented. Although interesting, this is not a measure of recognition sensitivity as traditionally conceptualized in studies of recognition memory.

POSSIBLE EFFECTS OF INSTRUCTIONAL WARNINGS

The results from previous studies indicate that instructional warnings can affect encoding processes in a manner that leads to increased sensitivity for critical words. However, this effect of warnings on encoding processes could take one of two different forms.

First, warnings may affect encoding in a manner that is specific to the processing of critical words only. Warnings may affect how critical words are encoded and stored, but not how related and unrelated words are encoded and stored. The reasoning is as follows. In the DRM paradigm, critical words are highly activated during encoding (even when they are not presented) through associative activation that spreads from the presented words to words related to the presented words. The spread of activation from related to critical words may be entirely automatic (c.f., Underwood, 1965; Roediger, Balota, & Watson, 2001), or alternatively, individuals may consciously bring the critical word to mind (Roediger & McDermott, 1995; McDermott & Roediger, 1998). Regardless of

Table 1. Summary of previous DRM warning experiments

	Critical words tested	Related words tested	Unrelated words tested	Sensitivity measure(s)
Experiments with warning after study:				
Anastasi, Rhodes, and Burns, 2000	New, USL	Old, USL	None	None
Gallo, Roediger, and McDermott, 2001*	New, old, USL	Old, USL	None	High threshold correction, A', d'; all measures compared recognition of words whose associated lists were presented at study (new, old) with false alarms to words whose associated lists were not presented at study (USL)
McCabe and Smith, 2002	New, USL	Old, USL	None	Hit rate for old related words was compared to false-alarm rate for new critical words
Neuschatz, Payne, Lampinen, and Toglia, 2001	New, USL	Old, USL	None	None
Experiments with warning before study:				
Gallo, Roberts, and Seamon, 1997	New only	Old only	None	None
Gallo, Roediger, and McDermott, 2001	New, old, USL	Old, USL	None	None High threshold correction, A', d'; all measures compared recognition of words whose associated lists were presented at study (new, old) with false alarms to words whose associated lists were not presented at study (USL)
McCabe and Smith, 2002	New, USL	Old, USL	None	Hit rate for old related words was compared to false-alarm rate for new critical words
McDermott and Roediger, 1998	New, old, USL	None	None	None
Neurschatz, Benoit, and Payne, 2003	New and old	None	None	None

New = word was not presented but associated list was presented at study.
Old = word was presented within context of associated list at study.
USL = word was not presented and associated list was not presented at study.
* = Gallo et al., 2001 also tested one group of participants with new and USL critical words only.

whether individuals are aware of this activation, it may be responsible for the high critical-word false-alarm rates and relatively low critical-word sensitivity typically observed in the (unwarned) DRM paradigm.

Gallo et al. (2001) suggested that when warnings similar to those used in previous experiments are given before the encoding phase, critical words are still associatively activated. Furthermore, participants are instructed (or may implement their own strategy) to infer and remember the critical word for each list. Thus, when the critical word is not presented, the critical-word representation is also 'tagged' as 'critical word not presented'. This enables warned participants to more effectively discriminate new from old critical words during the recognition test, compared with unwarned participants. Consistent with this possibility, Neuschatz et al. (2003) demonstrated that warnings are more effective at reducing false recognition when the critical words are highly identifiable than when they are not as easily identifiable. Importantly, according to this view, the ability to discriminate between old and new unrelated words should not be affected by warnings, because the instruction or strategy to 'tag' critical words as not presented is not applicable to word lists lacking any common theme. By this theory, sensitivity for critical words should be greater for warned participants compared with unwarned participants, but sensitivity for unrelated words should not differ as a function of warning. This prediction was tested in the following experiment.

Alternatively, instructional warnings may affect encoding processes in a more general manner. Warnings may affect how all word types are encoded and stored, not just critical words. For example, according to global matching theories of memory (see Clark & Gronlund, 1996, for a review), each word presented during encoding is stored as a vector of features. During the subsequent recognition test, each test word is compared in parallel against all traces in memory. If the sum of the similarities between the test word and all stored traces (i.e. the global similarity) exceeds some criterion, an 'old' response is given. According to one particularly applicable global matching theory (MINERVA2; Hintzman, 1988), in the DRM paradigm, at test, critical words would have many features in common with many stored traces, because of the associative similarity between the related words and their critical words. Therefore, even when critical words do not exactly match one of the stored traces in memory (i.e. when critical words are new at test), the global similarity between critical words and many stored traces may still be large enough to elicit 'old' responses to these words (e.g. Arndt & Hirshman, 1998; Westerberg & Marsolek, 2003). This could help to explain the typical observations of relatively numerous false alarms to critical words and relatively low sensitivity for critical words in the (unwarned) DRM paradigm.

In addition, warnings may cause participants to pay greater attention to each word actually presented during encoding. Thus, warned participants may encode a greater number of item-specific features for each word presented at encoding compared with participants who are not warned. (Item-specific features are those that are distinctive to a particular word and not shared with other words used in the experiment.) This should increase the impact of the exact matches that occur between old test items and their stored traces on global similarity, resulting in a greater ability to discriminate old from new test words. It is important to note that this should be a general effect that applies to all three word types. Critical, related, and unrelated words all have item-specific features; thus, all could benefit from the encoding of additional item-specific features in a manner that increases recognition sensitivity at test. By this theory, sensitivity should be greater when participants are warned compared with when they are not warned, and this should be true for critical words, related words and unrelated words. We also tested this prediction in the following experiment.

NEW EXPERIMENTAL TESTS

Although results from prior studies suggest that warnings can increase critical-word sensitivity, no previous experiments have included the necessary conditions to examine the effects of warnings on sensitivity for other word types (see Table 1). In the following experiment, we investigated the effects of warnings on all three word types and obtained crucial evidence regarding possible mechanisms through which warning instructions reduce false recognition.

Using the modified version of the DRM paradigm described above, the standard related-word lists and lists of unrelated words were presented at encoding, and recognition memory for old and new critical, old and new related, and old and new unrelated words was tested. Additionally, one group of participants was given warnings similar to those used in previous experiments before the encoding phase, and another group was not warned. This enabled us to calculate SDT measures of sensitivity and bias for critical, related and unrelated words for both warned and unwarned participants. Furthermore, the inclusion of critical words in some of the encoding lists eliminated the utility of adopting a strategy whereby all test words highly related to the list themes are rejected (Gallo et al., 2001). If warnings affect memory processes in a manner that is specific to critical words only (as predicted by the spreading activation/tagging theory), critical-word sensitivity should be greater for warned participants than for unwarned participants, but unrelated-word sensitivity should not differ between those groups, producing an interaction effect between instruction type and word type. Alternatively, if warnings affect how all words are encoded and stored (e.g. as predicted by global matching theory), sensitivity should be greater for warned participants than for unwarned participants as a main effect, but no interaction should be observed between instruction type and word type.

Note, however, that if warnings similar to those used in previous studies do increase sensitivity to all word types, it is possible that the spreading activation/tagging theory could be retained as an explanation for increased sensitivity for critical words, as long as an additional explanation can be offered for the accompanying increased sensitivity for unrelated words. One possibility is that in addition to causing participants to 'tag' unpresented critical words (thereby increasing critical-word sensitivity), warnings that encourage participants to identify the critical word for each list may also increase the level with which all words presented at encoding are activated. This would increase the ability to discriminate between old and new unrelated words at test.

To test that hypothesis, we also warned a third group of participants about the DRM false memory effect in a manner that does not encourage them to identify the critical word for each list. In previous DRM warning experiments (see Gallo et al., 2001; McDermott & Roediger, 1998; Neuschatz et al., 2003), warned participants were shown a sample study list and its associated critical word, and they were encouraged to note and remember whether the critical word for each list was presented during encoding. In the following experiment, in addition to warning one group of participants with 'list-theme encouraging warnings' similar to those used in previous experiments, a third group of participants were shown a sample study list and its associated critical word, but were then strongly discouraged from thinking about the themes of each list (i.e. the critical words), and instead were encouraged to think about the unique and distinctive aspects of each presented word, in an effort to help reduce false recognition.

From a spreading activation perspective, 'list-theme discouraging' warnings may increase the spread of activation to unique and distinctive associates of each presented word, and decrease the amount of activation that spreads to words associated with the list themes. Because critical words are highly associated with the list themes, unpresented critical words may receive much less activation following this type of warning compared with no warning, which may increase the ability to discriminate between old and new critical words. However, no such increase in sensitivity may be observed in the unrelated word condition. Because unpresented words are not associated with the presented words in an unrelated word list, a decrease in the spread of activation from presented words to associated words may not affect sensitivity for unrelated words. Therefore, list-theme discouraging warnings may increase critical-word sensitivity, but not unrelated-word sensitivity, producing an interaction between word type and instruction type.

Alternatively, according to global matching theory (e.g. Hintzman, 1988), list-theme discouraging warnings should affect all word types in the same manner. Specifically, warnings that encourage participants to focus on the unique and distinctive aspects of each presented word may cause a greater number of item-specific features to be stored for each presented word. Therefore, sensitivity for all word types may increase equivalently, because the encoding of additional item-specific features would increase the ability to recognize all old words as having been presented previously during encoding.

Method

Participants

One hundred ninety-two students from the University of Minnesota participated in this experiment in exchange for course credit (64 in the unwarned condition, 64 in the list-theme encouraging warning condition, and 64 in the list-theme discouraging warning condition). Half of the participants in each condition were male, and the other half were female. All participants had normal or corrected-to-normal vision.

Materials

Thirty-four lists of 16 words from the Stadler, Roediger, and McDermott (1999) norms were used in these experiments. Fifteen of the 16 total words comprising each list (15 related words and 1 critical word) were presented to a given participant. Thirty-two of the lists were used as the experimental lists. The two additional lists were fillers, one presented at the beginning of the encoding phase and one presented at the end of the encoding phase, to attenuate primacy and recency effects. Additionally, four lists of 15 unrelated words were created and used as experimental lists. The unrelated words were judged to be unrelated to each other and to the other experimental words (i.e. the related words and the critical words). They were also matched in frequency of occurrence in the English language (Francis & Kucera, 1982) and number of distinct dictionary entries (Softkey Multimedia Inc., 1995) to the critical words.

Stimulus presentation was controlled by an Apple Power Macintosh 7600/132. Words presented during the encoding phase were presented auditorily over headphones. Words presented during the recognition test were presented visually on a Mitsubishi DiamondPro 710 monitor. Each test word appeared in the centre of the screen, in 24-point bold black

Helvetica font against a white background. A 2-mm black dot in the center of the screen served as a central fixation point.

Procedure

The experiment consisted of an encoding phase immediately followed by a test phase. Experimental sessions were conducted individually.

Prior to the encoding phase, one-third of the participants were given list-theme encouraging warning instructions modelled after those used in past DRM warning experiments (Gallo et al., 2001; McDermott & Roediger, 1998). Before the study lists were presented, participants were shown a sample study list and its associated critical word, they were told that people frequently falsely recognize the critical words when given a recognition test, and they were instructed not to make this error. In addition, these participants were instructed to infer the critical word for each list presented during encoding and remember whether it was or was not actually presented. Another one-third of the participants were given list-theme discouraging warning instructions before the encoding phase. They were shown a sample study list and its associated critical word, they were told that people frequently falsely recognize unpresented critical words on the recognition test, and they were instructed not to make this error. Most importantly, to help achieve this goal, participants were strongly discouraged from thinking about the themes of the related-word lists, and instead were encouraged to think about the unique and distinctive aspects of each presented word. The remaining one-third of the participants were given the standard, unwarned DRM encoding instructions, encouraging them to remember all of the words that are presented because their memories would be tested.

Encoding phase. Participants listened to 34 word lists (32 related-word lists and 2 unrelated-word lists, randomly intermixed), presented at a rate of 1.5 seconds per word. Auditory presentations were used to ensure that robust false recognition effects would be observed (Smith & Hunt, 1998). During the presentation of each list, participants looked at a fixation point in the centre of the monitor while listening to the words. Immediately after each list was presented, the fixation point disappeared. Participants then attempted to complete a visual maze for 20 seconds, to minimize rehearsal of the presented words between list presentations. After 20 seconds, a tone sounded, indicating that participants should stop working on the maze and look at the screen, as the next list would be presented in 2.5 seconds.

As is standard for experiments using the DRM paradigm, within each related-word list, the words were presented in order of their association to that list's critical word. The first word was the highest associate of the critical word, and the 15th word was the lowest associate of the critical word. For half of the word lists, the critical word was presented during encoding, in place of one of the related words. The critical word replaced related words occurring in the odd-numbered 1–15 ordinal positions, and the replacement was counterbalanced across participants and across the 16 lists in which the critical word was presented for one participant. The replacement of a related word with a critical word provided us with a new related word to be presented during test and an old critical word to be presented during test. For the other half of the word lists, the critical word was not presented during encoding, and these lists provided us with an old related word to be presented during test and a new critical word to be presented during test. In addition, eight words from each of the two unrelated-word lists presented during encoding were presented during test as old unrelated words, and eight words from each of the two unrelated-word lists that were not presented during encoding were presented during test as new unrelated words. Words were counterbalanced across participants, assuring that all critical words, all odd-numbered related words, and all unrelated words were used equally often in the old and new test conditions.

Recognition test. The recognition test comprised 96 words. Thirty-two were critical words, 32 were related words, and 32 were unrelated words. Within each group of 32 words, 16 were old words and 16 were new words. The order of trial presentation was random, with the exception that the same experimental condition was represented in no more than three consecutive trials. For each of the 96 test trials, a central fixation point appeared on the monitor for 500 milliseconds, followed by a centrally presented word for 183 milliseconds. Participants were asked to judge whether the word had been presented previously in the encoding phase, and to indicate their decision by pressing a button on the button box. If they thought the word was presented previously (was 'old'), they pressed the left button, and if they thought the word had not been presented previously (was 'new'), they pressed the right button. Participants were asked to respond as accurately as possible, and to press a button as soon as they had made a decision. After each old/new decision, participants were asked to rate their confidence level regarding their decision on a 1–6 scale (1 = definitely old, 6 = definitely new), by pressing the corresponding number on the keyboard.

Results

For each participant, the observed hit and false-alarm rates were used to calculate the SDT measures of sensitivity and bias for each of the three word types. The hit and false-alarm rates are listed in Table 2. The calculations require that hit and false-alarm rates be transformed to z-scores, and because values of 0 and 1 are undefined in z-space, those proportions were converted to 1/(2N) and 1-1/(2N), respectively, as recommended by Macmillan and Creelman (1991). A five-point receiver-operating-characteristic (ROC) curve was generated for each participant for each condition, by plotting the z-transformed hit rate against the z-transformed false-alarm rate for each level of confidence. The first point of the curve was the hit rate versus the false-alarm rate when words were judged to be 'definitely old', and each successive point was the hit rate versus the false-alarm rate for the next confidence level and all previous confidence levels. Linear regression analyses conducted on the z-ROC curves indicated that the slopes of the lines of best fit were less than one for some of the experimental conditions, and differed across the three word types. This meant that the traditional measures of sensitivity and bias (d' and c) were not

Table 2. Hit and false-alarm rates by condition (standard deviations in parentheses)

	Instruction type					
	No warning		List-theme encouraging		List-theme discouraging	
Word type	Hits	False alarms	Hits	False alarms	Hits	False alarms
Critical	0.80 (0.14)	0.71 (0.18)	0.82 (0.12)	0.68 (0.20)	0.76 (0.10)	0.63 (0.19)
Related	0.60 (0.17)	0.42 (0.19)	0.68 (0.15)	0.49 (0.17)	0.59 (0.14)	0.38 (0.15)
Unrelated	0.49 (0.15)	0.29 (0.15)	0.55 (0.18)	0.29 (0.17)	0.54 (0.16)	0.28 (0.15)

appropriate for the data, because they assume a slope of 1.0. Therefore, d_a and c_2 were calculated to measure sensitivity and bias (respectively) for each participant for each word type (averaged across the five points of the *z*-ROC curve). These measures allow for appropriate and comparable calculations of sensitivity and bias when some of the slopes do not equal 1.0 (and when a slope does equal 1.0, the measures reduce to the traditional measures of d' and c; Simpson & Fitter, 1973; Macmillan & Creelman, 1991).

The sensitivity and bias measures (d_a and c_2) calculated for each condition for each participant were the dependent measures submitted to separate repeated-measures analyses of variance (ANOVAs), using participants as the random variable. Word type (critical, related or unrelated) was the within-participants independent variable, and instruction type (no warning, list-theme encouraging warning, or list-theme discouraging warning) was the between-participants independent variable.

Sensitivity

Figure 2 displays the results from the sensitivity measure. First, the main effect of instruction type was significant, $F(2, 189) = 4.07$, $MSe = 0.31497$, $p < 0.05$, indicating that sensitivity was greater for participants who received warning instructions of either type (list-theme encouraging warning = 0.56, list-theme discouraging warning = 0.55) than for participants who were unwarned (0.41), $F(1, 189) = 8.14$, $MSe = 0.314966$, $p < 0.01$, for the simple effect contrast. Sensitivity was not significantly different between the two warning types, $F < 1$, for the simple effect contrast. Second, as observed previously by Westerberg and Marsolek (2003), the main effect of word type was significant, $F(2, 378) = 7.89$, $MSe = 0.16655$, $p < 0.001$. A simple effect contrast indicated that sensitivity for critical words (0.41) was significantly lower than sensitivity for related words (0.57) and sensitivity for unrelated words (0.54), $F(1, 378) = 15.4$, $MSe = 0.166547$, $p < 0.001$.

Most important, the new finding is that the interaction between instruction type and word type did not approach significance, $F < 1$, indicating that the false recognition effect did not differ as a function of instruction type. A simple effect contrast revealed that in the

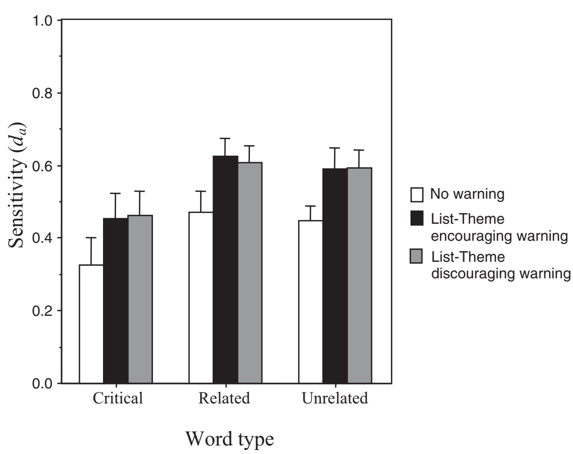

Figure 2. Sensitivity (d_a) plotted as a function of instruction type (no warning, list-theme encouraging warning, list-theme discouraging warning) and word type (critical, related, and unrelated). Error bars depict standard errors of the mean

unwarned condition, critical-word sensitivity (0.33) was significantly lower than related-word sensitivity (0.47) and unrelated-word sensitivity (0.45), $F(1,756) = 4.55$, $MSe = 0.16655$, $p < 0.05$. Similarly, simple effect contrasts indicated that in the list-theme encouraging warning condition, critical-word sensitivity (0.45) was significantly lower than related-word sensitivity (0.63) and unrelated-word sensitivity (0.59), $F(1,756) = 9.06$, $MSe = 0.16655$, $p < 0.005$, and in the list-theme discouraging warning condition, critical-word sensitivity (0.46) was significantly lower than related-word sensitivity (0.61) and unrelated-word sensitivity (0.59), $F(1,756) = 7.34$, $MSe = 0.16655$, $p < 0.01$.

Bias

Replicating previous experiments (Miller & Wolford, 1999; Westerberg & Marsolek, 2003), the main effect of word type was significant, $F(2,378) = 486$, $MSe = 0.04640$, $p < 0.001$. Simple effect contrasts revealed that bias to respond 'old' to critical words (-0.61) was significantly larger than bias to respond 'old' to related words (-0.15), $F(1,378) = 440$, $MSe = 0.04640$, $p < 0.001$, and bias for related words significantly differed from bias for unrelated words (0.06), $F(1,378) = 89.5$, $MSe = 0.04640$, $p < 0.001$. The main effect of instruction type also was significant, $F(2,189) = 3.44$, $MSe = 0.12853$, $p < 0.05$. Bias to respond 'old' was larger for participants warned with list-theme encouraging instructions (-0.29) than for participants warned with list-theme discouraging instructions (-0.21) and unwarned participants (-0.21), $F(1,189) = 6.87$, $MSe = 0.12853$, $p < 0.01$, for the simple effect contrast. Bias did not differ significantly between participants warned with list-theme discouraging instructions and unwarned participants, $F < 1$, for the simple effect contrast. The interaction between word type and instruction type was also significant, $F(4,378) = 3.48$, $MSe = 0.04640$, $p < 0.01$. Bias to respond 'old' was significantly greater for participants warned with list-theme encouraging instructions compared with participants warned with list-theme discouraging instructions and unwarned participants for both critical and related words, $F(1,567) = 5.07$, $MSe = 0.07378$, $p < 0.05$, and $F(1,567) = 44.9$, $MSe = 0.07378$, $p < 0.001$, respectively, for the simple effect contrasts. However, bias to respond 'old' to unrelated words was not significantly different for participants warned with list-theme encouraging instructions compared with participants warned with list-theme discouraging instructions and unwarned participants, $F < 1$, for the simple effect contrast. Participants warned with list-theme encouraging instructions may have applied a relatively liberal criterion for critical and related words because they expected many of the new test items to be tagged in memory as not previously presented (which may create a tendency for all other test items related to the list themes to be judged 'old').

In addition, to facilitate comparisons between the current results and results from previous DRM warning experiments, two separate ANOVAs were conducted on the hit and false-alarm rates observed for each participant for each word type. In these analyses, word type (critical, related or unrelated) was the within-subjects independent variable and instruction type (no warning, list-theme encouraging warning or list-theme discouraging warning) was the between-subjects independent variable.

False Alarms

Replicating previous DRM warning experiments, the main effect of word type was significant, $F(2,378) = 318$, $MSe = 0.02326$, $p < 0.001$. More false alarms were made to critical words (0.67) than to related (0.43) and unrelated words (0.28), $F(1,378) = 430$, $MSe = 0.02326$, $p < 0.001$, for the simple effect contrast. Also, more false alarms were made to related words than to unrelated words, $F(1, 378) = 87.0$, $MSe = 0.02326$, $p < 0.001$, for the simple effect contrast. The main effect of instruction type was also significant, $F(2, 189) = 4.04$, $MSe = 0.04081$, $p < 0.05$. False alarms were significantly reduced for participants given the list-theme discouraging warnings (0.43) compared with participants given the list-theme encouraging warnings (0.49) and unwarned participants (0.47), $F(1, 189) = 7.52$, $MSe = 0.04081$, $p < 0.01$, for the simple effect contrast. There was no difference in false alarms for participants given the list-theme encouraging warnings and unwarned participants, $F < 1$, for the simple effect contrast.

These main effects were modulated by the significant word type by instruction type interaction, $F(2, 378) = 3.02$, $MSe = 0.02326$, $p < 0.05$ (see Table 2). False alarms to critical words were not reduced for participants given the list-theme encouraging warnings compared with unwarned participants ($F < 1$), contrary to findings from previous warning experiments (e.g. Gallo et al., 1997, 2001; McCabe & Smith, 2002; McDermott & Roediger, 1998). However, critical-word false alarms were reduced for participants given list-theme discouraging warnings compared with unwarned participants and participants given list-theme encouraging warnings, $F(1, 567) = 5.91$, $MSe = 0.02911$, $p < 0.05$, for the simple effect contrast. For related words, false alarms actually increased for participants given the list-theme encouraging warnings compared with unwarned participants and participants given the list-theme discouraging warnings, $F(1, 567) = 12.7$, $MSe = 0.02911$, $p < 0.001$, for the simple effect contrast, but related-word false alarms did not differ for unwarned and list-theme discouraging warned participants, $F(1, 378) = 1.48$, $MSe = 0.02911$, $p > 0.2$, for the simple effect contrast. No differences were found in unrelated-word false-alarm rates across the three instruction types (all $F\text{s} < 1$, for the simple effect contrasts).

Hits

As expected, the main effect of word type was significant, $F(2, 378) = 241$, $MSe = 0.01444$, $p < 0.001$. The hit rate for critical words (0.79) was larger than the hit rate for related words (0.62), $F(1, 378) = 198$, $MSe = 0.01444$, $p < 0.001$, for the simple effect contrast, and the hit rate for related words was larger then the hit rate for unrelated words (0.53), $F(1, 378) = 57.1$, $MSe = 0.01444$, $p < 0.001$, for the simple effect contrast. The main effect of instruction type was also significant, $F(2, 189) = 4.71$, $MSe = 0.03813$, $p < 0.05$. The hit rate for participants given list-theme encouraging warnings (0.68) was larger than the hit rate for participants given list-theme discouraging warnings (0.63) and unwarned participants (0.63), $F(1, 189) = 9.43$, $MSe = 0.03813$, $p < 0.005$, for the simple effect contrast. There was no difference in the hit rate for participants given list-theme discouraging warnings and unwarned participants ($F < 1$), for the simple effect contrast.

The word type and instruction type main effects were modulated by the significant word type by instruction type interaction, $F(4, 378) = 3.70$, $MSe = 0.01444$, $p < 0.01$ (see Table 2). The critical-word hit rate was smaller for participants given the list-theme discouraging warnings than for participants given the list-theme encouraging warnings and unwarned participants, $F(1, 567) = 5.90$, $MSe = 0.02233$, $p < 0.05$, for the simple effect contrast, and the critical-word hit rate did not differ between participants given list-theme encouraging warnings and unwarned participants ($F < 1$, for the simple effect contrast). The related-word hit rate was greater for participants given the list-theme encouraging warnings than for participants given list-theme discouraging warnings and unwarned participants, $F(1, 567) = 13.0$, $MSe = 0.02233$, $p < 0.001$, for the simple effect contrast, and the related-word hit rate did not differ between participants given list-theme discouraging warnings and unwarned participants ($F < 1$, for the simple effect contrast). The unrelated-word hit rate was smaller for unwarned participants than for participants given either list-theme encouraging or list-theme discouraging warnings, $F(1, 567) = 5.93$, $MSe = 0.02233$, $p < 0.05$, for the simple effect contrast, and the unrelated-word hit rate did not differ between participants given list-theme encouraging warnings and participants given list-theme discouraging warnings ($F < 1$, for the simple effect contrast).

DISCUSSION

Previous experiments have demonstrated that when participants are warned about the false memory effect before the encoding phase in the DRM paradigm, the ability to differentiate old and new critical words improves (Gallo et al., 1997, 2001; McDermott & Roediger, 1998; Neuschatz et al., 2003). However, no previous studies have included the conditions necessary to determine whether warnings improve the ability to differentiate old and new words of other types. The goal of the current investigation was to determine whether warnings increase memory performance specifically for critical words or, alternatively, if warnings generally change the way all words are encoded and stored. The results indicate that warnings do not improve the processing of critical words in particular; rather warnings increase memory performance in general, across all word types.

One group of participants was given warnings that encourage participants to identify the themes of the study lists (similar to warnings given in previous experiments). The results indicated that these warnings increase memory performance for all word types, compared with no warning. According to spreading activation/tagging theory, because the warnings specifically encouraged participants to 'tag' in memory critical words that were not presented with their lists, these warnings should have increased recognition sensitivity for critical words but not recognition sensitivity for unrelated words. In at least some of the lists, the critical word may not have been highly identifiable (see Neuschatz et al., 2003). Given that our modified procedure involved testing related words that were sometimes omitted from their respective lists during study, it is possible that some of the unpresented related words may have also been tagged as 'not presented'. Note, however, that if unpresented related words were also tagged, both critical- and related-word sensitivity should have increased relative to unrelated-word sensitivity when list-theme encouraging warnings were given, which was not observed in this experiment.

It should also be noted that although the list-theme encouraging warnings used in the present experiment were similar to those used in previous DRM warning experiments, the pattern of results observed in the present experiment were slightly different than results obtained in previous warning experiments. In previous experiments, this type of warning resulted in a significant decrease in false alarms to critical words (Gallo et al., 1997, 2001; McCabe & Smith, 2002; McDermott & Roediger, 1998; Neuschatz et al., 2003). In the present experiment, list-theme encouraging warnings numerically decreased the critical-word false-alarm rate compared with no warning, but this decrease was not significant. One possible reason for this discrepancy is that the number of study lists presented to each participant was much larger in the present experiment (36 lists = 540 words) than in the previous experiments (e.g. Gallo et al. presented 12 lists = 180 words). The number of words that participants are asked to remember may influence the strategy used by participants to decrease false recognition. Tagging may not have been advantageous in the present experiment because there was a very large number of critical words (32) to keep track of. In previous experiments, tagging may have been more beneficial because there was a relatively smaller number of critical words to note and remember (e.g. 12 in Gallo et al., 2001). Thus, it is possible that when fewer lists are studied, participants are more likely to tag critical words, which in turn may help reduce false alarms to critical words.

Another group of participants was given warnings that strongly discouraged them from thinking of the list themes and instead encouraged them to focus on the unique and distinctive aspects of each presented word. The results indicated that these warnings also increase memory performance for all word types, compared with no warning. According to spreading associative activation theory, if the level of activation of words associated to the list themes is at least somewhat amenable to conscious control, these warnings should have increased sensitivity for critical words but not sensitivity for unrelated words. It is possible that the spread of activation is entirely automatic; thus, these warnings may not have effectively reduced the amount of associative activation received by critical words. However, accumulating evidence argues against this possibility. Many experiments have demonstrated that participants claim to remember specific details about the study episode while falsely remembering unpresented critical words (e.g. Norman & Schacter, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger & McDermott, 1995), which suggests that the critical words are consciously activated during study. Furthermore, perceptual priming has been observed for non-presented critical words in the DRM paradigm, which has been taken as evidence that critical words were consciously evoked during study (McDermott, 1997). Admittedly though, the extent to which associative activation is amenable to conscious control is still under debate and is likely to be dependent on multiple factors (Roediger, Balota, & Watson, 2001). Future research is necessary to explore this possibility in more detail.

The results from the present experiment can be accounted for from a global matching perspective. Both kinds of warnings improved memory performance for all word types (critical, related and unrelated words) and the magnitude of this improvement was the same across all conditions. This is consistent with the hypothesis that both kinds of warnings lead to storage of memory traces for the actually presented words that contain a greater number of item-specific features than occurs when no warnings are given. However, it is important to note that no extant theory of false memory, including global matching models, can account for the complete pattern of sensitivity results and receiver-operating characteristics found in DRM experiments (see Westerberg & Marsolek, 2003). Thus, further theoretical development is needed in this area.

Although the present discussion has focused on global matching and spreading activation accounts of warning-related reductions in false recognition, it should be mentioned that other theories make specific claims regarding processes underlying false recognition. For example, fuzzy trace theory (Reyna & Brainerd, 1995) suggests that recognition decisions can be supported by either gist or verbatim information; thus, false recognition of critical words may arise from erroneously relying on gist information to make recognition decisions. According to the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993), strong levels of retrieved semantic information may be used to incorrectly assign the source of a non-presented critical word to one of the presented lists. These theories suggest that the critical processes responsible for false recognition occur at the time of retrieval. However, as noted above, warnings presented between encoding and retrieval are not effective at reducing false recognition, as these theories would predict. Although these theories do not provide explanations for observed reductions in false recognition with instructional warnings, they can account for other false memory effects observed in this paradigm, and certain aspects of these theories may be crucial for any complete theory of false recognition in the DRM paradigm.

The results from this experiment also have implications for conceptualizing false recognition in the DRM paradigm. In past experiments, warnings given before encoding were shown to improve memory performance for critical words. If false recognition is conceptualized as high false-alarm rates or just low sensitivity for critical words, then it can be concluded that warnings do reduce false recognition. However, if false recognition is more specifically conceptualized as significantly lower sensitivity for critical words than for related and unrelated words, then the present results suggest that warnings actually do not reduce false recognition, at least when a relatively large number of lists are studied. In the present study, participants who were unwarned and participants who were warned in two different ways all exhibited significantly lower sensitivity for critical words than for related and unrelated words. Thus, by the latter conceptualization of false recognition, no reductions in false recognition per se were observed as a function of warning. Whatever the mechanism that is responsible for lower sensitivity for critical words than for related words and unrelated words in the modified DRM paradigm (for discussion, see Westerberg & Marsolek, 2003), it is a mechanism that apparently is not modified by warnings. However, it is possible that the number of lists studied is an important predictor. When fewer lists are studied, a tagging strategy could be implemented to improve critical-word sensitivity relative to related- and unrelated-word sensitivity, thereby reducing false recognition (Gallo et al., 2001). Future research is needed to explore this possible predictive factor.

The present findings also have implications concerning the accuracy of memory in everyday life. First, they suggest that warning people of the dangers of false recognition can be effective not only at reducing false alarms (as observed in previous warning experiments), but also at increasing memory sensitivity, regardless of the type of information that is encountered or the type of warning strategy that is used. For example, consider a witness to a theft who is asked to identify the crooks in a line-up. Without being warned, there is a high likelihood that the witness may incorrectly identify an innocent individual who is similar in appearance to the actual crooks as one of the individuals involved in the theft. Results from previous warning experiments suggest that warning the witness should decrease the likelihood that individuals who are similar to the criminals but are innocent are incorrectly identified. The results from the present experiment suggest that warnings may be useful to reduce other injustices in this situation as well. Consider also that one of the crooks may not be very similar in appearance to the rest of the crooks. When the witness is not warned, the witness will be less likely to implicate that dissimilar crook than the other, more similar looking crooks. However, according to the results of the present study, if the witness is warned, this will also increase the ability of the witness to identify the dissimilar crook as having been involved in the crime. Thus the present results are important because they demonstrate that the effect of warnings is not limited to just reducing false alarms to similar but unencountered items. Rather, warnings improve the ability to discriminate between old and new items of all kinds, which can take the form of improving the ability to identify items that have been previously encountered as well as improving the ability to avoid implicating items that *have* not been previously encountered. However, the persistence of relatively low sensitivity for critical words compared with sensitivity for other word types, even under conditions of implementing great caution, indicates that although the ability to discriminate between highly similar items can be improved by warnings, warnings will not improve this ability to a level that is comparable with the ability to discriminate between less similar items. This should be taken into account when assessing the accuracy of a memory.

In conclusion, the present experiment is the first DRM warning experiment to demonstrate that although warnings generally improve memory performance in the DRM paradigm, the improvements are not specific to memory for critical words, at least when a large number of lists are studied. This was observed with warning instructions that encourage participants to identify the list themes and with warning instructions that discourage participants from thinking about list themes. This finding is important because it helps to constrain theoretical interpretations of false recognition in the DRM paradigm, and it has implications for reducing false memories in everyday life. Future research should be directed at determining if and how memory performance for critical words can be improved to a level comparable to that of other word types.

ACKNOWLEDGEMENTS

This research was supported by Grant HD07151 from the Center for Cognitive Sciences in conjunction with the National Institute of Health and Human Development and a grant from the Office of the Vice President for Research and Dean of the Graduate School at the University of Minnesota. We thank Mirra Fine, Erin Hannigan, Matt Holt, and Vaughn Steele for their assistance with data collection. Parts of this research were reported at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO (November, 2002), and at the 10th Annual Meeting of the Cognitive Neuroscience Society, New York, NY (April, 2003).

REFERENCES

Anastasi, J. S., Rhodes, M. G., & Burns, M. C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *American Journal of Psychology, 113*, 1–26.

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: explanations from a global matching perspective. *Journal of Memory and Language, 39*, 371–391. DOI: 10.1006/jmla.1998.2581.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: how the models match the data. *Psychonomic Bulletin and Review, 3*, 37–60.

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

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage.* Boston, MA: Houghton Mifflin Company.

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, & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin and Review, 8*, 579–586.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

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

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

Kelley, C. M., & Jacoby, L. L. (1990). The construction of subjective experience: memory attributions. *Mind and Language, 5*, 49–68.
Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 584–589.
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
McCabe, D. P., & Smith, A. D. (2002). The effects of warnings on false memories in young and older adults. *Memory and Cognition, 30*, 1065–1077.
McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin and 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 and Language, 39*, 508–520. DOI: 10.1006/jmla.1998.2581.
Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: the role of criterion shift in false memory. *Psychological Review, 106*, 398–405.
Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott False-Memory paradigm: the role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 35–41.
Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*, 53–71.
Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: exploring the characteristics of illusory memories. *Memory and Cognition, 25*, 838–848.
Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 261–285.
Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy trace theory: an interim synthesis. *Learning and Individual Differences, 7*, 1–75.
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, 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. Supernant (Eds.), *The Nature of Remembering: Essays in Honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association.
Simpson, A. J., & Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin, 80*, 481–488.
Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin and Review, 5*, 710–715.
Snodgrass, J. G., & Corwin, J. (1998). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 177*, 34–50.
Softkey Multimedia Inc. (1995). American Heritage Talking Dictionary for Macintosh (Version 4.0) [Computer software]. Cambridge, MA: Author.
Stadler, M., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory and Cognition, 27*, 494–500.
Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122–129.
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, and Cognition, 29*, 747–759.
Wickens, T. D., & Hirshman, E. (2000). False memories and statistical decision theory: comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review, 107*, 377–383.
Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107*, 368–376.

