Journal of Experimental Psychology:
Learning, Memory, and Cognition
2001, Vol. 27, No. 2, 328–338

0278-7393/01/$5.00 DOI: 10.1037//0278-7393.27.2.328

Are Emotionally Charged Lures Immune to False Memory?

Bryan J. Pesta
Cleveland State University

Martin D. Murphy and Raymond E. Sanders
University of Akron

Using the Deese–Roediger–McDermott task and E. Tulving's (1985) remember–know judgments for recognition memory, the authors explored whether emotional words can show the false memory effect. Participants studied lists containing nonemotional, orthographic associates (e.g., *cape, tape, ripe; part, perk, dark*) of either emotional (e.g., *rape*) or nonemotional (e.g., *park*) critical lures. This setup produced significant false "remembering" of emotional lures, even though initially no emotional words appeared at study. When 3 emotional nonlure words appeared at study, emotional-lure false recognition more than doubled. However, when these 3 study words also appeared on the recognition test, false memory for the emotional lures was reduced. Across experiments, participants misremembered nonemotional lures more often than they did emotional lures, but they were more likely to rate emotional lures as "remembered," once they had been recognized as "old." The authors discuss findings in light of J. J. Freyd and D. H. Gleave's (1996) criticisms of this task.

In the Deese–Roediger–McDermott (DRM) task, participants hear a study list containing items that are strong associates of a nonpresented theme word, called a *critical lure* (Deese, 1959; Roediger & McDermott, 1995). For example, with *chair* as the critical lure, the study list might contain the following items: *table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, bench.* After hearing the study list, participants take an immediate recall or recognition test. This simple study–test format produces strong false memory.[1] With the best available DRM lists (e.g., the *chair* or *window* lists; see Stadler, Roediger, & McDermott, 1999), participants typically misrecall the lure almost as often as they correctly recall midlist target items (Roediger & McDermott, 1995). In recognition with these lists, false alarms to the lure usually approximate hit rates to the targets, and participants often report high confidence in their (false) memory of the lure (see Roediger, McDermott, & Robinson, 1998, for a review).

Many DRM-task studies also use Tulving's (1985) remember–know technique, where participants rate each word they recognize as being either "remembered" or "known." A remember response indicates a vivid, specific memory of the item's occurrence in the study list, whereas a know response indicates only that the item is familiar in the experimental context, but that no specific details about its presentation can be recollected. Remember–know ratings measure more than just people's confidence in their memory of a word (see, e.g., Gardiner & Java, 1993; Rajaram, 1993), and unique to the DRM task, most people rate critical lure false alarms as remembered rather than known (Roediger et al., 1998).

Clearly, in the DRM task, people often report specific recollections of words that never occurred in the study phase of the experiment. Although the memory illusion this task creates is striking, doubts exist about its external validity. For example, Freyd and Gleaves (1996) questioned the relevance of the DRM task to real world examples of potentially false memory (i.e., in eyewitness testimony or in recovered memories of abuse). Freyd and Gleaves raised many concerns, one being that the real world scenarios where false memory may be a problem typically involve highly emotional events (e.g., abuse). To date, however, researchers using the DRM task have studied false memory for just nonemotional critical lures, produced by having people study just their nonemotional associates.

Freyd and Gleaves (1996) proposed a challenge to memory researchers, in the form of a thought experiment. The experiment's goal was to see whether people would misremember something emotional after experiencing something unrelated and nonemotional. The example Freyd and Gleaves suggested was to "see how often the participant misidentifies the word *penis* as being on the [*foot*] list" (p. 812). Freyd and Gleaves believed such a demonstration to be critical, because contested memories of abuse often occur within the context of a "normal" childhood (see Freyd & Gleaves, 1996, p. 812).

With one exception, no one has explored how emotional words moderate DRM-task false memory. In a reply to Freyd and Gleaves (1996), Roediger and McDermott (1996) reanalyzed their original data set (Roediger & McDermott, 1995) and found that no

Bryan J. Pesta, Department of Psychology, Cleveland State University; Martin D. Murphy and Raymond E. Sanders, Department of Psychology, University of Akron.

We thank Andy Stone for help in data collection, and David Payne, Steve Lindsay, and Reed Hunt for helpful comments and suggestions on the article.

Correspondence concerning this article should be addressed to Bryan J. Pesta, 14724 Grapeland, Cleveland, Ohio 44111. Electronic mail may be sent to bpesta22@cs.com.

[1] We use the term *false memory* throughout the article. We define *false memories* as occurring when people systematically report specifically "remembering" (i.e., Tulving, 1985) words that were not present at study. Whether and how these data generalize to real world situations where people also use the term (i.e., in eyewitness testimony or in recovered memories of abuse) is an issue we do not address in this article.

participant falsely recalled *penis*, or any other emotional word, after studying the *foot* list. One explanation for the absence of false recall here is that the emotional distinctiveness of a word like *penis* leads few people to misremember its presence in a study list. A second possible explanation is that items in the *foot* list are simply not highly related to this lure, yet no experiments exist that directly test these issues.

Literature on distinctiveness shows that when the distinctive word is a studied target, memory is enhanced (the "von Restorff effect"; von Restorff, 1933). In the DRM task, one could predict that if *penis* were included in a study list, most people would correctly recognize it. The key issue here, however, is the distinctiveness of distractor items (i.e., lures). Distinctive distractor items would likely produce a "reverse von Restorff effect"—floor effects on false alarms to critical lures that are (emotionally) distinct from the study list items.

Distinctiveness may even help address why some DRM lists (e.g., *chair* or *window*) produce stronger false memory than other DRM lists (e.g., *king* or *fruit*). Critical lures in the "weak" lists may be orthographically distinct from their study items. That is, none of these study items sound like or share many letters with their lure. Orthographic distinctiveness may make these lures stand out at test, leading few people to misremember them as studied items. Recent work by Watson, Balota, and Roediger (2000) is consistent with these ideas. Watson et al. found that study lists containing both orthographic and semantic associates of their lures produced the strongest false memory effects. We find it interesting that these "hybrid" lists produced *superadditivity*—stronger effects than could be predicted by adding the separate false memory effects in the "semantic only" and "orthographic only" lists (Watson et al., 2000).

In Watson et al. (2000) orthographically related target items may have decreased the distinctiveness of the critical lures, thus increasing false memory. In the present study, distinctiveness may work in the reverse direction. Lures that are emotionally distinct from the study list items may not be falsely remembered, or perhaps may show significant but weak false memory. Again, however, which of these patterns might hold is an issue whose direct resolution requires an experiment.

Therefore, our purpose was to explore whether emotionally charged lures can show the false memory effect. We departed somewhat from conducting Freyd and Gleaves' (1996) exact thought experiment, which requires that the study-list items be both nonemotional and unrelated (or at most, weakly related) to the missing emotional lures. Instead, we created study lists that contained only nonemotional items (i.e., satisfying the former condition) that were nonetheless strong associates of emotional lures (i.e., violating the latter condition).

To compile study lists like these, we could not use semantic associates of emotional lures, as they would likely themselves be emotional words. Our method was to create study lists that contained orthographic neighbors of the critical lures. *Neighbors* are words that are spelled the same as the lure after changing only one letter and keeping the other shared letters in the same position (e.g., some nonemotional neighbors of *rape* are *cape*, *tape*, *ripe*, and *rope*). Recent research shows that both orthographic (Schacter, Verfaellie, & Anes, 1997; Watson et al., 2000) and phonemic (Sommers & Lewis, 1999) list–lure relationships can produce robust false memory. Hence, orthographic list–lure relationships are ideal because they allow us to create study lists that are built around emotional lures but that contain only nonemotional words. In four experiments, participants studied nonemotional list items that were orthographic associates of either emotional (e.g., *rape*) or nonemotional (e.g., *park*) lures.

General Method

This section highlights the general method used in each of our four experiments. Table 1 summarizes specific across-experiment differences in method or procedure. All experiments were modeled after Roediger and McDermott's (1995) Experiment 2. We started by dividing a (12-list) master stimulus set into two (6-list) subsets. Within subsets, half our lists consisted of orthographic neighbors of emotional lures, whereas the other half consisted of orthographic neighbors of nonemotional lures. All study items were typically nonemotional words (but see Table 1 for exceptions). Within each experiment, half our participants studied the first subset (making the study-list items and lures in the second subset the "unrelated items" for these participants), whereas the other half studied the second subset.

This method gave us two benchmarks for comparing false memory of emotional to nonemotional lures: (a) false alarms to emotional and nonemotional lures in the "related" conditions only and (b) the difference in false alarms across the related and unrelated conditions, for emotional and nonemotional lures. This last comparison constitutes a 2 (emotional lure vs. nonemotional lure) × 2 (related vs. unrelated) analysis of variance (ANOVA), using just the lure data.

Participants and Design

Each experiment contained 40 undergraduates who received extra credit for their participation. Table 1 shows the mean age of participants and the percentage of men in each experiment (no gender differences in false recognition of emotional lures appeared in any experiment). Within experiments, we collected data in small groups until we reached the desired sample size of 20 people in each of the two, counterbalanced study sets. Because of participant turnout in Experiment 4, 21 people studied the first set, and 19 people studied the second set. The design for each experiment was a 2 × 2 × 2 factorial with associate type (associates of nonemotional vs. emotional lures), item type (list vs. lures), and relatedness (items related to the study phase for each participant vs. items unrelated to the study phase for each participant) as within-subjects variables.[2]

Materials and Procedure

We selected these six emotional lures: *rape, bitch, slut, hell, whore,* and *penis,* mainly because they have many orthographic neighbors (we selected *penis,* however, because it is the example lure behind Freyd and Gleave's, 1996, thought experiment). For each lure, we compiled a 10-item study list by first including any nonemotional word that was a strict orthographic neighbor to the lure. After exhausting the supply of orthographic neighbors, we then filled in the remaining list items with associates that shared many phonemes with the critical lure. The Appendix lists these items for all six emotional lures.

For each emotional list and lure, we compiled a nonemotional list and lure counterpart, also presented in the Appendix. The list immediately next to each emotional lure is its nonemotional lure, control list. Within each

[2] The general design of our experiments is not actually a fully crossed factorial, as unrelated study-list items were never presented to participants. In all experiments, only items in the related cells correspond to DRM lists that each participant saw at study, and only the related, list items cells correspond to actual target items.

Table 1
Summary of Experimental Conditions

Exp.	*M* Age	% Male	Study phase	Test phase
1	27.2	35	Two sets of 6 DRM lists each. Participants studied one or the other set. Each DRM list had 10 nonemotional orthographic associates of an emotional or a nonemotional lure (within study sets, there were 3 lists with emotional lures and 3 lists with nonemotional lures). No emotional words appeared at study.	48 items (3 items each from the 6 lists in the studied set; 3 items each from the 6 lists in the nonstudied set; and 12 critical lures).
2	25.7	25	Same as Exp. 1, but included 3 orthographically unrelated emotional words in the study list, placed in serial positions 13, 34, and 50.	Same as Exp. 1.
3	23.3	28	Same as Exp. 2, but the 3 emotional study-list words appeared in red ink—all other study words appeared in black ink.	Same as Exp. 1.
4	23.0	40	Same as Exp. 3.	The same 48 items as in Exps. 1–3, plus the 3 emotional study-list items, appearing in positions 2, 17, and 32 on the test. Hence, people rated 1 emotional study-list item before rating any emotional lures. All items appeared in black ink.

Note. $N = 40$ for all experiments. After Experiment 1, we strengthened standard instructions against guessing by asking participants to rate any word they were unsure about as "new." Exp. = Experiment; DRM = Deese–Roediger–McDermott.

pair, the lists are matched on (a) lure word frequency, (b) lure letter length, (c) the number of strict orthographic neighbors in the study list, and (d) neighborhood density. The number of words that rhyme with the lure or that start with the same initial phoneme as the lure are also approximately matched within the list pairs. It is likely impossible to match two DRM lists on all dimensions that might influence false memory. However, for the simple demonstration that emotional words are subject to false remembering, the key comparison is the false-alarm rate to the emotional lure when its list is studied (i.e., the related condition), versus when its list is not studied (i.e., the unrelated condition).

We grouped the 12 study lists into two subsets of 6 study lists each, with each subset containing three emotional lists and their three matched, nonemotional lists. No lure appeared in its study list. Half the participants studied the first set, whereas the other half studied the second set. We ordered the items into each study set by placing two filler words first and last to control for serial position effects. The 60 items in between rotated through the six lists in blocks of five study items per list. For example, Set A started with the first five items listed in the Appendix for the *rape* list, followed by the first five items for the *hook* list, and ended with the last five items for the *peach* list.

We presented the list items visually, rather than verbally, using a standard overhead projector. Most DRM-task studies present the list items verbally (but see Gallo, McDermott, Percer, & Roediger, 2001; Smith & Hunt, 1998), but with the orthographic lists used here, all of the study items sound like the critical lure. Visual presentation eliminates the potential problem of people confusing pronunciation of a study list item for pronunciation of a lure. Incidentally, presenting the study lists visually is one way to increase the distinctiveness of the lures (see, e.g., Schacter, Israel, & Racine, 1999; Smith & Hunt, 1998), and most studies that use visual list presentation show weaker false memory effects, compared with verbal list presentation (but see Gallo et al., 2001).

For the test phase, we used a single, one-page recognition test for all participants. It contained three randomly selected list items from each study list and the 12 critical lures, for a total of 48 items (Experiment 4 contained three other items; see Table 1). We ordered the 48 test items randomly in four columns of 12 items each. Next to each item was an answer space that participants used for the old–new and remember–know ratings. Given the large number of items in the study list (64 items total, including the 4 fillers), we did not use a distractor task in any experiment.

For ethical reasons, our informed consent form included the following statement:

> WARNING: A few of the words you will see in this experiment are sexually-charged, vulgar, or offensive. If you would rather not be exposed to these types of words, please do not participate.

After everyone signed the consent form, we conducted the 25-min experiment by first telling participants that they would see a long list of words on the overhead projector and that we would test their memory for the list soon after its presentation. We then explained the difference between "old" versus "new," and between "remember" versus "know" judgments on the recognition test (adopting our instructions after those described in Roediger & McDermott, 1995). Next, we presented the stimuli from either Study Set A or Study Set B (determined pseudorandomly), one item at a time, at a presentation rate of approximately 1.5-s per item. After showing the last item in a study list, we circulated the untimed recognition test. Finally, at the experiment's end, we debriefed participants and thanked them for their participation.

Experiment 1

Experiment 1 comes closest to Freyd and Gleave's (1996) thought experiment. No emotional words were present at study, but unlike the conditions of the thought experiment, all of the list items were orthographic associates of an emotional or a nonemotional lure.

Results

Recognition. Table 2 lists the proportion "old" results for Experiment 1. We used $p < .05$ as the significance level for all analyses throughout the article. A 2 (associates of emotional vs. nonemotional lures) × 2 (list items vs. lure items) × 2 (related lists vs. unrelated lists) within-subjects ANOVA revealed main effects of associate type, $F(1, 39) = 53.4$, $MSE = 0.06$; item type, $F(1, 39) = 118.0$, $MSE = 0.06$; and relatedness condition, $F(1, 39) = 191.0$, $MSE = 0.05$. All interactions except Relatedness Condition × Associate Type ($F < 1.0$) were significant: Relatedness Condition × Item, $F(1, 39) = 20.7$, $MSE = 0.07$; Associate Type × Item, $F(1, 39) = 52.3$, $MSE = 0.05$; Relatedness × Item × Associate Type, $F(1, 39) = 12.9$, $MSE = 0.04$.

Post hoc tests (Tukey least significant difference tests in text; see also the Loftus & Masson, 1994, posttests presented in Table 2) in all of our experiments focused on two issues: (a) Did

Table 2
Mean Proportion Old and Signal Detection Analyses of Sensitivity (d′) and Bias (C) for All Experiments by Associate Type, Item Type, and Relatedness Condition

	Orthographic relationship between the study-list and recognition test items				
Associate type	Related	Unrelated	Difference	d'	C[a]
		Experiment 1			
List items					
Of nonemotional lures	.81	.39	.42*	2.13	−0.57
Of emotional lures	.86	.31	.55*	3.60	−0.52
Difference	−.05	.08	—	−1.47*	0.05
Critical lures					
Nonemotional lures	.64	.33	.31*	3.41	0.14
Emotional lures	.18	.05	.13*	1.36	3.73
Difference	.46*	.28*	—	2.05*	3.59*
		Experiment 2			
List items					
Of nonemotional lures	.81	.22	.59*	3.35	0.05
Of emotional lures	.81	.23	.58*	3.29	0.03
Difference	.00	−.01	—	0.06	0.02
Critical lures					
Nonemotional lures	.76	.31	.45*	3.95	−0.46
Emotional lures	.42	.13	.29*	2.86	2.11
Difference	.34*	.18*	—	1.09*	2.57*
		Experiment 3			
List items					
Of nonemotional lures	.78	.34	.44*	2.44	−0.04
Of emotional lures	.80	.33	.47*	2.70	−0.32
Difference	−.02	.01	—	−0.26	0.28
Critical lures					
Nonemotional lures	.78	.42	.36*	3.25	−1.04
Emotional lures	.43	.19	.24*	2.29	1.78
Difference	.35*	.23*	—	0.96	2.82*
		Experiment 4			
List items					
Of nonemotional lures	.75	.30	.45*	2.27	−0.02
Of emotional lures	.84	.29	.55*	3.06	−0.28
Difference	−.09	.01	—	−0.79	0.26
Critical lures					
Nonemotional lures	.73	.35	.38*	3.38	−0.28
Emotional lures	.30	.07	.23*	2.53	2.90
Difference	.43*	.28*	—	0.85	3.18*

Note. Only values in the "related, list items" cells correspond to words participants actually saw in the study phase.
[a] See text for how d' and C were computed for items that were not studied.
* $p < .05$, using Loftus and Masson's (1994) within-subjects confidence intervals.

emotional lures show significant false recognition? and (b) How did false memory of emotional and nonemotional lures compare? For the first issue, false alarms to emotional lures were significantly higher in the related condition (18%) than in the unrelated condition (5%), $t(39) = 3.40$. This comparison is the key to demonstrating DRM-task false memory, as emotion is controlled (i.e., the words in both cells are emotional), yet only the "related" cell corresponds to associates seen in the study phase.

Also relevant to the first issue above is that emotional lure false alarms in the related condition (i.e., 18%) were actually lower than false alarms to list items in the unrelated conditions (i.e., 31% and 39%, emotional and nonemotional lists; $t[39] = 2.21$ and 3.26, respectively). As previously mentioned, unrelated list items here are actually unrelated "distractors," in that participants did not see these specific items at study (due to our counterbalancing). People were more likely to misremember distractors unrelated to the study phase of Experiment 1 than they were to misremember emotional lures whose nonemotional neighbors had been studied. We conclude, therefore, that emotional lures showed modest but significant false recognition effects.

The second important issue concerns how the false memory effect differs for emotional and nonemotional lures. As previously mentioned, two separate analyses help answer this question. First, in the related conditions, false alarms were greater for nonemotional than emotional lures, $t(39) = 8.84$. Second, we conducted a 2 (emotional lure vs. nonemotional lure) × 2 (related vs. unrelated) ANOVA on just the lure data. A significant interaction would indicate stronger false memory effects for one type of lure over the other type of lure. This analysis revealed main effects of lure type, $F(1, 39) = 79.4$, $MSE = 0.07$; relatedness, $F(1, 39) = 29.7$, $MSE = 0.07$; and the Lure Type × Relatedness interaction, $F(1, 39) = 10.5$, $MSE = 0.03$. Therefore, both the ANOVA and the t-test results show stronger false recognition of nonemotional over emotional lures. Summarizing the proportion "old" data, emotional lures significantly reduced (but did not eliminate) false recognition in the DRM task.

Signal detection. Table 2 also provides signal detection measures of sensitivity (d') and criterion (C) for the Experiment 1 data. The measure of bias we selected produces negative numbers for liberal bias and positive numbers for conservative bias (see, e.g., Snodgrass & Corwin, 1988). Because signal detection measures cannot be computed whenever a hit rate equals 1.0 or whenever a false-alarm rate equals 0.0, we first transformed the proportion old data, as recommended by Snodgrass and Corwin (1988, p. 35). Each measure of sensitivity and bias in Table 2 applies to the row of data it appears in. For these analyses, we categorized the proportion of old values within the related cells to be hit rates, and the proportion old values (in the same row) within the unrelated cells to be the corresponding false-alarm rates (in the related cells, the proportion old values for critical lures are really false alarms). Therefore the d' values for the critical lure rows measure sensitivity to false memory, as the comparison treats lures related to the study phase as old items, and lures unrelated to the study phase as new items.

We conducted separate Item Type (list vs. lure items) × Lure Type (emotional vs. nonemotional) ANOVAs for the d' and C measures. The ANOVA on the d' data revealed neither main effect to be significant ($Fs < 1.0$), although the interaction was significant, $F(1, 39) = 19.4$, $MSE = 6.41$. The interaction indicates that false memory sensitivity to lures was stronger in the nonemotional than emotional cells, but that true memory sensitivity to targets showed the reverse pattern. We find it interesting that people were just as sensitive in judging nonemotional lures to be list items as they were in detecting target items.

The ANOVA on the C data revealed significant main effects of item, $F(1, 39) = 95.6$, $MSE = 2.57$; lure, $F(1, 39) = 64.8$, $MSE = 2.04$; and the Item × Lure interaction, $F(1, 39) = 57.1$, $MSE = 2.20$. The interaction shows neutral to somewhat liberal measures of bias for all conditions except emotional lures, which showed a strong conservative bias (i.e., participants were likely to call emotional lures "new").[3] The signal detection analyses, therefore, are consistent with the proportion old data in showing that false memory was significant for the emotional lures, though weaker than for the nonemotional lures.

Remember–know. Table 3 presents the proportion of remember and of know responses for all conditions in Experiment 1. False alarms in the unrelated cells were predominantly know responses, whereas false alarms in the related cells were mostly remember responses. Looking at just emotional lures, the proportion of remembered values was significantly higher in the related than unrelated conditions, $t(39) = 3.56$. It appears that when people do misremember an emotional lure, they are likely to rate it remembered, rather than known.

Discussion

As previously discussed, one explanation for how emotional words moderate false memory appeals to distinctiveness (see, e.g., Hunt & Eliott, 1980; Hunt & Mitchell, 1982; von Restorff, 1933). An item is distinct from other items in a list if it varies on some obvious dimension from the other items (e.g., color or case; see, Rundus, 1971). The effects of distinctiveness operate both at study and at test (Schmidt, 1985). In general, distinctiveness serves to increase memory accuracy for target items but to decrease memory error for distractor items. In Experiment 1, the six emotional lures were distinct from all other items (both at study and at test) by virtue of their emotionality. Because these items were distractors, their emotional distinctiveness made them relatively easy to reject as studied items. Hence, false recognition of the emotional lures was modest compared with false recognition of the less distinct, nonemotional lures.

We also believe the warning provided in the consent form further highlighted the distinctiveness of the emotional lures before the experiment even began. Specifically, the warning seemed to pique our participants' interest. Many were curious as to just what types of words they might see on the overhead. It seems possible that failing to meet their expectations—by not providing any emotional items at study—helped reduce false alarms to the emotional lures at test. In other words, participants were expecting

[3] The C parameter indicates the placement of the response criterion relative to the intersection of the two familiarity distributions on which d' is based. As such, when one of the two familiarity distributions used to compute d' for each of the two different conditions is not the same distribution (as seems likely in the present comparisons), the different C values do not necessarily indicate that people used more than one response criterion for the different conditions (see Wickens & Hirshman, 2000; Wixted & Stretch, 2000).

Table 3
Mean Proportion Remember and Know Responses for All Experiments by Associate Type, Item Type, and Relatedness Condition

	Orthographic relationship between the study-list and recognition test items					
	Remember responses			Know responses		
Associate type	Related	Unrelated	Difference	Related	Unrelated	Difference
	Experiment 1					
List items						
Of nonemotional lures	.60	.06	.54*	.21	.33	−.12*
Of emotional lures	.65	.08	.57*	.21	.23	−.02
Difference	−.05	−.02	—	.00	.10*	—
Critical lures						
Nonemotional lures	.35	.08	.27*	.29	.24	.05
Emotional lures	.15	.02	.13*	.03	.03	.00
Difference	.20*	.06	—	.26*	.21*	—
	Experiment 2					
List items						
Of nonemotional lures	.65	.10	.55*	.16	.12	.04
Of emotional lures	.72	.10	.62*	.09	.13	−.04
Difference	−.07	.00	—	.07	−.01	—
Critical lures						
Nonemotional lures	.55	.13	.42*	.21	.18	.03
Emotional lures	.33	.07	.26*	.09	.06	.03
Difference	.22*	.06	—	.12*	.12*	—
	Experiment 3					
List items						
Of nonemotional lures	.55	.09	.46*	.23	.25	−.02
Of emotional lures	.58	.15	.43*	.22	.18	.04
Difference	−.03	−.06	—	.01	.07	—
Critical lures						
Nonemotional lures	.41	.10	.31*	.37	.32	.05
Emotional lures	.28	.05	.23*	.15	.14	.01
Difference	.13*	.05	—	.22*	.18*	—
	Experiment 4					
List items						
Of nonemotional lures	.56	.06	.50*	.19	.24	−.05
Of emotional lures	.69	.09	.60*	.15	.20	−.05
Difference	−.13*	−.03	—	.04	.04	—
Critical lures						
Nonemotional lures	.52	.04	.48*	.21	.31	−.10*
Emotional lures	.22	.03	.19*	.08	.05	.03
Difference	.30*	.01	—	.13*	.26*	—

Note. Only values in the "related, list items" cells correspond to words participants actually saw in the study phase.
* $p < .05$, using Loftus and Masson's (1994) within-subjects confidence intervals.

some offensive words in the study phase but saw none. At test, when emotional words appeared for the first time, few people false alarmed to them.

Experiment 2

A simple way to test the above ideas is to include a few emotional words in the study phase of this experiment. These items should serve as "distinctiveness attenuators" by reducing the salience of the emotional lures at test, thereby making it harder for participants to decide whether these lures actually appeared at study. Also, adding these three emotional words to the study lists should allow us to satisfy participants' expectations about seeing offensive words in the study phase, while still exploring the role that emotion plays in false memory at the test phase. Hence, in Experiment 2, we added three emotional words to our study phase: *fuck, piss,* and *asshole.* We selected these emotional words to minimize the amount of orthographic and semantic overlap with the six emotional test lures.

As illustrated in Table 1, this experiment differed from the first in only two ways. First, we inserted the above three emotional words into the study lists. They appeared in serial positions 13, 34, and 50 respectively, for a total of 67, rather than 64, study items. These items did not, however, appear in the recognition test. Second, just before participants completed the recognition test, we strengthened our standard instructions against guessing by asking everyone "to rate any word [they were] unsure about as new." Given that we sought to increase false memory of emotional lures here, we wanted to raise the false memory standard, to further rule out guessing as a partial explanation for any high intrusion rates we may observe. We used this warning in this experiment and in the experiments that follow.

Results

Recognition. Table 2 also lists the recognition results for Experiment 2 (see also Footnote 2). A 2 (associates of emotional vs. nonemotional lure) × 2 (list items vs. lure items) × 2 (related lists vs. unrelated lists) within-subjects ANOVA revealed main effects of associate type, $F(1, 39) = 24.0$, $MSE = 0.05$; item type, $F(1, 39) = 28.8$, $MSE = 0.04$; and relatedness condition, $F(1, 39) = 237.0$, $MSE = 0.08$. As in Experiment 1, the Relatedness Condition × Item and the Associate Type × Item interactions were significant, $F(1, 39) = 23.5$, $MSE = 0.04$, and $F(1, 39) = 32.1$, $MSE = 0.04$, respectively. However, unlike in Experiment 1, the Relatedness Condition × Associate Type interaction did not quite reach significance, $F(1, 39) = 3.99$, $MSE = 0.04$, $p = .053$, nor did the Relatedness × Item × Associate Type interaction, $F(1, 39) = 1.29$, $MSE = 0.04$, $p = .10$.

The pattern of mean differences for Experiment 2 is close to that obtained in Experiment 1 for study-list items and nonemotional lures. Related emotional lures, however, showed a much higher false-alarm rate here (42%) than in Experiment 1 (18%), $t(78) = 3.36$, between subjects. Indeed, including emotional words in the study list more than doubled false-alarm rates to the related, emotional lures, compared with Experiment 1. The likely explanation is that the emotional study items lessened the distinctiveness of the emotional test lures, resulting in stronger false memory.

Even though Experiment 2 produced higher false-alarm rates for the related, emotional lures, this value was still smaller than for related, nonemotional lures, $t(39) = 4.50$. However, a 2 (emotional lure vs. nonemotional lure) × 2 (related vs. unrelated) ANOVA showed main effects of lure type, $F(1, 39) = 31.8$, $MSE = 0.08$, and relatedness, $F(1, 39) = 104$, $MSE = 0.05$, but the interaction was now only marginally significant, $F(1, 39) = 3.81$, $MSE = 0.07$, $p = .058$. Thus, the interaction, which tests the difference in false alarms in the related and unrelated cells for emotional and nonemotional lures, was slightly weaker than in Experiment 1. In sum, we found much higher false-alarm rates for emotional lures here as compared with Experiment 1.

Signal detection. The Item Type (list items vs. lure items) × Lure Type (emotional vs. nonemotional) ANOVA for the Experiment 2 d' data produced no significant effects: item, $F(1, 39) < 1.0$; lure, $F(1, 39) = 1.46$, $MSE = 9.13$; and Item × Lure, $F(1, 39) = 1.25$, $MSE = 8.23$. Hence, true memory sensitivity for targets did not differ across the emotion factor, nor did false memory sensitivity for lures (although consistent with Experiment 1, the smallest d' value in Table 2 is for the emotional critical lure row). The ANOVA on the C data, however, replicated that which was reported in Experiment 1, as all effects were significant: item, $F(1, 39) = 11.6$, $MSE = 2.13$; lure, $F(1, 39) = 65.0$, $MSE = 3.07$; and Item × Lure, $F(1, 39) = 39.3$, $MSE = 2.29$. Once again, participants displayed neutral to somewhat liberal biases for all items except emotional lures, for which they showed a conservative bias (see, however, Footnote 3).

Remember–know. In Table 3 for Experiment 2, most of the false alarms in the unrelated cells are again know, rather than remember, responses. Second, remember responses for related, emotional lures were more frequent than for unrelated, emotional lures, $t(39) = 6.37$, but less frequent than for related, nonemotional lures, $t(39) = 2.92$. Third, across-experiment comparisons show that remember responses to emotional and nonemotional lures in this experiment (33% and 55%, respectively) are higher than they are in Experiment 1 (15% and 35%), $ts(78) = 2.76$ and 2.70, respectively.

Discussion

In Experiment 1, the emotional lures were distinct from all other items both at study and at test. The result was modest but significant false recognition of these items. In Experiment 2, the distinctiveness attenuators present at study seemed to have lessened the salience of the emotional lures at test. Accordingly, this manipulation increased emotional lure false alarms in the related condition from 18% in Experiment 1 to 42% in Experiment 2.

Experiment 3

Decreasing lure distinctiveness increased emotional lure false recognition. Using the same logic, it might be possible to reduce again the false memory rate to emotional lures by using distinctiveness in reverse. If we could make the three emotional study words themselves distinct from the six other emotional test lures, participants might realize that the study-phase emotional items differed from the test-phase emotional items. Thus, in Experiment 3, we made a simple change to the study list presented in Experiment 2. Specifically, as outlined in Table 1, we printed the three emotional study words in red ink, with the other 64 study-list items and all items in the recognition test still appearing in black ink. In effect, we attempted to make the distinctiveness attenuators themselves distinct from the emotional test lures. The result should therefore be lower false recognition of emotional lures here, relative to Experiment 2.

Results

Recognition. Table 2 presents the recognition results by condition in Experiment 3 (see also Footnote 2). A 2 (associates of emotional vs. nonemotional lures) × 2 (list items vs. lure items) × 2 (related lists vs. unrelated lists) within-subjects ANOVA revealed main effects of associate type, $F(1, 39) = 16.8$, $MSE = 0.09$; item type, $F(1, 39) = 15.5$, $MSE = 0.06$; and relatedness condition, $F(1, 39) = 208$, $MSE = 0.05$. The results for the interactions were identical to those found in Experiment 1, as all interactions except Relatedness Condition × Associate Type, $F(1, 39) = 1.48$, $MSE = 0.03$, were significant: Associate Type × Item, $F(1, 39) = 18.1$, $MSE = 0.09$; Relatedness Condition × Item, $F(1, 39) = 13.1$, $MSE = 0.03$; and Relatedness Condition × Item × Associate Type, $F(1, 39) = 4.39$, $MSE = 0.02$. (In

Experiment 2, however, the three-way interaction was only marginally significant.)

Contrary to our prediction, printing the emotional study words in red ink had no effect on false alarms to related, emotional lures. The false-alarm rate for these items was 43% in this experiment and 42% in Experiment 2. As in Experiment 2, false alarms to related, emotional lures in this experiment were higher than to unrelated, emotional lures, $t(39) = 5.61$, but lower than to related, nonemotional lures, $t(39) = 4.44$. Also, the 2 × 2 ANOVA only on the lure data resulted in conclusions identical to those reached in Experiment 2—main effects of lure type, $F(1, 39) = 19.2$, $MSE = 0.17$; and relatedness, $F(1, 39) = 96.2$, $MSE = 0.04$; but a nonsignificant interaction, $F(1, 39) = 3.62$, $MSE = 0.04$, $p = .065$. Finally, false alarms to all unrelated items seem slightly elevated in this experiment, relative to Experiment 2. However, in the four pairwise comparisons of related versus unrelated items (i.e., across associate type and item type) within Experiment 3, the proportion "old" rate is significantly higher in the related condition.

Signal detection. In Experiment 3, the Item Type (list items vs. lure items) × Lure Type (emotional vs. nonemotional) ANOVA for the d' data also produced no significant effects: item, $F(1, 39) < 1.0$; lure, $F(1, 39) = 1.29$, $MSE = 3.73$; and Item × Lure, $F(1, 39) = 2.27$, $MSE = 6.47$, $p > .10$. As in Experiment 2, true memory sensitivity for targets did not differ across the emotion factor, nor did false memory sensitivity for lures. Likewise, the ANOVA on the C data replicated those reported in Experiments 1 and 2: item, $F(1, 39) = 5.32$, $MSE = 2.33$; lure, $F(1, 39) = 15.1$, $MSE = 4.26$; and Item × Lure, $F(1, 39) = 22.5$, $MSE = 4.27$. The emotional lure row in this experiment were the only data to show a strong conservative bias (see also Footnote 3).

Remember–know. As shown in Table 3, the remember–know data in this experiment are similar to those in Experiment 2. For example, there is only a small across-experiment difference in remember responses to related, emotional lures (i.e., 33% remember responses in Experiment 2, and 28% in Experiment 3), which was not significant ($t < 1.00$). The same pattern of parity in remember responses across Experiments 2 and 3 holds for items in the unrelated cells as well. In fact, summarizing the remember–know data leads to the same conclusion as for the proportion old data. Printing the three emotional study words in red ink did not reduce false recognition of the emotional lures.

Discussion

Participant expectations again offer a possible explanation for the failure to reduce emotional lure false alarms in Experiment 3. This time, expectations operated at test rather than at study, and they may have hampered our attempts to make the emotional study items distinct from the emotional lures. Specifically, in Experiment 3, the emotional words present at study were distinct along two dimensions—emotion and color. It is highly likely that participants noticed these items as they appeared on the overhead. The only emotional words in the recognition test, however, were lures. Moreover, half these lures (i.e., those in the related condition) had a high degree of familiarity at test, due to our presenting their nonemotional neighbors at study. Perhaps participants false alarmed to the familiar lures because they were the only emotional words available at test, and participants knew that emotional words were indeed present at study.

Experiment 4

To test this idea, we placed the three emotional study words in the recognition test. Our goal was to satisfy participants' expectations about emotional words appearing on the memory test and to allow participants to separate emotional words that actually appeared at study from emotional lures that did not (using the distinctiveness logic discussed in the preview to Experiment 3). As illustrated in Table 1, the three emotional red-ink study words appeared in black ink in the recognition test, for a total of 51, rather than 48, items. They appeared in positions 2, 17, and 32 in the recognition test (counting down the columns rather than across the rows). Thus, participants had to rate one of the emotional study words as old or new before rating any of the emotional lures. This applied whether a person completed his or her test by going down the rows or across the columns.

Recognition

Table 2 also presents the recognition results by condition in Experiment 4 (see also Footnote 2). A 2 (associates of emotional vs. nonemotional lures) × 2 (list items vs. lure items) × 2 (related lists vs. unrelated lists) within-subjects ANOVA revealed main effects of associate type, $F(1, 39) = 38.7$, $MSE = 0.05$; item type, $F(1, 39) = 48.7$, $MSE = 0.05$; and relatedness condition, $F(1, 39) = 268.0$ $MSE = 0.05$. The results for the interactions were the same as found in Experiments 1 and 3, in that all interactions except Relatedness Condition × Associate Type ($F < 1.0$) were significant: Associate Type × Item, $F(1, 39) = 87.8$, $MSE = 0.03$; Relatedness Condition × Item, $F(1, 39) = 17.2$, $MSE = 0.05$; and Relatedness Condition × Item × Associate Type, $F(1, 39) = 8.01$, $MSE = 0.04$. (The three-way interaction did not quite reach significance in Experiment 2.)

Printing the emotional study items in the recognition test did reduce false alarms to the related, emotional lures from 43% and 42% in Experiments 2 and 3, respectively, to 30% in this experiment, $ts(78) = 1.67$ and 1.80, $p < .05$, one-tailed. The reduction, however, was neither strong nor complete, as the false-alarm rate for emotional lures in this experiment was still above the baseline level found in Experiment 1 (18%), in which no emotional words were present at study.

As in all previous experiments, false alarms were higher for related nonemotional lures than for related emotional lures, $t(39) = 6.41$. Further, as in Experiments 2 and 3, the interaction in the 2 (emotional lure vs. nonemotional lure) × 2 (related vs. unrelated) ANOVA was not quite significant, although the main effects were: lure type, $F(1, 39) = 75.3$, $MSE = 0.07$; relatedness, $F(1, 39) = 67.9$, $MSE = 0.05$; and Lure Type × Relatedness, $F(1, 39) = 3.83$, $MSE = 0.06$, $p = .057$.

Signal Detection

Again, the Item Type (list items vs. lure items) × Lure Type (emotional vs. nonemotional) ANOVA for the d' data produced results consistent with those found in Experiments 2 and 3: Neither main effect was significant ($Fs < 1.0$), nor was the interaction,

$F(1, 39) = 3.03$, $MSE = 8.88$, $p < .10$. This same pattern of across-experiment consistency held for the ANOVA on the *C* data: main effects of item, $F(1, 39) = 37.5$, $MSE = 2.27$; lure, $F(1, 39) = 36.8$, $MSE = 2.31$; and Item × Lure, $F(1, 39) = 62.1$, $MSE = 1.91$. The interaction was stronger in this experiment than that which was found in Experiments 2 and 3, as the difference in bias between emotional lures and all other items is quite large in this experiment (as it was in Experiment 1; see also Footnote 3).

Remember–Know

The remember–know data in Table 3 also show a small reduction in remember responses to related, emotional lures in Experiment 4 relative to Experiment 2, but this difference was not significant, $t(78) = 1.58$, $p > .10$. In fact, remember responses to related items in this experiment were very similar to these values in Experiment 2 (e.g., ranking the means from high to low for the four related conditions results in the same values across Experiments 2 and 4). Therefore, including emotional items in the study list made the emotional lure false memory effect resistant to reduction because (a) printing the emotional study items in red ink had no effect on emotional lure false alarms in Experiment 3 and (b) putting the emotional study items in the recognition test only partially reduced emotional lure false alarms in Experiment 4.

Conditionalized Remember–Know Data for All Experiments

Table 4 shows the conditional probability of rating an item remembered or known, given that it was rated old. We collapsed these data across the 160 participants in all four experiments, and we calculated the conditional probabilities simply by dividing each participant's proportion remembered (or proportion known) value by his or her proportion old value. Conditional probabilities are useful because they adjust for baseline differences in item memorability, allowing a more direct comparison of how remember and know responses differ across conditions, only for items that participants have rated old. As reflected in Table 4, however, one problem with conditional probabilities is that we had to exclude an individual's data for any cell where the proportion old value equaled zero (to avoid division by zero). We therefore collapsed these data across all four experiments, as in many cases within a single experiment, the sample sizes in some cells were too small for meaningful statistical comparisons.

Somewhat unexpected in Table 4 was that the conditional probability of rating a related item remembered, given that it was rated old, was (a) higher for emotional versus nonemotional lures, $t(92) = 2.63$, and (b) comparable for emotional lures versus associates of emotional lures (i.e., targets), $t(96) = 1.82$, $p < .10$, and versus associates of nonemotional lures (also targets), $t < 1.0$. The three distinctiveness attenuators were almost always remembered when they appeared in the recognition test in Experiment 4 (conditionalized proportion remembered = 0.99).

The conditionalized remember–know data show high rates of remember responses to both emotional and nonemotional lures judged to be old. They also show that once a person false alarms to an emotional lure, he or she is likely to rate it as remembered, rather than known, even more so than if the lure were nonemotional. The high conditionalized remember rates found in this experiment are consistent with other DRM-task studies that employ the remember–know technique (see Roediger et al., 1998). Specifically, unique to the DRM-task, the majority of false alarms that participants made to critical lures corresponded to remember rather than to know responses, and interestingly, having these lures be emotional seemed only to increase this effect.

General Discussion

To review, Experiment 1 produced modest but significant false recognition of emotional lures when the study list contained no emotional words. Experiment 2, however, more than doubled false recognition of these items by including just three emotional words in the study lists. We offered distinctiveness as a potential expla-

Table 4
Conditionalized Remember–Know Responses Collapsed Across All Experiments by Associate Type, Item Type, and Relatedness Condition

	Orthographic relationship between the study-list and recognition test items					
	Remember responses			Know responses		
Associate type	Related	Unrelated	Difference	Related	Unrelated	Difference
List items						
Of nonemotional lures	.75 (160)	.27 (128)	.48*	.25 (160)	.73 (128)	−.48*
Of emotional lures	.80 (160)	.36 (128)	.44*	.20 (160)	.64 (128)	−.44*
Difference	−.05	−.09	—	.05	.09	—
Critical lures						
Nonemotional lures	.62 (154)	.26 (110)	.36*	.38 (154)	.74 (110)	−.36*
Emotional lures	.75 (97)	.35 (39)	.40*	.25 (97)	.65 (39)	−.40*
Difference	−.13*	−.09	—	.13*	.09	—

Note. Conditional probability = Proportion of remembered (or known) responses / proportion of old responses. We calculated these values subject by subject rather than by using the means listed in Tables 2 and 3. The *n* sizes (in parentheses) are unequal across cells because we had to exclude data whenever a person's proportion old value equaled zero (to avoid division by zero).
* $p < .05$, using Loftus and Masson's (1994) within-subjects confidence intervals.

nation for these data, with emotion being one dimension along which items can be distinct. When lures are (emotionally) distinct, few people false alarm to them (i.e., the reverse von Restorff effect found in Experiment 1). When lure distinctiveness is decreased, however, by including emotional items in the study list, emotional lure false recognition is increased (from 18% to 42% and 43% in Experiments 1, 2, and 3, respectively). And, in Experiment 4, when the emotional study items also appeared in the recognition test, emotional lure false memory is decreased again.

We can now answer several key questions raised in the introduction. First, can emotionally charged lures be falsely remembered? The answer is yes; we found false-alarm rates to emotional lures ranging from 18% to 43% across four experiments. These rates (a) were always higher than false alarms to the unrelated, emotional lures but (b) were also always lower than false alarms to the related, nonemotional lures.

Second, can nonemotional study items produce false memory of emotional lures? Experiment 1 satisfied this criterion in that its study list contained no emotional words at all, and it produced significant false recognition of emotional lures, together with a high rate of remember responses to those emotional lures judged to be old.

Third, can the emotional distinctiveness of the lures be reduced, thereby increasing their false-alarm rates? In Experiment 2, adding just three emotional words to the study list more than doubled false memory of the emotional lures at test. These three study items seemed to work as distinctiveness attenuators by making the emotional lures at test less salient, emotionally, than they otherwise would have been. This process then made the discrimination between target and lure items more difficult, which in turn increased the emotional lure false-alarm rate.

Fourth, can this reduction in distinctiveness, once achieved, be reversed? Experiments 3 and 4 were an attempt to make the distinctiveness attenuators themselves different from the emotional lures by varying the former on some other dimension (in this case, color). We found that printing the distinctiveness attenuators in red ink at study failed to have an effect when used alone (Experiment 3), but did result in a significant reduction of emotional lure false alarms when the distinctiveness attenuators themselves appeared in the recognition test (Experiment 4). However, the reduction in emotional lure false alarms achieved in Experiment 4 may well have happened even if we had printed the three emotional items in black (rather than red) ink at study.

Fifth, can people use the emotion-based distinctiveness of a word like *rape* as an aid to correct rejection of a nonstudied item? Given that nonemotional-lure false alarms were always higher than emotional lure false alarms, perhaps people did use emotion to indicate that a word was not present at study (i.e., "If *rape* had occurred in the study list, I would have remembered it."). The use of such a strategy, however, implies that participants changed their criteria for the different types of items (i.e., emotional vs. nonemotional) within a single experiment. Indeed, the issue of criterion shifts is one of current interest for DRM-task researchers. Miller and Wolford (1999), for example, argued that DRM-task false recognition results from participants using liberal criteria for recognition-test items consistent with the theme of the study list but conservative criteria for recognition test items inconsistent with this theme (i.e., for unrelated lures and targets). Not all researchers agree with this interpretation (see Roediger & McDermott, 1999), as trial by trial criterion shifts are both unlikely and difficult to demonstrate, empirically (Stretch & Wixted, 1998; Wickens & Hirshman, 2000; Wixted & Stretch, 2000). At any rate, we do not see the present data as bearing convincingly on these issues, as our experiments were not designed specifically to address them, nor did we include conditions where the lures actually appeared at study, as in Miller and Wolford (1999).

One difference between our data and Roediger and McDermott's (1995) is the higher false-alarm rates we observed for unrelated items across all experiments. Possible reasons for this difference are (a) the relatively large study sets used here (e.g., 64 or 67 items), (b) our intermixing items from different study lists into one master stimulus set, and (c) our use of orthographic rather than semantic list–lure themes. These methodological factors may have made our nonstudied or unrelated distractors (even though unrelated to the study phase items) less distinctive from our target items, relative to other DRM-task studies. Thus, distinctiveness may also partially account for the disparity in false-alarm rates to the unrelated distractors in Roediger and McDermott's (1995) work and ours.

Finally, how do our data relate to Freyd and Gleaves' (1996) criticisms of the DRM task? We did not run their exact thought experiment. In all experiments, the study items were strongly related (orthographically) to the emotional lures, and in the last three experiments, the study phase also contained some emotional words. Moreover, we used only a limited set of emotional lures, with most being sexually explicit or vulgar. Whether the present effects generalize to other types of emotional lures (e.g., *abortion* or *murder*), or to situations outside the laboratory, are issues we did not address. Nonetheless, we did show that emotional critical lures are subject to false remembering in the DRM task, and that this effect can occur even when the study phase contains no emotional words.

References

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

Freyd, J. J., & Gleaves, D. H. (1996). "Remembering" words not presented in lists: Relevance to the current recovered/false memory controversy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 811–813.

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 and Cognition, 27,* 339–353.

Gardiner, J. M., & Java, R. I. (1993). Recognizing and remembering. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 163–188). Hove, England: Erlbaum.

Hunt, R. R., & Eliott, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General, 109,* 49–74.

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 81–87.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review, 4,* 476–490.

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

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102.
Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.
Roediger, H. L., III, & McDermott, K. B. (1996). False perceptions of false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 814–816.
Roediger, H. L., III, & McDermott, K. B. (1999). False alarms about false memories. *Psychological Review, 106,* 406–410.
Roediger, H. L., III, McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memory. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, England: Psychological Press.
Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77.
Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1–24.
Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology, 11,* 331–342.
Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinct events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 565–578.
Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5,* 710–715.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.
Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40,* 83–108.
Stadler, D. L., Roediger, H. L., III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27,* 494–500.
Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1379–1396.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1–12.
von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld [On the effect of field formations in the trace field]. *Psychologische Forschung, 18,* 299–342.
Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2000). *The role of semantic and phonological activation in the creation of false memories.* Manuscript in preparation.
Wickens, T., & Hirshman, E. (2000). False memories and statistical design 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.

Appendix

Matched Study Lists Used in Experiments 1 through 4

Study Set A

RAPE (5)	HOOK (5)	BITCH (6)	SHAVE (6)	WHORE (2)	PEACH (3)
cape	book	ditch	slave	chore	beach
nape	look	hitch	stave	bore	leach
tape	cook	batch	shove	wore	teach
ripe	nook	pitch	share	more	reach
rope	rook	itch	have	tore	poach
race	took	botch	shade	pore	peak
rapt	hock	mitch	shake	sore	perch
rake	honk	butch	shale	horn	peace
rare	hood	birch	shame	shore	preach
raze	hoof	witch	shape	core	peal

Study Set B

SLUT (1)	RINK (2)	HELL (95)	PARK (94)	PENIS (0)	DIGIT (1)
slug	link	bell	bark	venus	widget
slum	mink	dell	dark	genus	midget
slur	sink	fell	hark	penal	bridget
slot	wink	jell	lark	peevish	fidget
slue	pink	sell	mark	penance	divot
shut	rank	tell	nark	venice	divvy
slit	risk	hall	pack	zenith	dimwit
smut	ring	hill	perk	pennies	digest
glut	rick	hull	pork	punish	gidget
scut	fink	shell	spark	pianist	dig

Note. Lures appear in capital letters. Next to each lure is its word frequency value using the Kučera and Francis (1967) norms.

Received May 11, 1999
Revision received August 2, 2000
Accepted August 7, 2000 ■