Journal of Experimental Psychology:
Learning, Memory, and Cognition
2006, Vol. 32, No. 4, 847–853

0278-7393/06/$12.00 DOI: 10.1037/0278-7393.32.4.847

On Rejecting Emotional Lures Created by Phonological Neighborhood Activation

Jeffrey J. Starns
Louisiana State University

Gabriel I. Cook
Claremont McKenna College

Jason L. Hicks
Louisiana State University

Richard L. Marsh
University of Georgia

The authors conducted 2 experiments to assess how phonologically related lures are rejected in a false memory paradigm. Some phonological lures were emotional (i.e., taboo) words, and others were not. The authors manipulated the presence of taboo items on the study list and reduced the ability to use controlled rejection strategies by dividing attention and forcing a short response deadline. The results converge on the idea that participants reduce false alarms to emotional lures by setting more stringent recognition criteria for these items based on their expected memorability. Additionally, emotional lures are less familiar than nonemotional lures because emotional lures have affective and semantic features that mismatch studied nonemotional items.

Keywords: false recognition, emotional lures, phonological lures, metacognition, recognition decision criteria

Most students of memory understand that there are at least three standard processes that can be used to reject distractor items on an old–new recognition test. First, a candidate memory can lack sufficient evidence or familiarity to pass a criterion for being labeled as studied (e.g., Donaldson, 1992; Ratcliff, Sheu, & Gronlund, 1992). That mechanism has widely been associated with single-process models of memory. Second, more consistent with dual-process models is a recall-to-reject strategy of one form or another (e.g., Gallo, 2004; Rotello, Macmillan, & Van Tassell, 2000). When studied items and distractors share certain properties such as being different only in noun plurality or verb tense, being able to recall the studied item (e.g., *dogs*, *help*) will provide a solid basis for rejecting the distractor (e.g., *dog*, *helped*). This strategy also works when small numbers of categorized lists have been studied and one can recall all of the studied items from that category to reject the candidate as not being among those few studied items. In the language of fuzzy trace theory, this process has been labeled recollection rejection (e.g., Brainerd & Reyna, 2002). Third, and finally, there is a class of closely related metacognitive strategies that basically posit that a distractor can be assessed to be quite memorable by virtue of its semantic, orthographic, perceptual, or other idiosyncratic properties, and a lack of those types of details in memory will lead to confident rejection of the test candidate. This mechanism has taken on many names over the years, including negative memory (e.g., Brown, Lewis, & Monk, 1977), a metacognitive strategy (e.g., Strack & Bless, 1994), and most recently, the distinctiveness heuristic (e.g., Dodson & Schacter, 2001; Gallo, Weiss, & Schacter, 2004; Israel & Schacter, 1997).

Against this backdrop, the purpose of the present study was to explore why Pesta, Murphy, and Sanders (2001) found that emotionally charged lures (i.e., taboo words) could be more readily rejected than otherwise comparable nonemotional lures (we use the terms *emotional* and *taboo* synonymously throughout). In their paradigm, they presented 10 phonological neighbors (e.g., *dark*, *lark*) that converged on 12 critical lures (e.g., *park*; also see Sommers & Lewis, 1999; Watson, Balota, & Roediger, 2003). In this manner each of the critical lures received activation during study from 10 phonological neighbors. Six of their critical lures were nonemotional, whereas the other 6 were emotional. Pesta et al. discovered that false alarms to the taboo items were vastly reduced as compared with the control-matched nonemotional critical lures. They also found that placing three additional emotional words on the study list increased false alarms from 18% to 42% for these taboo, emotional lures. Thus, not only were emotional items prone to being retrieved as false memories, but they appeared to be influenced by their distinctiveness from the encoding episode. Studying some emotional items appears to have attenuated the distinctiveness of the emotional lures at test.

Pesta et al. (2001) asserted that the mechanism underlying fewer false memories of emotional than nonemotional items is a metacognitive strategy based on the distinctiveness of emotional items.

Jeffrey J. Starns and Jason L. Hicks, Department of Psychology, Louisiana State University; Gabriel I. Cook, Department of Psychology, Claremont McKenna College; Richard L. Marsh, Department of Psychology, University of Georgia.

We thank Todd Lindsey, Christopher Coughlin, and Genna Jacobs from the University of Georgia and David Doub, Shannon Arnes, and Chad Campbell from Louisiana State University for their dedicated help in collecting the data.

Correspondence concerning this article should be addressed to Jason L. Hicks, Department of Psychology, Louisiana State University, Baton Rouge, LA 70803-5501. E-mail: jhicks@lsu.edu

However, Pesta et al. did not conduct any experiments that actually provided empirical support that their participants were using a metacognitive strategy to avoid accepting the taboo lures. In fact, the full pattern of results reported by these researchers can be readily explained in terms of all three rejection mechanisms summarized in the first paragraph herein. In the following paragraphs, we interpret Pesta et al.'s results in terms of each possible rejection mechanism.

Pesta et al.'s (2001) emotional lures may have inspired fewer false alarms because they were deficient in familiarity compared with the nonemotional lures. For the present study, we define familiarity in the same terms as global matching models; namely, familiarity is the global match of the features in a candidate memory to the features of an entire set of studied items (see Clark & Gronlund, 1996, for a review). Matching features generate activation that is interpreted as evidence that the test cue was studied. Both emotional and nonemotional lures match the phonological features of many items in the encoding set; however, emotional lures have affective and semantic features that mismatch all (or nearly all) of the studied items. These mismatching features may counteract the familiarity created by the phonological match, which would make emotional lures seem less familiar than nonemotional lures. Notably, the retrieving effectively from memory model (REM; Shiffrin & Steyvers, 1997) explicitly predicts that mismatching features decrease familiarity. Pesta et al. also showed that false alarms to emotional lures increased when emotional items were added to the study list. The familiarity-based account can explain this result as well. The emotional studied items match many of the affective and semantic features of the emotional lures, which increases the global familiarity of these lures. Thus, Pesta et al.'s data are consistent with a rejection strategy based on deficient familiarity.

Pesta et al.'s (2001) results are also consistent with a categorical rejection strategy in which participants use the category of the tested word to disqualify the possibility that it may have been studied. To apply this explanation, we assume that taboo or emotional words form a category that can be identified easily by participants. When no emotional words are presented at encoding, participants can use this knowledge to reject emotional words presented as test candidates. When a limited number of emotional words are presented, participants can reject emotional lures by recalling all of the studied emotional words and noting that the test candidate is not among the members of the set (i.e., the recall-to-reject strategy; Gallo, 2004; Rotello et al., 2000). Categorical rejection strategies can explain why there are fewer false alarms for emotional lures, because only these lures belong to a well-defined category with no (or only a few) members in the encoding set. This explanation is also consistent with an increase in false alarms to emotional lures when emotional words are encountered at study because, in this situation, lure rejection is predicated on the ability to exhaustively recall the studied taboo words. Although Pesta et al. did not consider this rejection strategy, based on the foregoing analysis we believe that it is a viable contender.

As noted by Pesta et al. (2001), their results can also be explained in terms of a metacognitive strategy. Participants may use the expected memorability of test items to adjust their standards for the type or amount of memory evidence that they require to claim that the item was studied (Johnson, Hastroudi, & Lindsay, 1993). If participants expect vivid memory for emotional words, then they should set stringent criteria for identifying these items as previously studied. Thus, there may be fewer false alarms for emotional lures because these items are evaluated against a higher standard. This rejection mechanism is similar to the distinctiveness heuristic in which participants reduce false memory after seeing a picture for each item at encoding (e.g., Israel & Schacter, 1997). Participants expect to recollect pictorial details for studied items and are able to reject lures based on the absence of these details. However, the metacognitive strategy investigated in the present study differs from the distinctiveness heuristic in an important way. In the distinctiveness heuristic, more stringent standards are applied to every item on a test because participants can expect *all* of the studied items to meet these standards. For instance, participants show fewer false memories after studying an entire list of pictures as compared with studying an entire list of words. In contrast, the shift in standards that could explain reduced false memory for emotional lures must occur on an item-by-item basis within a single recognition test. On the same test, participants must use relatively lenient standards for nonemotional lures and relatively stringent standards for emotional lures. We label this strategic metacognitive shift in standards an item-based criterion account because participants must evaluate the expected memorability of individual items to apply the strategy.

With an additional assumption, the item-based criterion account can also explain the effect of adding taboo items to the study list. Specifically, one may assume that participants evaluate expected memorability based not only on the characteristics of the item itself but also on the item's distinctiveness in relation to the other items on the study list. When taboo items are studied, the taboo test lures lose their distinctiveness against people's evaluation of the entire encoding experience. In other words, rejecting the word *penis* is not accomplished by assessing what recollective details about this particular word would have come to mind, but rather by assessing how likely would *penis* be remembered given what else was encountered during the entire study context. Thus, participants set more stringent standards for taboo lures when no taboo items are studied because they expect taboo words to be more distinctive and better remembered. By this account, the emotional lure false-alarm rate should be lower when taboo words are not studied. However, if participants do not consider distinctiveness from the studied items when evaluating expected memorability, then an item-based criterion shift should reduce false memories for emotional lures to the same extent regardless of their presence or absence on the study list.

The purpose of this study was to more definitively specify the rejection mechanism (or combination of mechanisms) that allows participants to make fewer false alarms to emotional versus nonemotional lures. We appealed to the hypothesis that if deficient familiarity is responsible for being able to avoid false memories to emotional items, then this ability should not decline under divided attention or with a response signal procedure. After all, such manipulations interfere with consciously controlled strategies but generally leave familiarity-based processes intact (e.g., Jacoby, Jones, & Dolan, 1998; Jacoby, Toth, & Yonelinas, 1993; Jennings & Jacoby, 1993). In each experiment, we also manipulated whether emotional items were presented on the study list to replicate Pesta et al. (2001) and to further elucidate the differences between memory processes for emotional and nonemotional lures.

Experiment 1

In this experiment and the one that follows, we manipulated in an entirely between-subjects design two independent variables, each of which had two levels. One variable was the presence of emotional words on the study list. The other variable manipulated in each experiment was designed to limit processing resources at retrieval. In Experiment 1, we achieved this by means of a response deadline manipulation (see Benjamin, 2001; Dodson & Hege, 2005). If a controlled rejection strategy is used to decrease false memories for emotional lures, then limiting resources at test should impair participants' ability to use the strategy, thereby increasing false alarms to emotional lures. Thus, there should be an interaction between type of lure and the processing manipulation. Specifically, the emotional lure false-alarm rate should be much lower than the nonemotional with no processing limitations. By contrast, the magnitude of the difference between the false-alarm rates of the two lure types should decrease when processing limitations are introduced because an inability to strategically reject emotional lures should bring them closer in the range of the response scale to the nonemotional lures. If emotional lures are simply less familiar than nonemotional lures (the deficient familiarity hypothesis), then the magnitude of the difference in false-alarm rates between emotional and nonemotional lures should not change when resource limitations are introduced because familiarity assessment is not a very resource-demanding process (e.g., Jacoby et al., 1998, 1993; Jennings & Jacoby, 1993).

In addition, limiting processing resources at retrieval will help us to specify whether adding taboo items to the study list affects the familiarity of taboo lures or one of the controlled strategies. If the presence of taboo items on the study list affects only familiarity, then adding taboo words to the study list should have the same effect regardless of whether the available processing resources at test are taxed versus not. By contrast, if the presence of taboo items influences a controlled rejection strategy, then adding studied taboo words should have a larger effect on emotional lures when processing resources are not taxed compared with when they are taxed. Processing limitations disrupt the ability to use a controlled rejection strategy; thus, they should also reduce the influence of any variable that affects memory performance via a controlled rejection strategy.

Method

Participants. University of Georgia undergraduate students volunteered in exchange for partial credit toward a course research requirement. Extra course credit was awarded to Louisiana State University students for volunteering. Each participant was tested individually in sessions that lasted approximately 25 min. Cell sizes are reported in Table 1; they were unequal because of some miscommunication between our laboratories at the end of an academic term.

Design, materials, and procedure. This experiment followed a 2 (self-paced vs. speeded testing) × 2 (zero vs. three taboo words on the study list) × 2 (emotional vs. nonemotional lures) factorial design. The response deadline variable and the presence of taboo words on the study list were manipulated between participants. Emotionality of the critical lures was manipulated within participants. The materials were the 12 lists of 10 words that converged phonologically on the 12 critical lures as found in the appendix of Pesta et al. (2001). Six of the lists converged on nonemotional lures and 6 converged on emotional lures. Nonemotional and emotional lures were matched for word frequency, letter length, and the number of orthographic neighbors in the study list (Pesta et al., 2001). We presented items at study in the center of the computer monitor for 3 s each. Presentation generally followed Pesta et al.'s method of blocking by list. In our case, we presented 4 items back to back from each list before cycling back to present another 4 items back to back from each of the 12 lists. In this fashion, participants studied 96 items of the original 120 items, thereby reserving 24 items to use as lures on the recognition test. For each participant, the order of list presentation was randomized anew before assigning blocks of 4 studied items to the study list. Participants who studied zero taboo words saw only the original 96-item study list. For those participants who studied three taboo words, we intermixed into the study list the words used by Pesta et al. for a total of 99 study trials. The taboo words appeared on Study Trials 28, 61, and 88. At test, 72 items total were presented for evaluation. For each list, three test words were studied (36 total), two were phonological lures (24 total), and one word was the critical lure for that list (6 emotional and 6 nonemotional total). When words were presented at study or test, they were always preceded by a fixation point for 250 ms and were accompanied by a short (100-ms) warning tone to assist participants in keeping up with the experimental procedures.

Table 1
Overall Recognition Memory and False Alarms (FAs) to Taboo and Nontaboo Distractors for Experiments 1 and 2

Experiment and condition	*N*	Hits	FAs	Hits − FAs	Nontaboo FAs	Taboo FAs
Experiment 1						
Unspeeded						
Study no taboo	30	.74 (.02)	.33 (.03)	.41 (.04)	.56 (.04)	.12 (.04)
Study three taboos	32	.74 (.02)	.39 (.03)	.35 (.04)	.62 (.05)	.22 (.03)
Response signal						
Study no taboo	37	.70 (.02)	.44 (.03)	.26 (.02)	.62 (.04)	.24 (.04)
Study three taboos	35	.72 (.02)	.45 (.03)	.27 (.03)	.68 (.04)	.34 (.04)
Experiment 2						
Full attention						
Study no taboo	32	.74 (.02)	.36 (.02)	.38 (.03)	.57 (.04)	.10 (.02)
Study three taboos	32	.73 (.02)	.35 (.03)	.38 (.03)	.55 (.04)	.18 (.03)
Divided attention						
Study no taboo	32	.65 (.02)	.40 (.03)	.25 (.03)	.52 (.04)	.28 (.04)
Study three taboos	32	.62 (.02)	.42 (.03)	.19 (.03)	.49 (.05)	.36 (.04)

Note. Standard errors are in parentheses.

For both the study and the test phases, participants first read the instructions from the computer monitor and then listened to a verbal reiteration by the experimenter. The consent form clearly indicated that some verbally offensive and vulgar words might be encountered during the course of the experiment; and the experimenter pointed this out before the beginning of the experiment. Across both experiments in this report, only 2 students declined to participate for this reason (and they were given credit for participating anyway). The study instructions asked that participants pay attention to the words for a later, unspecified memory test. A 4-min distractor activity was inserted between the end of the study list and the test instructions. One group of participants worked through the recognition test at their own pace, and another group of participants were compelled to respond very quickly with a response deadline. In the latter group, every test item was followed with a high-pitched tone 400 ms after the onset of the item. Participants were instructed that they should make their recognition decision immediately on hearing this unique tone. Although we did not directly enforce this instruction with feedback about their response latencies, the instruction was stressed both verbally and in the written instructions that preceded the test phase of the experiment. To familiarize them with the procedure, we asked participants to study a short list of 10 items followed by a recognition test consisting of 20 items. This practice phase preceded the real study list and test phase. There were no taboo items in this practice phase.

Results and Discussion

Unless otherwise specified with an explicit probability value, the probability of a Type I error did not exceed .05. For those under the response deadline, trials with response latencies longer than 1,400 ms were removed (1.96%) on the assumption that this was 1 full second past the signal, which should have provided ample time to respond. The results are summarized in the top portion of Table 1. In no experiment in this report did the hit rate or the false-alarm rate differ for lists that converged on emotional versus nonemotional critical lures. Consequently, for simplicity we have pooled over these hit and false-alarm rates. The data were first analyzed with a series of 2 (self-paced vs. speeded testing) $\times$ 2 (zero vs. three taboo words on the study list) between-subjects analysis of variance (ANOVA) models on each column presented in Table 1. The response signal procedure did not affect the hit rates, but it did increase the false-alarm rates to noncritical phonological lures (column 4 of Table 1), $F(1, 130) = 7.25$. When combined to calculate the two high-threshold correction (hits–false alarms), performance was worse with a deadline imposed as compared with when more leisurely responding was allowed, $F(1, 130) = 12.52$. In none of these models did the manipulation of placing three taboo items on the study list (vs. not) affect performance (and there were no significant interactions either). The response signal served to increase the false-alarm rate to taboo lures, $F(1, 130) = 8.98$, but not to the nonemotional lures, $F(1, 130) = 1.83$, *ns*. That outcome is consistent with nonemotional lures being accepted on the basis of familiarity and taboo lures being rejected using a controlled strategy. Moreover, only the false-alarm rate to the taboo lures was increased by placing three taboo items on the study list, $F(1, 130) = 6.94$, whereas the nonemotional lures were not affected, $F(1, 130) = 1.85$, *ns*. The response signal manipulation did not interact with number of taboo items for either the emotional or nonemotional lures.

To confirm these observations, we directly compared the emotional lures with the nonemotional lures by adding the within-subjects variable of emotionality of the lure to the previous 2 $\times$ 2 ANOVA (i.e., a three-way ANOVA). This expanded ANOVA produced a main effect of having a deadline (vs. not), $F(1, 130) = 8.32$, and a larger false-alarm rate to the nonemotional lures than the taboo ones, $F(1, 130) = 207.60$, but no interactions. Thus, replicating Pesta et al. (2001), we showed an increase in emotional false alarms when taboo items were added to the study list. The fact that the response deadline increased false memories for taboo lures without significantly influencing nonemotional lures suggests that participants without a deadline used a controlled rejection strategy to depress false responding to taboo lures. Stated differently, the interaction of lure type and processing resources predicted by the two accounts positing controlled rejection strategies is nominally apparent in the data; namely, the magnitude of the difference between emotional and nonemotional false-alarm rates decreased when participants had to respond on a deadline. However, this interaction did not reach significance. In Experiment 2, we used random-number generation as a concurrent task because this task has been shown to severely limit executive function (Baddeley, 1986). With a more powerful manipulation of the available resources at test, we hoped to establish the statistical significance of the apparent interaction of lure type and processing resources.

Experiment 2

Because Experiment 1 is the first demonstration that rejecting emotional lures requires conscious, controlled resources, we believed that it was incumbent on us to demonstrate that the effect is robust and generalizes to other ways of limiting available resources. According to dual-process model theorists (e.g., Jacoby et al., 1998), divided attention at test disrupts the assessment of recollective details and leaves familiarity assessments generally intact. Therefore, Experiment 2 is a replication of Experiment 1 except that rather than introducing a response deadline, we had participants concurrently perform random-number generation on the recognition test.

Method

Participants. Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a course requirement. Extra course credit was awarded to students from Louisiana State University for their participation. The 128 volunteers were quasi-randomly assigned to the four between-subjects conditions created by crossing numbers of taboo items on the study list (zero vs. three) with full versus divided attention at test.

Procedure. In all respects except one, this experiment was conducted in an identical manner to Experiment 1. We replaced the response signal procedure with a divided attention manipulation. Half of the participants performed the recognition test under full attention, whereas the other half had to generate random numbers concurrently. Random-number generation conformed to our previous uses of it (e.g., Hicks & Marsh, 2000; Marsh, Cook, & Hicks, 2006). In brief, participants were asked to generate numbers in time to a taped metronome that played a beep every second. They were given detailed instructions on what it meant to be random and were given practice after the study list and before the recognition test itself. The metronome was started 20 s prior to the recognition test so that participants were already generating numbers when the first test item appeared. Practice generating random numbers served instead of the distractor task for those participants whose attention was divided during the test.

Results and Discussion

The results are summarized at the bottom of Table 1 in a manner consistent with the previous experiment. Not surprisingly, in the 2 (full vs. divided attention) × 2 (zero vs. three studied taboo items) ANOVA, dividing attention reduced the hit rates, $F(1, 124) = 23.52$, while also increasing the false-alarm rates, $F(1, 124) = 3.44$, $p = .06$. When these measures are rolled into the two high-threshold correction metric, performance was worse under conditions of divided attention, $F(1, 124) = 27.48$. In none of these models was there an effect of presenting emotional items (vs. not) on the study list or any interactions. Therefore, concurrent random-number generation at retrieval lowered recognition performance just as Hicks and Marsh (2000) demonstrated (see Lozito & Mulligan, in press). The parallel analysis on the nonemotional lures showed no significant effects of attention, the number of taboo items on the study list, nor any interactions, all $Fs(1, 124) < 1.76$, *ns*. Those null outcomes are consistent with our argument that nonemotional lures are largely judged old (erroneously) on the basis of their phonetic similarity and familiarity whose assessment is relatively immune to manipulations of attention. In contrast, the taboo lures were claimed to be old significantly more often under divided attention as compared with full attention, $F(1, 124) = 24.40$. These same items were also more frequently claimed to be old after studying three taboo words on the list, $F(1, 124) = 4.93$, but there was no interaction. That outcome is consistent with taboo lures being rejected on the basis of a strategy that requires attentional control.

In the three-way ANOVA adding in the variable of emotionality of the lures, nonemotional lures were claimed to be old more often than emotional lures, $F(1, 124) = 127.27$. Moreover, emotionality of the lures interacted both with the manipulation of attention, $F(1, 124) = 19.15$, and with the presence of taboo items on the study list, $F(1, 124) = 3.87$. The first effect reflects the fact that taboo lures were sensitive to the manipulation of attention whereas the nonemotional lures were not. The second effect reflects the fact that the false-alarm rate to emotional lures increased when three other vulgar words were studied, but this manipulation did not affect nonemotional lures. All other significant effects were qualified by these two interactions. The pattern of results mirrors the first experiment, and the apparent interactions of lure type with processing resources and with the presence of studied taboo items reached significance in this experiment.

The most obvious implication of Experiments 1 and 2 is that some form of controlled rejection strategy allowed participants to "edit" false memories for emotional items under normal test conditions. When the use of this controlled strategy was disrupted by introducing processing limitations, the emotional lure false-alarm rate increased to a magnitude more similar to the nonemotional lure false-alarm rate. However, careful consideration of both experiments reveals that familiarity also played a role in the differences in false memory to emotional versus nonemotional lures. Adding items from the taboo category to the study list increased false alarms to emotional lures, and the magnitude of this increase was the same regardless of the level of processing resources available at test.[1] That is, in the analyses on the emotional lure false-alarm rate, the effect of adding taboo studied items did not interact with the processing-resource manipulations in either experiment. Moreover, the means in Table 1 reveal no hint of such an interaction. This outcome demonstrates that adding taboo items to the study list increased the familiarity level of emotional lures but did not influence the controlled rejection strategy that participants were trying to apply. If the presence of studied taboo items affected the controlled strategy, then adding studied taboo words should have had a smaller effect when the strategy was disrupted by imposing processing limitations. Obviously, this did not happen.

Overall, our results indicate that emotional lures begin as less familiar than nonemotional lures because they have affective and semantic features that mismatch all of the nontaboo items on the study list. However, emotional lure familiarity can be increased by adding words from the taboo category to the study list. Additionally, when normal processing resources are available, participants can reduce false alarms to emotional lures even further by applying a controlled rejection strategy. In support of the dual role of deficient familiarity and controlled rejection of taboo lures, in both experiments the false-alarm rate to emotional lures remained lower than nonemotional lures even when controlled processing was disrupted.

One remaining question is which of the two possible controlled rejection strategies, categorical rejection or item-based criterion shift, is responsible for reducing emotional false alarms when sufficient processing resources are available at test. A key finding answering this question is that the controlled strategy was not influenced by the presence of taboo items on the study list (this manipulation only influenced familiarity, as revealed in the lack of an interaction between processing resources and the presence of taboo items). This outcome is inconsistent with participants relying on a categorical rejection strategy because this strategy should be more difficult to apply when taboo items appear on the study list. As noted in the introduction, when no taboo items are on the study list, participants can reject emotional lures just by noting that the test candidate belongs to an unstudied category. When there are taboo items on the study list, participants can disqualify emotional lures only by recalling all of the taboo items (see Gallo, 2004). Thus, a categorical rejection strategy is inconsistent with the finding that the presence of taboo items on the study list had no influence on the controlled rejection of lures.

In contrast, the item-based criteria account is perfectly consistent with this finding. If participants base their expected memorability judgments on stable characteristics of the candidate item, then expected memorability (and thus recognition criteria) will not be influenced by the other items on the study list. When participants have the resources necessary to adjust criteria on an item-by-item basis, they will set high standards for emotional lures regardless of the number of taboo items on the study list. From this perspective, the characteristics of an emotional lure itself, not its relationship to the other encoded items, promote stringent criteria.

[1] In an unreported experiment conducted as part of the same project, we presented either three or eight taboo words on the study list and observed that the false-alarm rate for emotional lures was higher with eight than with three taboo items. This result bolsters our conclusion that adding taboo items to the study list increases the familiarity of emotional lures: With more studied taboo items, there are more memory traces that match the affective and semantic features of taboo lures, resulting in higher levels of familiarity for these lures.

Thus, our full pattern of results involved two independent effects: The emotionality of the lures influenced the memory criteria used by participants, and the presence of studied taboo items directly increased the level of familiarity experienced for emotional lures.

General Discussion

In the experiments reported herein, participants were able to avoid false memories of emotional lures more successfully than control-matched lures that lacked the emotional dimension. Our goal was to discriminate three theoretical accounts of this pattern of results: deficient familiarity, item-based criteria, and categorical rejection. The results showed that item-based criterion shift allowed participants to avoid false responding to emotional lures. That is, participants expected to have very good memory for emotional items and, consequently, they set stringent standards for the evidence required to claim that an emotional test candidate was studied. When the response deadline or concurrent number generation task disrupted people's ability to assess each item's memorability and adjust criteria on an item-by-item basis, false responding to emotional lures increased.

Our results show that emotional lures are also less familiar than nonemotional lures because they have features that mismatch the encoded items. False memories for emotional lures remained considerably lower than false memories for nonemotional lures even when controlled rejection was disrupted with processing limitations. Also consistent with this account, studying taboo items that shared features with emotional lures increased the familiarity of these lures. The fact that adding taboo items to the study list increased the emotional lure false-alarm rate to the same extent with normal or limited processing resources demonstrates that the increase was a familiarity-based effect.

Our participants were able to reduce false memory by flexibly adjusting response criteria within a recognition test on the basis of the expected memorability of individual items. It is interesting to note that several studies have shown participants are unable to make item-by-item shifts in criteria even when provided with ample support. In another false memory experiment, Starns, Hicks, and Marsh (in press) explored participants' ability to reduce false memory when accurate memory was enhanced through item repetition. When all of the lists in the study phase were repeated, participants were able to reduce false responding for unrepeated lists by setting a more stringent recognition criterion. However, when repeated and nonrepeated lists were mixed into the same test, participants actually displayed higher rates of false memory for repeated lists. Thus, item repetition induced participants to change global standards applied uniformly across a test list, but participants did not set separate criteria for repeated and nonrepeated words on an item-by-item basis within a test.

Similarly, Stretch and Wixted (1998) presented a study in which words on a single study list were presented either once or three times. In addition, these researchers presented all repeated items in red and all nonrepeated items in green at both study and test, allowing participants to be absolutely sure that any red test candidate would have been studied repeatedly if it were on the study list. Even when provided with this information, participants did not adjust their response criterion on the basis of the repetition status of the items: The false-alarm rates for lures presented in red and lures presented in green were equal. Thus, in many circumstances, participants seem unwilling or unable to flexibly adjust response criteria within a test based on item memorability. In our experiments, the item characteristics signaling the need for criteria adjustment were highly salient and distinctive compared with the other items on the test, the item classes markedly differed in memorability, and participants likely had acute metamnemonic awareness of how their memory for the two types of items would differ. Under these circumstances, participants were able to tailor their recognition criteria for individual items.

Some have claimed that the principles of false memory uncovered in the laboratory are not applicable to emotional items or events (Freyd & Gleaves, 1996). Pesta et al. (2001) demonstrated that emotional lures are more easily rejected than nonemotional lures, but also that emotional lures are quite susceptible to false memory. The experiments reported herein demonstrate that memory processes for emotional items do not violate basic memory principles. To the contrary, we were able to determine why participants display less false memory for emotional lures by exploring three rejection mechanisms derived from basic research. Our results showed that emotional lures are less familiar than nonemotional lures because fewer studied items match the features of emotional than nonemotional lures. In addition, participants set more stringent response criteria for emotional items when they have the resources needed to make those adjustments. Thus, our results confirm the intuition that people are more likely to falsely remember that they took a field trip to the zoo as a child than that they saw an armed robbery as a child. Less evidence should be available in memory that is consistent with a candidate memory of armed robbery as compared with a trip to the zoo. Moreover, people will require more evidence to assert that a candidate memory of armed robbery is veridical as compared with a candidate memory of a field trip. Nevertheless, our results suggest that falsely remembering emotional events may be more common than previously suggested, especially when the cognitive demands and distractions of everyday life are factored into the equation of recollection.

References

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 941–947.

Brainerd, C. J., & Reyna, V. F. (2002). Recollection rejection: How children edit their false memories. *Developmental Psychology, 38,* 156–172.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology, 29,* 461–473.

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

Dodson, C. S., & Hege, A. C. G. (2005). Speeded retrieval abolishes the false memory suppression effect: Evidence for the distinctiveness heuristic. *Psychonomic Bulletin & Review, 12,* 726–731.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review,* 8,155–161.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121,* 275–277.

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. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 120–128.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory & Language, 51,* 473–493.

Hicks, J. L., & Marsh, R. L. (2000). Toward specifying the attentional demands of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1483–1498.

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

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of knowledge judgments and exclusion errors. *Psychonomic Bulletin & Review, 5,* 705–709.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122,* 139–154.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8,* 283–293.

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

Lozito, J. P., & Mulligan, N. W. (in press). Exploring the role of attention during memory retrieval: Effects of semantic encoding and divided attention. *Memory & Cognition.*

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory, 14,* 148–160.

Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 328–338.

Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535.

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language, 43,* 67–88.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145–166.

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.

Starns, J. J., Hicks, J. L., & Marsh, R. L. (in press). Repetition effects in associative false recognition: Theme-based criterion shifts are the exception, not the rule. *Memory.*

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory & Language, 33,* 203–217.

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.

Watson, J. M., Balota, D. A., & Roediger, H. L., III. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory & Language, 49,* 95–118.

Received November 16, 2005
Revision received February 24, 2006
Accepted March 2, 2006 ■