MEMORY, 2006, 14 (6), 742–761

Psychology Press
Taylor & Francis Group

Repetition effects in associative false recognition: Theme-based criterion shifts are the exception, not the rule

Jeffery J. Starns and Jason L. Hicks
Louisiana State University, Baton Rouge, LA, USA

Richard L. Marsh
University of Georgia, Athens, GA, USA

Previous reports have demonstrated that false memory for the critical items of associative lists decreases when lists are studied multiple times (Benjamin, 2001). In three experiments, we explored two hypotheses that might account for false memory reductions with repetition. Under an identification hypothesis, repetition decreases false memory because participants realise that critical items are absent from the list at encoding and thus reject them at test. Under a criterion shift hypothesis, repetition decreases false memory because it increases the discriminability of studied words from lures, causing participants to set a higher response criterion for positive recognition responses. Results uniquely supported the criterion shift hypothesis. Furthermore, results showed that participants only changed their criterion on separate recognition tests, not on an item-by-item basis within a single recognition test. The failure to establish separate criteria within a test *increased* false memory for repeated lists.

Memory improvement is often discussed in terms of encoding more and/or forgetting less information. However, a wealth of empirical evidence reveals that accurate memory relies not only on the ability to reproduce previously encoded information, but also on the ability to avoid erroneously reporting details that do not objectively correspond to past experiences (see Roediger, 1996, for a review). The associative list paradigm (Deese, 1959; Roediger & McDermott, 1995) provides a dramatic example of people's susceptibility to falsely remembering inaccurate information. In this paradigm, participants encode lists of words that are all strongly conceptually related to a single non-presented word, called the critical item. Following the presentation of such lists, participants frequently falsely remember encountering critical items; indeed, participants are sometimes more likely to claim to have studied critical items than the words that were actually presented (e.g., Roediger & McDermott, 1995; Tussing & Greene, 1997). Thus, the associative list paradigm is well suited to investigate variables that influence people's ability to avoid false memories.

Prior research suggests that people are sometimes better able to reject highly related lures when memory for presented words is strengthened through repeated presentation (Benjamin, 2001; Brainerd, Reyna, & Kneer, 1995; Hall & Kozloff, 1970; Kensinger & Schacter, 1999; Schacter, Verfaellie, Anes, & Racine, 1998; Seamon, Luo, Schwartz, Jones, Lee, & Jones, 2002; Tussing & Greene, 1999). Some of these studies

Address correspondence to: Jason L. Hicks, Department of Psychology, Louisiana State University, Baton Rouge, LA 70803-5501, USA. E-mail: jhicks@lsu.edu
We thank Janina Mayeaux and Patrick Callahan for their assistance in collecting the data for this project.

© 2006 Psychology Press Ltd
http://www.psypress.com/memory DOI:10.1080/09658210600648514

involved lures that were related to only one studied item (Brainerd et al., 1995; Hall & Kozloff, 1970), which promotes a simple mechanism of false memory reduction. When tested with a related lure, participants might recall the item that was studied and reject the test item because it is different from the recalled item. This strategy was initially discussed by Tulving (1983), and Gallo (2004) has recently classified the strategy as disqualifying recall-to-reject. False recognition naturally decreases with repetition when this strategy is used, because participants are more likely to successfully recall repeated versus non-repeated items.

Disqualifying recall-to-reject cannot suppress false memory in the typical associative list paradigm because critical items are related to a number of words from their lists (up to 15 in some studies). Exhaustively recalling all studied items related to a test word before making a recognition decision is difficult or impossible, which eliminates the effectiveness of the strategy (Gallo, 2004). Consistent with this argument, Tussing and Greene (1997) found no effect of repetition on false memory with the associative list methodology. Nevertheless, other studies have demonstrated that list repetition reduces false memory in some conditions in this paradigm (Benjamin, 2001; Kensinger & Schacter, 1999; Schacter et al., 1998; Seamon et al., 2002). Benjamin's (2001) study provides a particularly relevant example. In this study, participants encoded sets of associative lists that were presented either one time or three times. For college-aged participants, presenting lists repeatedly reduced false recognition for critical items in addition to enhancing recognition of studied words. Interestingly, repetition increased false memory for elderly adults and for young adults when recognition responses were speeded. These collective results led Benjamin to conclude that repetition increased critical item familiarity, but also that young adults in unspeeded tests could exert cognitive control to oppose the added familiarity with some sort of rejection process to reduce false memories.

The purpose of this article is to differentiate several mechanisms that may create reductions in false memory for repeated lists in the associative list paradigm. We approach this issue from the standpoint of signal detection theory (SDT; Macmillan & Creelman, 2005). Applying SDT to recognition memory implies that there are two general influences on recognition performance: the information encoded into memory and the decision processes applied to this information. The amount of evidence in memory is represented as a unidimensional, continuous variable. We will simply refer to strength-of-evidence as *familiarity* without implying any specific theoretical assumptions as to its nature (cf. Pastore, Crawley, Berens, & Skelly, 2003). On a recognition test, both targets and lures produce varying amounts of familiarity, but targets are more familiar than lures on average as a result of the information encoded during the study list's presentation. SDT mathematically represents this situation by assuming that the familiarity values of targets and lures are both normally distributed and that the mean of the target distribution is greater than the mean of the lure distribution. Participants translate familiarity values into recognition responses by setting a criterion and claiming to recognise any test word that has a familiarity value that exceeds the criterion.

A depiction of SDT relevant to the current study is presented in the top panel of Figure 1. In the panel, there are three distributions corresponding to the three tested item types in a typical associative list experiment: targets, lures, and critical items. Lures are the least familiar items on average, so the lure distribution is the farthest to the left on the familiarity continuum. Critical items gain familiarity as a result of the presentation of highly related words in the study phase and the critical item distribution therefore lies to the right of the lure distribution (see Wixted & Stretch, 2000). The d'_c parameter represents the distance between the lure and critical item distributions. Targets gain familiarity as a result of their literal presentation in the study phase, and the specific representation in Figure 1 assumes that targets gain more familiarity from the study phase than do critical items (i.e., the target distribution is farther up the familiarity continuum than the critical item distribution). The d'_s parameter represents the distance between the lure and studied item distributions. The vertical line labelled "lambda" on the figure represents the familiarity value used as the response criterion—the proportion of each distribution falling to the right of this line is the proportion of positive recognition responses that would be made to that item type.

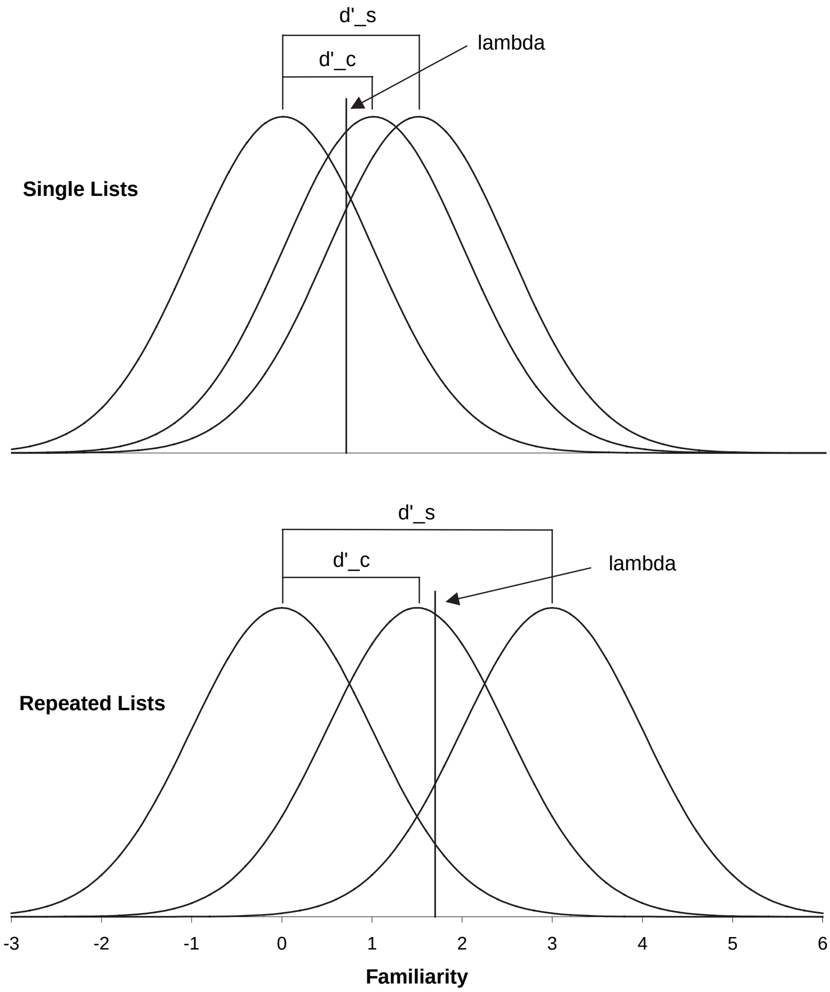

Figure 1. Signal detection models displaying a reduction in false memory with list repetition resulting from a criterion shift. The top panel displays a model for lists presented once, and the bottom panel displays a model for lists presented repeatedly. In both panels, the left-most distribution is the related lure distribution, the right-most distribution is the target distribution, and the middle distribution is the critical item distribution.

THE INFLUENCE OF REPETITION ON FAMILIARITY AND DECISION PROCESSES

All of the prominent theories of false memory—i.e., activation-monitoring theory (Roediger, Watson, McDermott, & Gallo, 2001), fuzzy trace theory (Reyna & Brainerd, 1995), and global activation models (e.g., Arndt & Hirshman, 1998)—agree that critical items could become more plausible as candidates for positive recognition responses when associative lists are repeated. Benjamin's (2001) results are consistent with this prediction, in that repetition increased false recognition under conditions that limited the role of controlled processing at test. Repetition increases the familiarity of critical items, so any reduction in false memory due to repetition requires an alternative mechanism that opposes the influence of that familiarity. We will consider two possible suppression mechanisms: false memory may decrease with repetition either because participants identify the critical item as a non-presented item when lists are repeated or because repetition induces a more stringent recognition response criterion.

When participants are told before encoding associative lists that each list has a missing but highly related word that they are likely to falsely remember, they dramatically reduce false memory by identifying the missing word at encoding, at least for critical items from lists that are high on an identifiability index (Neuschatz, Benoit, & Payne, 2003). List repetition may lead to a similar result. When participants study a repeated list, they may expect certain words to appear on subsequent list presentations based on their memory for prior presentations. Associative lists engender high rates of false memory, so it is likely that participants will expect to see the critical item in repeated presentations of the list. When the critical item does not appear, participants may realise that this word is missing from the list. Finally, when the word later appears at test, participants could remember noting its absence and use this as a basis to reject it.

Whereas noting the absence of critical items during list presentation is essentially an encoding process, repetition could also reduce false memory by encouraging a more stringent response criterion at retrieval (Benjamin, 2001). Participants may require more evidence to claim to have studied words from repeated lists than words from non-repeated lists because they have stronger memories for repeated items. A situation in which list repetition might affect false recognition within a signal detection framework is displayed in Figure 1. The top panel of this figure displays a hypothetical model for once-presented lists, and the bottom panel displays a model for repeated lists. In the figure, list repetition increases the familiarity of both critical items and targets (i.e., both of these distributions shift to the right). The figure also displays an effect of repetition on response criterion, such that the criterion is set at a higher, more conservative value when lists are repeated. This criterion shift occurs in response to the enhanced familiarity of the targets—the target distribution "pulls" the criterion with it as it moves to the right (Hirshman, 1995). As a result of this criterion shift, repetition has the effect of decreasing false memories for critical items. A larger proportion of the critical item distribution lies to the right of the response criterion in the top panel than in the bottom panel. This demonstrates how repetition can reduce overall false memory rates, even if it enhances the familiarity of critical items. This model of repetition effects assumes that repetition enhances the familiarity of targets to a greater extent than critical items, and that the response criterion increases when target familiarity increases (see Hirshman, 1995).[1]

A critical consideration in predicting the effects of repetition on false memory is the *specificity* of changes in criterion. Criterion shifts may either be applied globally to all items on a test or locally on an item-by-item basis (Stretch & Wixted, 1998). If criterion shifts are made at a global level, then repetition should only reduce false memory when repeated and non-repeated items appear on different tests. When repeated and non-repeated items are mixed in the same test, the same criterion is applied to critical items from both types of lists, so responding can only differ based on the location of the critical item distribution. As such, repeating lists should create more false memories. In contrast, if participants can change decision processes on an item-by-item basis, false memories may decrease for repeated lists even when they are mixed on the same test as non-repeated lists. This selectivity in criterion setting would imply that participants could use a higher criterion when tested on an item from a repeated list (i.e., lower panel of Figure 1) but a lower criterion for items associated with non-repeated lists (i.e., upper panel of Figure 1).

Benjamin (2001) showed a within-test reduction in false memory for repeated associative lists, so his results are consistent with local changes in decision processes, at least for younger participants on an unspeeded test. This mechanism requires that people can identify items that should be evaluated against the higher standard of occurring repeatedly as opposed to occurring once. People cannot make adjustments based on whether they think the test item itself was repeated or not—if they know the item's frequency of presentation, then the recognition decision has already been made. Instead, people must decide how many times the list associated

[1] In an effort to conceptually replicate many of the features of Benjamin's (2001) study, we did not include unrelated lures on the recognition test. Some readers will correctly infer that repetition could actually cause all three distributions in the lower panel of Figure 1 to increase (i.e., including related lures). However, most studies of false recognition in the DRM paradigm do not often show systematic variations in false recognition of weakly related lures, even when unrelated lures are included on the test. Because the associative relationship of related lures to the list is not as strong as the relationship of the critical item to the list, we are not concerned about the possibility of shifts in the related lure distribution. Most important, the use of related lures as opposed to unrelated lures does not change the fundamental predictions made by the signal detection model.

with a test item was presented when the presentation history of the test word itself is still in doubt. This could perhaps be achieved by recalling studied words related to the tested item and evaluating the memory evidence retrieved for these words. If memory for related items is strong, people could set a high criterion in order to decide that the tested item was studied. Gallo (2004) refers to this process of using recalled related items to inform the amount of evidence that should be experienced for a tested word as diagnostic recall-to-reject. This strategy may have allowed Benjamin's participants to alter decision processes on an item-by-item basis.

Stretch and Wixted (1998) reported a series of experiments suggesting that people fail to make item-by-item criterion shifts even when they can be absolutely certain which test words should be considered as repeated items and which should be considered as single presentation items. Participants studied red words multiple times and green words only once, and targets on the recognition test always appeared in the colour in which they were studied. Thus, all red words on the test would have been presented repeatedly if they were studied, and all green words would have been presented only once if studied. In this situation, participants could have applied separate response criteria for repeated and non-repeated items with complete certainty and improved their recognition performance accordingly. However, participants applied the same criterion for repeated and non-repeated items. For example, in Stretch and Wixted's Experiment 5, the false alarm rate to lures presented in red (the strong condition) was not different from the false alarm rate for lures presented in green (the weak condition). This result indicates that people do not strategically shift their criterion within a test, even when they are given a great deal of support in the testing situation. However, Stretch and Wixted noted that within-test criterion shifts are not necessarily impossible, and speculated that local criterion shifts may be more likely when repeated and non-repeated items are conceptually distinct (1998, p. 1394). In the associative list paradigm, words fall into obvious, well-defined groups (i.e., theme-based lists), and the repetition variable is applied across these groups. This affords semantic distinctiveness between repeated and non-repeated items, which may facilitate local criterion shifts.

In summary, the experiments reported in this article explored whether decreases in false memory due to repetition reflect identification of critical items as non-presented or reflect changes in recognition response criterion. Our experiments also tested whether such criterion changes are applied at a global or a local level. Moreover, our manipulations were designed to determine if a diagnostic recall-to-reject strategy can increase the specificity of criterion shifts.

EXPERIMENT 1

The first experiment explored whether repetition decreases false memory by allowing participants to note the absence of critical items or by inducing changes in decision processes. Participants studied a set of associative lists, completed a recognition test, and then repeated the process twice more with the same lists for a total of three study/test cycles. Each list was studied in each cycle. If list repetition decreases false memory, then the false alarm rate for critical items should decrease across cycles. There were important differences in the way in which items were tested across cycles. On cycle 1, not all studied items were tested and not all critical items were tested. On cycles 2 and 3, some of the tested words had appeared on the previous test(s), whereas other words were tested for the first time in that cycle (i.e., words not tested in cycle 1). This allowed us to explore whether testing items once versus more than once affected people's ability to identify critical items. If repetition decreases false memory by allowing participants to identify words that are missing from the study lists, the absence of a word that previously appeared on a test of the same list should be particularly salient. Thus, it should be easier for participants to avoid false memories for critical items that have been previously tested than for critical items that have not.

Method

Participants. A total of 53 Louisiana State University undergraduates participated in exchange for course extra credit.

Design and materials. The conditions in this experiment represent the incomplete factorial combination of study repetition and prior testing. The design was incomplete because it was impossible to have previously tested items on the first study/test cycle. Each participant completed three replications of the full design. That is, participants completed three sets of three study/test cycles, with different lists in each set. We did this to increase the number of observations in each condition.

Because the following description of stimulus assignment across cycles is complex, a schematic of the experimental design is depicted in Table 1. We developed 18 associative lists by selecting 12 high associates for each of 18 target words in the University of South Florida word association norms (Nelson, McEvoy, & Schreiber, 1998). The lists were divided into three sets of six lists each and we counterbalanced which set was used in the first, second, and third replications across participants. Table 1 represents only one hypothetical replication. Within each replication, lists were assigned to three subsets of two lists each. For a given participant, the critical items from one of these subsets (i.e., Lists 1 and 2 in Table 1) were tested in cycles 1, 2, and 3. Thus, these items were first-time tested critical items in cycle 1, and previously tested critical items in cycles 2 and 3. Critical items from the other two list subsets were tested either only on cycle 2 (i.e., Lists 3 and 4 in Table 1) or only on cycle 3 (i.e., Lists 5 and 6 in Table 1). These items were the first-time tested critical items in their respective cycles. The subset assigned to each condition was counterbalanced across participants.

The 12 words in each associative list were divided into two sets of six items each and we counterbalanced which set was used as targets versus lures. All six targets for each list were studied in each cycle. Within each study set, two items were tested in all three cycles, another two were tested only in cycle 2, and the remaining two were tested only in cycle 3. The six related lures for each theme were assigned to the test cycles in a similar fashion. Assignment of items to testing cycles was counterbalanced across participants.

Procedure. Participants were told that they would see lists of words, and that each list would be repeated three times. In the study phase, words appeared on the screen for 600 ms followed by 200 ms of blank screen. The words from a given list were presented in a blocked format and the order of the lists as well as the order of the items within the lists was randomised anew for each participant and for each study/test cycle. There was no gap between successive lists within a single study phase. Following each individual study

TABLE 1
Schematic depiction of Experiment 1

Cycle 1		*Cycle 2*		*Cycle 3*	
Study	*Test*	*Study*	*Test*	*Study*	*Test*
1–6 List 1	1–2 OLD 7–8 RL CI List 1	1–6 List 1	1–4 OLD 7–10 RL CI List 1	1–6 List 1	1–2, 5–6 OLD 7–8, 11–12 RL CI List 1
1–6 List 2	1–2 OLD 7–8 RL CI List 2	1–6 List 2	1–4 OLD 7–10 RL CI List 2	1-6 List 2	1–2, 5–6 OLD 7–8, 11–12 RL CI List 2
1–6 List 3	1–2 OLD 7–8 RL	1–6 List 3	1–4 OLD 7–10 RL CI List 3	1–6 List 3	1–2, 5–6 OLD 7–8, 11–12 RL
1–6 List 4	1–2 OLD 7–8 RL	1–6 List 4	1–4 OLD 7–10 RL CI List 4	1–6 List 4	1–2, 5–6 OLD 7–8, 11–12 RL
1–6 List 5	1–2 OLD 7–8 RL	1–6 List 5	1–4 OLD 7–10 RL	1–6 List 5	1–2, 5–6 OLD 7–8, 11–12 RL CI List 5
1–6 List 6	1–2 OLD 7–8 RL	1–6 List 6	1–4 OLD 7–10 RL	1–6 List 6	1–2, 5–6 OLD 7–8, 11–12 RL CI List 6

Schematic of target, critical item, and related lure assignments to study and test cycles in Experiment 1.
CI refers to the critical item for a given list; RL refers to a related lure for a given list; Items 1–6 are studied and items 7–12 are related lures.

phase, participants completed maths problems for 30 s before completing the test. For each test, participants were told that they would see words that did and did not appear in the last study phase and that they were to press an ''S'' label (placed on the ''D'' key) if they thought the word was studied or a ''NS'' label (placed on the ''J'' key) if they thought the word was not studied. The recognition test was self-paced. Upon completion of the first test, participants immediately moved on to the second cycle of learning the same group of words and they immediately moved on the third cycle after the second test. After all three study/test cycles, the next replication began. Participants were told that they would repeat the study/test cycles with a new group of words. Once participants completed three replications, they were debriefed and thanked for their participation.

Results and discussion

Throughout the study, we report analyses both on the proportion of items participants claimed to have studied and on signal detection parameters. We calculated adjusted recognition scores in all experiments by adding .5 to the number of items given ''yes'' responses and adding 1 to the total number of items, before dividing the former value by the latter to derive the proportion of ''yes'' responses (Snodgrass & Corwin, 1988). This adjustment eliminates proportions of 0 and 1, which is necessary because signal detection parameters cannot be calculated for these proportions. Although the adjustment biases the absolute value of signal detection parameters (Hautus, 1995), it does not compromise the validity of comparisons across conditions because it is applied to all observations. The proportions reported in all tables and analysed in all inferential tests are the adjusted values. These values were also used to calculate signal detection parameters. We used a Type I error rate of .05 for all statistical comparisons and we controlled the error rate of all follow-up analyses using the Bonferroni adjustment.

The recognition tests in Experiment 1 included three item types, so the appropriate signal detection model had three parameters. The first parameter, d'_c, is the distance between the related lure and critical item distributions, and reflects the amount of mnemonic evidence gained for critical items as a result of the presentation of related words in the study phase. The second parameter, d'_s, is the distance between the related lure and the studied item distributions, and reflects the amount of evidence gained as a result of a word's presentation on the study list. The difference between the d'_s and the d'_c parameters denotes how discriminable studied items are from critical items. The third parameter, lambda, is the position of the response criterion relative to the mean of the related lure distribution. We calculated these three parameters from the hit rate, related lure false alarm rate, and critical item false alarm rate for each participant and for each of the conditions. We compared the values across conditions using traditional inferential statistics (i.e., ANOVAs and *t*-tests).

TABLE 2
Recognition performance in Experiment 1

	Test condition and repetition cycle				
	First test appearance			*Tested before*	
Recognition measure	*1*	*2*	*3*	*2*	*3*
List HR	.72 (.02)	.76 (.02)	.76 (.01)	.84 (.01)	.87 (.01)
Related lure FAR	.20 (.02)	.14 (.02)	.11 (.01)	.36 (.02)	.37 (.02)
Critical item FAR	.46 (.03)	.44 (.03)	.35 (.03)	.57 (.02)	.51 (.03)
d'_s	1.56 (.08)	1.98 (.09)	2.16 (.09)	1.43 (.07)	1.61 (.10)
d'_c	.85 (.08)	1.01 (.09)	.98 (.08)	.53 (.08)	.42 (.08)
Lambda	.95 (.07)	1.21 (.07)	1.39 (.08)	.37 (.05)	.36 (.07)

Recognition performance in Experiment 1 across the study/test cycles for both previously tested items and items making their first test appearance.

HR = hit rate; FAR = false alarm rate. Standard errors are in parentheses.

Repetition effects. Table 2 presents the data for items tested first in cycles 1–3 and for items tested subsequently in cycles 2 and 3. The first set of analyses focused on the effect of repetition across cycles for items that were not repeatedly tested (i.e., the first three columns of Table 2). Repetition of study themes had a modest, but significant, effect on hit rate, $F(2, 104) = 3.41$, $p < .05$, $MSE = .009$. Hit rate increased across cycles, although no pairwise comparisons reached significance. The repetition effect was also significant for related lures, $F(2, 104) = 27.92$, $p < .001$, $MSE = .004$, as false alarms to these items decreased across cycles, with all pairwise comparisons significant. Critical items were also affected by repetition, $F(2, 104) = 5.42$, $p < .01$, $MSE = .031$. The critical item false alarm rate was significantly lower in cycle 3 than in either cycle 1 or cycle 2. Thus, we observed a decrease in false memory when lists were repeated.

Next, we evaluated the effect of repetition on the memory parameters from the signal detection model with a 3×2 ANOVA including repetition and type of memory parameter (d'_s or d'_c) as factors. In this analysis, an interaction indicates a change in the discriminability of studied items from critical items with repetition. The analysis showed a main effect of parameter type, $F(1, 52) = 204.17$, $p < .001$, $MSE = .353$, because d'_s (1.90) was higher than d'_c (.95). There was also a main effect of repetition, $F(2, 104) = 9.89$, $p < .001$, $MSE = .400$, and both effects were qualified by a significant interaction, $F(2, 104) = 11.19$, $p < .001$, $MSE = .128$. We followed up the interaction by conducting separate ANOVAs on the repetition effect for d'_s and d'_c. Repetition significantly increased d'_s across cycles, $F(2, 104) = 29.45$, $p < .001$, $MSE = .169$. There was a significant increase from cycle 1 to cycle 2 and a nominal, but non-significant, increase from cycle 2 to cycle 3. Repetition produced higher values of d'_c in cycles 2 and 3 as compared to cycle 1, but the effect was not significant, $F(2, 104) = 1.16$, *ns*, $MSE = .360$. Thus, repetition led to greater increases in memory evidence for studied items than for critical items (see Arndt & Hirshman, 1998, for similar effects resulting from increased encoding time). In symmetry with d'_s, the lambda parameter also rose steadily across cycles. An ANOVA showed that this effect was significant, $F(2, 104) = 23.04$, $p < .001$, $MSE = .114$, and all pairwise comparisons were significantly different. Thus, repetition resulted in not only increased memory for list items, but also in a higher response criterion. This higher criterion explains the reduction in false memory to once-tested critical items across test cycles. It also explains why the increase in hit rates to once-tested items was so modest across cycles.

Prior testing effects. A second set of analyses focused on the effects of prior testing on positive recognition responses. We compared previously tested versus first appearance items for cycles 2 and 3 (i.e., the last four columns in Table 2). Because this variable was designed to facilitate the process of realising that critical items were absent in their lists, we analysed only the critical item false alarm rate. The prior-testing effect was significant for cycle 2, $t(52) = 3.30$, $p < .01$, and for cycle 3, $t(52) = 5.31$, $p < .001$. In both cases people were *more* likely to false alarm to critical items that were previously tested.

Summary. The overall pattern of results from Experiment 1 is quite clear: study repetition increased participants' memory for studied items and concurrently increased the amount of evidence that participants required to claim that an item was studied. The increase in response criterion led to an overall decrease in false memory with repetition. This was true both for critical items and for related lures tested for the first time across test cycles. The results also indicated that the combined study and test repetition of some critical items did not help participants realise that those critical items were absent from the study lists. If participants were going through a process of figuring out which words were missing from the lists, this process should have been facilitated for items that repeatedly appeared on tests but not in the study lists, leading to a reduction in false memory for these items. In fact, false memory was inflated for items that were repeatedly tested compared to items tested only once. Thus, prior testing did not help people realise that lures were absent from the study list. Prior testing only made lures more familiar.

EXPERIMENT 2

In Experiment 1, participants reduced false memory for repeated lists by changing the decision processes that they used to evaluate the information stored in memory. More specifically, repetition induced an increase in the

TABLE 3
Schematic depiction of Experiment 2

Cycle 1		*Cycle 2*		*Cycle 3*	
Study	*Test*	*Study*	*Test*	*Study*	*Test*
1–6 List 1		1–6 List 1		1–6 List 1	1–6 OLD 7–12 RL CI List 1
1–6 List 2		1–6 List 2		1–6 List 2	1–6 OLD 7–12 RL CI List 2
1–6 List 3	1–6 OLD 7–12 RL CI List 3				
1–6 List 4	1–6 OLD 7–12 RL CI List 4				
		1–6 List 5	1–6 OLD 7–12 RL CI List 5		
		1–6 List 6	1–6 OLD 7–12 RL CI List 6		
				1–6 List 7	1–6 OLD 7–12 RL CI List 7
				1–6 List 8	1–6 OLD 7–12 RL CI List 8

Schematic of target, critical item, and related lure assignments to study and test cycles in Experiment 2. CI refers to the critical item for a given list; RL refers to a related lure for a given list; Items 1–6 are studied and items 7–12 are related lures.

recognition response criterion that decreased the frequency with which participants claimed to have studied critical items and related lures. An important aspect of Experiment 1 is that repeatedly studied items appeared on test cycles 2 and 3 separate from non-repeated items (i.e., test cycle 1). Thus, global shifts in criterion were sufficient to reduce false memory because the shifts occurred across tests. In Experiment 2, we examined whether participants can make local, item-by-item criterion shifts when repeated and non-repeated items are mixed in the same test. As in Experiment 1, participants completed three study/test cycles for a given set of lists. However, in cycle 3 of this experiment, only some of the studied lists had been presented in the previous cycles; other lists appeared for the first and only time in cycle 3. Thus, the test in cycle 3 included items from both the repeated and non-repeated lists. If participants make item-based criterion shifts, the false memory data should look the same as in Experiment 1. That is, there should be fewer false memories for critical items from repeated lists. If participants apply a single global criterion to all test items, there will be no reduction in false memory. In fact, the nominal increase of critical item d′ in the first experiment suggests that false memory may actually increase with repetition if repeated and non-repeated items are evaluated with the same global response criterion.

Method

Participants. A total of 63 Louisiana State University undergraduates participated to earn course extra credit.

Design, materials, and procedure. As in Experiment 1, participants completed three study/test cycles within each of three replications. However, only some associative lists repeated across cycles, while others were studied in only one cycle. Lists studied in only one cycle were tested in that cycle and never again. Lists presented in all three cycles were not tested until the third cycle. The key comparison involved the difference in performance for repeated and non-repeated lists tested in the third cycle. The first two study/test cycles were included to make this experiment comparable to Experiment 1.

Table 3 depicts a schematic for a hypothetical replication. We added 6 associative lists to the 18 used in Experiment 1. The 24 lists were divided into three sets of eight lists each, to counterbalance which lists appeared in each replication. Each replication set was further divided into four subsets of two lists each. One subset (i.e., Lists 1 and 2 in Table 3) appeared in all study cycles and was tested only in cycle 3. The other three subsets were studied and tested in cycle 1, 2, or 3 (i.e., the remaining lists in Table 3). We counterbalanced the assignment of list subsets to conditions across participants.

As in Experiment 1, each list was divided into two sets of six items that served as targets or related lures in a counterbalanced fashion. In cycle 1, participants studied four associative lists and completed a test with the 12 targets, 12 related lures, and 2 critical items from the once-presented subset (i.e., Lists 3 and 4 in Table 3). In cycle 2, participants again studied the non-tested associative lists from cycle 1 plus two new lists. The test in cycle 2 involved the 12 targets, 12 related lures, and 2 critical items from the new lists. In cycle 3, participants again studied the two non-tested lists from cycles 1 and 2 along with two other lists appearing for the first time. The test in cycle 3 included items from each list studied in that cycle, for a total of 24 targets, 24 lures, and 4 critical items. All procedural details matched those of Experiment 1.

Results and discussion

Effects of repetition. The analyses of interest compared repeated to non-repeated lists in cycle 3, and the relevant data appear in Table 4. The hit rate was higher for repeated than non-repeated lists, $t(62) = 9.68$, $p < .001$, and repetition also increased the false alarm rate for critical items, $t(62) = 2.61$, $p < .05$. The false alarm rate for related lures did not differ for repeated and non-repeated lists, $t(62) = 0.41$, *ns*. Thus, in contrast to Experiment 1, there was no evidence that repetition decreased false memory. Repetition actually increased false memory for critical items.

Concerning the signal detection parameters, we first analysed how repetition affected memory parameters with a 2 (repetition) $\times$ 2 (type of d′) ANOVA. This analysis revealed a main effect of type, $F(1, 62) = 132.76$, $p < .001$, $MSE = .316$, which arose because d′_s was greater than d′_c. There was also a significant effect of repetition, $F(1, 62) = 26.49$, $p < .001$, $MSE = .361$, and a significant interaction, $F(1, 62) = 11.87$, $p < .001$, $MSE = .151$. The interaction arose because repetition increased d′_s to a greater extent than d′_c, although both showed a significant increase with repetition, $t(62) = 7.67$, $p < .001$ and $t(62) = 2.11$, $p < .05$, respectively. In contrast to the first experiment, there was no effect of repetition on lambda, $t(62) = 0.57$, *ns*, indicating that participants used the same response criterion for repeated and non-repeated lists.

TABLE 4
Recognition performance in Experiment 2

Recognition measure	*Test condition*	
	Single	*Repeated*
List HR	.61 (.02)	.79 (.02)
Related lure FAR	.15 (.01)	.16 (.01)
Critical item FAR	.38 (.03)	.47 (.03)
d′_s	1.42 (.07)	1.98 (.07)
d′_c	.77 (.08)	.99 (.08)
Lambda	1.12 (.06)	1.08 (.05)

Recognition performance for repeated and non-repeated lists in the third repetition cycle of Experiment 2.

HR = hit rate; FAR = false alarm rate. Standard errors are in parentheses.

Summary. Repetition increased false memory for critical items in this experiment, a result that differs markedly from the decrease in false memory observed for repeated lists in the first experiment. The signal detection analyses reveal the critical difference between the two experiments. The effect of repetition on memory evidence in this experiment was very similar to the effect in Experiment 1. Repetition increased memory evidence for both critical items and studied items, with the increase for studied items more pronounced. However, the effect of repetition on the response criterion differed in the two experiments. In Experiment 1, the criterion moved higher as lists were repeated. In Experiment 2, there was no difference in criterion for repeated and non-repeated lists. These results show that our participants set a recognition criterion at a global level and did not make item-by-item shifts in criterion. Therefore, participants in this experiment could not (or did not) capitalise on the additional information available for repeated lists.

EXPERIMENT 3

In Experiment 3, we included a test condition that should make it easier for participants to identify the studied items from thrice-presented lists, which may encourage them to change decision processes on a more specific, item-by-item basis. We required some of the participants to recall a studied word related to the test word before they decided whether the test word was studied. This introduces the possibility that participants will use their memory for the related word to inform their response criterion for the test word. For example, if participants have a strong memory for the recalled word, they may require a great deal of memory evidence in order to claim that the test word was also studied. This diagnostic recall-to-reject strategy (Gallo, 2004) would allow participants to use a higher response criterion for repeated lists even if they are tested along with non-repeated items. Thus, forcing participants to recall related items before their recognition decision may lead to less false memory for repeated lists.

Gallo (2004) used a similar strategy to investigate diagnostic recall-to-reject in false recognition memory. He instructed one group of participants to recall as many words related to a test word as they could and to use this information to help them reject lures on the test. He compared this group to participants who received standard recognition instructions. The recall-to-reject instructions only reduced false memory when all of the related items could be recalled (disqualifying recall-to-reject). Recalling only one or a few list items did not induce a change in recognition decision processes. However, Gallo speculated that diagnostic recall-to-reject was not effective because recalling related list items did not yield any information about the test word that could be used to strategically shift response criterion. When some lists are repeated, recall can reveal an important piece of information about the test word, namely, the number of times that it would have been presented. Thus, diagnostic recall-to-reject could potentially play a larger role in this experiment than in Gallo's.

Method

Participants. A total of 77 Louisiana State University undergraduates participated for course extra credit. Of these participants, 34 were assigned to the no recall group, and 33 were assigned to the recall group.

Design and materials. Experiment 3 involves the factorial combination of repetition (lists presented once versus three times) and recall (participants do or do not recall a related word prior to the recognition decision). Repetition was manipulated within subjects and recall was manipulated between subjects. Participants completed two study/test replications in this experiment to increase the number of observations in each within-subject condition.

We used the same 24 lists from Experiment 2. For this experiment, lists consisted of only the first six associates to each critical item. Lists were first divided into two sets of 12 to counterbalance the lists used in each replication. Half of the 12 lists in each replication were presented three times, and the others were presented only once. We counterbalanced which lists appeared in each condition. Also, half of the lists did not contain the critical item, and in the other half the critical item replaced the sixth associate in the study phase. The test for each replication included only the critical items from each of the studied lists. As noted earlier, half of these critical items had been presented with their lists (targets) and the other half had not (lures). We chose to construct the test this way because each list theme is tested only once, eliminating the possibility that participants would be tested on a word that they earlier recalled. Also, because all the test words bear the same strong associative relationship to other list words, and because the items used as targets and lures are fully counterbalanced, targets and lures will be equally effective retrieval cues for the recall of related studied items.

Procedure. Participants were told that they would see lists of related words and that some of the lists would be presented multiple times. In the first replication, participants studied 12 associative lists of six items each and half of these lists were repeated three times. Each word remained on the screen for 800 ms, followed by 200 ms of blank screen. The order of the lists, as well as the order of items within each list, was randomised

TABLE 5
Recognition performance in Experiment 3

	Recall and repetition conditions			
	No recall		*Recall*	
Recognition measure	*Single*	*Repeated*	*Single*	*Repeated*
List HR	.50 (.04)	.77 (.03)	.49 (.04)	.76 (.03)
Critical item FAR	.31 (.03)	.45 (.04)	.31 (.03)	.47 (.04)
d′	.56 (.12)	1.04 (.12)	.54 (.12)	.90 (.12)
Lambda	.57 (.10)	.17 (.11)	.57 (.10)	.09 (.11)

Recognition performance in Experiment 3 for lists presented one or three times and for participants who did or did not recall related items preceding their recognition decisions.
HR = hit rate; FAR = false alarm rate. Standard errors are in parentheses.

for each participant. The number of intervening lists that occurred before a repeated list was re-presented was also random. After all lists were presented, participants completed maths problems for 30 s and were then given the test instructions. Participants assigned to the recall group were told that they would see test words related to the words they studied, and that some of the test words were themselves studied and some were not. They were told to recall a related word that they studied and then decide whether or not the test word was previously studied. During the test, participants in the recall group first saw the test word with instructions to type in a related word that they studied. They typed in their response using the keyboard and pressed the enter key when they were done. Following recall, they were re-presented with the test word and instructed to decide whether the word was previously studied. They made "yes" responses by pressing the "1" key and "no" responses by pressing the "2" key. Participants in the no recall group only made the recognition decision for each test word. Following the test, participants began the second replication, which was identical to the first except that a new set of lists was used.

Results and discussion

Recall and repetition effects. We conducted a series of 2 (repetition) × 2 (recall) ANOVAs for each stimulus type, and Table 5 displays the relevant data. The ANOVA on the hit rate for critical item targets showed an effect of repetition, $F(1, 65) = 78.10$, $p < .001$, $MSE = .031$. Participants recognised more targets from repeated lists. Neither the effect of recall nor the interaction of recall with repetition reached significance, $F(1, 65) = 0.03$, *ns*, $MSE = .051$, and $F(1, 65) = 0.01$, *ns*, respectively. Regarding the false alarm rate for critical item lures, there was a significant effect of repetition, $F(1, 65) = 33.11$, $p < .001$, $MSE = .022$, reflecting more false memory for repeated lists. The recall effect was not significant, $F(1, 65) = 0.08$, *ns*, $MSE = .055$, and recall and repetition did not interact, $F(1, 65) = 0.11$, *ns*. The fact that the effects of repetition were identical for participants who recalled related words before their recognition decisions and for participants who did not, demonstrates that recalling related words did not lead participants to adjust recognition decision processes on an item-by-item basis.

The signal detection model for this experiment differs from the previous experiments because there were only two item types on the test—critical item targets and critical item lures. Therefore, the model involves only two parameters: d′ (the distance between the target and lure distributions) and lambda (the placement of the criterion relative to the lure distribution). The analysis on d′ revealed a significant effect of repetition, $F(1, 65) = 11.13$, $p < .001$, $MSE = .527$. Targets were more discriminable from lures when lists were repeated. Recall had no significant main effect, $F(1, 65) = 0.40$, *ns*, $MSE = .487$, and did not interact with repetition, $F(1, 65) = 0.21$, *ns*. Repetition had a significant effect on lambda, $F(1, 65) = 29.08$, $p < .001$, $MSE = .228$. Neither recall, nor the interaction of recall with repetition, significantly affected lambda, $F(1, 65) = 0.09$, *ns*, $MSE = .538$, and $F(1, 65) = 0.24$, *ns*, respectively.

Inspection of Table 5 suggests that the response criterion was *lower* for repeated lists. However, this parameter must be interpreted with caution in the present experiment, because it is measured in relation to the position of the critical lure distribution. The position of the lure distribution also affects the d′ estimate. Therefore, a lower lambda can mean either that participants require less evidence to make positive recognition responses or that the critical lure distribution has shifted up. The results of the first two experiments can discriminate these possibilities with regard to critical item lures, because the tests included related lures, allowing the model to keep track of the movement of the critical item distribution independent of the criterion. Together, the results of Experiments 1 and 2 show that the critical item lure distribution for repeated lists is higher than the distribution for lists studied only once. Thus, it is most reasonable to assume that the lower criterion for repeated lists in Experiment 3 results from a shift in the critical item lure distribution, not a change in the amount of evidence participants required to claim that an item was studied.

Summary. As in the first two experiments, repetition increased the discriminability of studied information from critical item lures. Results were also consistent with Experiment 2, in that they showed that list repetition increased false memory for critical item lures when repeated and non-repeated lists are mixed on the same test. This increase in false memory indicates that participants did not make item-by-item shifts in decision processes. Rather, people applied decision processes uniformly throughout the test. Results also indicated that forcing the recall of related items before recognition did not induce a diagnostic recall-to-reject strategy guiding item-by-item criterion shifts. In all cases, the effect of repetition on recognition performance did not differ between participants who recalled before each recognition decision and those who did not.

GENERAL DISCUSSION

In the current study we explored potential false memory editing mechanisms for repeated associative lists. Identifying critical items as non-presented played no role in false memory decrements with repetition. When people were provided the opportunity to study a list after its critical item had already been tested in Experiment 1, they were no better at avoiding false alarms. Indeed, the added familiarity associated with repeated testing actually made people more likely to false alarm to previously tested critical items than to critical items tested for the first time. If participants were noting the absence of words related to the list associates, the absence of a word that repeatedly needs to be classified as studied or non-studied would be particularly relevant and notable. Moreover, there is no reason why an identification strategy would be less effective when repeated and non-repeated words appeared on the same versus a different test. Yet we observed decreases in false memory only when repeated and non-repeated items were tested separately.

The patterns of true and false memory observed across our experiments can be explained well by a global conservative criterion shift for repeated lists. In each experiment, repetition led to larger increases in target familiarity than in critical item familiarity, although the increase for critical items was always present and was significant in Experiments 2 and 3. When separate criteria could be used for repeated and non-repeated lists (i.e., when they appeared on separate tests in Experiment 1), the criterion for repeated lists moved up along with the target distribution, resulting in a decrease in false memory for critical items. When the same criterion was used for repeated and non-repeated lists (i.e., when they were mixed on the same test in Experiments 2 and 3), the increase in critical item familiarity due to list repetition resulted in greater false memory for critical items from repeated lists. Participants did not make local criterion shifts, even in Experiment 3 when the requirements of the test were tailored to assist identification and differentiation of repeated versus non-repeated study items. Participants could have realised that lures were from repeated lists by recalling a related study word and evaluating the memory strength of the recalled word. We gave participants the opportunity to use this strategy by forcing them to recall a related item before they made each recognition decision. If strong evidence was stored for the recalled item, then participants could have inferred that it came from a repeated list. Remarkably, these participants still used a single criterion throughout the test and their recognition performance was no different from participants who did not recall related words. Our results are in agreement with those of Gallo (2004) by showing that even when participants recall other studied words before a recognition decision, they do not use this information to strategically change recognition decision processes within a test.

Comparison with other studies

In this section, we compare and contrast our results with those of other studies that manipulated repetition within-subjects in the associative list paradigm. Kensinger and Schacter (1999) and Schacter et al. (1998) conducted studies comparing the effect of repetition on false memory for normal young adults versus older participants or versus certain types of amnesic patients, respectively. Their repetition procedure was similar to our first experiment in that participants completed multiple study/test cycles on the same associative lists, meaning that items from each level of the repetition variable were evaluated on separate tests. Just as in Experiment 1, these researchers found a steady decrease in false recognition for critical items across study/test cycles for their normal young adults (the group most comparable to our participants). These studies reinforce the conclusion that participants can reduce false memory when they have a chance to set separate global criteria for repeated and non-repeated items.

Seamon et al. (2002) reported a repetition study in which repeated and non-repeated associative lists were included on the same recognition test. These researchers presented associative lists in various repetition conditions (1, 5, or 10 presentations in their first experiment, and 1, 5, or 25 presentations in their second experiment) and different groups of participants studied each item for either 20 ms or 2 s at encoding. In the 20 ms encoding condition, false recognition of critical items increased steadily across the levels of the repetition variable. In the 2 s encoding condition, false recognition increased from 1 to 5 presentations, but decreased from 5 to 10 (Experiment 1) or 5 to 25 (Experiment 2) presentations. Thus, this study also found an increase in false recognition in the repetition condition that was most similar to our own (five presentations versus our three presentations). Interestingly, Seamon et al.'s results suggest that we may have observed within-test reductions in false memory had we repeated our lists a greater number of times. However, Experiment 1 here demonstrated that three presentations were sufficient to reduce false memory as long as participants could establish a separate criterion for thrice-presented and once-presented lists. If participants had flexibly adjusted their criterion on an item-by-item basis, we would have observed decreases in false memory for repeated lists in Experiments 2 and 3.

Benjamin (2001) also conducted a study where lists were presented either one or three times with repeated and non-repeated items on the same test, but his results contradict our results and those of Seamon et al.'s (2002) five-presentation conditions. In Benjamin's study, there was less false recognition for repeated lists as compared to lists presented once. Although his procedure was most similar to our Experiments 2 and 3, his results are most similar to our Experiment 1. Thus, it would seem that Benjamin's participants made local criterion shifts, whereas our participants made only global criterion shifts. The reason for this discrepancy is not immediately apparent, although it is true that our Experiments 2 and 3 differed from Benjamin's study in a number of ways, the most important of which was encoding time. Our encoding times were less than 1 second, whereas Benjamin's was 3 seconds.

However, it is unlikely that procedural differences fully account for the differences in results. In the Appendix, we report data from three experiments conducted as part of a separate project that motivated the present study. In each experiment, there was a single encoding phase in which some associative lists were presented once and others were presented three times. The specific procedural details are reported in the Appendix, but the reader should note that each experiment reflects a methodology that is progressively more similar to Benjamin's (2001). Indeed, Experiment A3 involved associative lists of a similar length (10 words per list compared to Benjamin's 8 words per list), each word was studied for a similar amount of time (2.5 s compared to Benjamin's 3 s), and list repetitions were presented back-to-back just as in Benjamin's study. Note that in every experiment false recognition was more likely for critical items from repeated lists.

These results are remarkable not only because the experiments were similar to Benjamin's (2001), but also because the experiments included an additional variable designed to assist participants in making local criterion shifts. In all experiments we presented repeated lists in red font and non-repeated lists in green font. In the colour test condition, all items for repeated lists were again presented in red and all items from non-repeated lists were again presented in green (cf. Stretch & Wixted, 1998). Critical items were also tested in the colour corresponding to their respective study lists. Thus, participants could be sure that all red test items would have been presented three times if they were studied, which gave them all the information needed to apply a separate criterion for repeated items. Nevertheless, we observed the same increase in false memory in the colour test condition as when the repetition status of test words was not cued by a test colour (i.e., when all test words appeared white against a black background).

To summarise, Benjamin reported two experiments in which, within a single test, repetition led to decreases in false memory. The present study reports five experiments in which repetition increased false memory (i.e., Experiments 2, 3, A1, A2, and A3). Seamon et al. (2002) reported two more experiments with the same result (at least in the repetition condition most comparable to Benjamin's and ours). Therefore, the overall evidence indicates that within-test criterion shifts are the exception, rather than the rule. What factors determine when participants will make local criterion shifts (e.g., number of list presentations, encoding time, blocking of lists) must remain an issue for future studies to address.

Nature and flexibility of decision process changes

Throughout the study, we have assumed for simplicity that repetition induces *quantitative* changes in recognition decision processes, namely that the overall response criterion becomes more conservative. People can also make *qualitative* changes in decision processes by altering the types of mnemonic information that are considered as evidence that an item was studied (Johnson, Hashtroudi, & Lindsay, 1993). People could change the types of retrieved evidence that are considered relevant to the recognition decision (Johnson et al., 1993) or change the types of information that are retrieved from memory by controlling internally reinstated retrieval cues (Humphreys, Dennis, Maguire, Reynolds, Bolland, & Hughes, 2003; Ratcliff, Van Zandt, & McKoon, 1995).

Qualitative changes in decision processes can reduce false memories if participants switch from relying on non-discriminative details such as conceptual information to discriminative details such as perceptual information (e.g., Hicks & Marsh, 1999). There is also evidence that people might spontaneously engage reality-monitoring mechanisms to reduce false memories. Dehon and Brédart (2004), for example, showed that young adults sometimes identify non-recalled critical items as having occurred to them internally during learning or retrieval, but not from the external source of list presentation (see also Brédart, 2000). Therefore, they reported purposefully withholding such responses during free recall.[2] Whether this sort of editing strategy generalises from free recall to recognition is unknown, but in this spirit, Multhaup and Connor (2002) showed that false recognition decreased when participants were given both a strong warning about the false memory phenomenon *and* the option to identify tested critical items as having not been presented but conceptually tying list words together. We manipulated variables in our study (e.g., repetition) that should affect the degree to which noticing or reality-monitoring processes could serve as a basis for critical item rejection in a standard recognition test. Yet there was no evidence for these processes in our study.

Repetition may induce such qualitative changes in decision processes by strengthening discriminative details for studied items, thus encouraging participants to rely more heavily on these details in their recognition decisions. The present experiments did not discriminate qualitative from quantitative decision process changes, and either version could explain reductions in false memory when people use different decision processes for repeated and non-repeated lists.

[2] The Brédart (2000) and Dehon and Brédart (2004) studies do not actually indicate whether participants realised the source of internally activated critical items during learning or during retrieval. Therefore, it is unclear whether people are truly noticing the absence of critical items during list presentation or realising their true source (and withholding recall) only during the recall phase.

However, the present experiments demonstrate that it is not the *nature* as quantitative or qualitative, but the *specificity*, of changes in decision processes that critically determines the presence or absence of a decrease in false memory induced by repetition.

The results of the current study echo those of Stretch and Wixted (1998) by demonstrating that participants either cannot or will not make local criterion shifts, even when the test procedure explicitly promotes selectivity. This strict adherence to a global criterion in recognition seems to contradict the flexibility that participants demonstrate in other retrieval tasks, such as source-monitoring decisions. In a source-monitoring task, participants are asked not only to decide if a given item was presented, but also to discriminate among various presentation formats they may have experienced for the item (e.g., a male or female voice). As mentioned earlier, Multhaup and Connor (2002) have identified conditions in which false recognition was reduced with a very stringent form of source-monitoring test. Source tasks often require that each item on the test be evaluated in terms of at least two separate criteria (Johnson et al., 1993). In tasks where sources differ in memorability, participants apparently set a higher criterion for the amount of evidence they need to claim the high-memorability source, and a lower criterion for the amount of evidence they need to claim the low-memorability source (e.g., Marsh & Bower, 1993). For example, when participants study items as words or pictures, they are more likely to claim that new items were studied as words than they are to claim that new items were studied as pictures, which is consistent with the use of a separate "picture" criterion that is set at a higher value than the "word" criterion (e.g., Gallo, Weiss, & Schacter, 2004). The fact that participants perform well on source-monitoring tasks suggests that they are perfectly capable of applying separate criteria on the same test. Yet this flexibility is difficult to achieve on a yes/no recognition test. It will be interesting for future research to explore the differences in memory processes on recognition and source-monitoring tasks that account for participants' ability to be much more flexible when deciding *how* an item was presented versus deciding *if* an item was presented.

One process or two?

Some may fault us for appealing to a single-process familiarity framework rather than appealing to dual-process principles in recognition. Most explanations of cases in which false recognition decreased due to repetition appeal to dual-process frameworks (e.g., Benjamin, 2001; Seamon et al., 2002). Others have also noted that the general consequence of repeating items at study is to increase both the familiarity *and* the recollection of those items (e.g., Jacoby, Jones, & Dolan, 1998). The corollary is that new items on recognition tests may often be rejected because of the lack of an expected level of recollection. Distinctiveness effects in false recognition are consistent with this view. When DRM lists are studied as pictures, false recognition of critical items tends to be lower as compared to cases in which DRM lists are studied as words (e.g., Israel & Schacter, 1997; Schacter, Israel, & Racine, 1999). In a similar manner, people in our study might have expected more distinctive recollections from thrice-presented lists as opposed to once-presented lists, having the potential effect of decreasing false recognition of critical items (Gallo, 2004).

One response to this criticism is that criterion setting in the SDT model is a descriptive way to characterise the manner in which false recognition might decrease. The mere use of an SDT model does not explain whether that change in criterion placement would have been caused by the expectation of distinctive recollections or by some other process. How a decision criterion is set might be determined by various factors, but that consideration could be separate from the type and amount of evidence one evaluates against such a criterion. The most important point we wish to make in this study is that it is possible to use a familiarity-based model to predict reductions in false recognition. That is, lower false recognition following the repetition of some DRM lists does not necessarily compel a dual-process explanation. As discussed in the Introduction, analysis of Figure 1 makes it apparent that the placement of the decision criterion in a SDT model has a powerful impact on the rate of false recognition. Any process, whether familiarity-based or recollection-based, that shifts the criterion as depicted at the bottom of Figure 1 could cause a reduction (or increase) in false recognition. Similarly, retrieval control factors such as time pressure during retrieval or the aging process (Benjamin, 2001) could viably affect criterion-setting mechanisms as opposed to, or in addition to, changing the basis for decisions from familiarity to recollection (cf. Hicks & Marsh, 2000). In this spirit, Hirshman (1995) discussed a range model of criterion setting within a SDT framework in which people recall the strongest few items from a study session to help establish the upper limit on the range of expected retrieval experiences. The range model is very similar to using expected recollection from items in thrice-presented lists to set the criterion for test items consistent with the themes of those lists. Examining other characteristics of recognition responses, such as ROC curves across confidence ratings or across payoff conditions, might allow for examination of whether a second, threshold-like process is operating in addition to a traditional familiarity-based process (see Lampinen, Odegard, & Neuschatz, 2004, for a recent example in the conjunction memory literature). One notable application of SDT to false recognition in the DRM paradigm, in fact, showed that ROCs for critical items reflected a continuous process, rather than a threshold-like process (Westerberg & Marsolek, 2003). ROC analysis would be particularly informative in addressing the effects of study repetition on false recognition.

Conclusions

We found that repetition of associative list themes had a different effect on true and false recognition. In each experiment, repetition led to a greater increase in discriminability of targets from lures as compared to discriminability of critical items from lures. Thus, repetition did not increase false memory as much as it increased true memory. However, in five reported experiments false memory increased due to repetition when tested items from repeated and non-repeated lists were mixed on the same test (Experiments 2, 3, A1, A2, and A3 herein). This result contrasts with that of others (e.g., Benjamin, 2001; Seamon et al., 2002) who showed that repetition decreased false memory. Only when repeated versus non-repeated lists were tested separately in Experiment 1 (i.e., first-tested critical items on cycle 1 vs cycle 3) did repeatedly studying associative lists reduce false recognition. This reduction was due to a conservative criterion shift across test cycles. At the same time, the results of Experiment 1 also showed that repeatedly *testing* critical items across study/test cycles increased false memory, indicating that people were not identifying the absence of critical items during encoding. Overall, our pattern of results is consistent with Stretch and Wixted's (1998) claim that people are either reluctant or unable to shift their recognition criterion based on different categories of stimuli mixed on the same test.

REFERENCES

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

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., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language*, *34*, 157–185.

Brédart, S. (2000). When false memories do not occur: Not thinking of the lure or remembering that it was not heard? *Memory*, *8*, 123–128.

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

Dehon, H., & Brédart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology & Aging*, *19*, 191–197.

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 and Language*, *51*, 473–493.

Hall, J. W., & Kozloff, E. E. (1970). False recognitions as a function of number of presentations. *American Journal of Psychology*, *83*, 272–279.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments Computers*, *27*, 46–51.

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

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

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts in the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 302–313.

Humphreys, M. S., Dennis, S., Maguire, A. M., Reynolds., K., Bolland, S. W., & Hughes, J. D. (2003). What you get out of memory depends on the question you ask. *Journal of Experimental Psychology: Learning, Memory Cognition*, *29*, 797–812.

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.

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

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology*, *16*, 399–415.

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 332–342.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 673–688.

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://www.usf.edu/FreeAssociation/

Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: The role of identifiability. *Journal of Experiment Psychology Learning, Memory and Cognition*, *29*, 35–40.

Pastore, R. E., Crawley, E. J., Berens, M. S., & Skelly, M. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review*, *10*, 556–569.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single process theories, and recognition memory. *Journal of Experimental Psychology: General*, *124*, 353–374.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.

Roediger, H. L. III (1996). Memory illusions. *Journal of Memory and Language*, *35*, 76–100.

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

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

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., & Racine, C. (1998). When true recognition suppresses false recognition: Evidence from amnesic patients. *Journal of Cognitive Nueroscience*, *10*, 668–679.

Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language*, *46*, 323–340.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.

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. (1983). *Elements of episodic memory*. London: Oxford University Press.

Tussing, A. A., & Greene, R. L. (1997). False recognition of associates: How robust is the effect? *Psychonomic Bulletin & Review*, *4*, 572–576.

Tussing, A. A., & Greene, R. L. (1999). Differential effects of repetition on true and false recognition. *Journal of Memory and Language*, *40*, 520–533.

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

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, *107*, 368–376.

APPENDIX A

In this appendix we present three experiments demonstrating an increase in false memory for repeated lists using procedures that are similar to the one used by Benjamin (2001). For each experiment, there is a brief description of the important methodological details followed by statistical analyses of repetition effects. Table A1 displays the data for all three experiments. Although the recognition tests for each experiment involved a number of item types, the Table presents data for only studied items, related lures, and critical item lures. We limited the Table to these three item types because they correspond to the item types reported in Experiments 1, 2, and 3. For brevity, we report statistical analyses only on the critical item false alarm rate for each experiment, but readers are encourage to contact the corresponding author for statistical results for other item types. However, the pattern of data in Table A1 is relatively obvious.

Experiment A1

This experiment involved 49 participants, 25 in the colour test group and 24 in the black/white test group. All participants were University of Georgia at Athens undergraduates who participated to fulfil a course requirement. Participants studied six associative lists of 10 items each, as well as 60 unrelated targets that had no conceptual grouping and no relationship to the associative lists. Items from half of the associative lists and half of the unrelated target items were presented three times, and the others were presented only once. Repeated items were presented in red font colour and non-repeated items were presented in green. Studied words appeared in a random order, which means that associative lists were not blocked by theme. Each word remained on the screen for 2.5 s at encoding. After the study phase, participants worked on maths problems for 5 minutes and then completed a recognition test. The test consisted of 30 unrelated item targets, 30 associative list targets (5 from each list), 6 critical items (one from each associative list), 6 related lures (one non-presented list word from each associative list), and 48 unrelated lures. Overall, there were 60 targets and 60 lures. For participants in the colour test condition, all repeated targets were tested in red, and all non-repeated targets were tested in green. In addition, the critical items and related lures from associative lists were tested in the colour corresponding to the targets from the same list. Unrelated lures were randomly assigned to the red or green test colour in equal numbers. In the black/white test condition, all test words were presented in white on a black background, regardless of repetition status.

A 2 (repetition) $\times$ 2 (test group) ANOVA showed a significant effect of repetition, F(1,

TABLE A1
Recognition performance for lists presented one or three times in each test condition

	Test condition and repetition status			
	Black/white test		*Colour test*	
Recognition measure	*Single*	*Repeated*	*Single*	*Repeated*
Experiment A1				
List HR	.61 (.03)	.86 (.03)	.65 (.04)	.89 (.03)
Related lure FAR	.13 (.04)	.18 (.05)	.19 (.05)	.11 (.04)
Critical item FAR	.44 (.06)	.58 (.06)	.45 (.06)	.60 (.06)
Experiment A2				
List HR	.70 (.04)	.92 (.02)	.78 (.03)	.92 (.02)
Related lure FAR	.26 (.06)	.19 (.05)	.16 (.05)	.19 (.06)
Critical item FAR	.74 (.06)	.83 (.06)	.72 (.06)	.84 (.05)
Experiment A3				
List HR	.67 (.04)	.86 (.03)	.71 (.04)	.91 (.02)
Related lure FAR	.19 (.03)	.19 (.03)	.20 (.04)	.19 (.03)
Critical item FAR	.68 (.05)	.75 (.07)	.76 (.05)	.82 (.05)

HR = hit rate; FAR = false alarm rate; Standard errors are in parentheses.

47) = 7.82, $p < .01$, $MSE = .064$, showing that there were more false memories for repeated (.59) than non-repeated (.45) lists. Neither the effect of condition nor the interaction was significant, $F(1, 47) = 0.03$, *ns*, $MSE = .129$, and $F(1, 47) = .01$, *ns*, respectively.

Experiment A2

A total of 37 participants from the same pool as Experiment A1 were included in this experiment, with 19 assigned to the colour test group and 18 assigned to the black/white test group. The methods matched those of Experiment A1, except that items from a given associative list were presented together in a blocked fashion, and repeated lists were presented three times in immediate succession. These changes make this study's procedure more similar to the one used by Benjamin (2001).

A 2 (repetition) × 2 (test group) ANOVA showed a significant effect of repetition, $F(1, 35) = 4.50$, $p < .05$, $MSE = .048$, showing that there were more false memories for repeated (.84) than non-repeated (.73) lists. Neither the effect of condition nor the interaction was significant, $F(1, 35) = 0.01$, *ns*, $MSE = .088$, and $F(1, 35) = 0.09$, *ns*, respectively.

Experiment A3

A total of 34 additional students participated in this experiment: 17 in the colour test group and 17 in the black/white test group. The methods used matched those of the second experiment, except that we removed all unrelated items from the study and test phases. Again, this was done to make this experiment more analogous to those conducted by Benjamin (2001).

A 2 (repetition) × 2 (test group) ANOVA found no significant effect of repetition, $F(1, 32) = 2.08$, *ns*, $MSE = .033$. Although the repetition effect failed to reach significance, results matched those of Experiments A1 and A2 in that there were more false memories for repeated (.78) than non-repeated (.72) lists. Neither the effect of condition nor the interaction was significant, $F(1, 32) = 1.70$, *ns*, $MSE = .070$, and $F(1, 32) = 0.01$, *ns*, respectively.