An Item Gains and Losses Analysis of False Memories Suggests Critical Items Receive More Item-Specific Processing Than List Items

Daniel J. Burns, Nicholas J. Martens, Alicia A. Bertoni, Emily J. Sweeney, and Michelle D. Lividini
Union College

In a repeated testing paradigm, list items receiving item-specific processing are more likely to be recovered across successive tests (item gains), whereas items receiving relational processing are likely to be forgotten progressively less on successive tests. Moreover, analysis of cumulative-recall curves has shown that item-specific processing produces a slower, but steadier rate of recall than relational processing. The authors relied on these findings to determine the type of processing that both list items and critical lures receive in the Deese–Roediger–McDermott false memory procedure. The first 2 experiments revealed that critical lures produced more item gains, but only the list items resulted in a decrease in item losses across successive tests. The critical lures also produced slower but steadier cumulative recall. In Experiments 3 and 4, the critical items were physically presented during study, which resulted in the lures producing progressively fewer losses across successive tests. The authors concluded that critical items receive more item-specific processing than list items but that unless they are presented in the list, they do not become part of participants’ organized retrieval scheme.

Keywords: false memories, cumulative recall, item-specific processing, relational processing, item gains and losses

Memory researchers have become intrigued with the phenomenon of illusory recollection—falsely remembering information that was never directly presented. An oft-used tool for studying such false memories is the Deese, Roediger, and McDermott (DRM) procedure, in which participants are typically presented a series of word lists, with the items on each list all associated with a critical item that is not presented on the list (Deese, 1959; Roediger & McDermott, 1995). For example, the list items PILLOW, DREAM, and SLUMBER are all related to the critical lure SLEEP. Although participants are never presented the critical lures, they often falsely recall or recognize the lures. Moreover, participants are confident in their belief that the falsely remembered items were actually presented (Read, 1996; Roediger & McDermott, 1995), often falsely recalling specific details about the presentation of the lures, including the gender of the voice and the sensory modality in which they were presented, as well as their list position and the list words presented near them (Gallo, McDermott, Percer, & Roediger, 2001; Hicks & Marsh, 1999; Matther, Henkel, & Johnson, 1997; Neuschatz, Payne, Lampien, & Toglia, 2001; Norman & Schacter, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996).

Although, the effects are typically modest, several variables have been shown to influence the occurrence of false memories in the DRM procedure. Roediger, McDermott, and Robinson (1998) pointed out that many of these variables are ones that are manipulated at time of list presentation, such as rate of presentation, order of presentation (blocked vs. random), modality of presentation, number of presentations, type of list items (pictures or words), and type of orienting task. Although all of these study variables influence the rate of false memories, they produce varying effects on retention of the actual list items, with some increasing retention, some decreasing it, and some having no effect.

Several theories have evolved to explain the reduction of false memories in the DRM procedure. Perhaps the most promising are fuzzy trace theory (e.g., Brainerd & Reyna, 1993, 1998; Brainerd, Reyna, & Brandse, 1995), Schacter and colleagues’ proposal of a distinctiveness heuristic (e.g., Israel & Schacter, 1997; Schacter, Israel, & Racine, 1999), and the activation/monitoring framework (Gallo & Roediger, 2002; McDermott & Watson, 2001; Roediger, Watson, McDermott, & Gallo, 2001). Briefly, fuzzy trace theory relies on the distinction between verbatim traces (ones with specific perceptual details) and gist traces (those that capture the meaning of events but lack perceptual detail) to explain illusory recollection. False memories are produced when retrieval is guided primarily by gist memory.

Schacter et al. (1999) proposed a distinctiveness heuristic to explain the occurrence of illusory recollection. They argued that when participants believe distinctive details of the list items were learned and should be accessible, they demand access to these details at time of retrieval before deciding whether an item was previously presented. However, when participants believe little perceptual detail should be available, they are less likely to use perceptual detail as a criterion in their decision process, relying more heavily on gist memory. The activation/monitoring framework combines elements of two earlier explanations, spreading activation (e.g., Roediger & McDermott, 1995) and source monitoring (e.g., M. K. Johnson, Hashtroudi, & Lindsay, 1993) to explain illusory recollection. Processing of the list items results in
activation of the related critical lures. At time of retrieval, this activation may cause problems in source monitoring such that the activation is falsely attributed to prior presentation.

Although there are clearly differences in the manner in which each theory explains the occurrence of false memories, there are also considerable similarities. Each approach assumes that list item receive some form of processing or activation. Moreover, they all further assume that item-specific processing of the list items tends to reduce false recollection of the critical lures, although the manner in which the item-specific processing exerts its influence is considerably different. Fuzzy trace theory assumes that item-specific processing strengthens verbatim memory, thereby reducing the likelihood that participants will rely on gist memory during retrieval. According to Schacter et al. (1999), item-specific processing will increase the likelihood that participants will invoke the distinctiveness heuristic to discriminate between old and new items at time of retrieval. The activation/monitoring approach assumes that item-specific processing will increase the source information encoded for the list items, allowing for better discrimination between the list and critical lures.

More interestingly, extant theories all assume that features of the critical lures are also activated or processed. Fuzzy trace theory assumes the gist or semantic features of the lures are processed. The account of Schacter and colleagues, although primarily focusing on retrieval conditions, also assumes gist or semantic features are activated (e.g., Israel & Schacter, 1997; Schacter, Norman, & Koutstaal, 1998), and of course, the activation/monitoring approach assumes the lures receive semantic processing via (spreading) activation.

Note that although each theoretical approach assumes item-specific processing of the list items will reduce false recollection, they are less specific about the type of processing the list items actually receive in the typical DRM procedure, or precisely how the list items are processed differently from the critical lures. It is tempting to assume that the list items tend to receive both item-specific and relational processing while the critical lures receive only relational processing (i.e., activation of semantic or gist features). This assumption, however, is not clearly stated for any of the theories. Because all conceptualizations of false recollection in the DRM procedure emphasize that reductions in false recollection result from emphasizing the difference between the critical lures and targets, a clear understanding of the exact nature of the type of processing or activation both types of items receive is critical.

Measures of Item-Specific and Relational Processing

In the present series of experiments we used the repeated-testing paradigm to assess the type of processing that both the critical lures and list items receive in a typical DRM experiment. In the repeated-testing paradigm, participants are presented with a (typically long) list of items and are then given any number of successive recall tests. Although repeated-testing has been used in conjunction with the DRM procedure, primarily to assess differences in forgetting between lures and list items, it has not been used specifically to assess the type of processing received by the lures and list items. To assess the type of processing, one must analyze differences in performance across repeated tests not only in terms of overall recall but also in terms of the number of item gains (items recalled on a later test that were not recalled on an earlier test) and item losses (items forgotten on a later test that were recalled on the preceding test).

Both Klein, Loftus, Kihlstrom, and Aseron (1989) and Burns (1993) demonstrated that manipulations known to increase item-specific processing produce a marked increase in item gains with negligible impact on item losses, whereas manipulations presumed to increase relational processing reduce item losses, with typically little or no impact on item gains. Hence, Burns suggested that an analysis of item gains and losses could be used to assess the extent of item-specific and relational processing, respectively. This technique has since been used with considerable success (e.g., Burns & Gold, 1999; Burns & Hebert, 2005; Engelkamp & Seiler, 2003; Klein, Loftus, Kihlstrom, & Schell, 1994; Mulligan, 2000, 2001, 2002; Mulligan & Duke, 2002).

Theoretically, relational processing has been suggested to decrease item losses because the encoded relational information is used as a retrieval plan to help generate the list items. The relational cues provide an organized search strategy that facilitates accessibility of the items (see Burns & Gold, 1999; Burns & Schoff, 1998; Hunt & McDaniel, 1993; McDaniel, Moore, & White, 1998). Presumably, the same (limited number of) relational cues are used across successive tests, resulting in a similar retrieval strategy that minimizes item losses. Moreover, the likelihood that participants will use the same retrieval routes across successive tests increases on later tests (McDaniel et al., 1998). This explains why item losses typically decrease across repeated tests (e.g., McDaniel et al., 1998; Mulligan, 2001, 2002; Mulligan & Duke, 2002).

Two explanations have been offered concerning the relationship between item-specific processing and item gains. Perhaps the simplest account is that item-specific processing produces numerous potential retrieval cues for each item. Because it is unlikely that all of the cues will be exhausted by the end of the first recall period, many of these cues may be accessed on later tests, producing previously unrecollected items. The second possibility is that following item-specific processing, the specific items that become accessible on any particular test are much more variable because recall is not guided by an organized retrieval strategy. Thus, items not accessed on the first test may be accessed on the second test, and vice versa.

Burns and Schoff (1998) pointed out that item gains are not always reflective of differences in item-specific processing. This is especially likely when either very short or very long recall tests are given. In the former case, item gains may be influenced by relational processing as well as by item-specific processing because the relatively few relational retrieval cues that were encoded have not been exhausted by the end of the first test. In the latter case, the item-specific retrieval cues may have been exhausted by the end of the first lengthy test, producing asymptotic recall levels. For these reasons, Burns and Schoff suggested that in addition to reporting item gains and losses, cumulative recall should also be plotted. They demonstrated that varying amounts of item-specific and relational processing produce specific patterns of cumulative recall.

To understand Burns and Schoff’s (1998) findings, the nature of cumulative-recall curves must first be discussed. Previous research has shown that cumulative-recall curves are well described by the following exponential equation:
\[ n(t) = n(\alpha)(1 - e^{-\lambda t}) \]  
where \( n(t) \) is the number of items recalled at time \( t \), \( n(\alpha) \) is asymptotic recall level, \( \alpha \) is the base of the natural logarithm, and \( \lambda \) is the rate of approaching asymptote (e.g., Bousfield & Sedgewick, 1944; Indow & Togano, 1970; Roediger, Stellon, & Tulving, 1977; Wixted & Rohrer, 1994). Moreover, research has shown that there is typically an inverse relationship between \( \lambda \) and \( n(\alpha) \) (e.g., Bousfield & Sedgewick, 1944; Hermann & Chaffin, 1976; Hermann & Murray, 1979; D. M. Johnson, Johnson, & Mark, 1951; Kaplan, Carmellas, & Metlay, 1969). In fact, until Burns and Schoff’s (1998) research no encoding variables had been shown to violate this inverse relationship.

Generally speaking, Burns and Schoff (1998) showed that the cumulative-recall curves produced following relational versus item-specific processing violate the inverse relationship between asymptotic recall and rate of approach to asymptote. Some of Burns and Schoff’s results are reproduced in Figure 1. Each panel depicts results from two groups of participants given different orienting tasks designed to produce differences in the amount of item-specific and/or relational processing performed. The orienting tasks used to induce item-specific or relational processing were selected because they had been used in previous studies. Item-specific tasks included pleasantness rating and forming individual images of each word. Relational tasks included category sorting and forming images of multiple list items interacting.

In the top panel, one group presumably processed more item-specific information, and the other group processed more relational information. The relational processing group produced higher initial recall but approached asymptote more quickly than the item-specific processing group.

The second panel of Figure 1 shows the results for two groups that both processed minimal relational information but differed in terms of the amount of item-specific processing performed. Both groups produced similar patterns of initial recall, but the group performing more item-specific processing recalled more items per minute later in the recall period. Estimates of \( \lambda \) and \( n(\alpha) \) showed that the group performing more item-specific processing approached asymptote slightly slower but had considerably higher asymptotic recall than the group performing less item-specific processing. The third panel compares a relational and item-specific processing group with a group processing only relational information. The former group produced superior recall starting very early in the recall period, producing a much higher estimate of asymptotic recall. Performing both item-specific and relational processing produces recall curves in which initial recall is very high compared to item-specific processing alone. Theoretically, relational cues are responsible for the initially high recall, and the item-specific cues sustain recall later in the cumulative-recall period. This general pattern of results has been successfully replicated (see Burns & Hebert, 2005).

### Experiment 1

The primary purpose of Experiment 1 was to use item gains and losses, as well as cumulative-recall functions, to assess the extent of relational and item-specific processing performed on both the critical lures and list items. We presented participants a long list of items made up of several successive DRM lists. Following auditory presentation, three recall tests were given. Although precise predictions derived from the current theories of illusory recollection in the DRM procedure are not entirely clear, most explanations assume that more item-specific information is encoded for list items than for critical lures. The predictions regarding relational processing are less obvious. Presumably, some relational information is encoded with the list items due to the semantic nature of the lists. Clustering scores support this contention in that the items are clustered by lists when participants are required to recall multiple lists simultaneously (Toglia et al., 1999). Moreover, when multiple lists are combined during presentation, both recall and clustering scores are higher following blocked presentation of the lists than following presentation in which the items from the different lists are randomly intermixed (Toglia et al., 1999). The relevant question here is whether the list items receive more or less relational processing than the critical lures.

There are several reasons to believe that more relational information may be encoded with the critical lures than with the list items. Certainly from a purely item-specific/relational processing framework (Hunt & Einstein, 1981; Einstein & Hunt, 1980) one would expect this outcome. If the list items receive more item-specific processing than the critical lures, as suggested above, and yet false recall of the lures is often as high as veridical recall of the list items, then more relational information must have been encoded with the critical lures to offset the memorial advantage for the list items resulting from more item-specific processing. Moreover, current explanations all assume that the lures receive as much, if not more, semantic activation or gist processing than the list items. Spreading activation and gist processing are often viewed as being correlated to, or synonymous with, relational processing. As will be shown, however, the data are entirely inconsistent with these loosely derived predictions.

### Method

Participants and materials. The 31 undergraduate students from Union College who participated in Experiment 1 were either paid $4.00 or
were given credit toward an out-of-class activity requirement of their introductory psychology course. They were tested in groups of 8 or fewer. All participants were presented the same long list of 50 words comprised of 10 sublists. The five words from each of the 10 sublists were taken from the lists used by Roediger and McDermott (1995). The items selected for each sublist were the strongest five associates. The associates from each sublist were blocked during presentation and the items within each sublist were ordered from strongest to weakest associative strength.

Procedure. All students were informed that they would hear a list of words being presented at a rapid rate and that they would be asked to recall the words afterward. No mention was made of the repeated tests. A tape recording was then played and participants heard the 50 items, which were presented at a 1.5-s rate. Following list presentation, the recall instructions were read aloud by the experimenter and lasted about 60 s. The students were told to write as many of the words they heard as they could remember, in any order they liked. They were told to be reasonably certain each word was presented before writing it down. To record cumulative-recall performance, participants were also told that they would be asked to draw a line under the last word they recalled after each min of recall. The first recall test lasted 4 min. The recall sheets were then collected and students were given a 4-min distractor task (recalling the names of the U.S. states), followed by the second 4-min recall test which was identical to the first. A second 4-min distractor task (recalling the names of state capitols) was then administered and was followed by the third and final recall test, which was identical to the first two tests.

Results

The various performance measures for both the list items and the critical items for all four experiments are shown in Table 1. Because there were 50 list items and only 10 critical lures in Experiment 1, all raw scores were converted to percentages to facilitate comparisons. As can be seen, the list items produced recall percentages that were approximately twice as high as the critical lures. A 2 (item type: list item or critical lure) × 3 (test number: 1, 2, or 3) within-subjects ANOVA was performed on the recall percentages and revealed that the advantage for the list items was significant, $F(1, 30) = 44.48,$ $MSE = 232.88,$ as was the interaction, $F(2, 60) = 4.51,$ $MSE = 21.79.$ Follow-up one-way ANOVAs revealed that whereas recall decreased across tests for the list items, $F(2, 60) = 11.31,$ $MSE = 9.46,$ there was no corresponding decrease for the critical lures ($F < 1$).

The percentage of items gained was calculated in two different ways. First, we divided the raw number of items gained from one test to the next by the total number of items presented (and then multiplying the quotient by 100). Second, we divided the number of item gains by the number of items not recalled on the previous test. The second procedure takes into consideration the fact that conditions with higher initial recall levels have fewer possible item gains on later tests. The two item-gain formulas produced an identical pattern of significant results in all four experiments, so we report only the results from the first formula. On the relatively rare occasions in which an item was recalled on the first test, forgotten on the second test, and then recalled again on the third test it was not considered as an item gain.

As can be seen in Table 1, the critical lures produced a greater percentage of item gains on the second test, but the list items produced more gains on the third test. The 2 (item type) × 2 (test number) within-subjects ANOVA verified that the interaction was significant, $F(1, 30) = 8.09,$ $MSE = 16.33.$ Follow-up tests re-
revealed that the advantage for the critical lures on the second test was not significant, $F(1, 30) = 16.24$, $MSE = 4.58$, although the latter comparison is contaminated by a severe floor effect for the critical lures, with only 1 participant producing any item gains.

The item-loss percentages were also calculated two different ways. First, we divided the raw number of items lost from one test to the next by the total number of items presented (and multiplied the quotient by 100). Second, we divided the number of items lost by the number of items recalled on the previous test. The second formula takes into consideration the fact that conditions with higher recall on the previous test have more items that can be potentially forgotten. The same pattern of significant results was obtained with the two formulas in all but the fourth experiment, so we present only the results from the first formula for Experiments 1–3.

The 2 (item type) × 2 (test number) ANOVA performed on the item-loss percentages produced a significant interaction, $F(1, 30) = 16.24$, $MSE = 4.58$, but that the advantage for the list items on the third test was significant, $F(1, 30) = 16.28$, $MSE = 4.58$, whereas the critical lures produced more item gains than the list items. The list items produced more item gains on the second recall test relative to the critical lures, but they also produced a decrease in losses across successive tests. Research has shown that participants tend to produce progressively fewer losses across successive tests, presumably because the items become more strongly entrenched in the participants’ organized retrieval plan with successive recall attempts. Hence, our results suggest that the critical lures were less likely to become part of the participants’ retrieval structure.

Analysis of the cumulative-recall curves revealed that while asymptotic recall levels were estimated to be much higher for the list items, approach to asymptote was much slower for the critical lures. Hence, although list items produce initially rapid recall, the critical lures produced slower but more steady recall. These

---

**Table 1**

*Mean Performance Measures (in Percentages) for the List and Critical Items in Experiments 1–4*

<table>
<thead>
<tr>
<th>Type of item</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>List Critical</td>
<td>List Critical</td>
<td>List Critical</td>
<td>List Critical</td>
</tr>
<tr>
<td>Test 1 recall, M</td>
<td>33.94 16.45</td>
<td>20.06 17.71</td>
<td>21.52 41.48</td>
<td>31.41 33.33</td>
</tr>
<tr>
<td>Test 2 recall, M</td>
<td>30.52 18.06</td>
<td>19.23 22.00</td>
<td>21.28 42.22</td>
<td>26.28 31.09</td>
</tr>
<tr>
<td>SD</td>
<td>9.82 13.76</td>
<td>4.75 16.94</td>
<td>7.74 14.50</td>
<td>10.98 12.76</td>
</tr>
<tr>
<td>Test 3 recall, M</td>
<td>30.97 16.12</td>
<td>19.23 22.29</td>
<td>20.82 42.22</td>
<td>26.60 31.41</td>
</tr>
<tr>
<td>SD</td>
<td>10.42 11.74</td>
<td>4.63 17.34</td>
<td>7.63 14.76</td>
<td>11.24 12.86</td>
</tr>
<tr>
<td>Test 2 gains, M</td>
<td>2.26 4.19</td>
<td>2.40 6.29</td>
<td>2.76 5.56</td>
<td>0.64 1.28</td>
</tr>
<tr>
<td>SD</td>
<td>3.00 7.65</td>
<td>2.05 8.08</td>
<td>2.47 5.77</td>
<td>1.53 3.49</td>
</tr>
<tr>
<td>Test 3 gains, M</td>
<td>2.52 0.32</td>
<td>1.63 3.42</td>
<td>0.95 2.22</td>
<td>0.64 0.64</td>
</tr>
<tr>
<td>SD</td>
<td>3.10 1.80</td>
<td>1.46 4.82</td>
<td>1.40 5.06</td>
<td>1.53 1.53</td>
</tr>
<tr>
<td>Test 2 losses, M</td>
<td>5.68 2.58</td>
<td>3.23 2.00</td>
<td>3.00 4.81</td>
<td>5.77 3.53</td>
</tr>
<tr>
<td>SD</td>
<td>4.38 6.31</td>
<td>1.75 4.06</td>
<td>2.46 6.43</td>
<td>3.92 3.67</td>
</tr>
<tr>
<td>Test 3 losses, M</td>
<td>2.06 2.90</td>
<td>1.74 3.14</td>
<td>1.40 2.22</td>
<td>2.56 1.60</td>
</tr>
<tr>
<td>SD</td>
<td>2.16 5.29</td>
<td>1.63 4.71</td>
<td>1.86 3.00</td>
<td>3.55 2.66</td>
</tr>
<tr>
<td>Approach to asymptote</td>
<td>1.82 0.48</td>
<td>0.74 0.38</td>
<td>0.76 0.83</td>
<td>1.03 1.17</td>
</tr>
<tr>
<td>Asymptote [n &gt;]</td>
<td>35.83 20.59</td>
<td>22.38 25.96</td>
<td>24.18 46.85</td>
<td>32.03 34.47</td>
</tr>
</tbody>
</table>

1 We used group means to aggregate data to estimate $\lambda$ and $n(\alpha)$ in the first three experiments, rather than individual scores, because several participants recalled only 0, 1, or 2 critical lures. With so few items recalled the estimates are either not calculable or the data are not well fit by Equation 1.
cumulative-recall curves suggest that the list items received more relational processing, whereas the critical lures received more item-specific processing. Hence, the combined results of Experiment 1 lead to the surprising conclusion that the list items received more relational processing, whereas the critical lures received more item-specific processing. This conclusion is explored after presenting the results of Experiment 2.

One difficulty with interpretation of Experiment 1, however, is that initial (Test 1) recall levels for the critical lures and list items were markedly different, making interpretation of the item gain and loss scores difficult (cf. Burns, 1993; Klein et al., 1994). For example, because twice as many list items were recalled as critical lures, there are twice as many items that can be potentially forgotten. Experiment 2 was designed to correct this problem of initial recall-level differences.

Experiment 2

Method

Experiment 2 was a direct replication of Experiment 1 with one exception. Previous research has shown that increasing the number of items in each DRM list decreases overall recall of the list items while improving recall of the critical lures (Robinson & Roediger, 1997). The number of items presented from each of the 10 DRM lists was doubled in Experiment 2, with the 10 strongest associates from each list used. As in Experiment 1, the items were ordered from strongest to weakest associative strength. There were 35 participants tested.

Results

As shown in Table 1, the huge recall advantage for the list items that was obtained in Experiment 1 was nearly eliminated in Experiment 2. A 2 (item type) × 3 (test number) ANOVA revealed that only the interaction was significant, $F(2, 68) = 48.29$, $MSE = 21.49$. Follow-up one-way ANOVAs revealed that whereas recall decreased across tests for the list items, $F(2, 68) = 3.30$, $MSE = 4.25$, it actually increased across tests for the critical lures, $F(2, 68) = 5.56$, $MSE = 41.29$. This hypermnestic recall effect for the lures but not the list items has been demonstrated elsewhere (Payne et al., 1996). In addition, analysis of Test 1 recall percentages showed no significant difference between item types ($F < 1$).

The 2 (item type) × 2 (test number) ANOVA performed on the item-gain percentages revealed that the critical lures produced significantly more gains than the list items, $F(1, 34) = 9.95$, $MSE = 28.42$, and that item gains decreased from the second to the third test, $F(1, 34) = 6.03$, $MSE = 19.12$. The interaction was not significant $F(1, 34) = 21.17$, $MSE = 17.56$. The $2 \times 2$ ANOVA performed on the item-loss percentages revealed that although neither main effect approached significance (both $Fs < 1$), the interaction was significant, $F(1, 34) = 4.21$, $MSE = 14.35$. Follow up tests revealed that the critical lures and list items produced a statistically equivalent percentage of losses both on Test 2, $F(1, 34) = 2.31$, $MSE = 10.53$, and on Test 3, $F(1, 34) = 2.82$, $MSE = 12.18$. However, losses decreased across tests for the list items, $F(1, 34) = 16.79$, $MSE = 2.31$, but remained relatively stable for the critical lures ($F < 1$).

The cumulative-recall curves are shown in Figure 3. The list items produced relatively rapid initial recall that more quickly approached asymptotic levels, whereas the critical lures produced poorer initial recall, but approach to asymptote was slower. In fact, although the critical lures were initially recalled at a lower rate than the list items, final cumulative recall was numerically higher for the lures. The estimates of $\lambda$ and $n(\alpha)$, presented in Table 1, showed that the list items produced a considerably higher estimate of $\lambda$ but a slightly lower estimate of $n(\alpha)$ than the critical lures. Although the estimates do not violate the inverse relationship rule, the fact that the critical items initially produced lower recall but then produced higher final cumulative recall clearly violates the typical findings.
Discussion

Increasing the number of items in each DRM list from 5 to 10 had the desired effect of eliminating initial (Test 1) recall level differences between the list items and critical lures, making interpretation of item gains and losses much less problematic. Despite the lack of initial recall differences, the critical lures produced improvements in recall across tests, whereas the list items did not. The critical lures also produced significantly more item gains than the list items, regardless of the test number, suggesting that they received more item-specific processing.

The item-loss scores produced a similar pattern of results to those of Experiment 1: Item losses decreased across successive tests for the list items but actually numerically increased across tests for the critical lures. As in Experiment 1, this pattern of results suggests that the critical lures are not part of the organized retrieval plan participants use to retrieve the list items.

The cumulative-recall curves showed that although initial recall levels were considerably higher for the list items, final cumulative recall was actually higher for the critical lures. In other words, the critical lures produced slower but steadier recall than the list items, suggesting the counterintuitive conclusion that more item-specific but less relational information was encoded for the critical lures.

How is it possible that more item-specific information is encoded with the critical lures, which were not physically presented, than with the list items? At first glance, this outcome seems implausible—clearly, more perceptual and contextual detail should be encoded with the list items. It is possible, however, that the critical lures receive more conceptual detail as a result of being activated by the list items. One possibility is that presentation of the list items not only activates critical lures but also activates items related to the critical lures that are themselves not related to the list items. For example, the list items HOT, SNOW, and WET activate the critical lure, COLD, which might then activate the concept BEER, which is not related to any of the list items but is related to the critical lure. BEER might constitute the item-specific information that facilitates later retrieval of the critical lure. The research on two-step priming provides support for this view. Two-step priming occurs when a word primes an unrelated word, presumably through a mediator that is related to the prime and target. For example, LION will prime STRIPES, via the mediator TIGER (e.g., Balota & Lorch, 1986).

A second possibility is that the critical lures do not actually receive more item-specific processing than list items. Consistent with extant theory, they may receive primarily relational processing, but because the encoded relational information apparently does not become part of participants’ strategic retrieval plan, it may impact item gains and cumulative recall in much the same way as item-specific information. For example, an item-item association might be formed between a list item and the lure, but that particular association may not be integrated into the organized retrieval plan. However, after the retrieval plan is selected and runs its course, participants may search for remaining cues. These remaining cues are usually item-specific in nature, but they may also include relational cues that were not already used. Because the relational cues were not part of the strategic retrieval plan, they would behave much like item-specific cues, increasing the likelihood of the lure being recalled later in the recall period (or on later tests).

Note that the two post hoc explanations presented above lead to different predictions about the effects of physically presenting the critical items in the DRM lists. If the critical items receive more item-specific processing, as suggested by the first explanation,

Figure 3. Mean cumulative-recall scores for both the list items and the critical lures in Experiment 2. Error bars signify the standard error of the mean.
then they should continue to receive item-specific processing when physically presented in the list. Of course, they would also receive considerable relational processing by virtue of their relationship with the other list items. Hence, the critical items would receive both item-specific and relational processing, whereas the other list items would receive primarily only relational processing. This should result in more item gains for the critical lures, and cumulative-recall curves similar to those shown in the third panel of Figure 1, in which the comparison groups differ only in terms of item-specific processing. On the other hand, if critical lures receive primarily relational processing, but the encoded relational information is not integrated into the retrieval scheme, then physically presenting the critical items in the list should result in the relational information being integrated into the retrieval plan, thereby eliminating the item-gain advantage.

Regardless of whether embedding the critical items in the DRM lists affects item gains, it should clearly impact item losses. If, as we suggest, the critical lures are not part of the organized retrieval scheme because they were not physically present and, therefore, were not consciously associated with the list items, then actually presenting the critical items in the list should result in the formation of interitem associations between the critical and other list items, allowing the critical items to be integrated into the retrieval plan. This should result in progressively fewer losses across successive tests for the critical items. To test these hypotheses we replaced one list item from each of the 10 lists with the critical lure.

Experiment 3

Method

Experiment 3 was identical to Experiment 2, with the following exception: One list item from each of the 10 lists was replaced with the critical item for that list. The serial position of each replaced list item was chosen by random selection without replacement, such that the critical items were presented in each serial position across the 10 lists. There were 27 participants in Experiment 3.

Results

The 2 (item type: critical item or other list item) × 3 (test number) ANOVA performed on the recall percentages, which are shown in Table 1, revealed only that the critical items produced significantly higher recall than the other list items, $F(1, 26) = 65.93, MSE = 264.94$ (both other $F$s < 1). The 2 × 2 ANOVA performed on the item-gain percentages showed that the critical lures produced significantly more gains than the other list items $F(1, 26) = 8.69, MSE = 12.89$, and that gains decreased across tests, $F(1, 26) = 9.96, MSE = 17.93$. The interaction did not approach significance, $F(1, 26) = 1.28, MSE = 21.04$, but the main effect of item type did not quite reach significance, $F(1, 26) = 3.24, MSE = 14.46$, and there was no evidence that the variables interacted ($F < 1$).

The cumulative-recall results, shown in Figure 4, are markedly different from those of Experiments 1 and 2. As can be seen, the critical items produced superior recall throughout the recall period. Estimates of $\lambda$ and $n(\alpha)$, shown in Table 1, demonstrated that the critical items approached asymptote at about the same rate as the other list items despite an asymptotic recall level that was nearly twice as high.

Discussion

Presenting the critical items in the DRM lists had a considerable impact on recall, with recall of the critical items nearly twice as
likely as recall of the list items on the first test. Despite the advantage in recall for the critical items, recall did not improve across tests for either item type. The item-gain advantage for the critical items found in Experiment 2, however, was replicated in Experiment 3. This finding suggests that the critical items received more item-specific processing than the other list items and argues against the notion that the item-gain advantage for the critical lures found in Experiment 2 was the result of more relational processing. The cumulative-recall results substantiate these conclusions. Whereas the recall curve produced by the other list items resembled those obtained in Experiments 1 and 2, the curve for the critical items showed initially very high recall levels with relatively steady growth. These curves resemble those displayed in the third panel of Figure 1, and suggest that although both item types received relational processing, the critical items also received more item-specific processing.

The item-loss results did not replicate the previous two experiments. Unlike Experiments 1 and 2, item losses decreased across successive tests for the critical items as well as for the list items, suggesting that relational information was encoded with the critical items and this information was included in participants’ organized retrieval plans.

Experiment 4

Experiment 4 addressed two important points. First, in Experiment 3, the critical items produced nearly twice the level of Test 1 recall as the other list items. Although this outcome was expected, it produces the same interpretational problems as those discussed for Experiment 1. Namely, it is difficult to compare item-loss differences when the number of potential items that can be forgotten differs. Thus, Experiment 4 was designed to study the memorial consequences of physically presenting the critical items under conditions in which initial recall levels were more closely equated. To this end, the list used in Experiment 4 consisted of 48 items, one list item from each of the 24 DRM lists used by Roediger and McDermott (1995) and the 24 critical items associated with those lists.

In addition to the goal of equating initial recall levels, a second objective was to determine whether the item-gain advantage for the critical items obtained in the first three experiments would be eliminated in Experiment 4. We assumed that if only one item from each DRM list was presented, then the amount of processing (e.g., spreading activation) the critical items received would be minimal. Assuming that the item-specific processing advantage for the critical items found in Experiments 1–3 was the result of the presentation of the numerous associatively related list items, then the critical items should receive little item-specific processing in Experiment 4. The result should be an elimination of the item-gain difference as well as similar cumulative-recall curves for the list and critical items. If, on the other hand, there is something inherently special about critical items that makes them inherently more distinctive, then the item-gain advantage (as well as a recall advantage) should be obtained under the conditions of Experiment 4.

Method

Experiment 4 was identical to Experiment 1 with the exception that the list consisted of the critical item and one list item from each of the 24 lists used by Roediger and McDermott (1995). The 24 list items were selected, such that three items came from list positions 4 and 5, and two items came from the remaining first 10 positions. In addition, because the list now consisted of relatively unrelated words, the presentation rate was decreased to 3 s per item. Twenty-six students participated in the experiment.

Results

The 2 (item type) × 3 (test number) ANOVA conducted on the recall percentages revealed that the only significant effect was the main effect of test number, \( F(2, 50) = 21.97, \, MSE = 10.72. \) The 2 × 2 ANOVA conducted on the item-gain percentages produced no significant effects (all \( F < 1 \)). A similar ANOVA performed on the item-loss percentages showed that the other list items produced significantly more losses than the critical items \( F(1, 25) = 4.77, \, MSE = 14.00. \) Losses also decreased across tests, \( F(1, 25) = 11.11, \, MSE = 15.39, \) but the interaction did not approach significance (\( F < 1 \)). No significant effects were found when the loss percentages were calculated by dividing the number of losses by the number of items recalled on the previous test (and multiplying the quotient by 100).

The cumulative-recall curves for the critical items and other list items, displayed in Figure 5, were very similar, suggesting that the two types of items received similar amounts of item-specific and relational processing. Moreover, both curves show initially rapid recall that approaches asymptote very quickly. This pattern of results indicates that neither type of item received much item-specific processing that could sustain recall throughout the 12-min period.

Discussion

The list composition used in Experiment 4 was effective in statistically eliminating recall differences between the critical and list items, as well as the item-gain difference obtained in Experiments 1–3. The critical items did produce fewer item losses (at least when the denominator in the item-loss percentage formula was based on the total number of items presented), but both critical and other list items produced fewer losses on Test 3 than on Test 2, suggesting that both item types were included in participants’ strategic retrieval plans. In addition, the cumulative-recall curves produced by the critical and other list items were nearly identical, suggesting that they received similar amounts of item-specific and relational processing.

The finding that despite equivalent levels of Test 1 recall performance, the critical items produced progressively fewer item losses across repeated tests demonstrates that the reduction in losses obtained in Experiment 3 was not simply the result of different levels of Test 1 recall performance. Rather, physically presenting the critical items seems to result in their becoming part of the retrieval plan, which reduces forgetting across tests. The other main conclusion is that the item-specific processing advantage for the critical items that presumably occurred in the first three experiments was apparently eliminated when the list items no longer converged on the critical items. This conclusion implies that the item-specific processing advantage for the critical items is not due to something inherent in the critical items themselves. The critical items are not simply more likely to receive item-specific processing than other items. Rather, it is the fact that the list items...
are all associatively related to the critical item that leads to the substantial item-specific processing.

**General Discussion**

The first two experiments revealed that nonpresented critical lures tended to produce slower but steadier recall than actual list items, which resulted in significantly more item gains across repeated tests in Experiment 2. In addition, whereas the list items showed progressively fewer losses across successive tests, the critical items showed a numerical increase in losses across tests. Taken together, these results were interpreted as evidence that critical items receive more item-specific processing than list items, but that they do not become part of the retrieval plan used during recall.

An alternative explanation, however, is that the critical items receive primarily relational processing, but the encoded relational information is not integrated into the organized retrieval plan, and affects item gains and cumulative recall in much the same ways as item-specific information. The logic of Experiment 3 was that actually presenting the critical items in the DRM lists should allow them to be included in participants’ retrieval strategies. The prediction derived from the alternative explanation is that since the relational information could become part of the retrieval plan, the item gain difference would be eliminated. Contrary to the prediction, the item gain difference occurred even when the critical items were physically embedded in the DRM lists, suggesting that they received more item-specific processing.

Experiment 4 demonstrated that the item-gain advantage (as well as the slow but steady recall across the lengthy recall period) was eliminated when the critical items were presented in a list of relatively unrelated words. Apparently the enhanced item-specific processing given to the critical items in the first three experiments was not due to something inherent in those items, but rather results from the activation (or gist processing) they typically receive via the related items in the list.

The picture that emerges from the results of Experiments 1–4 is that critical items receive more item-specific processing than the list items because of the additional activation they receive from the related items. However, the critical items do not become part of an individual’s retrieval plans unless they are physically presented.

It is possible that the critical lures may actually receive more relational processing during encoding than the list items, but that some type of information is available that helps participants filter out the critical lures at time of retrieval. For example, current theoretical approaches might suggest that participants rely on item-specific information (e.g., perceptual detail) encoded with the list items to filter the lures. The explanation cannot be that simple, however, because the item gain and cumulative-recall results suggest that critical lures actually receive more item-specific processing, not less. Perhaps a more promising explanation, one that is fairly consistent with extant theories, is that different types of item-specific information are at work here. More perceptual features may typically be encoded for the list items, whereas more conceptual features are encoded for the critical items. The result is that while the conceptual features lead to greater item gains and more steady cumulative recall for the critical items, the perceptual detail of the list items can provide a means for filtering out some of the critical lures at time of test.

**The Forgetting of False Memories**

Previous research has shown that false memories tend to be resistant to forgetting across repeated memory tests, a phenomenon

![Figure 5. Mean cumulative-recall scores for both the list items and the critical items in Experiment 4. Error bars signify the standard error of the mean.](image-url)
termed the persistence effect. In fact, research suggests that false memories are actually more resistant to forgetting than actual list items, especially for tests of free recall (e.g., McDermott, 1996; Seamon et al., 2002; Thapar & McDermott, 2001; Toglia et al., 1999). The results of the present experiments suggest that the situation is more complex. The results of Experiment 2 revealed that, across the three tests, the critical lures were forgotten (i.e., lost) at about the same rate as list items. However, the previously unrecalled critical lures were more likely to be recalled on a later test. Because previous research has not separated forgetting into its two components (gains and losses), the global analysis may have given the mistaken impression that recall of the lures was more stable. It is ironic that when both item gains and losses were analyzed in the present experiments, the critical lures were actually less stable: Critical lures that were not previously recalled on an earlier test were more likely to be recalled on a later test. Of course, the present research only used short retention intervals; it may be that critical and list items are lost at a different rate when longer retention intervals are used or when retention interval is manipulated between participants (see Seamon et al., 2002; Toglia et al., 1999, who showed greater forgetting of list items over longer retention intervals).

Are Critical Lures Consciously Processed During Study?

One recurrent issue in the literature is whether critical lures are brought to mind and consciously processed during the presentation of a DRM list. Although an activation account of false memories does not necessitate that the participant consciously retrieve the item, research suggests that conscious retrieval may sometimes occur. For example, research suggests that the critical items may actually be rehearsed at time of study to the same extent as list items (Mather, Henkel, & Johnson, 1997), and that they may be associated with other list items (Norman & Schacter, 1997). The present experiments suggest that this is not always the case. If the critical lures are brought to mind at time of study and form interitem associations with other list items, then it seems likely they would have become part of the retrieval plan. The results suggest, instead, that they only become part of the retrieval plan when physically presented in the list. Thus, although not definitive, our results suggest that the critical lures were not necessarily brought to mind and confused as list items. Either participants did not consciously process the items at time of list presentation or they did consciously process them but somehow managed to differentiate them from the list items, excluding them from their retrieval plan.

False Memories and Cumulative-Recall Curves

The present experiments are the first comprehensive attempt to study cumulative recall of critical items. Roediger and McDermott (1995) showed that critical lures tend to be recalled in later output positions than list items (see also McDermott, 1996), suggesting that the critical lures are accessed later in the recall period. However, they required recall of individual DRM lists, so only one critical lure could be recalled during the recall period. McDermott (1996) plotted cumulative recall of both the list items and critical lures following presentation of 24 DRM lists. Her results showed that the critical lures tended to be recalled better than the list items throughout the 15-min final recall period. In fact, her results closely resembled the pattern shown in the third panel of Figure 1 (with the critical lures representing the higher line). However, the final recall test was given 2 days after learning the lists, and participants had been required to recall most of the lists 2 days earlier. Hence, the final cumulative-recall scores were undoubtedly contaminated by performance on the earlier recall tests.

The results of Experiments 1 and 2 were free of prior testing effects, and they clearly showed that critical items were recalled at a slower rate than list items. Experiment 2 further showed that although recall of the critical lures was initially lower than recall of the list items, by the end of the cumulative-recall period it was considerably higher. Previous research has shown that plotting cumulative recall across several successive tests produces the same cumulative-recall pattern as plotting recall across one long test (e.g., Roediger & Thorpe, 1978). It remains to be seen whether this holds true for critical lures as well as list items, but if it does, this suggests that the length of recall test chosen by researchers could be an important determinant of the extent to which false memories are obtained.

Are Critical Lures Inherently Special?

It has been suggested that the DRM lists are constructed in such a way that the critical lures are somehow special, making them inherently more memorable. Whittlesea (2002) called this the life effect. Typically, this life effect is tested by comparing false recognition for critical lures from nonpresented DRM lists to false recognition for typical list items from nonpresented lists (e.g., Brainerd & Reyna, 1998; Israel & Schacter, 1997; Seamon et al., 2002; Whittlesea, 2002). Although the effect is not always significant, the critical lures consistently produce higher false recognition scores than the nonpresented list items. Experiment 4 provided an alternative test of the memorability of critical lures. When the critical items from each list were intermixed with a single item from each DRM list, they were not recalled significantly better than the list items. This suggests that, at least in tests of recall, there is no life effect. Close inspection of the recognition studies also suggests that the critical items may not be inherently more memorable. In those studies, both the critical lures and the list items from the nonpresented lists that were presented on the recognition test came from the same DRM lists. Moreover, usually several (e.g., three) list items were presented from each nonpresented list. It is possible that because the list items were related to the critical lure, participants were more likely to select the critical lure as previously seen (e.g., due to more test-induced priming or activation). Previous research has shown little test-induced priming for critical lures from presented DRM lists, but it has also shown large priming effects for lures from nonpresented lists (Marsh, McDermott, & Roediger, 2004). Therefore, on the basis of the results of Experiment 4, we conclude that there appears to be little reason to believe that there is anything inherently special about critical items.

Conclusion

It seems ironic that critical lures, which are not physically presented during study, receive more item-specific processing than the actual list items. Nonetheless, the item gain results, as well as
the cumulative-recall curves, seem to demand this conclusion. One important avenue for future research is to study the exact nature of the item-specific information encoded with the list items and critical lures. As mentioned previously, it is possible that list items and critical lures receive different types of item-specific processing (e.g., the encoding of perceptual vs. conceptual information). Perhaps the item-specific processing given to the critical items is more likely to enhance item accessibility, and the item-specific processing given to the list items is more likely to serve a discriminative function.

References


Read, J. D. (1996). From passing thoughts to false memories in 2 minutes:


Received February 22, 2005
Revision received August 16, 2005
Accepted October 23, 2005