Disregarding Familiarity During Recollection Attempts: Content-Specific Recapitulation as a Retrieval Orientation Strategy

Stephen J. Gray and David A. Gallo
The University of Chicago

People can use a content-specific recapitulation strategy to trigger memories (i.e., mentally reinstating encoding conditions), but how people deploy this strategy is unclear. Is recapitulation naturally used to guide all recollection attempts, or is it only used selectively, after retrieving incomplete information that requires additional monitoring? According to a retrieval orientation model, people use recapitulation whenever they search memory for specific information, regardless of what information might come to mind. In contrast, according to a postretrieval monitoring model, people selectively engage recapitulation only after retrieving ambiguous information in order to evaluate this information and guide additional retrieval attempts. We tested between these models using a criterial recollection task, and by manipulating the strength of ambiguous information associated with to-be-rejected foils (i.e., familiarity or noncriterial information). Replicating prior work, foil rejections were greater when people attempted to recollect targets studied at a semantic level (deep test) compared to an orthographic level (shallow test), implicating more accurate retrieval monitoring. To investigate the role of a recapitulation strategy in this monitoring process, a final test assessed memory for the foils that were earlier processed on these recollection tests. Performance on this foil recognition test suggested that people had engaged in more elaborate content-specific recapitulation when initially tested for deep compared to shallow recollections, and critically, this elaboration effect did not interact with the experimental manipulation of foil strength. These results support the retrieval orientation model, whereby a recapitulation strategy was used to orient retrieval toward specific information during every recollection attempt.

Keywords: recollection, familiarity, retrieval monitoring, source memory, false memory

Episodic memory errors originate from a variety of sources (Roediger, 1996), but people often are successful at using retrieval monitoring processes to avoid them. By retrieval monitoring, we refer to the strategic search and evaluation processes that shape the accuracy of memory (cf. Johnson, 2006; Koriat, Goldsmith, & Pansky, 2000). The concept of retrieval monitoring has been heavily influenced by dual-process theories, which differentiate between a decontextualized feeling of familiarity or “oldness” elicited by a retrieval cue, and the specific recollection or “reexperiencing” of associated contextual details from the past (for review of dual process theories, see Yonelinas, 2002). This research has revealed that recollection can be used in a strategic or controlled way to override the potentially erroneous influences of familiarity (Jacoby, Kelley, & McElree, 1999). Retrieval monitoring also is a central concept within the source monitoring framework, which in addition to familiarity, proposes that memory errors can be based on the retrieval of fragmented or recombined information and incorrect source attributions (i.e., false recollections, Dodson, Bawa & Slotnick, 2007; Lyle & Johnson, 2006).

Considerable research shows that memory errors are less likely when the study materials contain unique features that increase recollective distinctiveness, indicating that retrieval monitoring is influenced by the qualitative characteristics of to-be-recollected information. For example, people make fewer false recollection judgments when their memory is tested for words that were deeply processed at study (i.e., processing unique meanings) compared to shallowly processed words (i.e., processing orthographic features), ostensibly because people expect that deeply processed words are more distinctive and hence more likely to be successfully recollected (Gallo, Meadow, Johnson, & Foster, 2008). In this and other cases, the failure to recollect expected information can help to reduce memory errors that would otherwise be caused by the retrieval of familiarity or fragmented recollections (e.g., a distinctiveness heuristic, see Schacter & Wiseman, 2006). Qualitative differences in retrieval expectations also can affect errors on source memory tests (Johnson, Foley, Raye, & Foley, 1981; see Johnson, Hashtroudi, & Lindsay, 1993), suggesting that the use of expectations is a major factor in retrieval monitoring more generally.

While prior research shows that expecting more distinctive recollections can minimize memory errors, significant questions about the nature of this retrieval monitoring process remain. With the current research we considered the extent that two general models of cognitive control could apply to this kind of retrieval monitoring, inspired by a distinction drawn by Jacoby et al. (1999). According to the retrieval orientation model, people naturally adopt an elaborative or content-specific search strategy when attempting to retrieve target information. Jacoby et al. (1999) called this an early selection process, suggesting that this strategy could...
constrain retrieval to target information at the front-end, thereby avoiding the potentially erroneous influence of less relevant information. By less relevant information, we mean familiarity or noncriterial recollections that could influence the accuracy of memory decisions (cf., Toth & Parks, 2006). Applied to distinctiveness effects, adopting a global retrieval orientation would be more effective (and potentially easier) when the target information is more distinctive. Of course, engaging in a content-specific search strategy does not guarantee that retrieval would be successfully constrained, and also could occur after less relevant information is automatically retrieved (see Kantner & Lindsay, 2013). Either way, with a retrieval orientation model, people engage in a content-specific search strategy for all test items (i.e., at a global level) as a way of prioritizing the retrieval of targeted information. This prioritization also might help them disregard or avoid the erroneous influences of less relevant information.

In contrast to the retrieval orientation model, in which a content-specific monitoring strategy operates at a global level for all retrieval attempts, the postretrieval monitoring model proposes that people first engage in unconstrained retrieval, attempting to retrieve any information associated with a test item regardless of content. They then make a decision by evaluating all of the retrieved information and rejecting that which seems less relevant or potentially inaccurate. In this decision stage, people selectively engage in more elaborate or content-specific retrieval monitoring only when needed. Jacoby et al. (1999) referred to this as a late-correction strategy, in the sense that participants only engage in this strategy in reaction to the retrieval of ambiguous or vague information. Applied to distinctiveness effects, this postretrieval monitoring strategy would be more effective at overriding the effects of less relevant information on memory errors when the target information is more distinctive. Importantly, the notion of postretrieval monitoring does not deny the possibility that participants can engage in a content-specific monitoring strategy, but it does differ from the retrieval orientation model in terms of when and why people might use this strategy. With a postretrieval monitoring model, people selectively engage in a content-specific monitoring strategy only for those test items that initially elicit ambiguous information, such as familiarity or noncriterial recollections.

Each of these models gains some support from the neuroimaging literature (for review, see Rugg, 2004), but that literature has not resolved the role of these models in distinctiveness effects on memory errors. For example, Rugg and colleagues have referred to retrieval orientation in order to explain content-specific brain activity that occurs for both studied and nonstudied test items, ostensibly reflecting an attempt to constrain or “gate” memory search and retrieval to target recollections (e.g., Hornberger, Rugg & Henson, 2006; Morcom & Rugg, 2012; Rugg & Wilding, 2000; see also Dobbins & Han, 2006). In contrast, the concept of postretrieval monitoring has been used to explain relatively late-occurring dorsolateral prefrontal activity that is more likely when people need to evaluate more difficult test items (e.g., Hayama & Rugg, 2009; Schacter, Buckner, Koutstaal, Dale, & Rosen, 1997). In a series of fMRI studies that targeted retrieval monitoring processes, Gallo and colleagues found that both DLPFC activity and response latencies were greater when participants monitored memory for less distinctive recollections (words compared to pictures; Gallo, Kensinger & Schacter, 2006; Gallo, McDonough & Scimeca, 2010; McDonough, Wong & Gallo, 2013). These effects suggest that participants engaged more effortful retrieval monitoring for less distinctive information, but the important point for present purposes is that these results are relatively mute with respect to why the more distinctive test condition was more accurate. The distinctive test condition may have benefited from a more effective (and less effortful) postretrieval monitoring process, or it might have benefited from a global retrieval orientation strategy that reduced the need to engage in postretrieval monitoring, either of which could have reduced response latencies and DLPFC activity (for relevant discussion, see Hudson et al., 2005; Gallo et al., 2006).

As discussed by Gallo, Weiss, and Schacter (2004), either of these models could explain the reductions in memory errors associated with more distinctive recollections, and little evidence definitively supports one model over the other. For example, distinctiveness effects on memory errors are minimized on speeded recognition memory tests (Dodson & Hege, 2005; Gallo, Perlmutter, Moore, & Schacter, 2008). While these results potentially implicate a postretrieval monitoring process that takes more retrieval time, these results also might reflect the relatively slower speed of recollection compared to familiarity, regardless of which model of retrieval monitoring applies. As another example, Gallo et al. (2004) argued that distinctiveness effects on multiple-option source memory tests might reflect a postretrieval monitoring process, because all sources of information are potentially relevant on these tests. However, it remains possible that these effects are due to the use of a more global retrieval orientation strategy on these tests, whereby participants attempt to search memory for the more distinctive source for all trials and attribute items to the less distinctive source only if this initial attempt failed to retrieve targeted information. In general, while prior work clearly shows that testing memory for more distinctive recollections can reduce memory errors, the extent that these effects should be attributed to a global retrieval orientation process or a selective postretrieval monitoring process is difficult to determine.

**Memory-for-Foils Procedure**

Although these two models have been difficult to disentangle empirically, Jacoby, Shimizu, Daniels, and Rhodes (2005) developed a memory-for-foils procedure that can provide new insights into retrieval monitoring (see also Jacoby, Shimizu, Velanova, & Rhodes, 2005). In this procedure, people studied words in deep or shallow processing conditions for subsequent recognition memory tests, on which studied words were intermixed with nonstudied foils. After this initial recognition test, people were given a second recognition test for the foils that had occurred on the initial test (i.e., the foil recognition test). The assumption was that different kinds of processing on the initial recognition test should have affected people’s memory for the foils, and this processing would be reflected on the foil recognition test. Consistent with this interpretation, foils from the deep testing condition were better remembered on the foil recognition test than were foils from the shallow testing condition, suggesting that people engaged in more elaborative foil processing when initially tested for the deeply processed items than the shallowly processed items (i.e., a foil-elaboration effect). Further evidence that this effect resulted from elaborative foil processing comes from Marsh et al. (2009), who
found that deep foils were more likely than shallow foils to elicit subjective judgments of “remember” on the foil recognition test.

To explain the foil-elaboration effect, Jacoby, Shimizu, Daniels et al. (2005) suggested that people engaged in mental recapitulation when processing test foils (cf. Alban & Kelley, 2012; Halamish, Goldsmith, & Jacoby, 2012). According to this idea, people strategically simulate or mentally imagine the test item as having occurred in the target source when this item is encountered at test (i.e., they attempt to mentally reinstate encoding conditions). This recapitulation can be thought of as a self-generated retrieval cue that could potentially boost retrieval monitoring accuracy, either by helping to trigger target recollections or by helping to evaluate and reject irrelevant information. Critically, people can engage in qualitatively different kinds of recapitulation. The simulation of a deep judgment when tested for deep information would provide more elaborative foil processing relative to the simulation of a shallow judgment when tested for shallow information, leading to superior memory on the subsequent foil recognition test. Consistent with this recapitulation account, Dunckert, MacLeod, and Fernandes (2011) found that testing participants for deep judgments primed the speed of subsequent deep judgments, and Alban and Kelley (2012) found that participants spontaneously described a recapitulation strategy when instructed to think aloud during recognition tests. Importantly, using self-report judgments, Marsh et al. (2009) found that foils were most likely to be recognized on the final test if they were initially rejected based on a lack of deep recollections, thereby linking these foil-elaboration effects to a recollection-based retrieval monitoring process.

Although prior work has identified the use of a recapitulation process during retrieval monitoring, it is unknown how this process is strategically implemented. Jacoby, Shimizu, Velanova et al. (2005) argued that the elaborative processing of test items might be an attempt to constrain one’s search to the target source, analogous to the global retrieval orientation model described previously. By this interpretation, engaging in more elaborate recapitulation on the deep test relative to the shallow test could help to retrieve the targeted recollections and reduce the effects of less relevant information on memory errors. In contrast, it also is possible that elaborative retrieval monitoring could be used as part of a postretrieval monitoring strategy. By this interpretation, people might first attempt to retrieve all information that is associated with a cue, and then selectively engage in recapitulation at test only if the initial search yields less relevant information (e.g., familiarity or noncriterial recollection). In this model, recapitulation could help people evaluate this ambiguous information with respect to their retrieval expectations, and also help to search memory for additional information. The critical difference between the two models is whether the recapitulation process is globally engaged for all test items (retrieval orientation) or whether it is selectively engaged in response to the retrieval of ambiguous information (postretrieval monitoring). We know of no behavioral evidence that clearly supports one model of the recapitulation process over the other, either in the memory-for-foils literature or in the memory literature more generally.

The Current Research

The goal of the current research was to provide an empirical test between the retrieval orientation and postretrieval monitoring models of the content-specific recapitulation process. Is recapitulation used to maintain a specific retrieval orientation regardless of whether other information is retrieved, or is recapitulation initiated because ambiguous information is retrieved and needs to be monitored? To answer this question we applied the logic of the memory-for-foils paradigm to a modified version of the criterial recollection task that we have used to investigate retrieval monitoring processes (for review, see Gallo, 2013). Participants first studied words with deep or shallow processing, and then we selectively tested recollection for either deeply or shallowly processed words across different test blocks. Next, we administered a foil recognition test for the foils that earlier had been tested on these recollection tests, in order to measure the effects of content-specific retrieval monitoring on foil processing. Critically, we experimentally manipulated the strength of the foils on the recollection tests—or the extent that they could elicit familiarity or noncriterial recollection—by varying the times they had earlier been presented in a prestudy familiarization list. As described next, the two alternative models of retrieval monitoring make different predictions under these foil repetition conditions.

According to the global retrieval orientation model, participants use a recapitulation process to search their memory for the to-be-recollected information on each test, regardless of the amount of familiarity or noncriterial information that might be retrieved. For example, participants might use a deep recapitulation process to try to constrain recollection on the deep test and a shallow recapitulation process to constrain recollection on the shallow test, yielding greater subsequent memory for the foils on the deep test. Importantly, because participants would attempt to constrain their search for all test items, this model predicts that the foil-elaboration effect would not interact with foil strength (Figure 1a). In contrast, the postretrieval monitoring model predicts that participants would selectively engage in a recapitulation process only after retrieving ambiguous or nondiagnostic information, predicting an interaction with foil strength (Figure 1b). Foils that are unfamiliar (relative to the target information) could be easily rejected without engaging in more elaborate retrieval monitoring, so that memory for these foils might not differ as a function of the deep or shallow test. However, for foils that are more familiar or more likely to elicit noncriterial recollections, participants would need to engage in more elaborative retrieval monitoring to evaluate this information and determine whether the item also had been studied in the target source. To the extent that this additional processing for stronger foils would involve recapitulation that is more memorable for deep than shallow sources, it would increase the size of the foil-elaboration effect on the foil recognition test.

We report three experiments using this procedure. Experiment 1 investigated the aforementioned hypotheses under conditions where target items were studied the same number of times in the deep and shallow study lists, so that both recollection and familiarity would be greater for deep items. To anticipate, we found a foil-elaboration effect on the foil recognition test (deep > shallow), and this effect was similar in size regardless of foil strength, consistent with the use of a global retrieval orientation strategy. Experiments 2 and 3 further explored these effects by differentially repeating the target items presented in the deep study list (Experiment 2) or the shallow study list (Experiment 3) in order to determine how the relative strength of these two study sources
affected the use of a recapitulation process. We elaborate on the rational for these manipulations in the context of each experiment.

Experiment 1

Experiment 1 had two main goals. First, we wanted to test whether the foil elaboration effect previously observed with the memory-for-foils procedure would extend to a task where people were explicitly tested for the recollection of specific information. Based on our earlier work with this kind of task (Gallo, Meadow et al., 2008), we expected that people would make fewer false alarms on recollection tests when tested for deep items compared to shallow items, because expecting more distinctive recollections should increase retrieval monitoring accuracy. Finding a foil elaboration effect in this task would be theoretically important, because most prior memory-for-foils work used recognition memory tasks that did not explicitly require people to engage in recollection-based responding (but see Halamish et al., 2012, for a cued-recall procedure for words). Thus, it was unclear whether the content-based search process described in earlier work would actually be associated with the reduction of recollection errors, which we assume is one of its main functions. Our second goal was to directly test between the retrieval orientation and postretrieval monitoring models of the recapitulation process described in the Introduction, by experimentally manipulating the strength of the foils in the recollection task.

Method

Participants. A total of 40 undergraduates (25 female) ranging from ages 18–22 ($M = 20.025$, $SD = 1.03$) were recruited from the University of Chicago. All participants were given course credit for their participation.

Materials and design. The pool of stimuli consisted of 240 words used in Gallo et al. (2004). Words were between five and seven letters in length (mean length = 5.73, $SD = .80$) and varied between 1 (e.g., thicket) and 591 (e.g., house) in overall frequency of use as measured by the Kuëra–Francis written frequency (mean frequency = 37.11, $SD = 69.29$). All of the words were concrete nouns. One hundred fifty of the words were divided into 10 groups of 15 words each, to be presented in the different study and test phases of the recollection task, with the remaining 90 words used as nonstudied foils in the final block of the study. For counterbalancing, words were rotated through the different phases of the experiment across participants.

Procedure. The task consisted of four phases (see Figure 2). In the familiarization phase, participants studied 120 words with instructions to memorize the words for a subsequent memory test. During this phase, half of the words were presented one time, and half were presented three times in order to increase word familiarity (McElree, Dolan, & Jacoby, 1999). Within each of these repetition conditions, half of the words would later be studied in one of the criterial study blocks (i.e., targets on the recollection tests), whereas half would not be studied in the criterial study blocks (i.e., foils on the recollection tests). In this way, we were able to ensure that the recollection of an item from the familiarization phase would not help to decide whether it was a target or a foil on the subsequent recollection test phase (i.e., preventing a recall-to-reject or list-based exclusion strategy, cf. Gallo et al., 2004). Words were presented in a random order for 750 ms each, and there was a 500 ms interstimulus interval between each word. Two rest periods were provided for participants, each after 80 words had been presented.

After the familiarization phase, participants were given the study phase for the recollection task (i.e., the criterial study phase). In order to manipulate levels of processing, this phase was divided into two study blocks. During one of these blocks participants judged whether the words contained the letter $e$ (the “shallow processing” study block), and during the other they judged whether the words were pleasant (the “deep processing” study block). Approximately half of the words in each set contained the letter $e$. Each of these words was presented only once during one of these two study blocks (30 in each block), and all of these words were
earlier presented in the familiarization phase (half had earlier been presented once, and half had earlier been repeated three times). Participants were required to wait 500 ms before making each judgment, but were allowed as much time as necessary to do so. The order of these two study blocks was counterbalanced across participants.

After the criterial study phase, participants were given the criterial recollection test phase. This phase was divided into two different test blocks, allowing us to explicitly manipulate the retrieval demands on these tests. During the shallow recollection test, participants were instructed to determine whether the given word had been earlier presented in the shallow processing study block (i.e., if they earlier made a judgment as to whether or not the word contained the letter e). For the deep recollection task, they were instructed to determine whether the word had been presented in the deep processing study block (i.e., if they earlier made a judgment as to whether or not the word was pleasant). Responses were self-paced. For counterbalancing purposes, the order of these two study blocks was yoked to the order of the corresponding criterial study blocks.

Each recollection test block contained five kinds of items. On the deep test, there were 30 targets. Each target had been presented once in the deep study phase, with half also having been presented once in the familiarization phase, and half having been presented three times in the familiarization phase. There were 45 foils, with 15 presented once in the familiarization phase, 15 presented three times in the familiarization phase, and 15 nonstudied (not presented in any prior phase). Because both targets and foils had been presented in the familiarization phase, participants needed to recollect the study judgment in order to differentiate targets from foils, and they were explicitly instructed to do so. The structure of the shallow test was similar, with the exception that the targets originated from the shallow criterial study block. Note that the recollection task used by Gallo, Meadow et al. (2008) held the kinds of test items constant across test blocks and only varied the task instructions, so that different kinds of items served as foils on the different test blocks. In the current design we held the nature of the foils constant across the recollection test blocks, so that we could more systematically manipulate foil strength.

The final phase of the experiment was the foil recognition test, adapted from Jacoby and colleagues (Jacoby, Shimizu, Daniels et al., 2005). This test was administered after the recollection tests and a 20-minute delay (filled with unrelated personality questionnaires) to minimize ceiling effects. There were a total of 180 words in this task, half of which were the foils presented during the two recollection tasks (old foils), and half of which had not been presented in any prior phase of the experiment (new foils). Of the 90 old foils, half had been foils on the deep recollection test and half had been foils on the shallow recollection test. Moreover, within each of these subsets, a third had been presented one time on the familiarization phase, a third had been presented three times on the familiarization phase, and a third were not presented in the familiarization phase. Participants were instructed to make a recollect/familiar/new judgment on whether or not the word had been presented during any previous phase of the experiment, using a modification of the “remember”/“know” task (Tulving, 1985). Participants were instructed to respond “recollect” when they could retrieve specific contextual details associated with the test item’s earlier presentation. They were instructed to respond “familiar” when they recognized the item as studied due to a feeling of familiarity, but could not recollect specific details associated with its prior presentation. They were instructed to respond “new” if they believed the item had not been seen before in any phase of the experiment.

Results and Discussion

The results are divided into two sections. The first section reports results from the criterial recollection tests, which conceptually replicated the levels of processing effects obtained on hits (deep test > shallow test) and false alarms (shallow test > deep test) in Gallo, Meadow et al. (2008). We also found that repeating foils in the familiarization phase increased false alarms on these recollection tests, and this effect was attenuated on the deep test.

Figure 2. Order of the different task blocks in Experiment 1 with arrows representing the trajectory of different item types in the design. For the criterial recollection tests, both targets and foils were studied in the familiarization block but only targets were studied in the blocks with shallow or deep judgments. The foil recognition test included foils from each recollection test along with new foils. In the figure the shallow blocks proceeded the deep blocks, but the order of these blocks was counterbalanced across participants.
relative to the shallow test, implicating a recollection-based monitoring process. The second section reports the results from the foil recognition test, in order to find evidence for elaborative foil processing on the recollection tests, and also to test between the retrieval orientation and postretrieval monitoring models. Unless specified otherwise, all differences reported in this paper were significant using a two-tailed test and the $p < .05$ threshold.

**Criteria recollection tests.** Means for the recollection tests are presented in Table 1. Within each of the recollection tests, participants demonstrated significant discrimination between targets and foils ($p < .01$). A 2 (test block: deep, shallow) $\times$ 3 (familiarization repetition: 0×, 1×, 3×) ANOVA on foil false alarms revealed a significant effect of test, $F(1, 39) = 58.73$, $MSE = .020$, $\eta^2_p = .601$, demonstrating that false alarms were lower on the deep test compared to the shallow test (i.e., a distinctiveness effect). There also was a significant effect of repetition, $F(2, 78) = 29.70$, $MSE = .011$, $\eta^2_p = .432$, demonstrating that repeated foils were more likely than nonrepeated foils to be incorrectly endorsed on the recollection tests, and there was a significant interaction with test, $F(2, 78) = 6.99$, $MSE = .007$, $\eta^2_p = .152$, demonstrating that the effect of foil repetition was greater on the shallow test compared to the deep test. This interaction suggests that familiarity or noncriterial recollection was less likely to affect performance when the targeted source was more distinctive (deep judgments) compared to less distinctive (shallow judgments), although participants likely were engaged in recollection-based responding even on the shallow test, because they discriminated between targets and repeated foils that should have been similar in familiarity.

A 2 (test block: deep, shallow) $\times$ 2 (familiarization repetition: 1×, 3×) ANOVA on targets revealed a significant effect of test, $F(1, 39) = 202.17$, $MSE = .031$, $\eta^2_p = .838$, demonstrating the levels-of-processing effect (deep > shallow), but no effect of repetition and no interaction (both $F$s $< 1$). The lack of a repetition effect is consistent with the idea that target hit rates are less sensitive to manipulations of familiarity or noncriterial recollection than foil false alarm rates, potentially because targets are more likely than foils to elicit accurate recollection of criterial information (cf. Gallo, 2013). We also computed discrimination scores (target hits minus foil false alarms, matching targets and foils on repetitions) for each test condition. A 2 (test block: deep, shallow) $\times$ 2 (familiarization repetition: 1×, 3×) ANOVA revealed an effect of test (deep > shallow), $F(1, 39) = 345.94$, $MSE = .036$, $\eta^2_p = .899$, and an effect of repetition, $F(1, 39) = 4.53$, $MSE = .036$, $\eta^2_p = .104$, driven by the strong effect of repetitions on false alarms. The interaction was not significant for discrimination scores, $F(1, 39) = 3.16$, $p = .08$.

Analysis of response latencies revealed that participants were overall faster to respond to foils on the deep test (1,138 ms) than on the shallow test (1,239 ms), $t(39) = 2.25$. Consistent with Gallo, Meadow et al. (2008), this result indicates that monitoring retrieval for more distinctive recollections can lead to faster responding, and it suggests that participants had engaged in more retrieval monitoring effort (i.e., more search or decision time) on the shallow test than on the deep test. However, as discussed in the Introduction, this result does not speak to the qualitative nature of retrieval monitoring on either of these tests, such as whether a recapitulation process was used and how this process was implemented (i.e., a global retrieval orientation for all items or a selectively applied postretrieval monitoring process).

**Foil recognition test.** On the foil recognition test, all of the foils from the criterial recollection task were targets (old foils), and items that had not been seen in any prior session served as new foils (false alarm rate of .20). Overall recognition hit rates to old items are presented in Figure 3 (i.e., the proportion of foils in each experimental condition that was recognized with either a “Recollect” or “Familiar” judgment). To avoid item-selection effects, we analyzed all foils on these tests, regardless of the response that participants had given to these foils on the earlier criterial recollection tests, although analysis restricted to foils that were initially rejected yielded an identical pattern of statistical results and conclusions for recognition judgments and recollection judgments. On

### Table 1

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Condition</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep recollection test</td>
<td>No repetitions</td>
<td>.88 (.02)</td>
<td>.95 (.01)</td>
<td>.70 (.03)</td>
</tr>
<tr>
<td></td>
<td>Repeated deep</td>
<td>.90 (.02)</td>
<td>.95 (.01)</td>
<td>.76 (.03)</td>
</tr>
<tr>
<td></td>
<td>Repeated shallow</td>
<td>.02 (.01)</td>
<td>.04 (.01)</td>
<td>.03 (.01)</td>
</tr>
<tr>
<td>Shallow recollection test</td>
<td>No repetitions</td>
<td>.02 (.01)</td>
<td>.04 (.01)</td>
<td>.05 (.01)</td>
</tr>
<tr>
<td></td>
<td>Repeated deep</td>
<td>.09 (.02)</td>
<td>.09 (.02)</td>
<td>.09 (.02)</td>
</tr>
<tr>
<td></td>
<td>Repeated shallow</td>
<td>.50 (.03)</td>
<td>.50 (.03)</td>
<td>.84 (.02)</td>
</tr>
</tbody>
</table>

**Note.** In addition to familiarization phase repetitions (0×, 1×, 3×), deep and shallow targets were studied 1× each in the target study phase in Experiment 1, deep targets were studied 5× in the target study phase in Experiment 2, and shallow targets were studied 3× with the reverse-cuing procedure in the target study phase in Experiment 3. Standard errors of each mean are in parentheses.
overall recognition, a 2 (test block: deep, shallow) \times 3 \text{(familiarization replications: 0, 1, or 3)} ANOVA revealed a significant effect of test, $F(1, 39) = 37.09$, $MSE = .013$, $\eta^2 = .487$. This effect demonstrates that participants had better memory for foils presented on the deep test than on the shallow test, thereby extending the foil-elaboration effect initially demonstrated by Jacoby, Shimizu, Daniels et al. (2005; Jacoby, Shimizu, Velanova et al., 2005) to a recollection task. Not surprisingly, there also was a significant effect of foil repetition, $F(2, 78) = 36.01$, $MSE = .020$, $\eta^2 = .480$, such that increasing the number of times a word was presented in the familiarization phase increased the likelihood of it being recognized on the foil recognition test. Critically, there was no interaction, $F(2, 78) = 2.30$, $MSE = .010$, $p = .11$, $\eta^2 = .056$. The lack of an interaction is consistent with a global retrieval orientation model, in which participants engaged in the same kind of elaborative retrieval monitoring for all of the foils on the deep test, regardless of foil strength. In fact, the foil-elaboration effect was somewhat reduced with stronger foils, which is clearly contrary to the prediction that stronger foils would elicit more elaborate retrieval monitoring (i.e., a postretrieval monitoring model).

We next analyzed old foils that were given a “recollection” response, which also are presented in Figure 3 (i.e., the proportion of foils in each experimental condition that was recognized for a “Recollect” judgment). It has been argued that elaborative retrieval monitoring processing should be most likely to affect subsequent recollection (Marsh et al., 2009), so these responses potentially provided a stronger test of our retrieval monitoring hypotheses. As with overall recognition results, there was a significant effect of test, $F(1, 39) = 32.89$, $MSE = .024$, $\eta^2 = .458$, a significant effect of foil repetition, $F(2, 78) = 45.28$, $MSE = .031$, $\eta^2 = .537$, and no interaction, $F < 1$. Finally, we analyzed “familiarity” responses given to old foils, using the IRK procedure to estimate familiarity (see Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). Seven participants were excluded from this analysis because they made no “familiar” judgments. For deep items, mean IRK = .52 (0×), .66 (1×) and .64 (3×); for shallow items, IRK = .42 (0×), .53 (1×) and .59 (3×). These data revealed a significant effect of test, $F(1, 32) = 8.58$, $MSE = .032$, $\eta^2 = .214$, a significant effect of repetition, $F(2, 64) = 8.70$, $MSE = .059$, $\eta^2 = .214$, and no interaction, $F < 1$. As with the other analyses, these results were more consistent with the use of recapitulation to guide retrieval orientation on the criterial recollection tests, rather than to guide a postretrieval monitoring strategy.

**Manipulation check.** To confirm that our deep judgments were associated with more recollective detail than shallow judgments, we administered the same familiarization and study phase as in the current experiment to a new group of 20 participants from the same population. However, instead of administering the current series of tests, we presented the items from the criterial recollections tests for a standard recognition memory judgment, followed by subjective judgments of recollection and familiarity (see Gallo, Meadow et al., 2008 for more details on this procedure). The critical result from this procedure was that deep items were recollected more often than shallow items, and they also were rated as containing more uniquely recollected details than shallow judgments on a 1–5 scale (3.45 vs. 2.57), $t(17) = 6.01$, confirming that deep items were associated with more distinctive recollections.

**Summary**

Experiment 1 revealed that testing for deep recollections reduced false alarms relative to shallow recollections. These results are consistent with Gallo, Meadow et al. (2008), showing that expecting more distinctive recollections on the deep test enhanced the accuracy of retrieval monitoring. These retrieval monitoring effects also were associated with differences in foil memory on the foil recognition test (deep > shallow). This foil-elaboration effect cannot be attributed to more processing time, as participants were faster to respond to foils on the deep test than on the shallow test (cf. Jacoby, Shimizu, Velanova et al., 2005; Marsh et al., 2009). Instead, this foil-elaboration effect shows that retrieval monitoring on the deep test involved more elaborative processing of the foils (i.e., a deep recapitulation process) than retrieval monitoring for shallow test. Critically, the foil strength manipulation did not interact with the foil-elaboration effect, a pattern that was predicted by the global retrieval orientation model but not the postretrieval monitoring model.

**Experiment 2**

The purpose of Experiment 2 was to attempt to replicate and extend the key patterns of Experiment 1. Given that our interpretation of Experiment 1 hinges on the lack of an interaction, we
wished to replicate this pattern to ensure that is reliable. We also
wanted to test whether this pattern would persist under conditions
that might be even more favorable to a postretrieval monitoring
model. According to one version of that model, participants first
attempt to accept (or reject) test items based on a general feeling
of familiarity (or lack thereof), selectively engaging in more elabor-
ate recollection-based monitoring only if insufficient or ambigu-
ous information is initially retrieved (cf. Atkinson & Juola, 1974).
It could be argued that Experiment 1 was biased against this
strategy on the deep test because familiarity or noncriterial infor-
meanation was not perceived by participants to be a sufficiently
accurate basis for making a memory decision. That is, if some of
the strongest foils were perceived as similar in familiarity to the
weakest targets, then participants might have decided to use a
global recapitulation strategy to initiate recollection attempts for
test items, in order to avoid errors to any of the foils.
To explore this possibility, Experiment 2 further exaggerated the
familiarity difference between targets and foils on the deep test by
repeating all of the items in the familiarization phase five times, while
keeping all other aspects of the design identical to Experiment 1.
As a result of this manipulation, the total number of presentations
of deep items (1X or 3X in the familiarization phase, plus 5X in the
deep study phase) would be at least twice as many as the foils
(0X, 1X or 3X in the familiarization phase), so that familiarity
would have been even more likely to differentiate between these
sources compared to Experiment 1. (For relevant evidence that
participants can differentiate between these ranges of repetitions in
memory, see Rowe, 1974.) By making familiarity more diagnostic of
the study source, we reasoned that this procedure might further moti-
ivate participants to use familiarity (or the lack thereof) to inform
their memory decisions on the deep test.
According to the postretrieval monitoring model, participants
could quickly reject the weakest foils on the deep test due to a lack
of familiarity or recollection (e.g., those that were never presented
in an earlier phase), but for the stronger foils they should be more
likely to engage in more elaborate retrieval monitoring to help
determine if the item had been studied in the criterial source. If
participants adopted this strategy, then the size of the foil-
elaboration effect on the foil recognition test should interact with
foil strength. In contrast, obtaining the same size foil elaboration
effect regardless of foil strength would bolster the conclusion that
participants prefer to use recapitulation to guide retrieval orienta-
tion during all recollection attempts, effectively disregarding fa-
miliarity even when it is made more diagnostic of the study source.
Of course, repeating the deep items also should enhance the
likelihood that these items are successfully recollected relative to
shallow items, but deep items would continue to elicit more distinctiverecollected and so the same kind of elaborative
content-specific monitoring process (deep > shallow) should still
apply.2

**Method**

**Participants.** A total of 40 undergraduates (26 female) rang-
ing from ages 18–22 (M = 19.925, SD = 1.05) were recruited
from the University of Chicago. All participants were given course
credit for their participation.

**Materials and design.** The same stimuli and equipment were
used as in Experiment 1.

**Procedure.** The procedure of Experiment 2 was identical to
that of Experiment 1, with the exception of the deep study phase.
In the deep study phase, the words were repeated five times each
(nonconsecutively).

**Results and Discussion**
As with Experiment 1, we first report results from the criterial
recollection tests, followed by the foil recognition test.

**Criterial recollection tests.** Mean responses for the recollec-
tion tests are presented in Table 1, and the overall pattern of
responses replicated Experiment 1. A 2 (test block: deep, shallow) x
3 (familiarization repetitions: 0X, 1X, 3X) ANOVA on foil false alarms revealed a significant effect of test, F(1, 39) = 70.46, MSE = .029, \( \eta^2_p = .644 \), demonstrating that false alarms were lower on the deep test compared to the shallow test. There
also was a significant effect of repetitions in the familiarization
phase, F(2, 78) = 29.50, MSE = .013, \( \eta^2_p = .431 \), and a significant interaction with test, F(2, 78) = 10.27, MSE = .010, \( \eta^2_p = .208 \), demonstrating that foil repetition was less likely to affect false
alarms on the deep test compared to the shallow test. As with
Experiment 1, this interaction suggests that foil strength was less
likely to affect performance when the targeted source was more
distinctive (deep judgments) compared to less distinctive (shallow
judgments).

As with Experiment 1, a 2 (test block: deep, shallow) x 2
(familiarization repetitions: 1X, 3X) ANOVA on target hits reve-
al a significant effect of test, F(1, 39) = 286.12, MSE = .026,
\( \eta^2_p = .880 \), demonstrating the levels-of-processing effect (deep >
shallow). There was no effect of familiarization phase repetitions,
F(1, 39) = 1.62, p = .21, and no interaction (F < 1), as targets
again were less sensitive to this manipulation than foils. Analysis of
discrimination scores (target hits minus foil false alarms, match-
ing targets and foils on repetitions) revealed a significant effect of
test (deep > shallow), F(1, 39) = 391.59, MSE = .044, \( \eta^2_p = .909 \), and a significant effect of repetition, F(1, 39) = 4.30, MSE = .020,
\( \eta^2_p = .099 \), driven by the strong effect of repetitions in the false
alarms. The interaction was not significant for discrimination
scores (F < 1).

**Foil recognition test.** Recognition rates for foils are presented in
Figure 4, and overall these patterns replicated those of Experi-
ment 1 (the false alarm rate to new foils was .18). As with
Experiment 1, we analyzed all foils regardless of the original
response to these items on the criterial recollection test, and
alternative analysis of only the foils that were initially rejected did
not change the statistical results or conclusions for recognition or
recollection judgments. On overall recognition hits, there was a
significant effect of test, F(1, 39) = 11.90, MSE = .023, \( \eta^2_p = .234 \), a significant effect of repetition, F(2, 78) = 42.44, MSE = .016, \( \eta^2_p = .521 \), and no interaction (F < 1). These results again
demonstrate the foil-elaboration effect (deep > shallow), and this

---

2 A reviewer wondered whether increasing the recollection of deep items
might encourage participants to engage in a global retrieval orientation
strategy to target these recollections. We consider this unlikely, because
repeating the deep items should increase the likelihood that targets would
elicit recollections without an elaborative recapitulation strategy. Thus,
participants might see even less need to engage an elaborative recapitula-
tion when encountering an unfamiliar foil, although they might use this
strategy for stronger foils (i.e., the postretrieval monitoring model).
significant effect of repetition, observed in Experiments 1 and 2 (and in prior work) was due to tests. To the extent that the foil-elaboration effect on the foil test items (via study repetition), so that the levels-of-processing effect strategy guide a global retrieval orientation.

regardless of foil strength, consistent with the use of a recapitulation again did not interact with the foil-elaboration effect. These favorable to the use of familiarity and hence a postretrieval monitoring for all of the foils on the deep recollection test,

Summary

Experiment 2 replicated the foil-elaboration effect found in Experiment 1, even though there were more extreme differences in familiarity between the deep targets and foils in the current study. According to one view, these conditions should have been more favorable to the use of familiarity and hence a postretrieval monitoring strategy than in Experiment 1, but the foil strength manipulation again did not interact with the foil-elaboration effect. These findings suggest that participants engaged in more elaborative retrieval monitoring for all of the foils on the deep recollection test, regardless of foil strength, consistent with the use of a recapitulation strategy guide a global retrieval orientation.

Experiment 3

The goal of Experiment 3 was to further test the retrieval orientation model by selectively increasing memory of the shallow items (via study repetition), so that the levels-of-processing effect on studied items would be minimized on the criterial recollection tests. To the extent that the foil-elaboration effect on the foil test observed in Experiments 1 and 2 (and in prior work) was due to more elaborative recapitulation on the deep test than on the shallow test, minimizing the recollection advantage of deep over shallow items also might minimize the foil-elaboration effect. More specifically, in addition to increasing familiarity, repetition of the shallow items during study should increase the opportunity to bind these items to contextual details, thereby increasing the amount of contextual information that participants might expect to retrieve for shallow items. If participants attempt to retrieve this information for all test foils on the shallow recollection test, as in a global retrieval orientation model, then this additional processing would reduce or eliminate the advantage of deep over shallow testing on the subsequent foil recognition test.

To minimize the levels-of-processing effect we used a similar procedure as in Gallo, Meadow et al. (2008), which involves presenting the study words prior to the levels-of-processing cue (cf. Craik, 1977), and also repeating each shallow item during the study phase. Gallo, Meadow et al. (2008) found that this procedure reversed the levels-of-processing effect on target hits (shallow test > deep test) on criterial recollection tests, indicating that shallow items were more memorable than deep items. Nevertheless, an independent measure of distinctiveness indicated that deep recollections were still rated as containing more unique or distinctive details than shallow recollections, and on the recollection task false alarms on the deep recollection test were still significantly lower than on the shallow recollection test. Based on these and other results, Gallo, Meadow et al. (2008) concluded that repeating the shallow items increased familiarity and recollection frequency (e.g., the number of shallow items associated with a recollection experience), but that the quality of the recollection experience continued to be richer or more distinctive for deep study judgments compared to shallow study judgments. As a whole, their results indicated that expectations about recollection quality contributed more heavily to retrieval monitoring than did expectations about familiarity or recollection frequency (see also Scimeca, McDonough, & Gallo, 2011).

Based on these results, we reasoned that repeating the shallow items in the current study would make them more familiar and easier to recollect, but would not necessarily make shallow judgments more distinctive than deep judgments. Thus, if distinctiveness of the study judgments themselves were the only factor contributing to retrieval monitoring, then we should find the same foils effects as in our previous experiments (cf. Marsh et al., Experiment 3). However, in the current design, the repetition of items across the familiarization phase and the main study phase also should have increased the likelihood that participants would be recursively reminded of earlier presentations (cf. Hintzman, 2010; Jacoby & Wahlheim, 2013). This repetition-related information would increase the likelihood of binding these items to other contextual details (e.g., thoughts about the earlier presentation of the item, noting it was repeated, etc.), providing additional details that participants could attempt to retrieve on the recollection test. Critically, this effect would be greatest for the shallow items, which were associated with less distinctive judgments but more study repetitions than deep items. Because shallow judgments themselves are relatively nondistinctive, this repetition-related information might be the most salient aspect of shallow item presentation. Although this repetition-related information might not benefit retrieval monitoring accuracy on the recollection tests as much as more distinctive deep judgments, attempting to

Figure 4. Proportion of foils that were recognized (recollection + familiarity judgments) and recollected (recollection judgments) on the foil recognition test of Experiment 2. Bars represent standard error of the mean.
recollect this additional information on the shallow test might nevertheless increase the processing of shallow foils, thereby minimizing the difference between deep and shallow foils on the final foil recognition test.

Method

Participants. A total of 40 undergraduates (29 female) ranging from ages 18–23 (M = 19.9, SD = 1.4) were recruited from the University of Chicago. All participants were given course credit for their participation.

Materials and design. The same stimuli and equipment were used as in Experiment 1.

Procedure. The task consisted of the same four phases as Experiment 1 (see Figure 2). The familiarization phase was identical to that in prior experiments. After the familiarization phase, participants were given a criterial study phase. In order to reverse the levels of processing effect for hits, we used the reverse cuing procedure. With this procedure a word first appeared on the screen for 1 second, followed by a 1-s blank screen. During this time participants needed to hold the word in mind for the upcoming judgment, thereby allowing some degree of lexical processing in all conditions. Participants were then prompted to either determine whether or not the given word contained the letter e (shallow) or whether the word was pleasant (deep), in order to explicitly encourage either a shallow or a deep level of processing. Unlike Experiment 1, this procedure required that we randomly mix deep and shallow items within the same list (cf. Gallo, Meadow et al., 2008), so that participants could not anticipate the judgment for any item. We also repeated all of the shallow items three times in this list, to further increase their memorability relative to deep items, which were only presented once. As in Experiment 1, half of the items in each level of processing condition had been presented once in the familiarization phase, and half had been presented three times in the familiarization phase. Thus, the shallow items had been presented a total of four or six times prior to taking the recollection tests. After the criterial study phase, participants took two recollection tests and then a final foil recognition test, analogous to those given in Experiment 1.

Results and Discussion

As with Experiment 1 and Experiment 2, we first report results from the criterial recollection task, followed by the foil recognition test.

Criterial recollection tests. Mean responses from the criterial recollection tests are presented in Table 1. A 2 (test block: deep, shallow) × 3 (familiarization repetitions: 0×, 1×, 3×) ANOVA on foil false alarms revealed a significant effect of test, F(1, 39) = 44.05, MSE = .016, ηp² = .530, a significant effect of familiarization, F(2, 78) = 35.02, MSE = .012, ηp² = .473, and a significant interaction, F(2, 78) = 9.82, MSE = .010, ηp² = .201. These findings demonstrate that foil strength effects on false alarms were reduced on the deep test relative to the shallow test. Similar to the effects observed in Experiment 1 and 2, these effects suggest that participants expected more distinctive recollections on the deep test than on the shallow test, thereby increasing retrieval monitoring accuracy.

Although we replicated the false alarm effects in Experiments 1 and 2, we also were successful in minimizing the levels-of-processing effect on studied items. Analysis of targets revealed a significant effect of test, F(1, 39) = 17.35, MSE = .027, ηp² = .308, as hits to shallow items were greater than hits to deep items. As in the other experiments, there was no significant effect of repetition in the familiarization phase, F(1, 39) = 3.10, p = .09, and no interaction, F(1, 39) = 3.81, p = .06. We also analyzed discrimination scores (target hits minus foil false alarms, matching targets and foils on repetitions). A 2 (test block) × 2 (repetition) ANOVA revealed no effect of test (F < 1) or repetition, F(1, 39) = 3.68, p = .07, although there was a significant interaction, F(1, 39) = 11.20, MSE = .013, ηp² = .223. This interaction indicates that discrimination on the shallow test was matched with the deep test for items that were presented once in the familiarization phase (deep = .65, shallow = .68), but for items that were repeated in the familiarization phase, discrimination was greater on the deep test than the shallow test (deep = .67, shallow = .58).

Foil recognition test. As with Experiments 1 and 2, we analyzed all foils regardless of the original response to these items on the criterial recollection test, and alternative analysis of only the foils that were initially rejected did not change the statistical results or conclusions for recognition or recollection judgments. In sharp contrast to Experiments 1 and 2, there was no foil elaboration effect in this experiment (see Figure 5). A 2 (test block: deep, shallow) × 3 (familiarization repetitions: 0, 1, or 3) ANOVA on overall recognition rates revealed a significant effect of foil repetition, F(2, 78) = 39.93, MSE = .013, ηp² = .506, but no effect of test or interaction (both Fs < 1). Analysis of recollection responses also revealed an effect of foil repetition, F(2, 78) = 54.21, MSE = .027, ηp² = .582, but no significant effect of test (F < 1) or interaction, F(2, 78) = 3.06, p = .053. Analysis of IRK familiarity estimates, for deep items, IRK = .55 (0×), .58 (1×) and .62 (3×), and for shallow items, IRK = .59 (0×), .61 (1×) and .69 (3×), showed this same pattern of results, with an effect of repetition, F(2, 78) = 3.99, MSE = .042, ηp² = .093, but no effect of test, F(1, 39) = 3.06, p = .09, or interaction (F < 1).

Manipulation check. To confirm that the repetition of shallow judgments increased the number of recollected details relative to deep judgments, we computed the proportion of recollected items that were also recollected on a subsequent criterial recollection test. A 2 (test block: deep, shallow) × 3 (familiarization repetitions: 0, 1, or 3) ANOVA on these proportion revealed a significant effect of test, F(1, 39) = 13.03, MSE = .016, ηp² = .258, as recollection was greater on the deep test than on the shallow test (deep = .67, shallow = .53).

Figure 5. Proportion of foils that were recognized (recollection + familiarity judgments) and recollected (recolletion judgments) on the foil recognition test of Experiment 3. Bars represent standard error of the mean.
to deep judgments, we administered the same manipulation check procedure described in Experiment 1 to a new group of 20 participants from the same population. The critical result from this procedure was that shallow items were rated as containing more uniquely recollected details than deep judgments (2.78 vs. 2.52), t(19) = 2.54, confirming that differentially repeating the shallow items increased the amount of recollected details associated with these items, such as noticing more repetitions of the shallow items across the different study phases of the experiment.3

Summary

The elimination of a foil-elaboration effect in this experiment suggests that differentially strengthening the shallow items affected the processing of the foils on the shallow test, effectively increasing the memorability of these items to the level of the foils on the deep test. As discussed, repetition alone should not have made shallow judgments more distinctive than deep judgments, but it could have increased number of noncriterial details that could have been recollected in association with shallow items. By increasing the overall amount of information that participants could have attempted to recollect on the shallow test, repetitions minimized the advantage of deep processing over shallow processing on the final foil recognition test. This pattern was observed at all levels of foil strength, again consistent with the use of a recapitulation strategy guide a global retrieval orientation.

General Discussion

These experiments provide important insights into the nature of the content-specific recapitulation strategy that people use during recollection attempts. We hypothesized that people might implement this strategy either as part of a global retrieval orientation strategy or as part of a more selective postretrieval monitoring strategy. According to the retrieval orientation model, people naturally engage recapitulation to strategically search memory for target information, regardless of the familiarity or noncriterial information that might come to mind. According to the postretrieval monitoring model, people more selectively engage recapitulation only after an initially broad retrieval attempt yields insufficient information that requires additional monitoring. Our results were more consistent with the retrieval orientation model of the recapitulation process, as we found that the foil-elaboration effect (deep > shallow) in both Experiments 1 and 2 was similar in magnitude regardless of the experimental manipulation of foil strength. These findings indicate that participants had engaged in deeper or more elaborate retrieval monitoring for all of the foils on the deep test, regardless of the strength of these foils, in an attempt to orient retrieval toward the target information. We also found that increasing the amount of information that participants could recollect for shallow items eliminated the foil-elaboration effect in each of the foil strength conditions in Experiment 3, again consistent with a retrieval orientation model. These results do not rule out the possibility that people engaged in some kind of postretrieval monitoring in our task, but they do suggest that people’s use of the recapitulation process primarily was to guide a global retrieval orientation.

Based on prior work we assume that the foil elaboration effect observed in our task was due to the engagement of a recapitulation process (e.g., Alban & Kelley, 2012; Danckert et al., 2011; Marsh et al., 2009), but it is worth considering whether an alternative account based on familiarity-based monitoring could explain our results. According to this account, the evaluation of familiarity for test foils is itself memorable, and when deep items are more familiar than shallow items, participants are more likely to engage in familiarity-based monitoring on the deep test than the shallow test. We see this account as unlikely for two reasons. First, foil strength did not affect the size of the foil-elaboration effect in any of our experiments, even though stronger foils should have required more familiarity-based monitoring by this account. Second, the difference in familiarity between the study sources was greater in Experiment 2 than in Experiment 1, which should have caused differences in familiarity-based monitoring, but the size of the foil-elaboration effect was similar across these experiments. As a whole, our foil recognition effects (deep > shallow) were more consistent with the use of a content-based recapitulation process than differences in familiarity-based monitoring.

Although we did not directly investigate the strategic nature of the recapitulation process, Alban and Kelley (2012) reported evidence that this process is under strategic control. They found that initially testing participants on an easy recognition test reduced the likelihood that they would engage a recapitulation process on a subsequent test, ostensibly because the easy test biased participants to use familiarity-based responding as opposed to recollection (i.e., a “good enough” strategy as described by Alban & Kelley, 2012). We found no evidence that test block order interacted with the foil elaboration effects reported here.4 Of course, our test item composition and task instructions explicitly required participants to make use of recollection. Under these conditions, people apparently use more elaborative retrieval monitoring as their default strategy for guiding recollection attempts. This is not to say that all instances of successful recollection require a recapitulation strategy, as some targets might have automatically triggered the recollection of associated details before participants could engage a recapitulation strategy. Nevertheless, in those instances where retrieval cues did not automatically elicit the recollection of target information, as was the case for foils in our procedure, the default strategy seems to have been to engage in a recapitulation process to guide retrieval orientation.

3 Gallo, Meadow et al. (2008) found that the differential repetition of shallow items over deep items in a single study list did not overcome the levels of processing advantage on ratings of recollected detail (i.e., deep > shallow), suggesting that deep judgments were more distinctive than shallow judgments. As discussed, the reversal of this pattern in the current experiment (i.e., shallow > deep) is likely due to the more complex design, which repeated shallow items across different study lists.

4 In overall foil recognition, there were no significant interactions between criterial test block order and either of the effects of interest (levels of processing on the criterial tests or foil strength) in any of the three experiments. There were no interactions in recollection judgments for Experiments 1 and 3, although there was a three-way interaction with order in Experiment 2. To unpack this interaction, we separately analyzed data from subjects in each order condition. We again observed the expected effects for subjects in the deep first condition: main effect of levels of processing, F(1, 19) = 12.97, main effect of repetition, F(2, 38) = 22.78, no interaction, F(2, 38) = 1.47, p = .24, but for subjects in the shallow first condition, there was a significant Levels × Repetition interaction, F(2, 38) = 3.50, as the foil-elaboration effect was greatest for the weakest foils. We have no good explanation for this last effect, but clearly it is inconsistent with the postretrieval monitoring prediction.
Implications for Dual-Process Theories

Our task explicitly encouraged the recollection of targeted information. From a dual-process perspective, it may not seem surprising that participants attempted to disregard familiarity in such situations and instead focused on recollection. However, our results are surprising in light of the fact that participants could have used familiarity as a cue for deciding to engage in a recapitulation strategy. We presented target items in both a familiarization phase and the criterial study lists, so targets should have been more familiar than foils on average. This difference was most pronounced in Experiment 2, where each of the deep targets had been studied at least six times prior to the recollection test. As a result of these extreme familiarity differences, participants could have readily rejected nonstudied foils on the recollection tests based on a lack of familiarity alone, which is often assumed from dual-process theories to be a relatively faster and less effortful means of responding compared to recollection-based processes (cf. Atkinson & Juola, 1974; Yonelinas & Jacoby, 2012). Nevertheless, our results indicate that people preferred to engage in an elaborative recapitulation process for all foils, apparently because they considered this to be the most effective way to guide retrieval orientation and hence control retrieval.

The fact that our criterial recollection task encouraged the recollection of specific information contrasts with more traditional memory tasks, which can encourage participants to rely on familiarity or unspecified kinds of recollected information. For example, in standard recognition tasks, it is logical for people to attempt to retrieve any relevant information in order to differentiate between studied and nonstudied items, including familiarity and idiosyncratic recollected details. Similarly, although people are more likely to use recollection than familiarity on source memory tasks (e.g., Lindsay & Johnson, 1989), familiarity can be used to infer source in some situations (e.g., Hicks, Marsh, & Ritschel, 2002), and unlike the current procedure, these tasks typically query multiple sources of information at once. Under these conditions, it remains possible that people would choose to engage in an initially broad retrieval search, engaging in a recapitulation strategy only if this initial search yields insufficient information (i.e., a postretrieval monitoring strategy). As discussed in the Introduction, we know of no evidence that definitively favors a retrieval orientation versus a postretrieval monitoring model on these kinds of tests, and this remains an open question. It may be that the kind of retrieval strategy employed by participants will depend on the task demands, which is why we have focused on a task that targets specific recollections instead of a more general recognition task. We believe that recollection tasks more closely resemble situations outside the laboratory where people are interested in the recollection of specific information about past experience.

Implications for False Memory Theories

Our results indicate that people did not use familiarity or non-criterial information to guide their use of a recapitulation strategy, but this does not mean that the retrieval of familiarity or other nondiagnostic information did not affect performance. In all of our experiments, we found that recollection test false alarms increased as a function of foil repetitions, and these effects were greater on the shallow test than the deep test. These results suggest that the retrieval of nondiagnostic information increased false alarms, and they also speak to the nature of this effect. Specifically, in those instances where target recollection had failed, our results do not support the idea that participants strategically used familiarity or noncriterial information to make an inference or guess about the source of an item. If participants were using this kind of inference on the recollection tests, then they should have made more false alarms when the targets and foils were more similar in strength. This was not the case across our experiments. We found that the size of the repetition effect on shallow test false alarms was similar in Experiment 1 and 3, even though the average strength difference between targets and foils on the shallow test should have been much greater in Experiment 3 than Experiment 1. Similarly, Experiment 1 and 2 revealed similar sized repetition effects on deep test false alarms, even though there should have been large strength differences between the deep items across these two experiments.

Instead of inferences or guessing based on nondiagnostic information, it seems more likely that familiarity or noncriterial information affected false alarms in our experiments by creating an illusory or false sense of recollection. The retrieval of nondiagnostic information might have an automatic effect on false recollection, analogous to the idea of fluency misattribution (e.g., Kelley & Jacoby, 1998; Whittlesea & Leboe, 2000) or gist-based phantom recollection (e.g., Brainerd, Wright, Reyna, & Mo乔丹in, 2001). In fact, although the recapitulation process described in the current studies is assumed to reduce recollection errors overall, it also is possible that this process could backfire and cause recollection errors. In the context of the current studies, participants might have engaged in a recapitulation process regardless of the strength of the foils (i.e., the retrieval orientation model). However, if participants actively engaged in mental imagery as part of the recapitulation process, then the retrieval of nondiagnostic information might have enhanced processing fluency for some test foils, causing these imaginations to be experienced as a false recollection (cf. Gallo & Roediger, 2003; Jacoby, Kelley, & Dywan, 1989). Regardless of exact mechanism of false recollection, the current study adds to a growing literature to suggest that familiarity can sometimes contribute to false recollection (see Lampinen, Neuschatz, & Payne, 1999).

In conclusion, our results suggest that people naturally use an elaborative recapitulation strategy when monitoring memory for specific information, regardless of whether they can retrieve or utilize less relevant information. These results provide strong behavioral evidence for the use of a recapitulation process to guide a global retrieval orientation, similar to the early selection strategy discussed by Jacoby et al. (1999) and related ideas that have been proposed in the brain imaging literature (e.g., Dobbins & Han, 2006; Morcom & Rugg, 2012). The present results also highlight a key functional benefit for this kind of recapitulation process: In addition to guiding retrieval toward target information, more elaborative recapitulation also is associated with reduced recollection errors.

References


Received August 7, 2013
Revision received May 28, 2014
Accepted June 4, 2014