Source-constrained retrieval influences the encoding of new information

Stacey L. Danckert · Colin M. MacLeod ·
Myra A. Fernandes

Abstract Jacoby, Shimizu, Daniels, and Rhodes (*Psychonomic Bulletin & Review, 12*, 852–857, 2005) showed that new words presented as foils among a list of old words that had been deeply encoded were themselves subsequently better recognized than new words presented as foils among a list of old words that had been shallowly encoded. In Experiment 1, by substituting a deep-versus-shallow imagery manipulation for the levels-of-processing manipulation, we demonstrated that the effect is robust and that it generalizes, also occurring with a different type of encoding. In Experiment 2, we provided more direct evidence for context-related encoding during tests of deeply encoded words, showing enhanced priming for foils presented among deeply encoded targets when participants made the same deep-encoding judgments on those items as had been made on the targets during study. In Experiment 3, we established that the findings from Experiment 2 are restricted to this specific deep judgment task and are not a general consequence of these foils being associated with deeply encoded items. These findings provide support for the source-constrained retrieval hypothesis of Jacoby, Shimizu, Daniels, and Rhodes: New information can be influenced by how surrounding items are encoded and retrieved, as long as the surrounding items recruit a coherent mode of processing.

Keywords Memory · Recognition · Word recognition

S. L. Danckert (✉) · C. M. MacLeod · M. A. Fernandes
Department of Psychology, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada
e-mail: sdancker@uwaterloo.ca

C. M. MacLeod
e-mail: cmacleod@uwaterloo.ca

M. A. Fernandes
e-mail: mafernan@uwaterloo.ca

Memory researchers often discuss encoding and retrieval processes separately, as if they were independent. Yet we know that they are not: The transfer-appropriate processing principle—that memory retrieval will be best when the processes invoked during retrieval match those undergone during encoding—asserts their intimate connection (see Morris, Bransford, & Franks, 1977; Roediger, 1990). Clearly, then, the two stages are highly interdependent. Indeed, under the proceduralist perspective proposed by Kolers (1973; Kolers & Roediger, 1984; for a review, see Roediger, Gallo, & Geraci, 2002), distinguishing encoding from retrieval is not really meaningful, because every encoding event is a retrieval event, and every retrieval event is an encoding event. For example, if we are at a grocery store trying to recall a grocery list, along with items from that list we also retrieve other information, such as what we plan to cook for dinner. We are also concurrently encoding new information, such as the layout of the store or what is on sale. As a result, it would be unreasonable to label this as purely an encoding event or purely a retrieval event. *Encoding* is simply a convenient shorthand for the first processing of an event; *retrieval* is the shorthand for that event's subsequent processing.

In light of these considerations, it is important to bear in mind the effects that these processing instances (encoding and retrieval) have on each other. Jacoby, Shimizu, Daniels, and Rhodes (2005; see also Jacoby, Shimizu, Velanova, & Rhodes, 2005; Marsh et al., 2009; Shimizu & Jacoby, 2005) have provided insight into this issue with the introduction of what we will call their "memory-for-foils" paradigm. In their Experiment 1, the researchers first had participants study two sets of words, one under deep-encoding instructions (pleasantness judgments) and the other under shallow-encoding instructions (vowel judgments). Then, participants performed a two-part recognition test, one part

containing deep targets and new foils and the other containing shallow targets and new foils. Finally, participants performed another recognition test in which the foils from the first test were now the targets, mixed with new foils. What they observed was that words seen on the first test as foils among the deeply encoded words were better recognized than words seen on the first test as foils among the shallowly encoded words. The encoding of the new words during the first recognition test was significantly influenced by the differential initial processing of the old words with which they were mixed.

To explain their finding, Jacoby, Shimizu, Daniels, and Rhodes (2005) proposed that participants reentered the original encoding mode (i.e., deep encoding or shallow encoding) during the initial recognition test for studied items, so as to constrain the memory search. As such, during the initial test, people assessed each item for evidence of having recently been processed either deeply or shallowly. This reengagement of the initial encoding task led to deep processing of both old and new items when the deeply studied words were the targets, and presumably to shallow processing of all items when the shallowly studied words were the targets. Because the type of processing used in recognizing the old words affected the encoding of the new words, Jacoby, Shimizu, Daniels, and Rhodes proposed the *source-constrained retrieval* hypothesis, claiming that the depth of encoding of new items during a recognition test is a by-product of using an efficient search process as a means of increasing success and decreasing the resources required during retrieval (i.e., constraining memory search to information [e.g., semantic features] consistent with the original encoding mode). Constraining memory searches when possible would be the most efficient way to conserve valuable resources yet still allow newly encountered items to be encoded for future access.

There is much remaining to be learned about how and why this new paradigm produces the effect initially reported by Jacoby, Shimizu, Daniels, and Rhodes (2005). If source-constrained retrieval is indeed responsible, the memory-for-foils effect should be apparent under different encoding modes as long as those modes can be reinstated at the time of retrieval—and not when those modes cannot be reinstated (see Marsh et al., 2009, Exp. 3). This mode reinstatement is the essence of the source-constrained retrieval account, which should therefore apply more generally than only to the levels-of-processing manipulation. In Experiment 1 below, we provide the first demonstration of the generalizability of the effect to another manipulation during encoding. To accomplish this, we used two types of imagery orientation at encoding: (1) pictorial imagery, in which the task is to imagine the referent picture corresponding to the word (see, e.g., Paivio, 1969, 1971) and (2) a relatively novel letter-based imagery task, in

which the participant must imagine a lowercase word in uppercase on each trial. When Hourihan (2008; see also Hourihan & MacLeod, 2011) used this manipulation in her dissertation, pictorial imagery led to much better memory; indeed, she used the letter task as a control condition. We selected imagery because it ranks with levels of processing as the encoding modes that produce the best retention.

Given its distinctiveness, imagery encoding should be readily reenacted during retrieval on the first test. Participants should reinstate the original encoding mode, which should again lead to better subsequent memory for those foils that have been tested among targets encoded via pictorial-based imagery than for those that have been tested among targets encoded via letter-based imagery. If this imagery manipulation failed to produce the memory-for-foils effect, this would call into question the generalizability of the source-constrained retrieval account and, indeed, of the memory-for-foils phenomenon itself. As set out by Jacoby, Shimizu, Daniels, and Rhodes (2005), nothing about the proposed mechanism (i.e., source-constrained retrieval) requires that encoding must involve a levels-of-processing manipulation.

Also important—indeed, critical—to understanding the effect is investigation of qualitative differences during the final test. Marsh et al. (2009) demonstrated, using the remember/know procedure introduced by Tulving (1985), that words that had been initially tested in the context of deep encoding not only were subsequently remembered better, but also had more detail-based memories than their shallowly encoded counterparts. In our Experiment 1, participants were also asked to make a remember/know distinction on the final test. While we make no claims about the relation of these responses to dual-process theories, we agree with Marsh et al. that this discrimination provides a means of separating more detail-based from less detail-based recognition. Following Marsh et al., we would expect—as they in fact found—that more detail-based (or "remember") responses should be given to words initially encoded during the test of deeply encoded words, as compared to those initially encoded during the test of shallowly encoded words. There should be more contextual details available for words encoded in the deep rather than the shallow context.

We then go on, in Experiment 2, to introduce a novel technique to provide more direct evidence for source-constrained retrieval during the initial recognition test. In so doing, we switched back to the levels-of-processing encoding manipulation, to connect more directly to the previous research. Instead of the final recognition test, however, we had participants perform speeded responses to the same question that had been used during encoding of the original targets. If, as source-constrained retrieval would predict, participants are reentering the context of the original encoding mode while they attempt to retrieve items during the initial

recognition test, then they should make that same judgment faster on the words encoded within that context. That is, judging the pleasantness of a foil from the recognition test containing deeply studied items should benefit from this same pleasantness judgment having taken place "behind the scenes" on the initial recognition test, which invoked that mode of processing. This should be evident primarily for foils from the deep test because, relative to the shallow mode, the deep mode provides greater benefit from having more unique associated details. Indeed, we are not convinced that there really is a shallow mode in the same sense that there is a deep mode.

Finally, in Experiment 3, we provide evidence that the results from Experiment 2 are not simply a by-product of some kind of association of the foils with deeply processed items. It might be argued that priming is generally more powerful for deeply associated items than for shallowly associated items. If this were true, foils from the test of deeply encoded items should be responded to more quickly, regardless of the type of judgment performed on those items, and should not be restricted to the same judgment as had been made during the original study. In Experiment 3, therefore, we examined whether there would be any difference between "deep" foil and "shallow" foil response times on an unrelated task: lexical decision. We expected not to see any difference— that is, there should be equivalent priming for the deep and shallow foils—which would provide additional support for the source-constrained retrieval hypothesis. The source constraint should be specific, not general.

These three experiments demonstrate that processes engaged during encoding and reinstated during retrieval can have substantial effects on the encoding of new information. More specifically, the processes invoked when a previously encoded set of items is retrieved can lead to differential encoding of new items encountered during that retrieval. Subsequent memory for these new words derives from them being remembered in terms of their context (i.e., their encoding mode), achieved by the separate groupings on the first test. Such groupings not only make remembering the studied items easier by limiting the types of search engaged, but also cause new items to be more likely to be considered in terms of that context and, as such, to accrue a corresponding benefit. More broadly, such a pattern would fit with the idea that encoding is ongoing within the retrieval process and that the two are intimately linked.

Experiment 1

We first set out to test the generality of the memory-for-foils effect by substituting a different encoding–retrieval mode in place of levels of processing. To accomplish this, we turned to the venerable encoding task of visual imagery (see

Paivio, 1971, 1995, 2007), and specifically to a variant used by Hourihan (2008; see also Hourihan & MacLeod, 2011) in her dissertation. The goal was to create two distinct imagery modes. The first was the standard pictorial imagery task, in which participants are instructed to form a mental picture of the word's referent object; we will refer to this as *deep imagery*. The second was a letter case imagery task, in which participants are instructed to imagine the presented lowercase word all in uppercase; we will refer to this as *shallow imagery*. Deep imagery should result in better memory than shallow imagery, as indeed it did in Hourihan's dissertation.

We expected these two encoding tasks to form coherent processing modes readily invoked again on the separate subtests in Test 1. Consequently, we should see the memory-for-foils pattern observed by Jacoby, Shimizu, Daniels, and Rhodes (2005). This outcome would therefore confirm the robustness and extend the generalizability of the phenomenon, specifically testing whether levels-of-processing tasks are required during encoding or whether any coherent, reproducible processing mode can also generate the effect. A further conceptual replication is provided by including a remember/know judgment in the final recognition test. This type of decision is included to assess the quality of the judgments: More "remember" responses would be linked to recognition associated with greater detail, which we would expect to see associated with foil words from the deep imagery test (similar to the findings of Marsh et al., 2009).

Method

Participants A total of 25 undergraduate students from the University of Waterloo (21 female, 4 male) participated for credit or remuneration ($5). After 1 female participant was removed for failing to comply with instructions on the final test, the mean age was 20.3 years (SD=3.3).

Materials The stimuli consisted of 247 words 5–8 letters in length obtained from the Thorndike and Lorge (1944) norms. The words had an average length of 5.7 letters and an average frequency of 22.1 per million. (Note that these stimuli were from the MRC database, which provided the additional information of word imageability.) All words had moderate to high imageability ratings between 550 and 800. In all phases, the words were presented in lowercase letters. Words were randomly assigned to six lists of 36 words each, with unique randomizations for each participant. In addition, each phase began and ended with three-word "buffers" to discount primacy and recency; these words were not included in any analyses.

Procedure A schematic of the experimental procedure is displayed in Fig. 1. All three phases were participant paced.

Fig. 1 Experiment 1: Schematic of the procedure. Every participant performed the two study sessions, then the two corresponding components of Test 1, and then Test 2
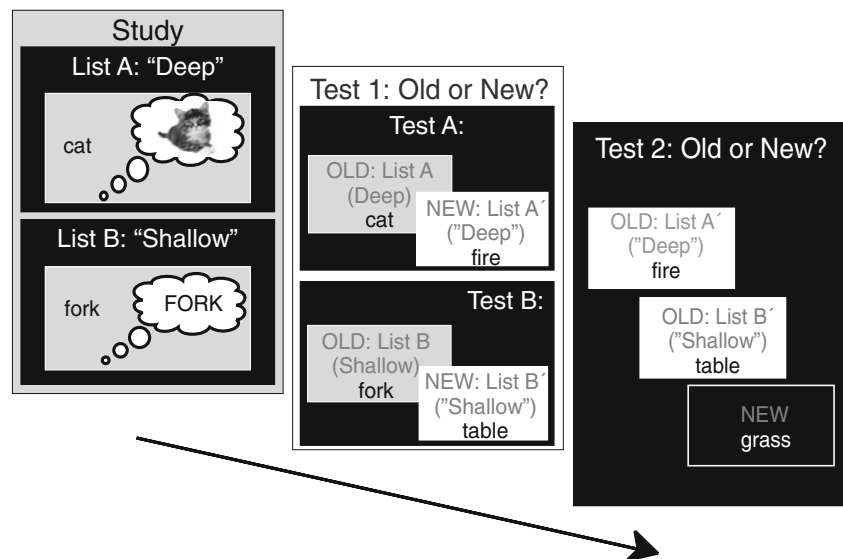
Participants were tested individually and completed the entire experiment in approximately 30 min. All stimuli were displayed in white font on a black computer screen.

The study phase involved two encoding tasks, with their order counterbalanced across participants. In the deep imagery task, 36 words were presented one at a time on a computer screen, and participants were to form a mental picture representing the referent of each item. In the shallow imagery task, participants viewed a different 36-word list and were to form a mental image of each word in capital letters (e.g., for cake: CAKE). Once a participant had created an image, they pressed a key; following this, a fixation cross was presented for 500 ms.

Next came the first recognition phase, Test 1. On two separate 72-item subtests, the 36 deeply imaged words were intermingled with 36 new words, and the 36 shallowly imaged words were intermingled with 36 other new words. The order of the two subtests was counterbalanced across participants, who were explicitly informed which list the old items were drawn from (e.g., "All old words are from the list for which you formed images in your head related to the words"). Participants were asked to press 1 for an old item (target) or 0 for a new item (foil) on the numeric keypad.

Finally, there was the second recognition phase, Test 2. Here, the targets were all of the former foils from the first recognition phase—from both the deep imagery and shallow imagery recognition tests (i.e., no deeply or shallowly encoded items from the study phase were included on Test 2). Intermixed with these newly defined targets was a completely new set of previously unseen words, such that there were 72 old words (36 deep-imagery foils and 36 shallow-imagery foils) and 72 new words. Participants were asked to respond based on the quality of their memories, saying either "remember," "know," or

"new." They were given very careful instruction and practice on deciding whether the words were new, or were old and accompanied by detailed memories (i.e., "remember" response), or were old and not accompanied by any detailed memories (i.e., "know" response). The instructions closely followed those used by Gardiner (1988, p. 311), including the examples they provided.

Results and discussion

Recognition Test 1 Participants were able to recognize the study lists very well across imagery conditions (overall hits=.78, overall false alarms=.11), as is shown in Fig. 2a. A paired-samples t test demonstrated that participants had considerably better overall memory for pictorially imaged words than for words imaged in capitals, $t(23)=6.74$, $p<.001$. This was true for hits, $t(23)=7.44$, $p<.001$, and showed a complementary pattern for false alarms—more false alarms for capital-imaged than for pictorially imaged words, $t(23)=2.19$, $p<.05$. Therefore, participants were effectively using the different encoding techniques, resulting in better encoding for words imaged as pictures than for words imaged in uppercase.

Recognition Test 2 Most importantly, a paired-samples t test demonstrated a significant effect of type of imagery, with better memory for old pictorial foils than for old capital foils, $t(23)=3.41$, $p<.005$. Thus, the memory-for-foils effect generalized to an entirely different form of encoding manipulation. In line with the levels-of-processing finding in Jacoby, Shimizu, Daniels, and Rhodes (2005), there were more hits for foils that had initially been pictorially imaged during the test than for foils that had initially been imaged

Fig. 2 Experiment 1: Manipulating type of imagery at encoding. (a) Recognition data from Test 1, demonstrating enhanced memory following pictorial imagery as compared to capital-letter imagery. (b) Recognition performance for Test 1 foils on Test 2, demonstrating a clear memory-for-foils effect. Error bars represent the standard errors of the corresponding means
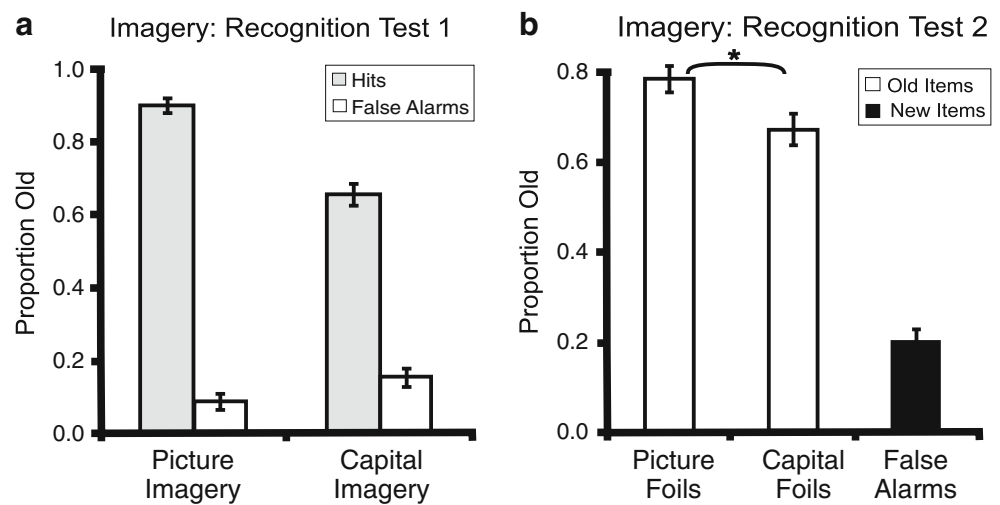
as capitals (Fig. 2b). Further, a two-way ANOVA of the "remember" responses showed a significant interaction, $F(1, 23) = 12.55$, $MSE=.007$, $p < .005$, $\eta_p^2 = .353$. There was a significant main effect of item depth, $F(1, 23)=10.27$, $MSE = .004$, $p<.005$, $\eta_p^2 = .309$, but no effect of response type, $F(1, 23) = 1.65$, $MSE = .053$, $p > .20$, $\eta_p^2 = .067$. Subsequent tests demonstrated significantly more "remember" responses to words from the test of deeply imaged items, as compared to shallowly imaged items, $F(1, 23) = 31.36$, $MSE = .004$, $p < .001$, $\eta_p^2 = .577$. There was, however, no significant effect for items given "know" responses, $F(1, 23) < 1$. These remember/know data are shown in Table 1. When the independent remember/know procedure (Yonelinas, 2002) was applied to the "know" responses, the contribution of these responses was higher overall (.35 and .34 for "deep" and "shallow" foils, respectively), but did not different across foil types.

Experiment 1 demonstrated that the memory-for-foils effect also occurs with an encoding task different from the only one that had previously been used to produce this effect. Participants showed enhanced subsequent recognition for new words tested among words that had been imaged pictorially as compared to new words tested among words that had been imaged in uppercase. This finding is in line with that of Jacoby, Shimizu, Daniels, and Rhodes (2005). In contrast, no memory-for-foils effect was found when encoding was strengthened using repetition, where items were presented once versus three times (Marsh et al., 2009; see

Replication 2 in the Appendix). Clearly, the important requirement is that the mode of encoding be sufficiently coherent that it can be reenacted at the time of retrieval. This mode at retrieval then "spills over" onto the foils, producing an encoding benefit for those that accompanied items that had been deeply encoded in the preceding study phase.

The results of Experiment 1 support the source-constrained retrieval hypothesis of Jacoby, Shimizu, Daniels, and Rhodes (2005), and for the first time demonstrate the generalizability of the memory-for-foils effect. In addition, we confirmed that deep foils make up a higher proportion of "remember" responses than do shallow foils, consistent with Marsh et al. (2009). This suggests that an increase in detail is associated with the foils from the test of deeply encoded words relative to those from the test of shallowly encoded words. This is further supported by research by Gallo, Meadow, Johnson, and Foster (2008), who demonstrated that typical levels-of-processing effects are based on recollective distinctiveness from the extra details that are available for items due to deep encoding. Our argument is that such detail is related to imagery of the items as a consequence of reentry into the picture imagery encoding mode.

Experiment 2

Thus far, there has been no direct evidence for mode reinstatement in any of the reported studies (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005; Marsh et al., 2009). Although better recognition of foils that accompany deeply processed targets is consistent with deeper processing of those foils, which in turn is consistent with a deeper mode of processing, that logic is indirect. In Experiment 2, our goal was to provide a more

Table 1 Experiment 1: Proportions of hits assigned "remember" and "know" responses following imagery-based processing

Response Type	Deep Foils	Shallow Foils
"Remember"	.27	.17
"Know"	.27	.29

direct index of processing mode reinstatement at the time of test. We reasoned that having prior experience at processing an item in a particular way (or in a particular context) should promote faster processing of that same item within that same context as compared to within a different context.

To test this idea, we returned to the typical levels-of-processing study manipulation (i.e., pleasant/unpleasant and "a"/no "a" decisions), for optimal connection to the previous literature, but we changed the final test. In place of the usual recognition test of former foils—Test 2—we substituted a speeded judgment test that involved repetition of the initial encoding question from the study phase, but carried out now on the foil items from Test 1. Half of the foils from the test of deep items and half of the foils from the test of shallow items were presented together with new items for a pleasantness judgment; the same was done for the letter "a" judgment. We predicted that if the foils that had accompanied deep targets had been processed deeply (i. e., for pleasantness), whereas the foils that had accompanied shallow targets had not been processed deeply, then only the deep foils would be faster to judge on the pleasantness judgment task, because only they had effectively already been processed deeply in terms of their pleasantness. Participants were not informed that some test item on this pleasantness judgment task would be old and some would be new, so effectively this was an indirect test, unlike the direct recognition test previously used.

If the memory-for-foils effect were a consequence simply of the former foils having been associated with deeply encoded items, it is unlikely that those items would be faster on a subsequent speeded performance test involving the original deep-encoding question. If, however, the deep foil items undergo processing within the same context as their old counterparts during Test 1, they should be faster to process with respect to pleasantness (the basis of the original deep judgment) than should the shallowly encoded items.

We did not expect a complementary benefit on the shallow judgment task favoring foils that had accompanied shallowly encoded items on the first test because of their relatively weak encoding, and also because we suspected that shallow encoding would not have been sufficient to produce a unique encoding mode that could be successfully reinstated. Nevertheless, to test the alternative hypothesis that accompanying deeply processed items on a prior test always leads to improved memory for foils, we did examine this context by having half of the deep and shallow foils appear on a "contains the letter a" judgment task.

Method

Participants A total of 41 undergraduate students from the University of Waterloo (24 female, 17 male) participated for credit or remuneration ($5). The mean age was 20.8 years

($SD=3.3$). The data of 3 participants were discarded from all analyses due to performing more than two standard deviations slower in the final phase than the mean response time performance for that phase.

Materials The stimulus words were identical to those used in Experiment 1. Two raters rated approximately half (53%) of the items as being pleasant. Of course, due to the subjective nature of such a rating task, there likely would be high variability in such ratings. Similarly, half of the items contained an "a," and the remaining half of the items did not.

Procedure Participants were tested individually and completed the entire experiment in approximately 30 min. Words were randomly assigned to six new lists of 36 words each for each participant. In addition, each task began and ended with an additional three words to minimize primacy and recency effects; these words were not included in any analyses. The order of the tasks within each of the phases was counterbalanced across participants.

In the study phase, participants performed deep- and shallow-encoding tasks on separated word lists. In the deep judgment task, 36 words were presented one at a time on a computer screen, and participants were asked to indicate whether each word represented something pleasant or unpleasant. In the shallow judgment task, participants viewed a different 36-word list and indicated whether each word contained the letter "a." The keyboard responses were 0 for "pleasant" or "a" or 1 for "unpleasant" or "no a." Following the classification response for each word, which was participant paced, a 500-ms fixation cross was displayed before the next word. In the recognition phase, participants performed a recognition test precisely as in Experiment 1, in which they were again provided with instructions describing the source of target foils (e.g., "All old words have come from the list for which you made pleasant/unpleasant decisions").

In the judgment phase, there were two subtasks: pleasantness judgment and letter "a" judgment, which were counterbalanced across participants. For pleasantness judgment, participants repeated the original deep-encoding question used at study ("Is the item pleasant or unpleasant?") for half of the foil items from each of the recognition test lists (18 from the test of deeply encoded items and 18 from the test of shallowly encoded items) intermingled with 36 new items (72 words in total). The remaining deep and shallow foil items from the first recognition phase were mixed with another set of new items, and for these participants responded to the same shallow-encoding question used during study ("Does the word contain an 'a' or no 'a'?"). Thus, both "deep" and "shallow" foils were tested with each judgment task. Additional instructions

requested that participants respond as quickly as possible while performing as accurately as they could. As before, they responded by pressing 1 or 0 on the keyboard. Participants were never instructed as to the nature of the words; that is, they were never told that old words would be appearing among the items during these decision tasks.

Results and discussion

Recognition test Participants performed well on the recognition test of the initially studied lists (overall hits = .75) and readily discriminated these studied words from new words (overall false alarms = .21). These results are displayed in Fig. 3a. A paired-samples *t* test showed that participants had better overall memory for deeply encoded as compared to shallowly encoded words, $t(37) = 10.91$, $p < .001$. This was true for hits, $t(37) = 8.86$, $p < .001$, and showed a mirror effect for false alarms—a greater number of false alarms for shallowly than for deeply encoded words, $t(37) = 4.31$, $p < .001$. Therefore, participants were effectively using the two encoding techniques, resulting in the typical levels-of-processing effect reported by Jacoby, Shimizu, Daniels, and Rhodes (2005) and by Marsh et al. (2009) and replicated in our Appendix.

Judgment task Following one-way ANOVAs, planned contrasts were conducted on the means of the participant median response times for each of the judgment tasks, which together formed the final phase of the experiment. For each judgment task, there were two contrasts, the first examining priming for the previously seen foils, and the second examining whether priming differed between the two types of previously seen foils.

Shallow judgment task On the shallow judgment task, the three conditions—deep foils, shallow foils, and new words—did not differ from each other, $F(2, 74) < 1$. Not surprisingly, therefore, neither planned comparison was significant, both *F*s < 1 (for shallow vs. deep; Fig. 3b). Therefore, priming did not occur either overall, for old versus new words, or differentially, for shallow versus deep test foils. We suspect that the processing carried out in judging whether words contain the letter "a" is so limited that participants cannot benefit from reinstating the vowel-based shallow mode, if indeed there actually is such a mode. We included this condition just for completeness, but did not expect any differential priming of items from the different test lists.

Deep judgment task The task of principal interest was the deep judgment task, since the findings of Jacoby, Shimizu, Daniels, and Rhodes (2005) and Marsh et al. (2009) had suggested that this mode of processing can be reinstated. If the foils presented among deep targets on the recognition
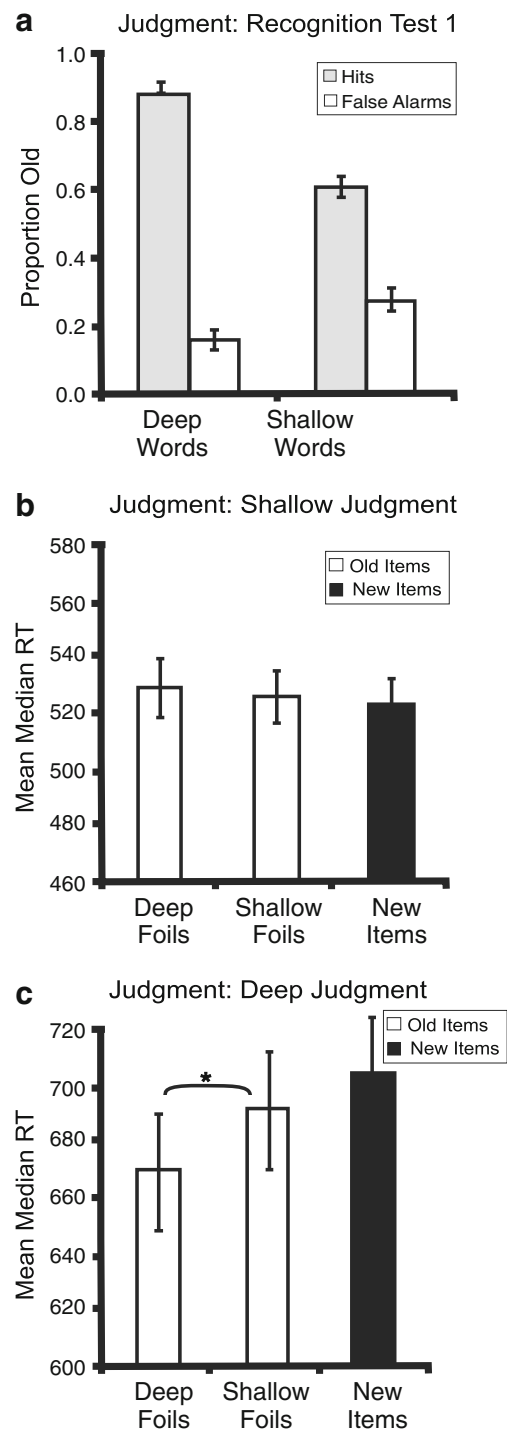

Fig. 3 Experiment 2: Evaluating processing of the foils following deep versus shallow study. (**a**) Recognition data from Test 1, demonstrating the levels-of-processing effect. (**b**) Performance from the shallow judgment final test, showing no difference in judgment times for foils from the deep versus the shallow prior recognition test. Error bars represent the standard errors of the corresponding means. (**c**) Performance from the deep judgment final test, showing faster judgment times for foils from the deep prior recognition test than for those from the shallow test

test were processed like the deep targets had been during study (i.e., for pleasantness), this should result in more priming of that same judgment for the deep foils relative to the shallow foils. There was a significant main effect across the three conditions—deep foils, shallow foils, and new words, $F(2, 74) = 7.46$, $MSE = 1,666.4$, $p < .001$, $\eta_p^2 = .167$. The first contrast showed an overall priming effect: Old words were responded to more quickly than new words, $F(1, 37) = 13.0$, $MSE = 7,410.9$, $p < .001$, $\eta_p^2 = .260$. The second planned contrast was the crucial test and did indeed demonstrate that participants were faster at making the pleasantness judgment for the foils from the test of deeply encoded words relative to the foils from the test of shallowly encoded words, $F(1, 37) = 4.11$, $MSE = 4,195.2$, $p < .05$, $\eta_p^2 = .100$; see Fig. 3c.

In sum, words that had been experienced as foils among target words that had been deeply processed at study benefited on a subsequent judgment task that required the same deep processing. This was not simply general priming from prior experience, because words experienced as foils among target words that had been shallowly processed at study showed reliably less priming. The benefit for the deep foils was specific, consistent with these items having been processed in the same way as their target counterparts. This provides direct evidence in support of the idea of source-constrained retrieval because, for such a benefit to occur, the words would have to have been associated with that relevant type of processing in a prior encounter—through reentry into the encoding context during the prior recognition test.

Experiment 3

The results of Experiment 2 provide direct support for the source-constrained retrieval explanation of the memory-for-foils effect. But one might still ask whether foils on a test of deeply encoded items could benefit in some way that is unrelated to the form of processing or judgment task that is subsequently performed. If so, deep foils should show that benefit not just on the judgment task used at study, but on other measures as well. While this deep priming advantage is unlikely, since deep foils showed no advantage in Experiment 2 during the shallow judgment task, to ensure that some sort of general advantage is not the source of the benefit, we repeated Experiment 2 and substituted a lexical decision task for the deep/shallow processing tasks in the final phase. If foils from the test of deeply encoded items have a general advantage relative to those from the test of shallowly encoded items, then word decisions should be faster for the deep foils as compared to the shallow foils. However, if deep foils were processed under the same encoding mode as their target counterparts, there should be no such benefit here, in sharp

contrast to Experiment 2. Like the judgment task of Experiment 2, this lexical decision task was indirect.

Method

Participants A total of 26 undergraduate students from the University of Waterloo (17 female, 9 male) participated for credit. Their mean age was 20.1 years ($SD = 1.67$). Four of the participants were removed from all analyses due to performing in the final phase more than two standard deviations slower than the mean response time for that phase.

Materials The stimulus words were identical to those used in Experiment 1. Nonwords were compiled using the ARC nonword database (www.maccs.mq.edu.au/~nwdb/nwdb.html; Rastle, Harrington, & Coltheart, 2002). Nonwords were 4–8 letters long and matched with the words on letter length frequency.

Procedure Participants were tested individually and completed the entire experiment in approximately 30 min. Words were randomly assigned to four new lists of 36 words each for each participant. Similarly, nonwords were randomly assigned to two lists of the same size. In addition, each task began and ended with an additional three words (or nonwords) to minimize primacy and recency effects; these items were not included in any analyses. The order of the tasks within each of the phases was counterbalanced across participants.

In the study phase, participants performed deep- and shallow-encoding tasks on separate word lists, identical to the procedure used in Experiment 2. In the recognition phase, participants performed the test precisely as in Experiments 1 and 2.

In the judgment phase, the participants performed a lexical decision task (i.e., "Is the item a word?") for half of the foil items from each of the recognition test lists (18 from the test of deeply encoded items and 18 from the test of shallowly encoded items), intermingled with an equal number of nonwords (72 items in total). To parallel as closely as possible the procedure of Experiment 2, this task was repeated in exactly the same way, with the remaining words from the deep and shallow test lists and a new set of nonwords. Participants were instructed to respond as quickly and as accurately as they could by pressing 1 or 0 on the keyboard. Because there were no methodological differences between the two lexical decision blocks, the data were combined.

Results and discussion

Recognition test As before, participants performed well on the recognition test of the initially studied lists (overall hits=.76) and readily discriminated these studied words from new words

(overall false alarms = .13). These results are displayed in Fig. 4a. A paired-samples *t* test showed that participants had better overall memory for deeply encoded as compared to shallowly encoded words, $t(21) = 10.66$, $p < .001$. This was true for hits, $t(21) = 12.60$, $p < .001$, and showed a mirror effect for false alarms—a greater number of false alarms for shallowly than for deeply encoded words, $t(21) = 2.71$, $p < .01$. Therefore, participants were effectively using the two encoding techniques, resulting in the typical levels-of-processing effect reported by Jacoby, Shimizu, Daniels, and Rhodes (2005) and by Marsh et al. (2009) and seen in our previous two experiments.

Lexical judgment task Following one-way ANOVAs, planned contrasts were conducted on the means of the participant median response times for the lexical decision task, which formed the final phase of the experiment. Two contrasts were performed, the first examining priming for the previously seen foils and the second examining whether priming differed between the two types of previously seen foils.

There was a significant main effect across the three conditions—deep foils, shallow foils, and new nonwords, $F(2, 42) = 11.80$, $MSE = 885.8$, $p < .001$, $\eta_p^2 = .360$. The first contrast demonstrated that old words were responded to more quickly than new nonwords, $F(1, 21) = 12.77$, $MSE = 9,672.6$, $p < .01$, $\eta_p^2 = .378$. As expected, the second planned contrast resulted in no difference between foils from the test of deeply encoded words and foils from the test of shallowly encoded words, $F(1, 21) = 2.00$, $MSE = 318.9$, $p = .17$, $\eta_p^2 = .087$; see Fig. 4b. Indeed, the observed difference was in the wrong direction, with respect to the hypothesis that deep items should always outperform shallow items. Therefore, foils first encountered among deeply encoded words did not incur any benefit over those first encountered among shallowly encoded items.

In sum, words that had been experienced as foils among target words that had been deeply processed at study were

not responded to any faster on a subsequent judgment task that did not require the same deep processing. Thus, it is only when processing is the same during the initial encounter with a word and on the final judgment task that deeply encoded items accrue a benefit. Therefore, the results of Experiment 2 were not the result of some form of general benefit for items processed within a deep context; instead, those results provide strong evidence for reprocessing of old items during recognition as if they had actually undergone the deep encoding in the initial phase. That is, the benefit for the deep foils was specific to the judgment during the study phase, entirely consistent with those foil items having been processed in the same way as their target counterparts. This provides direct evidence in support of the idea of source-constrained retrieval.

General discussion

Jacoby, Shimizu, Daniels, and Rhodes (2005; see also Jacoby, Shimizu, Velanova, & Rhodes, 2005; Marsh et al., 2009; Shimizu & Jacoby, 2005) demonstrated that the way in which targets are processed on a recognition test can influence subsequent memory for the accompanying distractors. Specifically, distractor words that appeared among target words that had been semantically encoded during an initial study phase were subsequently recognized better than distractor words that had appeared among targets that had been encoded nonsemantically during initial study. In Replication 1 in the Appendix, we report a faithful replication of this basic finding. In our Experiment 1, we generalized this memory-for-foils effect from the levels-of-processing manipulation used previously to a novel imagery encoding manipulation. Words that people did not intend to learn nevertheless benefited on a later memory test when they were experienced among other, previously elaborated, words; we now know that this effect occurs

Fig. 4 Experiment 3: Evaluating processing of the foils with a lexical decision task. (a) Recognition data from Test 1, demonstrating the levels-of-processing effect. (b) Performance on the lexical decision final test of deep foils, shallow foils, and nonwords. Priming was equivalent for foils from the deep and shallow prior recognition tests. Error bars represent standard errors of the corresponding means
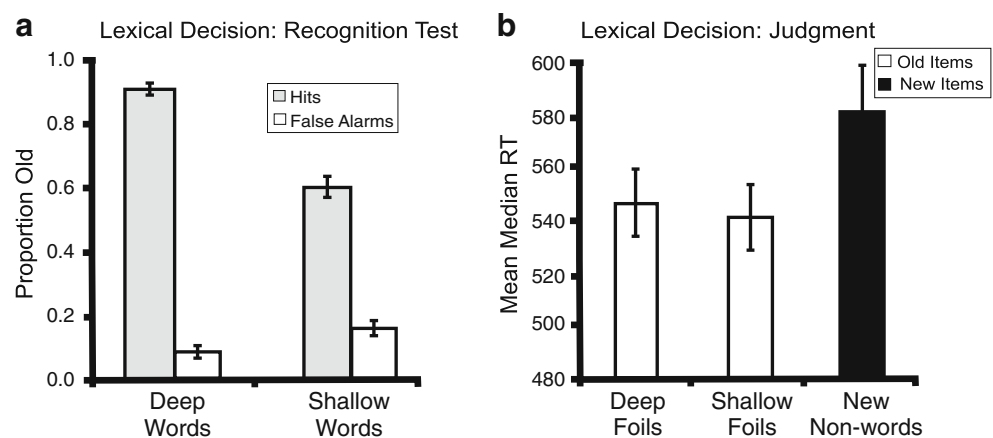
using two of the most widely studied modes of elaboration: levels of processing and imagery. We also know from the work of Marsh et al. (2009) that this effect is not simply the consequence of differential strength of encoding, which they showed by manipulating number of presentations in their Experiment 3. In Replication 2 in the Appendix, we report a faithful replication of this finding as well.

Whereas the levels-of-processing mode is based on semantic versus nonsemantic analysis of information—as opposed to perceptual processing—the elaboration brought about by imagery as an encoding mode certainly appears to have a different basis. Imagery is not equivalent to semantic processing, invoking as it does perceptual elements of what is imaged (see Paivio, 1971, 1995, 2007). But imagery is a coherent mode of processing, in the same sense that deep semantic processing is: Both are readily engaged ways of thinking about what is presented. This is why we reasoned that deep versus shallow imagery should also be capable of inducing and reinducing a beneficial mode of processing.

Our Experiment 2 fits a key piece to the puzzle. Here, we addressed the question of whether it would be possible to obtain more direct behavioral evidence of reentering the original encoding mode. If during Test 1 the foils were reprocessed with respect to the original mode of processing of the accompanying targets, that would be evident when the foils subsequently must be processed in terms of that original mode. To test this hypothesis, we ended not with a recognition test but with the same judgment task that had been used during study. By showing that people were faster to respond on a pleasantness judgment task to foils from the test of deeply encoded words, we demonstrated that these words were indeed encoded within the same deep context. Further, we know from Experiment 3 that this benefit for deep foils is not due to a general processing benefit for items associated with deeply encoded words but, instead, only occurs within the context of the original encoding task. Therefore, during retrieval, participants do appear to reenact the encoding task.

What seems to be essential to benefit memory for the foils is that encoding involve differentiable modes of processing being applied to the two sets of words during study, and that the reinstatement of these same modes of processing—separately—be accomplished at the time of the first recognition test. If both conditions are met, and if encoding was initially done more elaborately, the foils also receive more elaborative encoding—the *same* more elaborative encoding—and are better remembered. In the framing of Jacoby, Shimizu, Daniels, and Rhodes (2005), the beneficial encoding mode is reinstated on the first recognition test, in accordance with the transfer-appropriate processing principle (Morris et al., 1977; Roediger, 1990). This is what Jacoby, Shimizu, Daniels, and Rhodes referred to as "source-constrained retrieval."

It appears, then, that we unintentionally actively process items on a second occasion in much the same way as we processed them on the first occasion, even without any explicit requirement to do so. This is not surprising: It is in accord with the idea of transfer-appropriate processing (Morris et al., 1977), which meshes well with the proceduralist view of memory (Kolers, 1973; Kolers & Roediger, 1984). It is worth noting, however, that the benefit of transfer-appropriate processing stems from processing during *encoding*, whereas the benefit of source-constrained retrieval results from reprocessing of items during *testing*. Such processing reinstatement optimizes retrieval success when it provides a coherent encoding mode, "spilling over" onto other items processed contiguously, even without any intention to learn them. This highlights that there is indeed a mode of processing that is active across trials during retrieval. We agree with the proceduralist analysis that there is very substantial overlap of the processes involved in encoding and retrieval. Instead of thinking of retrieval as separate and distinct from encoding, retrieval could more parsimoniously be regarded as another encoding event.

In the present study, we have demonstrated a link between encoding and retrieval processes: The way that old items are retrieved has a direct and measurable influence on the success of encoding of new items. We have also shown that this influence is not restricted to a single mode of processing. Reinvoking the encoding processes (or modes) during retrieval permits all items on the recognition test (including the new items) to undergo that processing, with the same benefits to memory for the new items as had been observed for the originally studied items. The new items are thus encoded using a retrieval process that increases the likelihood of richer encoding and produces measurable facilitation in the speed of subsequent processing. Importantly, our results demonstrate that the mode of processing engaged during encoding, and reinstated during retrieval, has substantial effects on the encoding of new information, thereby helping to specify how encoding and retrieval are linked.

Appendix

Here, we report two experiments that replicate circumstances under which the memory-for-foils effect does (Replication 1) and does not (Replication 2) occur, both consistent with the source-constrained retrieval hypothesis.

Replication 1

Before we began addressing our specific questions about the central finding in the Jacoby, Shimizu, Daniels, and

Rhodes (2005) study, we undertook a faithful replication of their Experiment 1, in which the memory-for-foils effect was first reported. We note that Jacoby, Shimizu, Velanova, and Rhodes (2005) had previously replicated the phenomenon, as had Marsh et al. (2009) independently, but we wished to be certain that we obtained the same basic pattern at the outset.

Method

Participants A total of 24 undergraduate students (20 female, 4 male) from the University of Waterloo participated for credit or remuneration ($5). Their mean age was 20.9 years (SD=1.5).

Materials The stimuli consisted of 247 words, 5–8 letters long, taken from the MRC Psycholinguistic online database (www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm). The words had an average length of 5.3 letters and an average frequency of 37 occurrences per million (Kučera & Francis, 1967). Words were randomly assigned to six lists of 36 words each, such that each participant had a unique assignment of items. Each phase began and ended with an additional three-word buffer; these "buffer" words were not included in any analyses.

Procedure Participants were tested individually and completed the entire experiment in approximately 30 min. In the study phase, they performed two encoding tasks, with the order counterbalanced across participants. This phase was identical to that of Experiment 2 and of Jacoby, Shimizu, Daniels, and Rhodes (2005). Next came the first recognition phase, Test 1, which was identical to that described in Experiment 2. Finally, there was the second recognition phase, Test 2. Again, this consisted of targets

that were all former foils from the first recognition phase—from both the deep and the shallow recognition tests (i.e., none of the originally studied items were included on Test 2). The composition of the test was identical to that of Experiment 2, except that, in this replication, participants were asked only to respond "old" or "new" by a buttonpress.

Results and discussion

Recognition Test 1 Participants readily distinguished between old and new items from both the deep and shallow encoding tasks (overall hits=.71, overall false alarms=.23). We used a paired-samples t test to compare the recognition scores (hits minus false alarms) for words that had been categorized deeply versus shallowly during study. A typically robust levels-of-processing effect was found, $t(23) = 4.24$, $p < .001$. Figure 5a presents the hit and false alarm rates. Separate paired-sample t tests for hits and false alarms revealed a difference only for the hits, $t(23) = 3.76$, $p = .001$, and not for the false alarms, $t(23) = 1.06$, $p > .20$. Clearly, participants were effectively using the different encoding techniques during study.

Recognition Test 2 To determine whether there was a levels-of-processing effect for the foils from Test 1 when they became the targets on Test 2, we conducted a paired-samples t test comparing proportions of correct responses across word types. As shown in Fig. 5b, memory was superior for foils that had initially been tested among deeply encoded words than for foils initially tested among shallowly encoded words, $t(23) = 2.48$, $p < .05$, providing a clear replication of Jacoby, Shimizu, Daniels, and Rhodes's (2005) pattern.

Fig. 5 Replication 1: Manipulating levels of processing at encoding. (**a**) Recognition performance on Test 1, demonstrating the levels-of-processing effect. (**b**) Recognition performance for "old" responses on Test 2, demonstrating the memory-for-foils effect. Together, these data provide a complete replication of Jacoby, Shimizu, Daniels, and Rhodes (2005, Exp. 1). Error bars represent the standard errors of the corresponding means
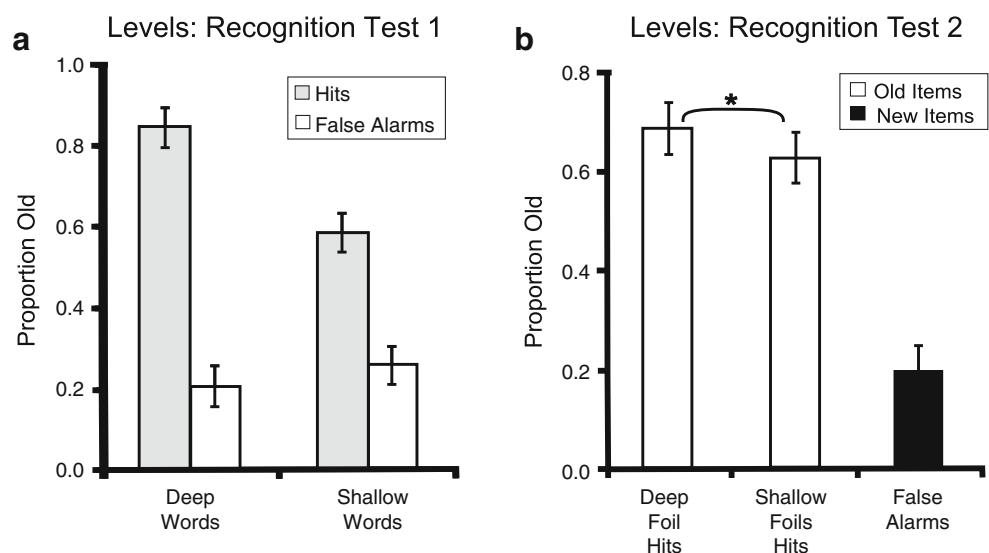
Replication 2

Marsh et al. (2009, Exp. 3) examined whether the memory-for-foils effect was a by-product of testing words that had been strongly versus weakly encoded during the study phase. In their study phase, words in one list had been studied only once each, whereas those in the other list had been studied three times each. Repetition is well established as influencing strength of encoding (e.g., Hintzman, 1976). Yet it is difficult to imagine how participants could reinstate the number of study presentations as a mode that would constrain retrieval, which means that the memory-for-foils effect should not occur. That is precisely what Marsh et al. observed. We conducted almost exactly the same experiment, and with the same outcome.

Method

Participants A total of 24 undergraduate students from the University of Waterloo (13 female, 11 male) participated for credit or remuneration ($5). Their mean age was 21.0 years (*SD*=3.2).

Materials The stimulus words were the same as in Replication 1.

Procedure All three of the phases were participant paced. Words were randomly assigned to six lists of 36 words each, with different randomizations for each individually tested participant. Again, three-word primacy and recency buffers were not included in any analyses.

In the study phase, participants studied two 36-word lists—one in which items were each presented once, and one in which items were each presented three times. For the latter list, the entire list was randomized three times, with an untested filler word inserted after each of the first two completely randomized

list presentations to prevent the unlikely possibility of repeating a word on consecutive trials. For both lists, each word was presented for 1,500 ms, followed by a fixation cross for 500 ms, with no response required. The order of studying the two lists was counterbalanced across participants.

Next came the first recognition phase, Test 1, conducted precisely as in Replication 1, including instructions to participants concerning the source of the items (e.g., "All old items are from the list in which you saw each item only one time"). The second recognition phase, Test 2, was identical to that of Replication 1, with targets again consisting of all of the former foils from the first recognition phase—from both the "triple" and the "single" recognition tests—among an equal number of new items. Responding was done as on the first recognition test.

Results and discussion

Recognition Test 1 Participants readily distinguished old from new items for both repeated- and single-presentation lists (overall hits=.74, overall false alarms=.08). The data appear in Fig. 6a. We used a paired-samples *t* test to compare the recognition scores (hits – false alarms) for words that had been studied either once or three times. In line with the extensive repetition literature (e.g., Bentin & Moscovitch, 1988; Murnane & Shiffrin, 1991; Ratcliff, Clark, & Shiffrin, 1990; Stretch & Wixted, 1998), participants recognized items that had been studied three times significantly better than items studied only once, $t(23)=5.76$, $p<.001$. Separate *t* tests demonstrated that there were (1) more hits, $t(23)=5.59$, $p<.001$, and (2) fewer false alarms, $t(23)=2.19$, $p<.05$, for words presented three times than for words presented once.

Recognition Test 2 This time, because repetition appeared unlikely to be a processing mode that could be reinstated,

Fig. 6 Replication 2: Manipulating strength via repetition at encoding. (a) Recognition performance on Test 1, demonstrating enhanced recognition for words studied three times as opposed to once. (b) Recognition performance for "old" responses on Test 2, showing no evidence of a memory-for-foils effect. Together, these data provide a complete replication of Marsh et al. (2009, Exp. 3). Error bars represent the standard errors of the corresponding means
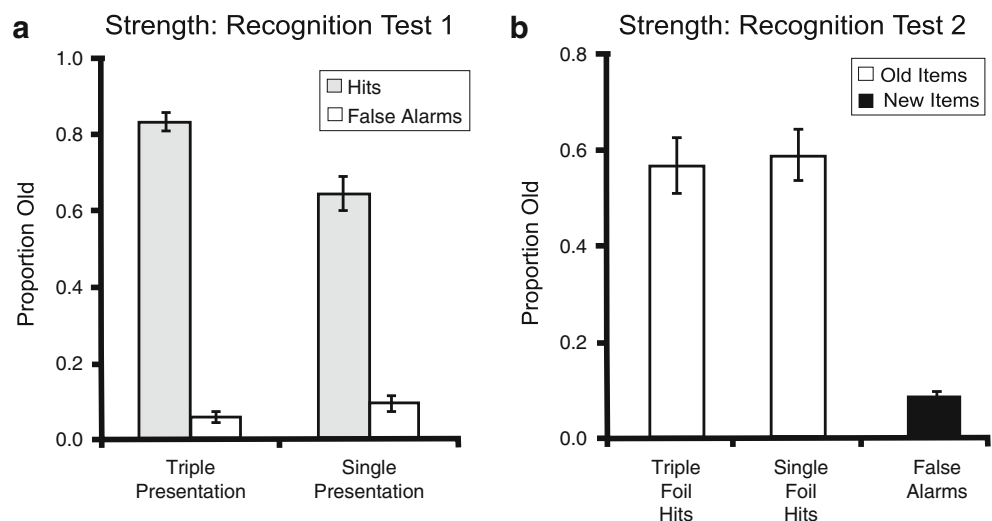
we expected not to observe the memory-for-foils effect. As shown in Fig. 6b, there was in fact no difference in recognition of former foil words as a function of whether they were encountered on Test 1 among targets presented once or targets presented three times, $t(23)=0.84$, $p>.20$. Further, the effect sizes for this test had quite adequate power to correctly reject the null hypothesis (power=.87). This absence of a memory-for-foils effect provides a clear replication of the Marsh et al. (2009, Exp. 3) pattern and is consistent with the idea that, for new items to be encoded with the same memory benefits as their old counterparts, there must be a coherent mode during the initial encoding session that can then be reapplied to items during a subsequent retrieval session.

References

Bentin, S., & Moscovitch, M. (1988). The time course of repetition effects for words and unfamiliar faces. *Journal of Experimental Psychology. General, 117,* 148–160.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language, 58,* 1095–1111.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Hintzman, D. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47–91). New York: Academic Press.

Hourihan, K. L. (2008). *The power of optimal encoding: Distinctiveness and differentiation defeat directed forgetting.* Doctoral dissertation, University of Waterloo, Waterloo, ON, Canada.

Hourihan, K. L., & MacLeod, C. M. (2011). *The power of optimal encoding: Distinctiveness effects in directed forgetting.* Manuscript submitted for publication.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005a). Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychonomic Bulletin & Review, 12,* 852–857.

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005b). Age differences in depth of retrieval: memory for foils. *Journal of Memory and Language, 52,* 493–504.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition, 1,* 347–355.

Kolers, P. A., & Roediger, H. L., III. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 425–449.

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

Marsh, R. L., Thadeus Meeks, J., Cook, G. I., Clark-Foos, A., Hicks, J. L., & Brewer, G. A. (2009). Retrieval constraints on the front end create differences in recollection on a subsequent test. *Journal of Memory and Language, 61,* 470–479.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 17,* 855–874.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76,* 241–263. doi:10.1037/h0027272.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Paivio, A. (1995). Imagery and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 977–986). Cambridge: MIT Press.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach.* Mahwah: Erlbaum.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: the ARC Nonword Database. *Quarterly Journal of Experimental Psychology, 55A,* 1339–1362.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16,* 163–178.

Roediger, H. L., III. (1990). Implicit memory: retention without remembering. *American Psychologist, 45,* 1043–1056.

Roediger, H. L., III, Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels of processing framework. *Memory, 10,* 319–332.

Shimizu, Y., & Jacoby, L. L. (2005). Similarity-guided depth of retrieval: constraining at the front end. *Canadian Journal of Experimental Psychology, 59,* 17–21.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24,* 1379–1396.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York: Columbia University, Teachers College.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12. doi:10.1037/h0080017.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Author note