

Top-down constraint on recognition memory

Justin Kantner · D. Stephen Lindsay

Published online: 21 November 2012
© Psychonomic Society, Inc. 2012

Abstract Can recognition memory be constrained “at the front end,” such that people are more likely to retrieve information about studying a recognition-test probe from a specified target source than they are to retrieve such information about a probe from a nontarget source? We adapted a procedure developed by Jacoby, Shimizu, Daniels, and Rhodes (Psychonomic Bulletin & Review 12:852–857, 2005) to address this question. Experiment 1 yielded evidence of source-constrained retrieval, but that pattern was not significant in Experiments 2, 3, and 4 (nor in several unpublished pilot experiments). In Experiment 5, in which items from the two studied sources were perceptibly different, a pattern consistent with front-end constraint of recognition emerged, but this constraint was likely exercised via visual attention rather than memory. Experiment 6 replicated both the absence of a significant constrained-retrieval pattern when the sources did not differ perceptibly (as in Exps. 2, 3 and 4) and the presence of that pattern when they did differ perceptibly (as in Exp. 5). Our results suggest that people can easily constrain recognition when items from the to-be-recognized source differ perceptibly from items from other sources (presumably via visual attention), but that it is difficult to constrain retrieval solely on the basis of source memory.

Keywords Recognition · Memory · Top-down · Constraint

Using a clever three-stage “memory-for-foils” procedure, Jacoby and colleagues convincingly showed that the way people process recognition memory test probes is affected by how they had encoded the to-be-recognized items during

the study phase (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005; Shimizu & Jacoby, 2005). First, subjects studied a list of words with either a deep, semantic processing task (e.g., rating the pleasantness of each word) or, for other subjects, a shallow processing task (e.g., counting the number of vowels in each word). Second, the subjects were tested on a random mix of studied and nonstudied probes presented one at a time for standard yes/no recognition judgments. Third, they took a second recognition memory test, in which nonstudied items from the first test were mixed with new foils, with instructions to recognize the words that had been presented as foils on the first test. The robust finding across several studies was that hit rates on this second test of memory for foils were substantially higher for subjects who had initially studied items with a deep processing task than for subjects who had initially studied the items with a shallow processing task (see also Alban & Kelley, 2012; Bridger, Herron, Elward, & Wilding, 2009; Danckert, MacLeod, & Fernandes, 2011; Marsh et al., 2009).

The memory-for-foils results are convincing evidence that how subjects interrogate or search memory for recognition test probes depends on how the to-be-recognized items were studied. People query recognition memory differently when the task is to recognize deeply processed items than when the task is to recognize shallowly processed items. As Alban and Kelley (2012) succinctly put it, on a recognition memory test “people query memory by mentally reinstating encoding operations” (p. 681). We call this “source-constrained search” of recognition memory.

Jacoby and his coauthors did not attribute their memory-for-foils findings solely to source-constrained search of recognition memory during the first test. Rather, they proposed the provocative thesis that recognition memory can be constrained “at the front end” (Shimizu & Jacoby, 2005, p. 17), such that when a probe is presented at test “only sought after information comes to mind” (Jacoby, Shimizu, Daniels, &

J. Kantner · D. S. Lindsay (✉)
Department of Psychology, University of Victoria,
P.O. Box 1700 STN CSC, Victoria,
British Columbia V8W 2Y2, Canada
e-mail: slindsay@uvic.ca

Rhodes, 2005, p. 852). That is, they interpreted their findings not only as evidence of source-constrained *search*, but also as evidence of source-constrained *retrieval* during the first test.

The notion of source-constrained retrieval in recognition memory is appealing for a number of reasons. For one thing, it is analogous to the idea of top-down constraint in perception. You can, for example, constrain visual search such that you are more likely to see what you are looking for than to see other things (e.g., Downing, 2000; Potter, 1975). If you lose track of your friend Jane in the shopping mall, while scanning the crowd for her you might look directly at your friend Don without “seeing” him. Similarly, you can source-constrain the recall of memories of prior episodes; if you are asked to recall a high-school experience involving a car, memories of driving to work this morning are unlikely to come to mind. The mind/brain can be configured such that memories from the desired source (e.g., high school) are more likely to come to mind than are memories from other sources (Jacoby, Kelley, & McElree, 1999). The phenomenon of recognition failure of recallable words (Tulving & Thomson, 1973), which suggests that targets may be missed if they are not processed in a transfer-appropriate way, is also consistent with the notion of a constrained retrieval process in recognition. Thus it is reasonable to speculate that in recognition memory, as in recall, top-down constraint can modulate the probability that a probe will cue recollections of a prior encounter with that item. In support of that idea, Alban and Kelley (2012) pointed to evidence that recognition memory can be enhanced by recapitulating encoding operations (see Nairne, 2002, for evidence that recapitulating encoding processes does not always enhance recognition).

Although the memory-for-foils procedure provides strong evidence of source-constrained *search* in recognition memory, it does not speak directly to the hypothesis that recognition memory can be constrained at the front end, such that sought-for evidence of oldness will be more likely to come to mind than non-sought-for evidence of oldness (i.e., source-constrained *retrieval*). After all, the foils were not studied, so there was nothing (deep or shallow) to be retrieved about them on the first test. While constrained search likely entailed deep processing of foils, augmenting subsequent memory for those items, constrained search may not have had any effect on the retrieval of study episodes per se: Retrieval of memories of a studied item might occur automatically when that probe is presented, regardless of whether subjects are trying to recognize items from Source A or B.

The constructs of constrained search and constrained retrieval in recognition memory are related to the concept of “retrieval orientation.” That phrase has sometimes been used in reference to the distinction between implicit/unintentional and explicit/intentional uses of memory.

Nelson, Canas, Bajo, and Keelean (1987), for example, compared the effects of various manipulations on cued recall versus word-fragment completion (with both tasks using the same cues). Performance on the two tasks was generally highly comparable, but Nelson et al. observed some dissociations that led them to note that “subjects given completion instructions may not be as likely as cued-recall subjects to recover episodically encoded meaning” (p. 546), suggesting constrained retrieval. Rugg and collaborators (e.g., Herron & Rugg, 2003; Rugg, Allan, & Birch, 2000) have used the term “retrieval orientation” to refer to “a tonically maintained retrieval strategy, which influences the cognitive operations engaged in response to each cue” (Rugg et al., 2000, pp. 664–665). We take this to be the same as the notion of source-constrained search in recognition memory. Rugg and coworkers have reported a number of clever event-related potential (ERP) and fMRI experiments evidencing differential processing of new items on a recognition memory test as a function of how the targets were encoded, which provides strong support for the operation of different retrieval orientations (a.k.a. source-constrained searches) but does not speak to the question of whether source-constrained searches yield source-constrained retrieval.

Bridger et al. (2009) and Bridger and Mecklinger (2012) found correlations across subjects between (a) ERP indicators of constrained search on new recognition probes and (b) recognition accuracy. They described that finding as “strong evidence that this class of retrieval processing operations benefits the accuracy of memory judgments” (Bridger et al., 2009, p. 1175). Thus, Bridger et al. spoke directly to the issue of central interest here, and claimed that their results demonstrate source-constrained retrieval. Their findings are both exciting and consistent with that claim, but it might also be that more motivated subjects (a) do a better job of encoding items and (b) are more likely to engage in source-constrained search at test. This could lead to a correlation, even if source-constrained search does yield source-constrained retrieval. Thus, although the findings of Bridger and coauthors support the claim that the retrieval of memory information in response to recognition probes can be source-constrained, we do not think that they compel that conclusion.

Alban and Kelley (2012, Exp. 2) found that if what was normally Test 1 in Jacoby’s memory-for-foils procedure was preceded by a very easy recognition memory test, then the memory-for-foils effect was eliminated. Apparently, the easy pretest led subjects to forego effortful source-constrained search on the first main test, such that their processing of foils was not affected by the level of processing with which they had studied the targets. This is consistent with the idea that constrained search is an effortful strategic process that subjects can choose to undertake. Interestingly, however, Alban and Kelley found that hit rates on the first main test were no higher among subjects in the

group who subsequently showed the memory-for-foils effect (i.e., those who evidently conducted source-constrained searches of memory) than among the group that did not. If constrained search successfully led to constrained retrieval, one would have expected higher hit rates in the group that more often used constrained search. Thus, Alban and Kelley's findings evidenced source-constrained search but not (in contrast to the Bridger et al. results) source-constrained retrieval, as we have defined it.

We developed a new procedure designed to test for source-constrained retrieval in the context of recognition memory. Our procedure did not emphasize memory for foils. Rather, we compared hit rates for studied items from two equally salient sources that were presented on a preliminary test and that were versus were not defined as targets on that preliminary test. In our procedure, subjects studied words from two sources, A and B. The two sources were blocked and differed along several dimensions (e.g., where on the screen the items appeared, the orienting-task judgment that subjects made, whether the subjects were standing or sitting, and whether they responded orally or via key-press). On Test 1, the subjects were informed that some of the A words and some of the B words, along with new words, would be presented one at a time on a CRT screen and that we were interested only in their ability to recognize words from Source A (or, for other subjects, Source B). Each item appeared on the screen briefly, and subjects were to hit the space bar if they recognized that item as one from the to-be-recognized source. No response was made to nontargets. We explicitly asked subjects to constrain recognition: to query memory only for evidence of having encountered items from the target source, and not to try to retrieve evidence of having encountered items from the nontarget source (see McDuff, Frankel, & Norman, 2009, and Pierce & Gallo, 2011, for related procedures).

We presumed that when subjects who were trying to recognize items from Source A were shown a test probe that had been presented in Source A, they would sometimes recollect their study-list encounter with that item. Furthermore, we presumed that recollecting the study encounter with a probe would strengthen memory for that encounter, over and above any strengthening effect occasioned by processing a probe without recollecting studying it. That is, processing a test probe and recollecting studying it should have a greater strengthening effect than merely processing a test probe without recollecting studying it. This assumption gains support from the testing effect literature, which has generally shown that an interpolated test has a greater beneficial effect on a subsequent memory test than does an interpolated study opportunity, and that this "testing effect" is larger when subjects answer the interpolated test questions than when they do not answer them (e.g., Kornell, Bjork, & Garcia, 2011; see also Hintzman, 2004, and MacLeod, Pottruff, Forrin, & Masson,

2012, on the recognition-enhancing effects of "reminding," and Raye, Johnson, Mitchell, Greene, & Johnson, 2007, on "refreshing").

The key question in our procedure was whether subjects would be more likely, on Test 1, to recollect studying words from the source that they were trying to recognize on Test 1 than to recollect studying words from the other source on Test 1. That is, if subjects were trying to recognize items from Source A, would they be more likely to recollect studying A-studied words than to recollect studying B-studied words on Test 1? This was our central question: Can subjects constrain retrieval on Test 1, so that they are more likely to recall studying items from the specified target source than from the other source?

If subjects do constrain retrieval on Test 1, then Test 1 presentation of studied words should strengthen memory for words from the to-be-recognized source (tested targets) more than it would strengthen memory for words from the other source (tested nontargets). That is, if subjects are more often reminded, during Test 1, of their study encounters with words from the to-be-recognized source than with words from the other source, their memories of the former should receive a bigger boost. To assay for such a difference, on Test 2 subjects were shown all of the studied words mixed with novel words and were asked to judge whether or not each item had appeared on the original study lists (regardless of source). Of the old items on this test, half had been tested on Test 1 and half had not (we refer to these as "T1-tested" and "T1-nontested" items, respectively), and half of each of these types of items were from the to-be-recognized source in Test 1 (henceforth, "targets") and half were from the other source ("nontargets"). To the extent that retrieval was constrained on Test 1, Test 2 recognition should be better for tested targets than for tested nontargets; if constraint were perfect, Test 2 hit rates would be no higher for tested nontargets than for nontested nontargets.¹

A critical feature of this experimental design is that constraining search during Test 1 cannot enhance Test 2 recognition of tested targets unless retrieval is also constrained during Test 1. Suppose that you are a subject taking Test 1, and you are looking for items that you had judged for value (on List 1, while standing, etc.). The word "fish" comes up on the screen, and you ask yourself "Did I judge fish for value?" Doing that sort of constrained search would impart no special memorability to "fish" as a function of whether or not "fish" was one of the words on List 1; constrained search, by itself, would not benefit targets over nontargets. Only if constraining search actually affected retrieval, such

¹ As an anonymous reviewer noted, Test 2 hit rates might be higher on tested targets than on tested nontargets even if retrieval were unconstrained, if subjects more often engaged in some memory-enhancing postretrieval process for targets than for nontargets.

that you would be more likely to recall your study list encounter with “fish” if it were a List 1 item than if it were a List 2 item, would a difference between targets and nontargets be expected to emerge. Thus, a Test 2 advantage for tested targets over tested nontargets, coupled with the lack of such an effect for nontested items, would indicate that retrieval (not just search) was constrained on Test 1.

We conducted a substantial number of experiments using this basic design, and we will begin by reporting three of them. Experiment 1 yielded a significant source-constrained retrieval pattern. Experiment 2 was an exact replication of Experiment 1, and Experiment 3 was similar, except that we imposed a response deadline on Test 2. These and the other experiments that we conducted using this design—unlike Experiment 1—did not yield a significant constrained retrieval pattern, although the pattern of means was in the predicted direction. We then report an experiment using a variant of our procedure that increased the relevance of source information on Test 2 but did not yield a significant constrained retrieval pattern. In a final pair of experiments, we obtained substantial evidence of constraint when items from the to-be-recognized target source were perceptibly different than items from the nontarget source (presumably due to selective attention during the first test), but no evidence of constrained retrieval when items from the two sources could be differentiated only via memory. We conclude that although it may be possible to constrain retrieval solely on the basis of memory source, it is not easy to demonstrate such an effect. In contrast, it is easy to obtain evidence of front-end-constrained recognition when the sources are perceptibly different.

Experiment 1

Method

Subjects We tested 32 University of Victoria undergraduates, who participated either for optional extra credit in a psychology course or for a \$5 payment. Three of the subjects were dropped for failure to follow the instructions.

Materials We used words from lists created by Gruppuso, Lindsay, and Kelley (1997). These were 124 common English concrete nouns that could easily be judged in either of two medium-level-of-processing tasks: monetary value (worth less or more than \$25) and frequency of encounter in the last month (less or more than ten times). These judgments were to be made with respect to the words’ referents, and items were selected so that about half were likely to be judged as low versus high in value and about half were likely to be

judged as common versus rare in frequency of encounter. For each subject, 20 of these words were randomly assigned anew to each of the following conditions: tested targets, nontested targets, tested nontargets, nontested nontargets, new “trick” items (i.e., foils on Test 1 that appeared again on Test 2), and new items (foils on Test 2 that had not previously been presented). The remaining four words were used as two primacy and two recency buffers for the study phase. Each word was converted into a computer voice (.wav file) using TextAloud software. As a filler task, we tested materials for an unrelated experiment. Subjects studied 32 pictures of faces in an initial filler task and were tested on recognition of those faces in a later filler task.

Procedure Subjects were tested individually in a quiet lab room by a research assistant. The experiment was run on a PC using E-Prime software. Phase 1 consisted of two blocks, in each of which a different set of 40 words was presented one at a time for a binary judgment. At the start of each trial, the judgment to be made and the two response options appeared on the screen for 1,250 ms, and then the word to be judged appeared. Half of the subjects judged the first block on value and the second on frequency, whereas the opposite was true for the remaining subjects. To further differentiate the two blocks, subjects stood during one block and sat during the other (counterbalanced across subjects). Also, in one block, each word appeared in a large, red, nonitalicized font in the upper left-hand corner of the screen and was not accompanied by a computer voice reading the word, whereas in the other, each word appeared in a small, blue, italicized font in the lower left-hand corner of the screen and was accompanied by a computer voice reading the word aloud. Subjects made judgments aloud, and the research assistant entered their responses via the keyboard.

After Phase 1, subjects studied 32 pictures of faces (a filler activity that took 2–3 min), and then the instructions for Test 1 were presented. Subjects were informed that the test consisted of studied words from the two sources mixed with nonstudied words, presented one at a time on the CRT, and that their task was to try to recognize the words from one of the sources, referenced in terms of the orienting task (i.e., half of the subjects were told to try to recognize items they had judged for value, and the others to recognize items they had judged for frequency). Subjects were to press the spacebar if they recognized a word from the designated source, otherwise they should just wait for the next item. At the start of each trial, the words “Judged for frequency?” or (for other subjects) “Judged for value?” appeared, according to the dimension of constraint, for 1 s. The item to be judged then appeared. Subjects pressed the spacebar to endorse the item as having been judged for the trait probed. A press of the spacebar caused the immediate start of the

next trial. If the spacebar was not pressed in 2 s, the next trial began automatically. The test consisted of 60 trials: 20 studied targets, 20 studied nontargets, and 20 new items.

Subjects then completed a second filler activity, consisting of a recognition test for the previously studied faces, which took about 4 min. Then they received the instructions for Test 2. In Test 2, all 120 items were presented, one at a time in a random order. On each trial, the question “Presented in Phase One?” and the yes/no response options appeared for 1 s before the probe appeared. Subjects pressed “1” to indicate that an item had been presented in the study phase and “0” to report that it had not. Responses were not speeded. Upon making a response, subjects were asked to indicate their confidence in the judgment on a 1–6 scale (1 = *completely guessing*, 6 = *100% certain*). Entering the confidence rating initiated the next trial.

Results

Our interest was focused on Test 2, and, in particular, on the question of whether Test 1 presentation of to-be-recognized target items would confer greater benefits to Test 2 performance than would Test 1 presentation of nontarget items. We could only hope to find such an effect if performance on Test 1 indicated that subjects were able to differentiate reasonably well between targets and nontargets and between studied and nonstudied items on that test. Therefore, we will begin each Results section with a brief summary of performance on Test 1.

Test 1 In Test 1, the subjects were to identify studied items from a designated source and to make no response to items from the other source or to new items. The mean “yes” rates for studied targets, studied nontargets, and new items were .84, .30, and .02, respectively. The within-subjects 95 % confidence interval based on the error term from a one-way analysis of variance (ANOVA) on these data (as per Masson & Loftus, 2003) was .07. Thus, subjects were quite good at discriminating studied targets from studied nontargets and from nonstudied items.

Test 2 In Test 2, the subjects were to say “yes” to every item that they recognized as something presented on the study list, and “no” to all other items. Our interest was focused on the hit rates for studied items, and those means are displayed in Fig. 1.² The error bars represent the within-subjects 95 % confidence intervals using the error term for the target-nontarget comparison (.009); if the bars overlap by less than 50 %, the difference is significant at the .05 level.

² As expected (and of no particular interest here), false alarms were more common on “trick” items (foils from Test 1, $M = .17$) than on new foils ($M = .04$), $t(28) = 3.994$, $p < .001$.

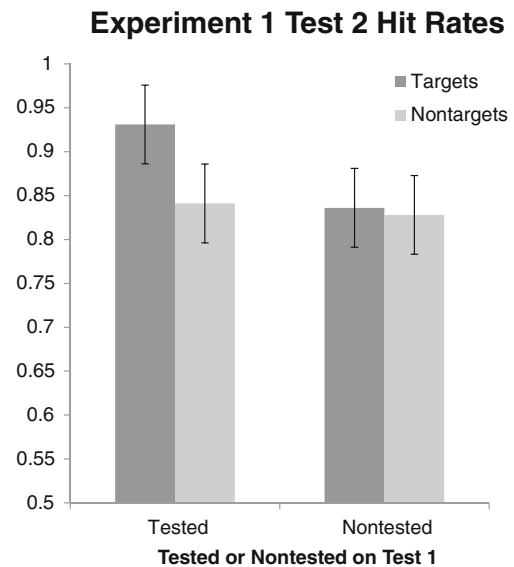


Fig. 1 Experiment 1: Mean hit rates on Test 2 for studied items that were versus were not presented on Test 1, as a function of whether they were from the source to be recognized on Test 1 (targets) or from the other source (nontargets). Error bars represent within-subjects 95 % confidence intervals (see the text for details)

The hit rate data were analyzed in a mixed-models ANOVA, with the Orienting Task factor (i.e., which orienting task had defined the targets in Test 1) as the between-subjects variable and Target Status (target vs. nontarget) and Test 1 Presence (tested vs. nontested on Test 1) as repeated measures. Orienting task was not significant as a main effect and did not enter into any significant interactions (all $ps > .07$). We did find two “near significant” tendencies toward effects involving orienting task. One was a nonsignificant tendency toward a main effect of orienting task, $F(1, 27) = 3.418$, $p = .075$, $\eta_p^2 = .112$, with the monetary value condition tending to have higher hit rates than the frequency condition. That tendency was subsumed by a nonsignificant tendency toward a Target Status \times Orienting Task interaction, $F(1, 27) = 3.105$, $p = .089$, $\eta_p^2 = .103$, reflecting that the aforesaid tendency toward a main effect of orienting task was primarily driven by nontargets—that is, people in the value condition recognized nontargets on Test 2 better than did people in the frequency condition. We suspect that this was merely due to error variance; regardless, these tendencies do not compromise interpretation of the key effects, to which we now turn.

As expected, the studied items presented on Test 1 were recognized more often on Test 2 than were those not presented on Test 1, $F(1, 27) = 9.946$, $p < .01$, $\eta_p^2 = .269$. The effect of target status was also significant, $F(1, 27) = 4.695$, $p < .05$, $\eta_p^2 = .148$. Critically, however, this advantage was only observed for items presented during Test 1; targets not tested during Test 1 were recognized no better than nontargets. This pattern was captured by a significant Target

Status \times Test 1 Presence interaction, $F(1, 27) = 7.944$, $p < .01$, $\eta_p^2 = .227$, and represents evidence of source-constrained retrieval.

Experiment 2

As we noted in the introduction, despite its name, Experiment 1 was not our first attempt to obtain evidence of source-constrained retrieval. In fact, we had “chased” the effect for some time, tweaking the procedure. Thus, it was essential to replicate the present Experiment 1. Experiment 2 exactly replicated Experiment 1, except that a random half of the subjects in Experiment 2 studied items with shallow orienting tasks; here we report data only from those subjects who used the orienting tasks from Experiment 1.³

Method

Subjects We tested 29 University of Victoria undergraduates, two of whom were dropped for failure to follow instructions: One endorsed 100 % of the studied targets and studied nontargets on Test 1, and the other endorsed 50 % of the studied targets and 80 % of the studied nontargets on Test 1.

Materials and procedure The materials and procedure replicated Experiment 1.

Results

Test 1 The mean “yes” rates for studied targets, studied nontargets, and new items were .81, .23, and .02, respectively. The within-subjects 95 % confidence interval was .06. As in Experiment 1, subjects did well at selectively endorsing studied targets.

Test 2 Mean hit rates are displayed in Fig. 2. The same mixed-factor ANOVA conducted for Experiment 1 showed that items that had been presented on Test 1 were more often recognized on Test 2 than were items that had not been presented on Test 1, $F(1, 25) = 25.62$, $p < .001$, $\eta_p^2 = .506$. Unlike in Experiment 1, that benefit of testing was not significantly greater for targets than for nontargets, evidenced by a null Test 1 Presence \times Target Status

³ We had the idea that shallow encoding might reduce the likelihood that subjects would spontaneously recognize nontargets on Test 1, and thereby increase the evidence of constrained retrieval on Test 2. But, in this case and in one other experiment not reported here, we found that shallow processing led to poor performance on Test 1, with subjects often failing to endorse studied targets and often erroneously endorsing studied nontargets, and to no evidence of constrained retrieval.

Experiment 2 Test 2 Hit Rates

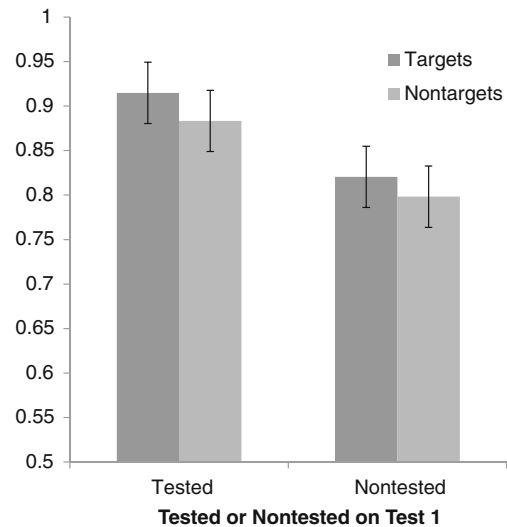


Fig. 2 Experiment 2: Mean hit rates on Test 2 for studied items that were versus were not presented on Test 1, as a function of whether they were from the source to be recognized on Test 1 (targets) or from the other source (nontargets). Error bars represent within-subjects 95 % confidence intervals for the target–nontarget comparison

interaction, $F < 1$, $\eta_p^2 = .003$. As in Experiment 1, subjects more often false alarmed to nonstudied items that had been used as foils on Test 1 compared to those that had not, $t(26) = 4.322$, $p < .001$.

Experiment 3

We speculated that imposing a deadline on Test 2 responses might make it easier to detect evidence of source-constrained retrieval. For one thing, in our earlier experiments Test 2 hit rates on nontested items were fairly high, limiting the extent to which hit rates on tested targets could outstrip them. For another, we thought that variations from trial to trial in the extent to which subjects used deliberative, strategic processes when making their Test 2 recognition judgments might muddy the waters. Experiment 3 thus included a tight deadline on Test 2 responses.

Method

Subjects We tested 40 University of Victoria undergraduates, of whom two were dropped for failure to follow the task instructions.

Materials and procedure The materials and procedure were the same as in Experiments 1 and 2, except in the following respects. First, during the filler tasks we collected norms for materials for a second unrelated experiment: Subjects gave

numerical answers to general knowledge questions (e.g., “In what year did the Napoleonic wars end?”). Second, a time limit of 1 s was placed on Test 2 responses to encourage subjects to make quick, strength-based recognition judgments. When a response was not made in 1 s, the message “Time’s Up!” appeared for 1,250 ms, followed by the next trial. Otherwise, the intertrial interval was 1,250 ms.

Results

Test 1 The mean “yes” rates for studied targets, studied nontargets, and new items were .81, .23, and .04, respectively. The within-subjects 95 % confidence interval was .04. Thus, subjects were again quite good at selectively endorsing studied targets.

Test 2 The mean hit rates are displayed in Fig. 3. The same mixed-factor ANOVA as in the previous experiments revealed that items that had been presented on Test 1 were again recognized much more often on Test 2 than were items that had not been presented on Test 1, $F(1, 36) = 128, p < .001, \eta_p^2 = .781$. Unlike in Experiment 1, but as in Experiment 2, that benefit was not significantly greater for tested targets than for tested nontargets, evidenced by a nonsignificant Test 1 Presence \times Target Status interaction, $F(1, 36) = 2.478, p = .124, \eta_p^2 = .064$. As in Experiments 1 and 2, false alarms were more common on “trick” items ($M = .21$) than on new foils ($M = .07$), $t(37) = 7.375, p < .001$.

Experiment 4

In our procedure, the source (A vs. B) of the targets was not relevant on Test 2, because items from both sources appeared on the test and subjects were to make the same response (“Yes”) to items from either source. In Experiment 4, we tested a between-subjects design that makes source relevant on Test 2 (as it is on Test 1).⁴ The study and Test 1 procedures were the same as those used in Experiments 1 and 2. For Test 2, the subjects were divided into two conditions. In one condition, Test 2 contained tested targets and nontested targets (i.e., items from only one of the two studied sources) intermixed with new words not presented previously in the experiment, and subjects were informed that all of the old words were from one source. For example, some subjects who had attempted to recognize frequency-encoded items on Test 1 were given a Test 2 that consisted only of items encoded with the frequency task (some of which had been on Test 1, some of which had not) mixed with new items; they were told that their task on Test 2 was to recognize items studied with the frequency task. This

⁴ We thank Larry L. Jacoby for suggesting this experiment.

Experiment 3 Test 2 Hit Rates

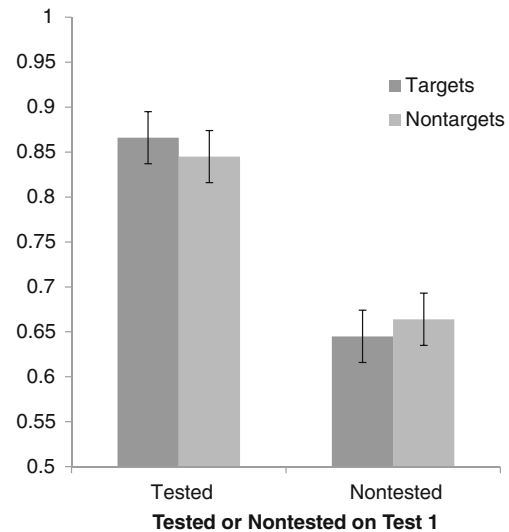


Fig. 3 Experiment 3: Mean hit rates on Test 2 for studied items that were versus were not presented on Test 1, as a function of whether they were from the source to be recognized on Test 1 (targets) or from the other source (nontargets). Error bars represent within-subjects 95 % confidence intervals for the target–nontarget comparison

afforded a comparison between tested targets and nontested targets while subjects were trying to recognize targets from a designated source. The second condition was the same, except that the old items on Test 2 came from the other source. For example, some subjects who had attempted to recognize frequency-encoded items on Test 1 were given a Test 2 that consisted only of items encoded with the monetary value task (some of which had been on Test 1, some of which had not) mixed with new items; they were told that their task on Test 2 was to recognize items studied with the monetary value task. This afforded a comparison between tested nontargets and nontested nontargets while subjects were trying to recognize nontargets. If constrained retrieval occurred on Test 1, recognition memory on Test 2 should be better when the Test 2 old items were from the Test 1 target source than when they were from the Test 1 nontarget source. Each subject also completed a Test 3 under the conditions complementing their Test 2 condition (e.g., if Test 2 was tested and nontested targets mixed with new items, then Test 3 was tested and nontested nontargets mixed with new items). This enabled us to make both within- and between-subjects comparisons of recognition of Test 1 targets versus nontargets.

Method

Subjects We tested 33 University of Victoria undergraduates in Experiment 4. One subject was removed for failure to follow the task instructions. Of the remaining 32, 17 were quasirandomly assigned to look for targets on Test 2 and

nontargets on Test 3, whereas the other 15 were assigned to look for nontargets on Test 2 and targets on Test 3.

Materials A set of 30 words with characteristics similar to the 120 used in the previous experiments was added to the stimulus set. The allocation of words to the conditions at study and during Test 1 was the same as in the prior experiments. Test 2 was populated with the 40 words that had been judged for either frequency or value at study (half T1-tested, half not), plus 20 completely new words. Test 3 consisted of 40 words judged for whatever dimension had not been covered in Test 2 (half T1-tested, half not), plus 20 completely new words. Not only did this segregate the items by studied dimension across the two tests, it omitted the “trick” items (T1-tested items that had not been studied) that had been present in Experiments 1, 2 and 3.

Procedure The procedure was the same as in Experiment 1, except for the addition of Test 3, a 2-s time limit on Test 2/3 responses, and Test 2/3 instructions that explained the composition of the test list (for one of these tests, these were words that had been studied for value mixed with new words; for the other test, these were words that had been studied for frequency mixed with new words).

Results

Test 1 Subjects were again good at discriminating targets, nontargets, and new items. The mean “yes” rates for studied targets, studied nontargets, and new items were .81, .21, and .03, respectively (95 % within-subjects confidence interval = .04).

Test 2 Hit rates (see Fig. 4) were analyzed in a 2 (Test 1 Presence: tested vs. nontested) \times 2 (Target Status: target vs. nontarget) \times 2 (Orienting Task: frequency vs. value) mixed factor ANOVA, with Test 1 Presence as a within-subjects factor and Target Status and Orienting Task as between-subjects factors. As in Experiments 1, 2 and 3, Test 2 hit rates were higher for items tested on Test 1 than for items not tested on Test 1, $F(1, 28) = 13.0, p = .001, \eta_p^2 = .316$. No other significant main effects or interactions emerged (largest $F = 1.517$, smallest $p = .23$); of particular interest was that subjects recognizing tested targets were no more accurate than those recognizing tested non-targets (main effect of item status, $p = .23$) and that we found no tendency toward a Test 1 Presence \times Target Status interaction, $F(1, 28) = 1.180, p = .84, \eta_p^2 = .001$.

Test 3 As in Test 2, there were no significant main effects or interactions, except the expected main effect of Test 1 presence, $F(1, 28) = 34.8, p < .001, \eta_p^2 = .554$. The critical main effect of target status did not approach significance ($p = .71$),

Experiment 4 Tests 2 and 3 Hit Rates

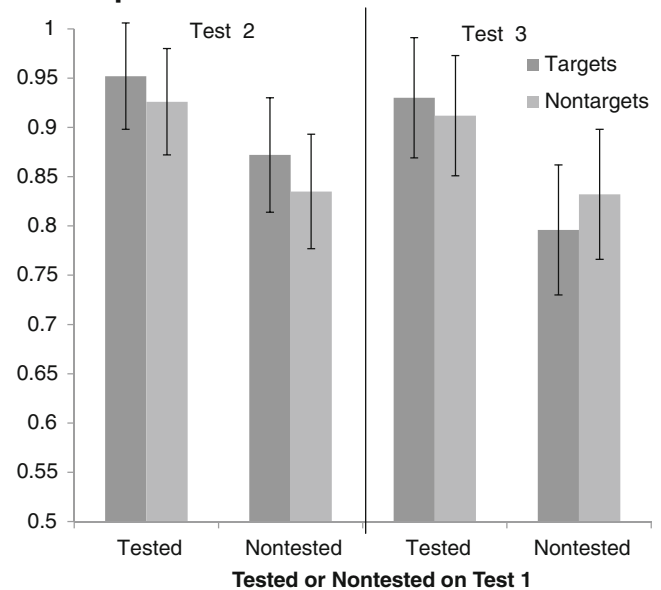


Fig. 4 Experiment 4: Mean hit rates on Tests 2 and 3 for studied items that were versus were not presented on Test 1, as a function of whether they were from the source to be recognized on Test 1 (targets) or from the other source (nontargets). Error bars represent between-subjects 95 % confidence intervals for the target–nontarget comparison, calculated for each condition and test

indicating that tested targets were no more recognized than tested nontargets in the between-subjects comparison. A nonsignificant tendency toward a Test 1 Presence \times Target Status interaction, $F(1, 28) = 2.566, p = .120, \eta_p^2 = .084$, and a nonsignificant tendency toward a three-way interaction with those two variables and orienting task, $F(1, 28) = 3.693, p = .065, \eta_p^2 = .117$ did emerge; both of these nonsignificant interactions were driven by relatively low hit rates on nontested targets among subjects for whom value-encoded items had been the targets on Test 1.

Within-subjects comparison As expected, presentation during Test 1 greatly improved the recognition of studied items on Tests 2 and 3, $F(1, 30) = 42.5, p < .001, \eta_p^2 = .586$. There was no main effect of target status ($p = .46$), no interaction of target status and Test 1 presence ($p = .45$), and no effect of nor any interactions involving orienting task (all $ps > .21$).

Discussion

The high hit rate for tested targets on Test 2 ($M = .952$) might, in principle, have obscured a tendency toward the Test 1 Presence \times Target Status interaction that would evidence constrained retrieval, but we think that this is unlikely. For one thing, the tested targets’ hit rates were at least as high in Experiment 1 (and, to anticipate, in Exp. 6), in which that interaction was significant. Also, the Test 3 tested-target

hit rate was less close to ceiling ($M = .93$), and still the interaction did not approach significance. More fundamentally, the reason for the lack of an interaction in Experiment 4 was not that there was no room for an increase from nontested to tested targets; quite the contrary, there was a substantial increase across those conditions. The reason for the null interaction was that the increase was just as great from nontested to tested nontargets as it was from nontested to tested targets. To put this differently, if constrained retrieval had occurred, its effect would have been to lower hit rates on tested nontargets (as it did in Exp. 1), not to have further increased hit rates on tested targets.

Experiment 5

In the experiments described above (and in others conducted in our lab that are not reported here), the bases for discriminating targets from nontargets on Test 1 inhered exclusively in memory for the study episodes in which those items had been encountered. That is, nothing about the test probes, as stimuli, differentiated the targets from nontargets. Suppose that a subject was looking for items judged for frequency of encounter (vs. monetary value) and the word “table” appeared as a probe; the only way that the subject could determine whether “table” had been judged for frequency of encounter would be to query memory. Our results suggest that it is difficult to query memory for having recently judged a word for frequency of encounter without also, spontaneously, cuing memory for having recently judged that word for monetary value (i.e., it is difficult to constrain recognition retrieval solely on the basis of memory source, at least under these conditions).

We speculated that constrained recognition might be more robust if the targets and nontargets differed perceptibly. Consider again the example of visual search mentioned in the introduction, in which, while visually searching for a designated target (your friend Jane), you fail to “see” (identify, recognize) a nontarget (your friend Don). Presumably, in such a situation the visual system uses something like a template, selecting for further processing stimuli that match that template to some criterial degree, and aborting further processing of other stimuli (Downing, 2000; Porter, 1975). In the foregoing experiments, the analogue for that template would be Memory Source A, but nothing in the test probe itself would indicate its source (i.e., there were no perceptible cues to the source). Thus, in our prior experiments, subjects could not use selective visual attention to stimuli from the target source as a means of constraining retrieval.

Experiment 5 was similar to Experiments 1, 2 and 3, except that all of the words from one source were five letters long and all of the words from the other source were seven

letters long. During Test 1, some of the studied and non-studied words of both lengths were presented, along with new words of each length, and subjects were instructed to detect only studied items of one specified length (five letters for some subjects, seven letters for others). Note that the perceptible difference between the targets and nontargets would afford front-end constraint of recognition via selective attention. Because subjects could reject studied nontargets on the basis of a superficial analysis of word length, they might do so without having processed such items enough to remember their prior occurrence. Precluding recognition of studied nontargets by truncating visual attention would constitute front-end constraint on search and retrieval (i.e., people would be more likely to remember aspects of their study-list encounters with targets than with nontargets), although the mechanism through which such constraint was implemented would more naturally be described as attention than memory.

On Test 2, all of the studied items were presented, along with new items of both lengths. As in Experiments 1, 2, 3 and 4, constrained search in the absence of constrained retrieval at Test 1 should not benefit Test 2 memory for targets: Searching memory for evidence that each probe was a studied seven-letter item, for example, should by itself confer no greater memorability on targets than on nontargets. If subjects are able to constrain recognition at the front end, such that they are more likely to recognize the targets on Test 1 than to recognize the other items, then Test 2 recognition performance should be better for tested targets than for tested nontargets, while nontested targets should show little or no advantage over nontested nontargets.

Less crucially, Experiment 5 also differed from Experiments 1, 2 and 3 in the nature of the foils on Test 2. In Experiments 1, 2 and 3, Test 2 included two types of foils: new words not previously presented in the experiment, and re-presentations of the foils used in Test 1. The re-presented foils may have complicated Test 2 for subjects, because they could not simply endorse words that they recognized as familiar from the experiment, but rather had to discriminate items presented at study (some of which had also been presented on Test 1) from items presented only on Test 1. We speculated that this complexity might have discouraged subjects from making the sort of strength-based judgments that we thought might be most likely to yield evidence of constrained retrieval. In Experiment 5, the foils used on Test 1 did not reappear on Test 2, and all of the foils on Test 2 were novel to the experiment (note that this had also been the case in Experiment 4).

Method

Subjects We tested 44 University of Victoria undergraduates, three of whom were dropped for failure to follow the task instructions.

Materials We selected 72 five-letter and 72 seven-letter nouns from the MRC Psycholinguistic Database (Coltheart, 1981). The five- and seven-letter sets were matched on Kučera and Francis frequency and concreteness. Subjects studied 40 words of each length (plus a primacy and recency buffer of each length) in blocks during the study phase. Test 1 targets were five-letter words for half of the subjects and seven-letter words for the remaining subjects. For each subject, 20 words from the appropriate set were randomly selected to be the tested targets on Test 1, and 20 words from the other set were selected to be the tested nontargets on Test 1. An additional ten words from each set were randomly selected to be new items on Test 1 that were not used on Test 2, and 20 from each set to be new items on Test 2 that had not been presented previously.

Procedure The procedure was the same as in Experiments 1, 2, and 3 with the following exceptions. The judgments made in Phase 1 were of item commonness (common vs. rare) and concreteness (concrete vs. abstract) instead of monetary value and frequency.⁵ Instructions at the beginning of the experiment mentioned that about half of the words were five letters long and the rest seven letters long. The words were blocked by length as well as by all of the other dimensions from Experiment 1 (e.g., whether subjects stood or sat). On Test 1, the intertrial interval was 2 s rather than 1 s, and during Test 2, subjects had 2 s rather than 1 s to respond.

Results

Test 1 Test 1 performance was excellent. Subjects rarely falsely endorsed a word of the wrong length (i.e., a nontarget), whether it had been studied ($M = .04$) or not ($M = .01$); they usually correctly endorsed studied words of the target length ($M = .89$), and they endorsed nonstudied words of the target length relatively rarely ($M = .20$).

Test 2 Figure 5 displays the Test 2 hit rates. Items presented during Test 1 were again better recognized than those that were not, $F(1, 39) = 52.7, p < .001, \eta_p^2 = .575$. There was a near-significant Test 1 Presence \times Condition interaction, $F(1, 39) = 3.983, p < .06, \eta_p^2 = .093$, reflecting a tendency for a greater effect of Test 1 presence when the target words were seven letters long than when they were five letters long (a tendency for which we lack an explanation). Targets were recognized significantly more often than nontargets, $F(1, 39) = 4.449, p < .05, \eta_p^2 = .102$. The central finding of interest was that the beneficial effect of having encountered an item on Test 1 was bigger for studied target-length items than for studied non-target-length items [for the interaction, $F(1, 39) = 9.24,$

⁵ We changed the orienting tasks merely because the new word sets did not lend themselves to frequency and value judgments.

Experiment 5 Test 2 Hit Rates

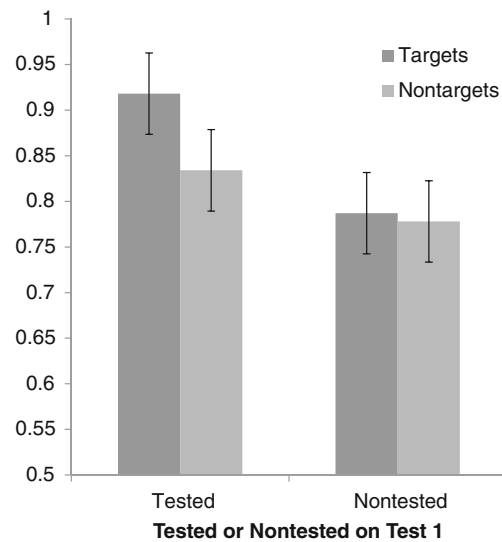


Fig. 5 Experiment 5: Mean hit rates on Test 2 for studied items that were versus were not presented on Test 1, as a function of whether they were from the source to be recognized on Test 1 (targets) or from the other source (nontargets). Error bars represent within-subjects 95 % confidence intervals for the target–nontarget comparison

$MSE = .006, p < .005, \eta_p^2 = .192$]. Hit rates were significantly greater for tested targets than for nontested targets, $t(40) = 7.44, p < .001$. Constraint on recognition was not complete, however, because hit rates were also higher for tested nontargets than for nontested nontargets, $t(40) = 3.06, p < .005$. Nonetheless, the fact that targets benefited from being tested twice as much as did nontargets supports the notion of constrained recognition. This was not simply a bias effect, as the false alarm rates were equivalent for foils of the target length ($M = .13, SD = .13$) versus foils of the other length ($M = .12, SD = .12$), $F < 1$.

Experiment 6

The results of Experiment 5 suggest that subjects can constrain recognition at the front end when items from the to-be-recognized source are perceptibly discriminable from items from the other source (presumably via selective visual attention to words of the target length), whereas the prior experiments suggested that it is difficult if not impossible for subjects to constrain retrieval when the two sources can be differentiated solely on the basis of episodic memory. But that gloss rests on across-experiment comparisons, with the experiments differing in several ways. In Experiment 6, we directly compared the evidence for constraint when the sources were discriminable solely on the basis of memory source versus when the sources were also perceptibly discriminable. Half of the subjects studied and were tested on

materials like those used in Experiment 1, 2, 3, and 4 (i.e., on Test 1, they were to recognize only items that they had studied with a particular orienting task), whereas the remaining subjects studied and were tested on materials like those used in Experiment 5 (i.e., five- and seven-letter words, with Test 1 constraint defined in terms of word length). For all subjects, Test 2 included re-presentations of the foils from Test 1 along with novel foils as in Experiments 1, 2, and 3, but, to simplify the Test 2 task, subjects were instructed to call any item previously presented in the experiment “old” (including words presented as foils on Test 1). Thus, subjects did not have to distinguish between items familiar from study and Test 1 and those familiar only from Test 1.

Method

Subjects We tested 52 University of Victoria undergraduates. Of these, four were dropped for failure to follow the task instructions, leaving 48 in the analysis. These were divided evenly across length- versus source-based constraint and the to-be-constrained dimensions (five vs. seven letters or frequency vs. value).

Materials The materials for subjects in the source-based constraint condition were essentially the same as those in Experiments 1, 2, 3, and 4, and those for subjects in the length-based condition were essentially the same as in Experiment 5, except that in both conditions 30 additional words were added to Test 2. Because the Test 2 task was to recognize any previously seen item (not just those from the study phase), we added these 30 new items to maintain the same 2:1 ratio of items for which the correct answer was “yes” versus “no” as in the previous experiments. These new items were selected from the same source and met the same criteria as the other items (e.g., for subjects in the source-based constraint condition, the words were all medium-frequency concrete nouns that could sensibly be judged on both monetary value and frequency of encounter).

Procedure The procedure was the same as in Experiment 5, with the following exceptions. The Test 1 probed dimension was orienting task (memory-source-based constraint: frequency of encounter vs. monetary value) for a random half of the subjects, and word length (perceptible-cue-based constraint: five vs. seven letters long) for the other subjects. Test 2 instructions were to endorse anything seen in the study phase and/or Test 1. Test 1 instructions were identical for the source-based and length-based conditions (with the exception of the dimension of constraint called for). Where Experiments 1, 2, 3, and 5 had differed procedurally (e.g., their timing parameters), Experiment 5’s parameters were adopted. Also, the study and Test 1 instructions of both conditions explicitly

reminded subjects of the “kitchen sink” manipulations (e.g., that the to-be-recognized items were presented silently in a blue font while the subject was standing) to increase the chances that subjects could use them as effective aids in constraining retrieval. Finally, whereas in Experiments 1, 2, 3, 4, and 5 the Test 1 probes had disappeared immediately if endorsed by a spacebar press, in Experiment 6 they remained on the screen for 2 s whether they were endorsed or not. This modification ensured that the targets and nontargets were presented for equivalent amounts of time in Test 1.

Results and discussion

Test 1 As in the previous experiments, performance on Test 1 was quite good. In the source-constrained condition, the mean endorsement rates for studied targets, studied nontargets, and new items were .87, .22, and .03, respectively. The 95 % within-subjects confidence interval was .04. Subjects in the length-constrained condition did at least as well in terms of hits ($M = .88$) and correctly rejected studied nontargets ($M = .01$). False alarms were more common to nonstudied targets (i.e., new items of the same length as the targets; $M = .17$) than to nonstudied nontargets (for which no subject made a false alarm).

Test 2, memory-source-constrained condition The key findings are depicted in Fig. 6. The memory-source-constrained condition replicated Experiments 2, 3 and 4. While items presented on Test 1 enjoyed the expected recognition advantage on Test 2, $F(1, 22) = 36.1$, $p < .001$, $\eta_p^2 = .621$, there was no significant advantage for tested targets over tested nontargets ($p = .21$), nor a Test 1 Presence \times Target Status interaction ($p = .88$). That is, presentation on Test 1 boosted Test 2 hit rates just as much for items from the not-to-be-recognized source as it did for items from the to-be-recognized source. Thus, there was little indication of constrained retrieval when constraint could be exercised solely on the basis of episodic memory.

Test 2, length-constrained condition The length-constrained condition replicated Experiment 5. Items appearing on Test 1 were recognized better on Test 2, $F(1, 22) = 26.4$, $p < .001$, $\eta_p^2 = .545$; targets were recognized better than nontargets, $F(1, 22) = 9.477$, $p < .01$, $\eta_p^2 = .301$; and these two factors interacted significantly, suggesting an effect of constrained recognition, $F(1, 22) = 6.736$, $p < .02$, $\eta_p^2 = .234$. Thus, when constraint could be exercised on the basis of word length, presentation on Test 1 increased hit rates for items of the to-be-recognized length much more than it increased hit rates for items of the other length. False alarms to foils of the target length ($M = .21$) were not significantly more frequent than false alarms to foils of the other length

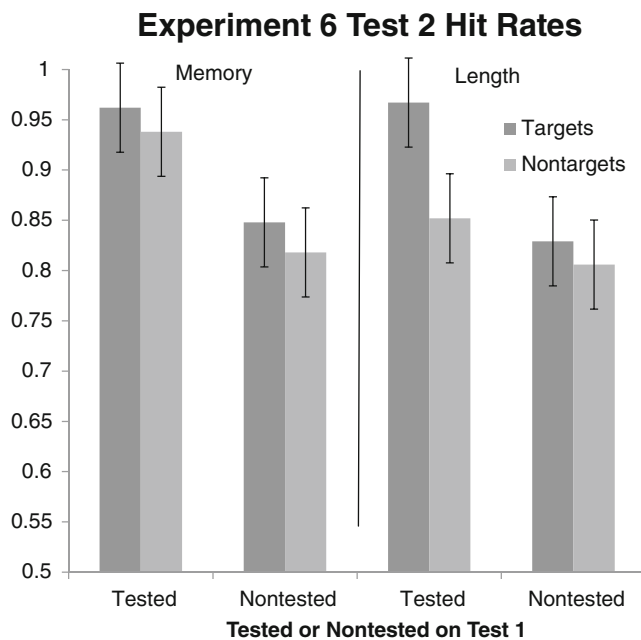


Fig. 6 Experiment 6: Mean hit rates on Test 2 for studied items that were versus were not presented on Test 1, as a function of whether they were from the source to be recognized on Test 1 (targets) or from the other source (nontargets) for subjects for whom the two sources differed only in terms of memory source or also in terms of length. Error bars represent within-subjects 95 % confidence intervals for the target–nontarget comparison

($M = .17$), $t(23) = 1.09$, $p = .29$. Of less interest, there was a near-significant main effect of condition, indicating that subjects for whom five-letter words were the targets identified more old items than did the seven-letter target group, $F(1, 22) = 3.926$, $p = .06$, $\eta_p^2 = .151$.

Comparison of source- and length-constrained groups An omnibus ANOVA was conducted on all of the Test 2 hit rates from this experiment. Of central interest in that analysis was the Test 1 Presence \times Target Status \times Constraint (memory source vs. perceptible) interaction, which fell just short of statistical significance, $F(1, 46) = 4.012$, $p = .051$, $\eta_p^2 = .080$. In an analysis of Test 2 hit rates for tested targets versus tested nontargets as a function of perceptible versus memory-source constraint, the interaction was significant, $F(1, 46) = 7.177$, $p = .01$, $\eta_p^2 = .135$. As is indicated above, the tendency for hit rates to be greater for tested targets than for tested nontargets was significant in the perceptible constraint condition, but not in the memory-source constraint condition.

General discussion

We sought evidence that recognition memory can be constrained at the front end, such that subjects are more likely to retrieve memories of encountering a probe item from a

designated source than to retrieve memories of an equally memorable encounter with a probe item from another source. The rationale for our procedure rests on the assumption that recollecting information about studying an item on Test 1 (“reminding,” to use Hintzman’s, 2004, 2011, term) will powerfully increment that item’s memorability, such that items queried and recognized on Test 1 will later be recognized more often than items queried but not recognized on Test 1.

Figure 7 depicts the key results of our experiments in the form of two difference scores: the difference in hit rates for targets that were versus were not on Test 1 (i.e., benefit of testing for targets) and the difference in hit rates for nontargets that were versus were not on Test 1 (i.e., benefit of testing for nontargets). To the extent that subjects exercised front-end constraint during Test 1, the difference score should be greater for targets than for nontargets. If subjects exercised no constraint at all, such that they were just as likely to retrieve memories of nontargets as of targets on Test 1, the two difference scores should be equivalent.

Overall, our results hint that people may be able to constrain retrieval to a designated memory source even when nothing perceptible differentiates the items from the two sources. The source-constrained retrieval pattern was significant in Experiment 1, and there were nonsignificant hints of such an effect in Experiments 2, 3, 4, and 6. An analysis combining the data from Experiments 1, 2, 3, 4, and the memory-source-constrained conditions of Experiment 6 (with experiment as an “independent” variable) yielded a

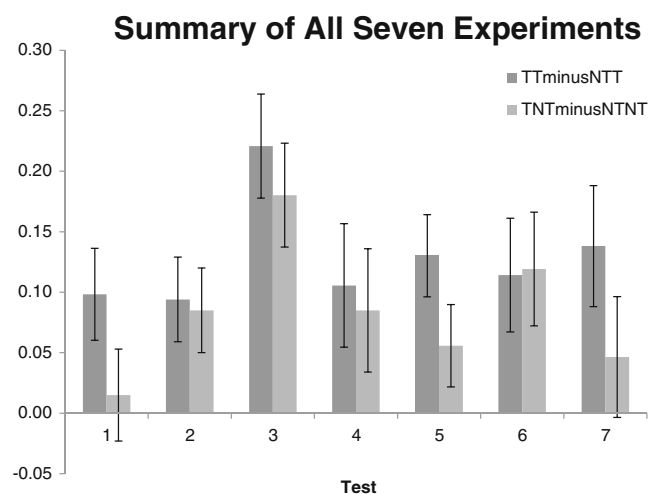


Fig. 7 Summary of the seven tests of the constrained-retrieval effect reported in this article. The darker bars represent the difference in hit rates for targets that were versus were not presented on Test 1, whereas the lighter bars represent the difference in hit rates for nontargets that were versus were not presented on Test 1. Error bars represent within-subjects 95 % confidence intervals calculated for the contrast between the two difference scores in each experiment. In the labels on the abscissa, 1 = Experiment 1; 2 = Experiment 2; 3 = Experiment 3; 4 = Experiment 4 (using the within-subjects comparison across Tests 2 and 3); 5 = Experiment 5, 6 = the memory-source constraint condition of Experiment 6, 7 = the perceptible-source constraint condition of Experiment 6

significant constrained-retrieval interaction (i.e., greater benefit for testing targets than for testing nontargets), $F(1, 146) = 5.211, p = .024, \eta_p^2 = .034$. The hit rate was significantly higher for tested targets ($M = .919, SD = .095$) than for tested nontargets ($M = .882, SD = .113$), $t(150) = 3.91, p < .001$, and the three-way interaction with experiment did not approach significance, $F = 1.326, p = .26$. An analysis of the same data excluding Experiment 1 yielded a significantly higher hit rate for tested targets ($M = .916, SD = .101$) than for tested nontargets ($M = .891, SD = .099$), $t(120) = 2.674, p < .01$, but the critical interaction was no longer significant, $F(1, 117) = 1.223, p = .27$. Taken together, these comparisons suggest the existence of a small constrained-retrieval effect, but one that was carried to a large extent by Experiment 1. We also failed to get the predicted effect in several studies not reported here. We are not asserting the null hypothesis, but nor can we make a strong case for having demonstrated memory-source-based constrained retrieval.

Why was the evidence of source-constrained retrieval so weak? One possibility is that although subjects could have constrained retrieval on Test 1, they chose not to do so. On that test, both items from studied sources and new items were randomly intermixed and presented one at a time, and subjects were instructed to focus exclusively on recognizing items from a designated source. We specifically told subjects that we wanted them to query memory solely for evidence that each probe was from the designated source. To foster that orientation, subjects responded overtly only when they judged an item to be from the designated source. Nonetheless, it is possible that subjects disregarded the instructions and deliberately tried to recognize items from both sources, using a recall-to-reject strategy when they recollected information indicating that an item was from the nondesignated source. Indeed, some subjects reported in debriefing that they had used this sort of strategy. An anonymous reviewer of an earlier version of this article suggested that the use of a recall-to-reject strategy could be discouraged by presenting some filler items in both sources and making sure that subjects understood that the fact that an item had appeared in one source did not necessarily mean that it had not also been in the other source, as in Pierce and Gallo (2011) and Bridger and Mecklinger (2012). We agree that this would be a good procedural improvement.

Even if subjects did follow instructions and tried to constrain retrieval on Test 1, many situational factors likely fostered spontaneous retrieval of information about the not-to-be-recognized (nontarget) studied items on Test 1, thereby undermining constraint. Despite our kitchen sink manipulations intended to differentiate Sources A and B (different orienting tasks, different colors and positions on the screen, etc.), items from both sources had been encountered in the same context, as part of the same episode, with the same researcher, and so forth. Those multiple kinds of similarity might have made it

difficult for subjects to query memory for evidence of having encountered items from one source without spontaneously retrieving information about studying items from the other source (cf. Lindsay, Allen, Chan, & Dahl, 2004).

Our results indicate that even in this difficult situation, constraint can robustly be exercised if items from the two sources perceptibly differ in an obvious way. Our results indicate that subjects can look directly at a recently studied word on a memory test with sufficient visual acuity to identify it as being of the nontarget length without being reminded of their study-list encounter with that item (as suggested by lower Test 2 hit rates for tested nontargets than for tested targets), if they have configured themselves to recognize items from a different source and if an obvious perceptible feature differentiates items from the target source from items from other sources. But our results suggest that it is difficult to exercise such constraint unless items from the to-be-recognized source are perceptibly different from items from other sources.

What was the mechanism of constraint in Experiment 5 and the perceptible-difference condition of Experiment 6? We think that it is reasonable to claim that recognition was constrained, because the data from Test 2 indicate that during Test 1 subjects had recognized tested nontargets less often than tested targets. But the mechanism that brought about that constraint was probably attentional or perceptual rather than mnemonic. When the targets and nontargets perceptibly differed, subjects presumably used something like a visual template, selecting stimuli that matched that template for further processing and aborting further processing of other stimuli at an early stage (perhaps before semantic identification). Such a strategy is a reasonable one if the goal of constraining recognition is to preempt extended processing of a probe on the basis of the absence of a sought-after feature. For a test probe to act as a cue and evoke memories of the study episode, that test probe must be processed to some considerable degree. We speculate that by truncating processing of nontarget stimuli at a relatively primitive level of analysis, subjects were able to reject those items as nontargets without being reminded of their study-list encounters with the items. This is like looking at a familiar person and not recognizing him or her because you are looking for someone else. That is, recognition is constrained through preemptive cessation of nontarget processing, guided by top-down influences as to what defines targets and nontargets, such that attentional/perceptual constraint subserves front-end-constrained recognition. Given longstanding claims as to the automaticity of word reading (see MacLeod's 1991 review), we think that it is impressive and interesting that subjects can filter out non-target-length words without recognizing them from the study list. But we would not describe that pattern as evidencing memory-based constrained retrieval.

We showed evidence of front-end constraint on recognition when words from two sources differed in length. It may be that recognition can also be constrained on the basis of stimulus characteristics that are more subtle and symbolic than word length. For example, if the items in Source A were verbs and those from Source B were nouns (a more abstract sort of perceptible difference), perhaps subjects could configure themselves selectively to recognize items from one of those sources. We speculate that as the nature of the “perceptible” difference between to-be-recognized and not-to-be-recognized items becomes more abstract, the ability to constrain recognition will be challenged. This is because differentiating to-be-recognized and not-to-be-recognized stimuli in such a case would require the subject to engage with and process the stimulus to a greater extent and a deeper level than when the two sources can be differentiated by a crude, physical characteristic such as length.

Can people exercise source-constrained retrieval without any perceptible source cues? The significant effects in Experiment 1 and in the mega-analysis suggest that they can, but the null effects in our other tests of this hypothesis imply that such constraint is difficult if not impossible. Perhaps if Sources A and B were widely separated in time and presented in qualitatively different contexts, then people could more easily selectively probe recognition memory for records of A items without recognizing B items, even if the two sources did not differ in any perceptible way. We speculate, for example, that if you encountered a distant in-law you might recognize that person in the context of a family reunion but fail to recognize him or her in the context of a psychology conference (even if your relatives look like psychologists). Answers to such questions will help add specificity to theoretical proposals as to the mechanisms underlying top-down constraint on recognition memory.

What is the theoretical import of this article? With respect to source-constrained retrieval, our data are ambiguous, and ambiguous data are less pleasing than clear data. But several years of concerted effort have yielded only tantalizing hints of memory-source-constrained retrieval. It may be that those hints are Type I errors, especially given that we sometimes indulged in data-peeking. We conclude that either (a) recognition memory cannot be constrained solely on the basis of source memory or (b) constraint is possible but (at least under the conditions that we have explored) difficult, such that the effects are small and inconsistent. We lean toward the latter explanation, but distinguishing between noneffects and small effects is notoriously difficult. The main contribution of this article, then, is to make the point that there is only ambiguous evidence for the claim that recognition memory can be constrained “at the front end,” in such a way that people are more likely to retrieve information

about items from a designated source than to retrieve information about items from another equally accessible source. There is value in the revelation of ambiguity, and we hope that the work reported here will inspire other researchers to develop methods capable of resolving the question of whether or not recognition memory can be constrained solely on the basis of memory source.

Author note This research was supported by an Erich and Shelley Mohr Fellowship to J.K. and a Discovery grant from the Natural Sciences and Engineering Research Council to D.S.L. We thank Robert Greene, Larry L. Jacoby, Colleen M. Kelley, Jeffrey P. Toth, and five anonymous reviewers for astute and insightful comments on earlier versions of the manuscript, and Brian Buchan, Priya Rosenberg, and Joseph Savidge for their assistance with data collection.

References

- Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory and Cognition*, *40*, 681–692. doi:10.3758/s13421-012-0185-5
- Bridger, E. K., Herron, J. E., Elward, R. L., & Wilding, E. L. (2009). Neural correlates of individual differences in strategic retrieval processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1175–1186. doi:10.1037/a0016375
- Bridger, E. K., & Mecklinger, A. (2012). Electrophysiologically dissociating episodic preresolution processing. *Journal of Cognitive Neuroscience*, *24*, 1476–1491. doi:10.1162/jocn_a_00152
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505. doi:10.1080/14640748108400805
- Danckert, S. L., MacLeod, C. M., & Fernandes, M. A. (2011). Source-constrained retrieval influences the encoding of new information. *Memory and Cognition*, *39*, 1374–1386. doi:10.3758/s13421-011-0117-9
- Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*, 467–473. doi:10.1111/1467-9280.00290
- Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 259–278. doi:10.1037/0278-7393.23.2.259
- Herron, J. E., & Rugg, M. D. (2003). Retrieval orientation and the control of recollection. *Journal of Cognitive Neuroscience*, *15*, 843–854. doi:10.1162/089892903322370762
- Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory and Cognition*, *32*, 336–350. doi:10.3758/BF03196863
- Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the “coordinates of truth.” *Perspectives on Psychological Science*, *6*, 253–271. doi:10.1177/1745691611406924
- Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–400). New York: Guilford Press.
- 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 and Review*, *12*, 852–857. doi:10.3758/BF03196776

- 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. doi:10.1016/j.jml.2005.01.007
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*, 85–97. doi:10.1016/j.jml.2011.04.002
- Lindsay, D. S., Allen, B. P., Chan, J. C. K., & Dahl, L. C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of Memory and Language*, *50*, 96–111.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203. doi:10.1037/0033-2909.109.2.163
- MacLeod, C. M., Pottruff, M. M., Forrin, N. D., & Masson, M. E. J. (2012). The next generation: The value of reminding. *Memory and Cognition*, *40*, 693–702. doi:10.3758/s13421-012-0182-8
- Marsh, R. L., Meeks, J. T., 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. doi:10.1016/j.jml.2009.06.005
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*, 203–220. doi:10.1037/h0087426
- McDuff, S. G. R., Frankel, H. C., & Normal, K. A. (2009). Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *Journal of Neuroscience*, *29*, 508–516. doi:10.1523/JNEUROSCI.3587-08.2009
- Nairne, J. S. (2002). The myth of the encoding–retrieval match. *Memory*, *10*, 389–395. doi:10.1080/09658210244000216
- Nelson, D. L., Canas, J. J., Bajo, M. T., & Keelean, P. D. (1987). Comparing word fragment completion and cued recall with letter cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 542–552. doi:10.1037/0278-7393.13.4.542
- Pierce, B. H., & Gallo, D. A. (2011). Encoding modality can affect memory accuracy via retrieval orientation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 516–521. doi:10.1037/a0022217
- Potter, M. C. (1975). Meaning in visual search. *Science*, *187*, 965–966. doi:10.1126/science.1145183
- Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex*, *43*, 135–145.
- Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, *12*, 664–678. doi:10.1162/089892900562291
- Shimizu, Y., & Jacoby, L. L. (2005). Similarity-guided depth of retrieval: Constraining at the front end. *Canadian Journal of Experimental Psychology*, *59*, 17–21. doi:10.1037/h0087455
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. doi:10.1037/h0020071