Journal of Experimental Psychology:
Learning, Memory, and Cognition
2009, Vol. 35, No. 1, 205–220


0278-7393/09/$12.00 DOI: 10.1037/a0014275

# The List-Strength Effect in Recall: Relative-Strength Competition and Retrieval Inhibition May Both Contribute to Forgetting

Michael F. Verde
University of Plymouth

According to the principle of relative-strength competition, stronger items in memory block the retrieval of weaker items. This principle, integral to many theories of forgetting over the years, derives much of its support from the list-strength effect (LSE), in which strengthening some items in a study list makes it more difficult to recall other items. Work in the retrieval-induced forgetting literature has challenged the existence of relative-strength competition, 1st by offering many examples of a null LSE and 2nd by proposing that extant observations of the LSE can be explained by retrieval inhibition. In the present study, a series of experiments produced a robust LSE in cued recall under conditions meant to control the contribution of retrieval inhibition. Simulations of the SAM-REM model of recall (K. J. Malmberg & R. M. Shiffrin, 2005) showed that a model based on relative-strength competition can accommodate both the presence and absence of an LSE. The empirical results and model simulations together make a case for the role of strength-based competition in forgetting.

*Keywords:* list strength, retrieval-induced forgetting, cued recall, retrieval inhibition, forgetting

For the better part of a century, the study of forgetting has focused on the detrimental influence of other memories. A classic demonstration of memory interference comes from paired-associate learning: Subjects study item pairs such as *apple-dog*, *apple-lake*, and *frog-stone* (designated *A-B*, *A-C*, *D-E*) and later must recall one item of a pair when given the other as a contextual cue (*apple-_?*). Typically, greater forgetting is observed for *A-B* and *A-C* than for *D-E*. While most theories attribute this to interference that arises when multiple items (*B*, *C*) are associated with a single retrieval cue (*A*), the specific mechanism of forgetting has been a matter of frequent debate over the years (M. C. Anderson, Bjork, & Bjork, 1994; Postman & Underwood, 1973).

When a cue is associated with several items, all become potential candidates for retrieval. The retrieval set may be quite broad: In free recall, the experimental context itself serves as a cue that is episodically tied to all recently studied items. A more specific cue like *apple-_?* defines a narrower set of candidates (*dog, lake*). According to the principle of competitor interference, multiple memories recruited by the same cue interfere with the retrieval of any specific memory, with interference increasing as a function of the number of competitors. It is a principle widely accepted among theories of recall and supported by evidence such as the list-length effect, the finding that increasing the length of study lists has a detrimental effect on recall accuracy and latency (Roberts, 1972; Wixted & Rohrer, 1994).

Michael F. Verde, School of Psychology, University of Plymouth, Plymouth, England.

I wish to thank M. C. Anderson and J. G. W. Raaijmakers for their insightful comments on a previous version of this article.

Correspondence concerning this article should be addressed to Michael F. Verde, School of Psychology, University of Plymouth, Plymouth PL4 8AA, England. E-mail: michael.verde@plymouth.ac.uk

More controversial is the principle of relative-strength competition, according to which interference depends not only on the number of retrieval candidates but also on their relative strength. A cue and an item become episodically associated when they are studied together, and the strength of this association depends on the length, frequency, or quality of study. Strengthening the cue-item association increases the likelihood of retrieving that particular item but also increases its ability to interfere with the retrieval of competitors that are also associated with the cue. For example, relative-strength competition predicts that repeatedly studying the pair *A-B* will make it increasingly easy to recall *A-B* and increasingly difficult to recall the studied competitor, *A-C*. Although details of the competitive process vary by model, recall is often described as a stochastic sampling process in which one item at a time is sampled from the pool of competitors. Strong competitors are more likely to be sampled, effectively blocking the sampling of weaker competitors (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980, 1981; Rundus, 1973; see also J. R. Anderson, Bothell, Lebiere, & Matessa, 1998). The principle of relative-strength competition has been applied to a range of phenomena and is integral to a number of influential models.

## List-Strength Effect

Key evidence for relative-strength competition comes from the *list-strength effect* (LSE), the finding that rates of recall are influenced by the relative strength of items in a study list. Although the term is used in this article in this broadest sense, it has most often been used specifically in the context of studies in which several types of lists are compared: mixed lists containing both strong and weak items, pure weak lists containing only weak items, and pure strong lists containing only strong items. In these studies, strength is typically manipulated via study duration, repetition, or depth of processing. Relative-strength competition predicts that memory for weak items should be worse in mixed lists, where competitors are stronger on average than in pure weak lists. Similarly, memory for strong items should be better in mixed lists, where competitors are weaker on average than in pure strong lists. Such a pattern, referred to as a positive LSE, has been observed often in free recall (Malmberg & Shiffrin, 2005; Ratcliff, Clark, & Shiffrin, 1990; Tulving & Hastie, 1972; Wixted, Ghadisha, & Vera, 1997). Somewhat puzzling is the fact that evidence for an LSE in cued recall has been less consistent (Ratcliff et al., 1990).

In recent years, the study of retrieval inhibition as a core mechanism of forgetting has led to a reevaluation of many standard accounts of recall phenomena, including retroactive interference, part-list cueing, output interference, and, notably, LSE (M. C. Anderson, 2003; M. C. Anderson et al., 1994). Retrieval inhibition has been most extensively studied with the retrieval-induced forgetting paradigm, in which subjects study several sets of pairs, each set composed of items related to a single cue. For example, one set might include the pairs *fruit-apple* and *fruit-orange* (*A-B*, *A-C*), another set *birds-robin* and *birds-sparrow* (*D-E*, *D-F*). Some of the sets are assigned to the retrieval-practice condition: Following study, half of the pairs (*A-B*) from these sets are presented in a cued-recall format (*fruit-a—?*). Finally, a recall test is given for all of the originally studied pairs. The critical comparison is between unpracticed pairs from the retrieval-practice condition (*A-C*) and pairs from the unpracticed condition (*D-F*). The standard finding is that pairs whose competitors were included in a recent test of recall (*A-C*) show increased forgetting relative to baseline pairs (*D-F*). According to the modal theory of retrieval inhibition (e.g., M. C. Anderson, 2003; M. C. Anderson, Green, & McCulloch, 2000), retrieving a target memory requires actively inhibiting other memories that share features or that are associated with the same retrieval cue, making those memories less accessible. Thus, recall of *A-B* during the retrieval practice phase inhibits *A-C*, making it subsequently more difficult to retrieve. This theory adheres to the principle of competitor interference in suggesting that the presence of multiple retrieval candidates makes it more difficult to retrieve a specific target. Unlike with relative-strength competition theories, however, forgetting is not thought to be the result of competition among memories per se but rather a byproduct of the inhibition mechanism that helps to resolve competition. The idea that the LSE may be largely a function of retrieval-induced forgetting poses a serious challenge to any model of recall based on the principle of relative-strength competition.

One way to distinguish between the theoretical viewpoints is to determine whether selectively strengthening a memory item is sufficient to induce forgetting of its competitors. Suppose that *A-B* is studied to a greater degree than is *A-C*. Relative-strength competition theory predicts that when *A-_?* is later presented as a cue to recall the studied items, the preferential retrieval of *A-B* will block or make less likely the retrieval of *A-C*. Inhibition theory, on the other hand, predicts no selective forgetting of *A-C* because there has been no prior attempt to retrieve its competitor *A-B*. A number of studies have explicitly tested these distinct predictions within the retrieval-induced forgetting paradigm, in which a retrieval practice phase tests recall of selected study items prior to the final test. Replacing the retrieval practice trials with trials in which the same items are simply presented again for additional study has produced consistent results: Unpracticed competitors are impaired following retrieval practice but not following additional study (M. C. Anderson & Bell, 2001; M. C. Anderson, Bjork, & Bjork, 2000; Ciranni & Shimamura, 1999). Note that M. C. Anderson and Bell (2001) initially observed impairment in the additional-study condition, but posttest questionnaires revealed that the impairment was primarily observed in subjects who claimed to have engaged in covert retrieval during the added study.

The inhibition theoretic view has been applied to other paradigms in which the effect of list strength on recall has been observed. In retroactive interference, memory for an initial list of items is impaired by the learning of subsequent lists. Barnes and Underwood (1959) showed that the degree of impairment is related to the strength of the subsequent lists, a finding used to support relative-strength competition models (e.g., Mensink & Raaijmakers, 1988). However, M. C. Anderson et al. (1994) pointed out that many early studies incorporated retrieval practice into the learning trials, confounding retrieval and strengthening. In an attempt to replicate earlier studies, Bäuml (1996) manipulated the strength of study items interpolated between an initial study list and a final memory test by varying their presentation duration; interpolated items were each studied for 5 s in the strong condition and 2 s in the weak condition. He found that when the opportunity for retrieval practice of the interpolated items was controlled, recall of the initial list was the same in both conditions. Thus, the strength of competing items did not influence the rate of forgetting (but see Delprato, 2005).

In part-list cueing, presenting some of the studied items as retrieval aids during the final memory test impairs recall of the remaining items (for review, see Nickerson, 1984). This phenomenon has sometimes been interpreted as a form of LSE: Presenting items as cues during recall strengthens those items, thereby increasing their ability to interfere with competitors (Rundus, 1973). M. C. Anderson et al. (1994) offered an alternative possibility: When items are supplied as cues, subjects covertly recall their earlier presentation, thereby inhibiting other studied items. In fact, this is consistent with M. C. Anderson and Bell's (2001) observation that subjects who expect to have their memory tested may engage in covert retrieval during study, presumably as a form of strategic rehearsal. Bäuml and Aslan (2004) manipulated this tendency in the part-list cueing paradigm. Prior to the final recall test, subjects were reexposed to a subset of studied items under one of three conditions. In the additional study condition, subjects were instructed to treat the trials as another chance to learn the items. In the cueing condition, the items were described as useful cues to use in the memory test to follow. In the retrieval condition, the items were targets in a cued-recall test. During the final test, in which subjects were to recall all studied items, items that had not been reexposed in the preceding phase were impaired in the retrieval and cueing conditions but not in the additional study condition. Bäuml and Aslan interpreted this as evidence of covert retrieval in the cueing condition.

Finally, Bäuml (1997) examined retrieval inhibition in the standard LSE paradigm. Study lists consisted of items belonging to several categories. In the pure strong and pure weak categories, each item was studied for 6 s and 2 s, respectively. In mixed categories, half of the items were studied for 6 s and half for 2 s. In previous studies with a similar design, the LSE has been observed most strongly and consistently in free recall. In free recall, however, subjects tend to produce strong items early during the test period (Wixted et al., 1997). According to inhibition theory, items retrieved early on will inhibit the recall of items not yet retrieved. Thus, weak items in mixed lists should be more susceptible to forgetting as a result of retrieval inhibition from strong items recalled earlier. Strong items, on the other hand, should be less susceptible to retrieval inhibition because fewer items should be recalled before them. Using cued recall to control output order, Bäuml found no evidence of an LSE for items tested at the beginning of the list. However, as predicted by the inhibition hypothesis, memory was poorer for items tested at the end of the list. Bäuml concluded that the LSE in free recall may be nothing more than an artifact of confounding strength and retrieval order.

The repeated failure to observe a connection between forgetting and competitor strength in the retrieval-induced forgetting literature suggests that, contrary to many established theories, relative-strength competition is not a mechanism of forgetting. Such a conclusion draws on two lines of argument. First, retrieval inhibition may be sufficient to account for recall phenomena previously attributed to relative-strength competition, making the latter unnecessary in a parsimonious theory of memory. Second, the principle of relative-strength competition seems to suggest that increasing the ability to recall a particular memory must produce an observable impairment in the ability to recall competing memories. Thus, strengthening without competitor impairment is proof against relative-strength competition. The present study addresses both of these arguments. To begin, a series of experiments are reported in which an LSE in cued recall is observed that cannot be easily attributed to retrieval-induced forgetting. Given the problematic nature of previous evidence for an LSE in recall (either weak and inconsistent findings or the confounding of strength with retrieval inhibition), a solid empirical basis for relative-strength competition needs to be established. Following this, I suggest that the theoretical treatment of relative-strength competition prevalent in the retrieval-induced forgetting literature is incomplete. The relationship between strength and competitor inhibition described above holds true for a basic model of competition often seen in theories of recall. However, more complex depictions of competition such as that found in Malmberg and Shiffrin's (2005) SAM-REM model of recall can predict both the presence and absence of an LSE.

## Experiment 1

A number of studies have observed an LSE in free recall. However, the inability to control the output order of strong and weak items makes it difficult to independently evaluate the retrieval inhibition and relative-strength competition accounts of these results. Malmberg and Shiffrin's (2005) study is no exception, although their finding of an LSE with some manipulations and not others leads to complications for the retrieval inhibition account. An explanation based solely on retrieval inhibition would suggest that spaced repetition makes it more likely that strong items are output early in the test, making weak items more vulnerable to retrieval-induced forgetting. Why would increased study duration not have the same effect? One might point to the fact that repetition produced more strengthening than was accomplished through the other methods in their study, in some cases substantially so. Thus, the other methods may not have sufficiently affected the output order of strong and weak items. Alternatively, there may be intrinsic reasons that some manipulations and not others influence output order. Whatever the merits of these possible explanations, the contribution of retrieval inhibition to the LSE that was observed in the study remains unknown.

In cued recall, the inconsistent and weak LSE observed by Ratcliff et al. (1990) is not strong support for relative-strength models. On the contrary, it has sometimes been used to argue against them (e.g., Bäuml, 1997). A recent study by Delprato (2005) has provided some additional evidence for an LSE with cued recall in the retroactive interference paradigm. In Delprato's initial experiment, subjects studied and recalled a first list, studied a second list some number of times, and recalled the first list again. In the final test, first list recall decreased as a function of the number of times the second list was repeated. From the retrieval inhibition perspective, a weakness of the study is that it failed to control for covert retrieval practice during repeated study of the second list. Delprato acknowledged that testing the first list immediately made subjects aware of forthcoming memory tests and may have encouraged strategic rehearsal (including covert retrieval) during the subsequent study list. To rectify this, he conducted a second experiment in which he removed the first recall test and did not warn subjects about the final test. Unfortunately, the repetition manipulation was weaker than in the initial experiment, so that the strength-related impairment of first list recall was not statistically significant.[1] Finally, as is discussed below, the confounding of the strength manipulation with time leads to additional problems of interpretation.

In Experiment 1, stimuli were randomly paired nouns formed into overlapping sets (*A-B*, *A-C*, etc.). Six pairs were created per set. For baseline sets, all pairs were shown once during study. For experimental sets, half of the pairs were shown once (weak items) and half were shown three times (strong items). Both weak and baseline conditions occurred within the same study list, allowing for a within-subjects test of the LSE. Only weak and baseline items were tested, and these conditions were probed on alternating trials within the test list. This design made it possible to separate the contributions of relative-strength competition and retrieval inhibition. An LSE is observed if recall is lower in the weak condition compared with the baseline condition. Because the factors that influence relative-strength competition are present at encoding, an LSE should be observed even at the beginning of the memory test, before any retrieval-induced forgetting can have occurred. However, an LSE observed at the end of the test can also be tied to relative-strength competition. Output order is controlled such that the number of items from the same set appearing earlier in the test is equivalent for weak and baseline conditions. Because retrieval-induced forgetting must therefore be equivalent in both conditions at each point in the test, any observed LSE cannot be attributed to retrieval inhibition.

If the goal is to show that strength-based competition can exist alongside inhibition, then it is desirable to have some measure of inhibition within the same task. To this end, performance at different points in the test was examined for evidence of output interference, the progressive decline in recall over the course of the

[1] In this experiment, studying a second list interfered with first-list recall compared with a control condition in which no second list was present. However, this is evidence for a list-length effect, which is not useful for discriminating between inhibition and relative-strength competition.

test. Inhibition theory suggests that output interference occurs when each successive retrieval trial leads to the suppression of items that have yet to be tested (M. C. Anderson et al., 1994; Bäuml, 1997). Of course, there are alternative explanations for output interference. Relative-strength competition predicts that if each retrieval trial is an additional encoding opportunity, this would strengthen the probed item and increase its ability to interfere with other items (Raaijmakers & Shiffrin, 1980). There is also the possibility of a host of factors correlated with time—such as decay, general cognitive interference, and contextual fluctuation—that might be broadly termed *nonspecific interference*. While it is not a primary goal of this study to test competing explanations of output interference, the alternatives are considered later. For the moment, output interference is used as an ostensible measure of inhibition that can be contrasted with the LSE, which is uniquely predicted by relative-strength competition.

A concern sometimes voiced in the LSE literature is that strong and weak items may be differentially rehearsed during study (Murnane & Shiffrin, 1991; Ratcliff et al., 1990). Borrowing rehearsal time from strong items to give to weak items may hide an LSE, while the reverse may falsely suggest that one exists. To deal with this concern, in Experiment 1b (which was otherwise identical to 1a) I provided a cue after each study pair to indicate whether or not the pair would appear in the memory test. Because subjects knew that all of the strong pairs and half of the baseline pairs would not be tested, there should be no reason to rehearse them at all (although the encoding task required some degree of initial processing).

### Method

*Subjects.* Undergraduates from the University of Massachusetts Amherst participated for extra credit in their psychology course. Forty-four participated in Experiment 1a, and 23 participated in Experiment 1b.

*Materials and design.* Stimuli were drawn from a pool of 180 low-frequency (<100 per million; Kucera & Francis, 1967), five- to eight-letter nouns. Each of nine blocks consisted of a study list followed by a test list. For each block, two critical sets were each created by pairing one word with six other words (*A-B, A-C, A-D, A-E, A-F, A-G*), producing six unique pairs in each set. For the baseline set, each pair appeared once (baseline condition) in the study list. Only three of the pairs would later appear in the memory test. For the strong–weak set, three of the pairs appeared once (weak condition) and three appeared three times (strong condition) in the study list. Only the weak pairs would later appear in the memory test. The right-side words (*B*, *C*, etc.) of the strong and nontested baseline pairs were matched between sets, such that each right-side word in the baseline set shared the first two letters with a word in the strong–weak set. This was to make clear that in order to perform accurately in the cued-recall test, subjects could not simply memorize the right-side words but had to learn the pairs. For the weak and baseline pairs included in the recall test, every right-side word began with a different pair of letters so that all word-stem cues in the test were unique. Finally, a third set of pairs was created as filler items. Each block contained a unique set of words. For each subject, right-side words for a block were drawn randomly (with the above constraints) from a subset of words whose initial letters fell within a range of up to six places on the alphabet (e.g., letters *G–L*).

Each study list contained 22 pairs. The first two (list positions 1–2) and last two (positions 21–22) pairs were filler items. The three strong pairs appeared in random order in positions 3–5 and again in positions 18–20. All 12 critical pairs appeared in random order in positions 6–17. Each test list contained eight pair fragments for cued recall (the left-side word and the first two letters of the right-side word from a studied pair; e.g., *umbrella-ab—?*). The first two were created from filler items. The remaining six fragments were created from three baseline (*H-I, H-J, H-K*) and three weak (*A-E, A-F, A-G*) pairs. Baseline and weak fragments alternated in list positions 3–8 (e.g., *H-I, A-E, H-J, A-F, H-K, A-G*). The first item appearing after the two filler items in each block alternated between the baseline and weak conditions, with the order in the initial block counterbalanced between subjects.

*Procedure.* At the beginning of the 40-min session, each subject was seated at a computer and fully instructed on the procedures of the study and test phases of the experiment. All materials were presented on the computer screen. Subjects were given a practice block before continuing on to the eight experimental blocks. On each trial of the study list, a word pair appeared in the center of the screen for 3,500 ms, then was replaced by a fixation line of "?" characters. Subjects were to decide if the two words "seem to go together and easily conjure up a vivid image," pressing the *z* key for low imagery or "/" key for high imagery as soon as the fixation line appeared. A 500-ms blank interval followed the keyboard response.

Each trial of the test list began with a fixation line of "+" characters in the center of the screen for 1,500 ms. This was replaced by a pair fragment (e.g., *umbrella-ab—?*).Subjects were to press the space bar as soon as they recalled the right-side word in the pair, then type the word on the keyboard. They were instructed not to guess; if they could not remember the missing word, they were to type *forgot* as their answer.

Experiment 1b differed from Experiment 1a in only one respect. During the period when subjects studied the list, when each of the eight pairs to appear in the subsequent test came on-screen, it was followed by a beep and the prompt *remember.* Subjects were told that they would have to remember only these pairs for the memory test.

### Results and Discussion

The weak and baseline conditions each consisted of different overlapping sets. Evidence of an LSE depends on the comparison between weak and baseline conditions, while evidence of output interference depends on the comparison between the first and last items tested within a set. The means from these comparisons are shown in Figure 1 (see also Table 1). Each experiment was analyzed separately with a 2 (condition) × 2 (test position) repeated measures analysis of variance (ANOVA). For Experiment 1a, a main effect of condition, $F(1, 43) = 10.55$, $MSE = 0.03$, $p = .002$, was due to lower recall for weak items compared with baseline. A main effect of position, $F(1, 43) = 35.13$, $MSE = 0.02$, $p < .001$, was due to lower rates of recall in the last test position compared with the first. There was a significant interaction between condition and test position, $F(1, 43) = 35.13$, $MSE = 0.02$, $p < .001$. Planned comparisons showed that fewer weak than

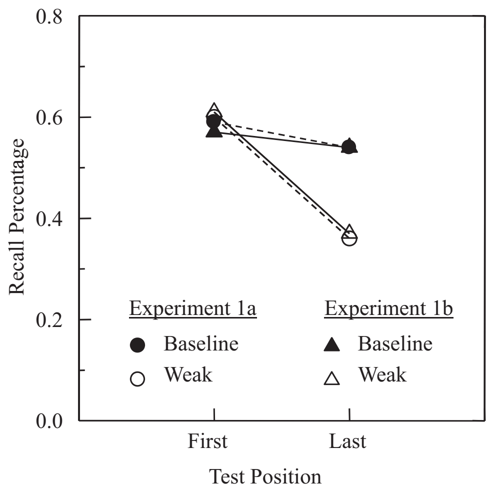


*Figure 1.* Experiment 1.

baseline items were recalled at the last test position, $t(43) = 5.26$, $p < .001$, but the weak and baseline conditions did not differ at the first test position, $t(43) = 0.23$, $p = .818$, *ns*. Similar results were obtained for Experiment 1b. There was a marginal main effect of condition, $F(1, 22) = 3.52$, $MSE = 0.03$, $p = .074$, a main effect of position, $F(1, 22) = 16.33$, $MSE = 0.03$, $p = .001$, and a significant interaction between the two, $F(1, 22) = 4.68$, $MSE = 0.06$, $p = .042$. Fewer weak than baseline items were recalled at the last test position, $t(22) = 3.35$, $p = .003$, but the weak and baseline conditions did not reliably differ at the first test position, $t(22) = 0.55$, $p = .585$, *ns*.

In Experiment 1b, subjects were cued immediately after seeing an item in the study list if that item would later appear in the memory test. Despite this, the results were identical to those in Experiment 1a, allowing the conclusion that subjects were not using differential rehearsal strategies. However, the observation of an LSE only at the end of the test was unexpected. The conditions at test for weak and baseline sets were identical in every way. There were an equal number of tested items from each set, and the items were of equal strength (one study repetition) and equally spaced within the test list. This combined with the fact that recall rates at the beginning of the test were also equivalent indicates that no retrieval-related factor could have produced an LSE at the end of the test. Therefore, the LSE must be attributed to the strength manipulation at study. However, the LSE should then have been evident at the start of the test as well. A second and perhaps related puzzle is the much greater decline in recall for the weak condition over the course of the test. Again, there is nothing obvious in the structure of the test that should have produced greater output interference for the weak condition.

The data provide mixed evidence for both competition and inhibition. A possible explanation for the curious results lies in the fact that little (two study fillers and two test fillers) separated the critical items in the study and test lists. The strong items, which belonged to the same set as the weak items, were presented at the end of the study list just before the fillers and may have still been available in short-term memory at the beginning of the test. Recall impairment has sometimes been tied to retrieval cues that impose a nonoptimal or inappropriate search path for the targets (Allen, 1969; Basden, Basden, & Galloway, 1977). Having items from the opposing set still prominent in short-term memory may have impeded the retrieval of baseline items by serving as inappropriate cues, masking the advantage that the LSE should have produced at the beginning of the test. This problem would have disappeared later in the test, when both search sets were activated to the same degree. The LSE observed in the final portion of the test is consistent with this.

## Experiment 2

The second set of experiments included materials and procedures similar to those in the first. However, the study and test lists were doubled in size by taking what had previously been two sets of lists and combining them into a single set. This meant that there were twice as many filler items separating the critical items in the study and test lists. Doubling the number of pair sets also made it less likely for any one of them to have a strong influence on defining retrieval paths at the beginning of the test. Experiment 2b was identical to Experiment 2a save that it further guarded against the influence of short-term memory by interpolating an unrelated memory-span task between study and test phases.

### Method

*Subjects.* Undergraduates from the University of Massachusetts Amherst participated for extra credit in their psychology course. Twenty-five participated in Experiment 2a, and 23 participated in Experiment 2b.

*Materials and design.* These were identical in most respects to Experiment 1. However, the study and test lists were twice as long.

Table 1
*Mean Percentage Recall and Standard Deviation for the First and Last Items Tested in a Set*

| Experiment and condition | *n* | First item *M* | First item *SD* | Last item *M* | Last item *SD* |
|---|---|---|---|---|---|
| 1a | | | | | |
| Baseline | 44 | .59 | .20 | .54 | .18 |
| Weak | 44 | .60 | .17 | .36 | .18 |
| 1b | | | | | |
| Baseline | 23 | .57 | .25 | .54 | .24 |
| Weak | 23 | .61 | .20 | .37 | .19 |
| 2a | | | | | |
| Baseline | 25 | .50 | .18 | .33 | .19 |
| Weak | 25 | .39 | .20 | .20 | .16 |
| 2b | | | | | |
| Baseline | 23 | .42 | .20 | .29 | .21 |
| Weak | 23 | .31 | .16 | .13 | .14 |
| 3 | | | | | |
| Baseline | 40 | .45 | .23 | .37 | .26 |
| Weak | 40 | .36 | .23 | .26 | .22 |
| 3 (Rep–mixed) | | | | | |
| Baseline | 20 | .41 | .23 | .38 | .25 |
| Weak | 20 | .34 | .25 | .24 | .17 |
| 3 (Rep–first) | | | | | |
| Baseline | 20 | .49 | .22 | .36 | .27 |
| Weak | 20 | .38 | .21 | .29 | .26 |

*Note.* Rep = repetition.

The eight experimental blocks used in Experiment 1 were reduced to four blocks by combining each pair of original blocks into a single, longer block. Thus, each of the four blocks included two baseline, two strong–weak, and two filler sets for a total of 36 unique study pairs. The relative positions of the various types of pairs within the study and test lists were unchanged. The initial practice block was the same length as the practice block in Experiment 1. For Experiment 2b, a pool of 140 surnames were drawn at random from a U.S. phonebook.

*Procedure.* In Experiment 2a, the procedure was identical to that in Experiment 1a. Experiment 2b added a short-term memory task between the study and test phases of each block. At the conclusion of the study list, a list of 15 surnames was shown at a rate of one surname every 1,500 ms. Following this, subjects had 30 s to recall the surnames, writing their responses on a provided answer sheet. A timer appeared onscreen to show the time remaining. At the end of 30 s, a beep and onscreen prompt instructed subjects to stop immediately and proceed to the cued-recall test phase of the block.

### *Results and Discussion*

The results from both experiments are shown in Figure 2 (see also Table 1). Each experiment was analyzed separately with a 2 (condition) × 2 (test position) repeated measures ANOVA. For Experiment 2a, there was a main effect of condition, $F(1, 24) = 16.52$, $MSE = 0.02$, $p < .001$, and a main effect of position, $F(1, 24) = 23.24$, $MSE = 0.03$, $p < .001$. Condition and position did not interact, $F(1, 24) = 0.11$, $MSE = 0.02$, $p = .742$, *ns*. Fewer weak than baseline items were recalled at both the beginning and end of the test. For Experiment 2b, there was a main effect of condition, $F(1, 22) = 15.09$, $MSE = 0.03$, $p < .001$, and a main effect of position, $F(1, 22) = 17.34$, $MSE = 0.03$, $p < .001$. Condition and position did not interact, $F(1, 22) = 0.43$, $MSE = 0.03$, $p = .518$, *ns*. Again, fewer weak than baseline items were recalled at both the beginning and end of the test.

Experiments 2a and 2b were designed to remove the confounding influence of study items still present in short-term memory. The result was that an LSE was observed even at the beginning of the test, and the size of the effect did not change as a function of test position. This is exactly the pattern predicted of a relative-strength competition mechanism. A retrieval inhibition mechanism, on the other hand, has little to say about differences between conditions early in the test phase, when there have not yet been any retrieval attempts within a set. Retrieval inhibition does predict the decline in recall from the first to the last test positions. The rate of decline, nearly identical in both experiments, was most similar to the rate of decline seen in the weak condition of Experiments 1a and 1b. This is consistent with the idea that suppression of baseline item recall was responsible for the null LSE observed at the beginning of the test in those experiments.

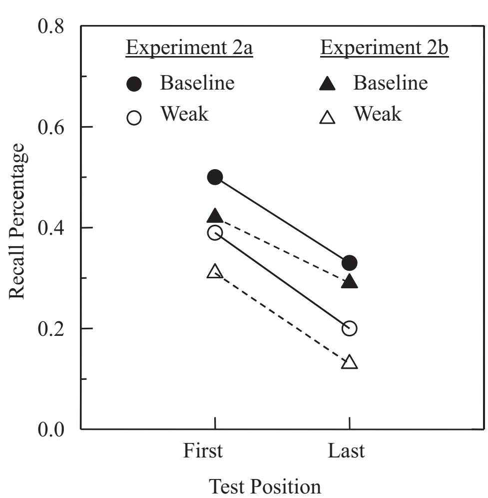


*Figure 2.* Experiment 2.

Experiments 2a and 2b differed primarily in that the latter produced generally poorer recall. Apparently, an interpolated memory span task using unrelated materials and procedures added to forgetting. At first glance, this seems to be evidence of what was earlier described as nonspecific interference. Experiment 4 examines the possibility that nonspecific interference is solely responsible for the decline in recall within the main test. As for the strength-based competition account of output interference, presenting items at test may also strengthen them, creating more interference for items in the set that have not yet been tested. However, note that the impairment of weak items relative to baseline observed at the beginning of the test is due to interference from two extra study repetitions of three set members. Items at the end of the test suffer from interference because of one extra repetition of two previously tested set members. If one assumes that study and test repetitions produce equivalent amounts of strengthening, then the latter should produce only a fraction of the forgetting, yet the difference in recall between the first and last test positions is nearly twice as large as the difference between weak and baseline conditions. In order for relative-strength competition to explain the decline in recall over the course of the memory test, additional assumptions are necessary. This is explored in the section on theory to follow.

## Experiment 3

It was suggested earlier that an LSE observed at the beginning of a test cannot be attributed to retrieval inhibition because of the absence of prior retrieval attempts. However, it has sometimes been noted that subjects may engage in covert retrieval practice during the study phase. In M. C. Anderson and Bell's (2001) study, where subjects were aware of an upcoming memory test, some subjects reported doing this spontaneously. Bäuml and Aslan (2004), on the other hand, found that subjects engaged in covert retrieval practice only when specifically encouraged to do so. Thus, it is not at all clear that this is a common occurrence. In fact, the many studies in the retrieval inhibition literature arguing that additional study does not increase forgetting make the assumption that covert retrieval does not normally take place. In light of this, the suggestion that the positive LSE findings in the current study are simply the product of covert retrieval practice does not seem persuasive. Nevertheless, the purpose of Experiment 3 was to deal with the possibility of strategic processing that might occur during the study phase. Presumably, subjects engage in covert retrieval practice as a way to rehearse for the upcoming memory test, and they may do this more often for the strong–weak set than the baseline set. The previous experiments began with a practice block followed by eight (Experiment 1) or four (Experiment 2) experimental blocks. Experiment 3 was reduced to two experimental blocks with no practice. At the outset, subjects were led to believe that the experiment was interested in only mental imagery so that the first block involved a surprise memory test. However, subjects would expect the memory test in the second block and could prepare for it strategically. To minimize any possibility that they might guess the purpose of the experiment despite the initial misdirection, I recruited subjects from the community at large rather than the School of Psychology. The critical questions were whether an LSE would still emerge in the first block and whether recall patterns would differ between the blocks.

As a final means of dealing with possible covert retrieval, the format of the study list was varied between subjects. In the repetition–mixed condition, the study list format was identical to that in the previous experiments: Strong items were presented at the beginning, middle, and end of the study list. In the repetition–first condition, strong item repetitions that had previously been at the end of the study list were moved to the beginning. Thus, following two repetitions of the strong items, all study items were presented once. Because the word pairs were random and novel, retrieval practice during the initial strong-item repetitions could not have inhibited weak items, because episodic links between the weak items and the context word did not yet exist. During the remainder of the study list, the number of presentations from each pair set, and thus the number of covert retrieval opportunities, was the same. Any retrieval inhibition owing to covert retrieval should be equivalent for weak and baseline conditions.

### Method

*Subjects.* Forty members of the Plymouth community were recruited from outside the School of Psychology and were paid for their participation. Subjects were equally divided between the repetition–mixed and the repetition–first study format conditions.

*Materials and design.* Save for having fewer blocks, the repetition–mixed condition was identical to that in Experiment 2b. The repetition–first condition differed only in that the strong-item repetitions previously at the end of the study list were moved to the beginning of the study list, so that all extra repetitions of the strong items occurred before presentation of the critical weak and baseline items.

*Procedure.* This was similar to Experiment 2b but with a shortened session that included only two study–test blocks. Subjects were told at the outset that the purpose of the study was to investigate mental imagery: They would view pairs of words and rate them for ease of imagery. There was no warning of a coming memory test, and there was no initial practice block. Subjects were informed at the beginning of the second block that it would proceed in the same manner as the first.

### Results and Discussion

An ANOVA with study format as a between-subjects factor and study–test block, test position, and condition as within-subject factors yielded two significant effects. There was a main effect of condition, $F(1, 38) = 10.54$, $MSE = 0.08$, $p = .002$, and a main effect of position, $F(1, 38) = 6.13$, $MSE = 0.10$, $p = .018$. As shown in Table 1, fewer weak than baseline items were recalled at both the beginning and end of the test. There were no significant effects of block or study format and no significant interactions (see Appendix B for the full results). Figure 3 shows the results by test block, collapsed over test position. A strong LSE was observed even in the first block, when subjects were not aware of an upcoming memory test and thus would not be expected to prepare for one. The size of the LSE did not change significantly over the two blocks, suggesting that test awareness did not encourage strategies that had a bearing on the LSE.

The observation of an LSE even in the repetition–first format can be taken as further evidence that covert retrieval could not have produced the selective impairment of weak items. However, an alternative interpretation is possible.[2] The repetition–first condition reflects a classic proactive inhibition design in which initial study of the strong items results in poorer recall of the subsequently studied weak items. The perspective taken so far has been that of retrieval interference theories, which hold that forgetting generally, and proactive inhibition specifically, is the product of difficulties at retrieval. An alternative perspective is one of encoding interference, the view that the initial items cause subsequent items to be more poorly encoded, perhaps because being reminded of the former while encoding the latter reduces their encoding time. One might argue that reduced encoding and covert retrieval both contributed to the LSE in the repetition–mixed condition and that the removal of covert retrieval in the repetition–first condition was made up for by an increase in the ability of the strengthened initial items to rob encoding time from later items. The plausibility of this interpretation rests on the plausibility of encoding reduction as a theory of proactive interference. Arguments against it include the finding that proactive interference often grows larger over time (Houston, 1967; Keppel & Underwood, 1968), inconsistent with an encoding deficit but consistent with reduced discriminability of competing items at retrieval (for more discussion, see Mensink & Raaijmakers, 1988; Postman & Underwood, 1973). However, the issue remains open to debate, and an appropriately conservative conclusion is that the test format manipulation rules out covert retrieval from the perspective of retrieval interference theories only.

The results of Experiment 3 replicated those of the previous experiments while adding two manipulations intended to diminish the possible contribution of covert retrieval to the LSE observed at the beginning of the test. The use of a surprise memory test assumes that covert retrieval is a strategic rather than an automatic occurrence, and the test format manipulation reflects a retrieval-based rather than an encoding-based view of proactive interference deficits. Thus, the interpretation of these results depends on the plausibility of these assumptions. That said, these pieces of evidence should be viewed alongside converging evidence that covert retrieval may not be a common strategy (Bäuml & Aslan, 2004) as well as the logical point that the many studies showing no effect of study presentations on forgetting rely on the assumption that such presentations do not normally invoke covert retrieval.

## Experiment 4

The goal of the present study is to show that relative-strength competition plays a role in forgetting independent of retrieval

[2] Thanks to M. C. Anderson for raising this possibility.

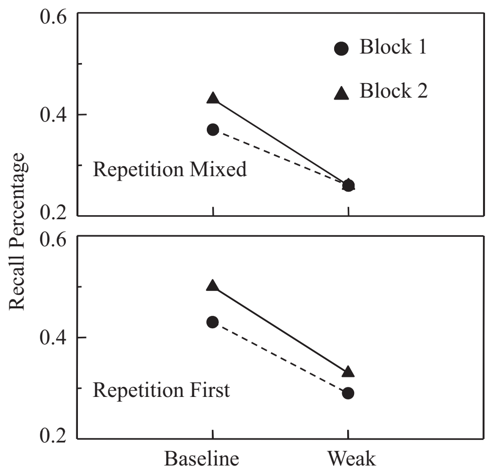


*Figure 3.* Experiment 3. Study list format was manipulated between subjects. In the repetition–mixed condition, strong-item study repetitions occurred at the beginning and the end of the list. In the repetition–first condition, strong-item study repetitions occurred only at the beginning of the list.

inhibition. This has been accomplished by juxtaposing the LSE, which is uniquely predicted by strength-based competition, with output interference, which has been interpreted in the retrieval-induced forgetting literature as being the product of inhibition. However, the finding in Experiment 2b that placing an unrelated filler task between the study and test phases increases forgetting raises the possibility that decline in recall over time may be due simply to nonspecific factors like decay, general cognitive interference, temporal fluctuations in the match between encoding and test context, and so on. The purpose of Experiment 4 is to confirm that specific interference consistent with retrieval inhibition contributes to forgetting independent of nonspecific factors.

## Method

*Subjects.* Thirty-four undergraduates from the University of Massachusetts Amherst participated for extra credit in their psychology course.

*Materials and design.* Stimuli were drawn from the pool of words used in Experiment 1. Like Experiment 1, each of the blocks consisted of a study list and a test list. For each block, four sets of word pairs were created; these were designated the filler A, filler B, no-delay, and delay sets. As before, each set consisted of six unique pairs created by pairing one word with six other words, none of which appeared in any other set. In order to encourage the learning of pairs rather than just the right-side words, I had each right-side word in the filler B, no-delay, and delay sets begin with the same first letter as a right-side word from each of the other two sets (in other words, the same six initial letters were used in each set). Each study list contained 24 pairs. The first and last three positions in the study list were pairs from the filler A set. The 18 pairs from the remaining three sets were randomly ordered in the center of the list.

Each test list contained 20 pair fragments (the left-side word and the first letter of the right-side word from a studied pair; e.g., *umbrella-a—?*). The first two pairs tested were from the filler A set. Three pairs each from the baseline and filler B sets alternated in list positions 3–8 (e.g., *A-B, H-I, A-C, H-J, A-D, H-K*). Three pairs each from the no-delay and delay sets alternated in list positions 9–14. Three pairs each from the delay and filler B sets alternated in list positions 15–20. In summary, each portion of the test list contained alternating pairs from two different sets. The set that appeared first within each portion was alternated between blocks.

*Procedure.* The procedure was identical to that in Experiment 1a.

## Results and Discussion

As shown in Figure 4, the memory test list was divided into three equal parts: beginning, middle, and end. For the no-delay set, three items were presented at the beginning and three in the middle of the test. For the delay set, three items were presented in the middle and three at the end of the test. The influence of test position on the decline of recall within a pair set was examined with a 2 (first half of the set vs. second half) × 2 (delay condition) repeated measures ANOVA. There was a main effect of set half, $F(1, 33) = 5.78$, $MSE = 0.01$, $p = .022$, owing to a decline in recall as more items from a set were tested. For the no-delay set, recall of the first three items ($M = 0.33$, $SD = 0.14$) was greater than that for the last three items ($M = 0.24$, $SD = 0.12$). Results were similar for the delay set, with recall of the first three items ($M = 0.28$, $SD = 0.14$) being greater than that for the last three items ($M = 0.22$, $SD = 0.11$). There was also a main effect of delay, $F(1, 33) = 44.61$, $MSE = 0.004$, $p < .001$, owing to overall lower rates of recall for the delay condition. There was no interaction of set half and delay, $F(1, 33) = 0.60$, $MSE = 0.01$, $p = .446$, *ns*.

If forgetting over the course of the test were caused only by nonspecific interference unrelated to set membership, then recall for all pair sets should have been the same at the middle position. This was not the case: Recall was more difficult in the no-delay condition, in which items from the same set had been probed earlier in the test. Thus, the increasing difficulty of retrieving items

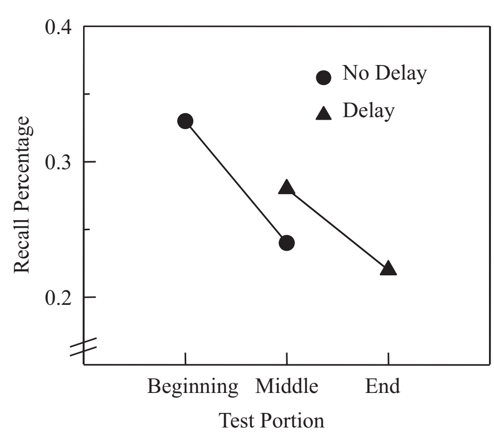


*Figure 4.* Experiment 4. The test list was divided into three equal portions. For the no-delay condition, half of the set was tested in the beginning and half in the middle portion of the list. For the delay condition, half of the set was tested in the middle and half in the end portion of the list.

from the same set, attributed in earlier experiments to retrieval inhibition, could not have been solely because of nonspecific factors. In what way did nonspecific factors contribute to forgetting? Unfortunately, the results allow little insight into this question. Clearly, lower overall recall for the delay set shows that not all forgetting is due to the retrieval of related items. However, this could be attributed to either competition or inhibition. The focus has been on competition among items that share the same left-side word. However, the fact that all items were encoded in the same general context produces some nominal amount of relatedness and thus interference across sets. Testing items strengthens their association with the general context, leading to increased competition for items tested later (Raaijmakers & Shiffrin, 1980). Similarly, one might argue that relatedness among items across the boundaries of the predefined sets—owing to shared context, overlapping orthography, and so on—produces a buildup of nominal amounts of inhibition as the test progresses (M. C. Anderson, 2003).

This series of experiments provides an empirical basis for the claim that relative-strength competition exists independent of retrieval inhibition. Is it possible to reconcile this with the more common finding of a null LSE in the retrieval-induced forgetting literature? It is true that in a relative-strength competition model, strengthening one item must reduce selection of its competitors. However, complexity lies in how one defines *strength*. Basic competition models have typically defined *strength* as being a direct function of an item's recallability. In such a model, the null LSE is indeed a problem. However, it is not necessarily a problem if *strength* is defined in other ways. In the next section, Malmberg and Shiffrin's (2005) SAM-REM model of recall serves to illustrate how a different way of thinking about strength can produce a relative-strength competition mechanism that is able to accommodate both the presence and absence of an LSE in cued recall.

## SAM-REM and the LSE

Malmberg and Shiffrin (2005) examined the effect of different methods of strengthening on the LSE in free recall. While comparing mixed and pure strength lists, they found that strengthening items with spaced repetition produced an LSE, but strengthening items by presenting them only once and increasing study duration or depth of processing did not. Malmberg and Shiffrin did not address the question of retrieval inhibition, but their account of these findings is relevant in that it demonstrates how a relative-strength competition model is able to cope with both the presence and absence of an LSE in free recall. REM is a Bayesian decision model in which the qualities of a retrieval probe are compared globally with the contents of memory to calculate the likelihood of a match (Shiffrin & Steyvers, 1997). The model has been implemented in a variety of forms, and Malmberg and Shiffrin's variation integrates the matching-likelihood mechanism of REM with SAM, a model widely applied to the retrieval process in recall (Raaijmakers & Shiffrin, 1981). In order to examine in greater depth the relationship between strength and competitor interference, I present an adaptation of SAM-REM for cued recall here. Core concepts of the model are discussed below, and a more detailed treatment is presented in Appendix A.

Malmberg and Shiffrin's (2005) analysis of the LSE requires a deconstruction of the term *strength*. In SAM-REM, stimuli and their resulting memory images are represented as vectors of features. Encoding into memory involves copying feature values of the stimulus vector onto the episodic memory vector. This process is gradual and error-prone, but strengthening manipulations that increase the time or quality of attention available for encoding produces images that are more like the original stimulus and thus are more readily retrievable when parts of the stimulus are presented again as a cue. Strengthening is usually thought of simply in terms of the memory image as a whole. However, an image contains information about both the stimulus item and its study context. Appreciating that these two types of information play different roles during the retrieval process is the key to explaining dissociations between item strength and competitor interference.

Recall in SAM-REM involves the stages of sampling and recovery. At the onset of the process, the retrieval cue is compared with each of the images resident in memory (as a standard simplifying assumption, this is limited to $k$, the set of studied items). The cue vector and a given memory vector are compared feature by feature, noting the matches and mismatches. The final result yields a likelihood ratio, $\lambda_j$, that represents how well the retrieval cue matches the given memory image $I_j$ relative to other images. Memory images are then sampled serially with replacement. The probability of sampling memory image $I_j$ given retrieval cue $Q$ on a sampling trial is expressed in the following:

$$P(I_j|Q) = \frac{\lambda_j}{\sum \lambda_k}. \tag{1}$$

Simply put, this is the ratio of the match of image $I_j$ to the summed matches of all $k$ images in the memory set. Relative-strength competition is captured in this stage of the model. Sampling probability depends critically on the degree to which competitors also match the probe; a strengthening manipulation that increases competitor match will make it more difficult to sample the target image. After an image is sampled, its contents must then be recovered to a degree sufficient to support a response. The probability of successful recovery is a function of the proportion of correctly stored feature values in image $I_j$. Cycles of sampling and attempted recovery continue until an appropriate response is produced or the process ends in failure after some criterion number of attempts.

The LSE is a product of the sampling mechanism. This is most easily illustrated in the free-recall task, where one must produce the study items given nothing more than the experimental context as a retrieval cue. The memory image of each study item is made up of both the item's features and the features of the context. The configuration of item features is unique to each image. However, images encoded within the same experimental context possess similar contextual features, and this is the reason for competition and interference during sampling. If all of the studied items are encoded in the same way, then all of the images contain, on average, the same amount of context information. Thus, all images match the retrieval cue to more or less the same extent, and all interfere with the sampling of other images to the same extent. Suppose, however, that a manipulation strengthens one of the images by increasing its storage of context features. This improves the match between the cue and the strengthened image. In Equation 1, the result is an increase in the numerator of the sampling ratio for the strengthened image (making it more likely to be sampled) and an increase in the denominator of the sampling ratios of all other images (making them less likely to be sampled). Malmberg and Shiffrin (2005) suggested that spaced repetition produces an LSE in free recall because each repetition is another opportunity for the storage of contextual information. Incidentally, it is also another opportunity for the storage of item information, which further improves the recovery of repeated items. Why is no LSE observed when depth of processing or study duration is varied? Malmberg and Shiffrin hypothesized that only a fixed amount of context is stored with each repetition. Thus, deep processing or long study improves recall by increasing the storage of item features, which helps during the recovery stage. However, both strong and weak items are left with an equivalent amount of stored context, so both compete equally during the sampling stage.

Manipulations that differentially affect the storage of item and context features are one way to produce dissociations between the effect of strengthening on item retrieval and its effect on competitor interference. Several other factors can also lead to dissociations, and these are better illustrated in the cued-recall task. Whereas a single, global context controls retrieval in free recall, in cued recall the context changes with each memory probe. Suppose that a study list consists of word pairs such as *apple-dog*, *apple-lake*, and *frog-stone* and that probes in the cued-recall test consist of the first word of the pair and the first letter of the second word (*apple-d—?*). In this case, the features of the first word are the "context," and features of the second word are the "item" that needs to be retrieved given this context.[3] Strengthening a pair by increasing storage of its first-word features (i.e., the context) will affect recall in several ways. Increasing the context stored in the memory image *apple-dog* will make it more likely to be sampled given the *apple-d—?* probe. The context features of that image now also yield a better match to the probe *apple-l—?*, thus interfering with sampling of the proper target, *apple-lake*. However, increasing the storage of context features makes an image more distinct from images that differ in context because SAM-REM takes into account mismatching as well as matching features when calculating probe-image match. The image *apple-dog* becomes a poorer match to the probe *frog-s—?* and thus competes less with the proper target, *frog-stone*. In sum, encoding contextual information has several interacting effects on retrieval: In addition to making the item itself more retrievable, it also increases interference with items that share the same context and decreases interference with items that differ in context.[4]

This example suggests that the emergence of an LSE depends on the composition of the list: Strong items will interfere with weak items to the extent that they share context features. Context provided by the retrieval cue determines competitor interference, and its influence is located in the sampling stage. Therefore, the LSE in recall can be understood by examining what occurs in the sampling stage. Simulations of the sampling process were conducted with an adaptation of the free-recall model described by Malmberg and Shiffrin (2005). Rather than global context, the features of the first word (as well as several of the second word) were used as the memory probe, simulating word-and-letter-stem cued recall. Both list composition and item strength were varied systematically. A set of six pairs was created, half assigned to the *strong* condition (bold) and half to the *weak* condition. In low overlap lists, all pairs had unique first-word contexts (***A-B,*** *C-D,* ***E-F,*** *G-H,* ***I-J,*** *K-L*). In medium overlap lists, each pair shared a first-word context with one other pair (***A-B,*** *A-D,* ***E-F,*** *E-H,* ***I-J,*** *I-L*). In high overlap lists, all pairs shared the same first-word context (***A-B,*** *A-D,* ***A-F,*** *A-H,* ***A-J,*** *A-L*). Weak pairs were studied once. Strong pairs were studied once, twice, or four times (the one-presentation case, where both conditions were actually identical in strength, serves as a baseline). Additional repetitions had the effect of increasing the storage of both context (first word) and item (second word) features. Details of the model can be found in Appendix A. No effort was made to find model parameters specifically for these simulations; rather, the parameters were taken directly from Malmberg and Shiffrin wherever possible.

Figure 5 shows the probability of sampling the target image (i.e., the image that correctly matches the test probe). By definition, an LSE occurs when sampling probability increases for strong images but decreases for weak images relative to baseline. The simulations indicate that the LSE is mediated by both contextual overlap and degree of strengthening. A positive LSE is absent in the low overlap condition. In fact, the sampling probability for weak items increases with that of strong items because strengthening serves to differentiate images that differ in context, making them less likely to interfere with one another. The medium overlap condition reveals a dissociation between strength and interference: Despite large improvements in sampling probability for strong images, impairment of weak images is negligible at low levels of strengthening and only readily apparent at higher levels. The impairment of weak items with repetition is slightly stronger in the high overlap condition. To summarize, the LSE arises from a disparity in stored contextual information between memory images that share the same context. During the sampling stage, images that more completely and accurately match the contextual retrieval cue are sampled to the exclusion of other images. However, retrieval impairment of weak items by strong items is significant only when there is sufficient contextual overlap among images and the disparity in degree of strengthening between strong and weak images is sufficiently large.

An LSE emerges in the sampling stage when strong images have a more complete representation of the context than do weak images. What happens when the storage of contextual features is improved for both strong and weak images? A second simulation

[3] It would be more accurate to say that context has two components: the specific context provided by the first (cue) word and the global experimental context. However, for the conditions being discussed, global context is more or less equivalent for all images and does not contribute to the critical effects. The global component of context is ignored hereafter in order to streamline the discussion and simulations; doing so does not change in any fundamental way the behavior of the model.

[4] This description of interference is simplified; the item-stem cue is yet another factor to consider. With pairs that share a context such as *apple-dog* and *apple-lake*, increased encoding both increases interference due to the context (*apple*) match and decreases interference due to the item (*dog*, *lake*) mismatch. When the retrieval cue contains sparse item relative to context information (*apple-d—?*), the detrimental effect of increased encoding will be dominant. This might explain why the LSE is robust in tasks where the cue has sparse item information such as free recall yet is often nonexistent in tasks where the cue has a great deal of information such as recognition (Ratcliff et al., 1990). The influence of having a small amount of item information is included in the SAM-REM simulations, but it is not necessary to understanding the critical patterns and has been left out of the exposition for simplicity's sake.

was run with the high overlap, four-repetition list from the first simulation and with the model parameter representing the probability of feature storage varied ($u* = .02, .15, .25$). Increasing the value of $u*$ leads to more complete contextual representations in both strong and weak items. As Figure 6 shows, this increases the probability of sampling the target image but also diminishes the difference in proportion of stored context features between strong and weak images, reducing the LSE until it all but disappears. An equivalent effect would be produced by pretraining items (studying them to a high degree) prior to a strength manipulation. Increasing the starting level of stored context diminishes the effect of additional strengthening, making it harder for an LSE to emerge. In other words, this second simulation shows that a null LSE can occur when the storage of context features is rapid (as might be expected when there are preexisting associations between the context and the item) or has reached a high degree of completion due to pretraining the context–item association. When the context–item association is well learned, it becomes difficult to create a significant disparity between nominally weak and strong items.

The basic competition model often criticized in the retrieval inhibition literature has been influential in theories of recall, and revealing its inadequacy is an important step forward. However, SAM-REM is an example of a complex model that incorporates relative-strength competition yet is able to accommodate dissociations between item strength and competitor interference. It illustrates the point that the null LSE does not allow us to dismiss altogether the principle of relative-strength competition. Of course, it is fair to ask a further question: Does SAM-REM offer any insight into empirical findings? The model does predict the LSE observed here. According to the simulations, an LSE is most likely to be evident when the context–item associations are novel and not well learned, when there is a high degree of contextual overlap between strong and weak items, and when the strengthening manipulation is of sufficient magnitude. These were precisely the conditions used in the present experiments. Conversely, a case can be made that various null findings in the literature occurred under conditions that SAM-REM suggests are not optimal for the emergence of an LSE.

Two null reports by Bäuml (1996; 1997) used study duration as a means of strengthening. According to Malmberg and Shiffrin's (2005) account, this manipulation should not produce an LSE because increased exposure does not encourage additional storage of context.[5] Notably, Delprato (2005) carried out a follow-up to the Bäuml (1996) retroactive interference study, manipulating strength via spaced repetition rather than duration. He did observe an effect of strength-related interference. Ratcliff et al.'s (1990) inconsistent findings of an LSE in cued recall have been cited as evidence against relative-strength competition. The nature of the inconsistency is revealing: Manipulating study duration produced no LSE, but manipulating spaced repetition did. It has also been noted that Ratcliff et al.'s LSE was surprisingly weak, but consider

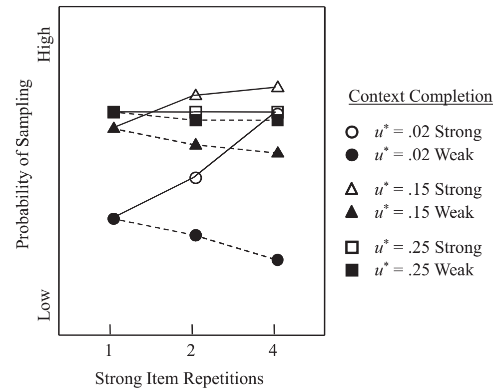


*Figure 6.* Simulation 2: Rate of context encoding. Simulation of the SAM-REM sampling stage following a study list of 18 word pairs in the high overlap, four times, strong-repetition condition (see Simulation 1). The $u*$ parameter, which controls the probability that some contextual information is stored at each attempt, was varied between simulated groups. Increasing $u*$ is analogous to increasing the ease of encoding contextual information and is also analogous to increasing the amount of training or exposure to context–item pairs. (Although the $u* = .02$ condition is identical to the high-overlap condition in Figure 5, the vertical axis has been rescaled here.)

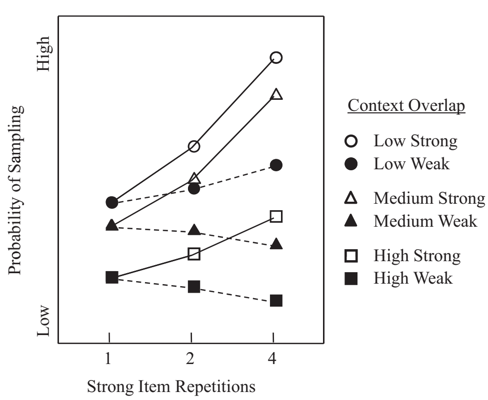


*Figure 5.* Simulation 1: Context overlap and repetitions. Simulation of the SAM-REM sampling stage following a study list of 18 word pairs. Strong condition items were presented one, two, or four times in the study list. Weak condition items were presented once. Context overlap (number of pairs in the list sharing the same cue word) and number of strong-item repetitions were systematically varied between simulated groups. *Probability of sampling* represents the probability that the item correctly matching the cue is successfully sampled on a given attempt. The actual value produced by the model depends on a scaling parameter and is not as important as the relative sampling probabilities among conditions and the patterns produced by varying overlap and repetition.

[5] Malmberg and Shiffrin (2005) proposed that when the presentation duration is increased, feature values corresponding to the target image continue to be stored over the additional time. However, feature values corresponding to the background context are stored only at the start of presentation. The authors attributed this to the fact that the context does not command additional attention or effort beyond an initial encoding period. Bäuml (1997) used category–exemplar pairs drawn from common categories (e.g., *vegetable-tomato*). Although the category label was nominally part of the study probe rather than the background context, it seems unlikely that subjects would have attended to the labels in the same way as to the exemplars. The memory task required learning the exemplars, not the already well-learned category labels. Thus, the category labels may have been processed in much the same way as background context, failing to accumulate additional storage with increasing study duration.

that they used lists of unique study pairs that did not share words across pairs. Thus, the contextual overlap of memory images can be assumed to have been moderate or low, which should result in a weak or null LSE.

The moderate overlap condition in Figure 5 shows that even a large increase in the sampling probability of strong items can be accompanied by little or no weak-item impairment. Thus, it can be difficult to rule out the possibility that a null LSE is due to an insufficiently strong manipulation. Delprato (2005; Experiment 1) demonstrated that this problem is real and not merely theoretical. Subjects studied an initial list of paired associates (*A-B*) followed by one, four, or eight repetitions of a second list of overlapping pairs (*A-C*). Although second-list recall increased from 47% to 67% when repetitions were increased from one to four, first-list recall did not statistically differ between the two conditions. Only the eight-repetition condition managed to produce a significant impairment in first-list recall relative to the one-repetition condition.

Two factors may have contributed to the null result reported by Ciranni and Shimamura (1999). Study items varied along three dimensions: color, shape, and location. Every item had a unique shape and location, but several shared the same color. Subjects were cued with a color and location and asked to recall the shape. Although the shared color cue introduced an element of contextual overlap among images, the unique location cue produced a strong match to the target image alone (including target-specific features in the probe served to differentiate it from nontarget images). The opposing effects of contextual overlap and contextual distinctiveness on sampling probability would have reduced or nullified the LSE. In addition, study items were learned to criterion before the strengthening manipulation, which should have reduced their ability to produce an LSE. To continue this last point, some studies reporting a null LSE (M. C. Anderson, Bjork, & Bjork, 2000; Bäuml, 1997; Bäuml & Aslan, 2004) have used category–exemplar pairs such as *fruit-apple*, *fruit-banana*, etc. The existence of well-learned, preexisting associations between the context and target may have either facilitated storage of context or served as a form of pretraining, conditions predicted to reduce the LSE (see Figure 6).

The SAM-REM model offers a way to understand the less-than-straightforward relationship between item strength and competitor interference. Another aspect of the present data is the pattern of output interference that has been attributed to retrieval inhibition. SAM-REM has no mechanism for inhibition, and to propose one would be a significant conceptual departure for the model. Are there other ways that the model might naturally account for the decline in recall over the course of a memory test? First, consider relative-strength competition. Presenting items as test probes constitutes additional study opportunities that should strengthen subjects' memory images, leading to increased sampling competition for items that have not yet been tested. As noted in the discussion of Experiment 2, the decline in recall over the course of the test appears to be too large if one assumes that study and test presentations result in equivalent amounts of stored contextual information. However, if more context is stored during a retrieval trial than during a study trial, then it is possible for test presentations to produce more interference than would the equivalent number of study presentations, as was the case in Experiment 2. There is no a priori reason to reject this possibility, but it does seem to contradict Malmberg and Shiffrin's (2005) contention that normally a set amount of context is stored per presentation. An obvious way in which encoding at study and retrieval might be thought to differ is in depth of processing. Malmberg and Shiffrin observed no LSE when "strength" was manipulated via depth of processing, which they interpreted in SAM-REM as showing that trials differing in processing depth were associated with the same amount of stored context. Of course, their study examined general experimental context, and perhaps other assumptions should apply to the more local forms of context examined in the present study.

Another mechanism of forgetting, explored in similar models (e.g., Mensink & Raaijmakers, 1988) and readily implemented in SAM-REM, is context fluctuation. Context might be assumed to fluctuate over time, so that the context reinstated at retrieval is similar but not identical to the original encoding context. As the time between encoding and test increases, so does the disparity between the original and reinstated context, leading to a decline in recall that increases as a function of time. Thus, recall at the end of the test should be lower than at the beginning. The results of Experiment 4 show that context fluctuation cannot be the sole source of declining recall. If it were, there should be no difference between the no-delay and delay conditions in the middle portion of the test. Context fluctuation also has limited explanatory power in the standard retrieval-induced forgetting paradigm, where cued recall is often used to hold test position constant when comparing inhibition and noninhibition conditions.

Can relative-strength competition combined with context fluctuation produce the output interference observed in the present study? This is a viable possibility and one that cannot be dismissed with these data alone. It is certainly a possibility worth exploring, although this would require a formal elaboration of Malmberg and Shiffrin's (2005) version of SAM-REM that is beyond the scope of this article. However, dismissing the contribution of inhibition to output interference brings into question the role of inhibition in forgetting more generally, and this does call for addressing other phenomena that have been linked to inhibition. Two in particular are at odds with a relative-strength competition mechanism. First, retrieval competition by definition depends on the identity of the retrieval cue. Retrieval-induced forgetting has often been observed to be cue-independent, a property in keeping with the concept of inhibition (M. C. Anderson & Spellman, 1995). Second, retrieval competition predicts interference for all competitors associated with a cue. The failure to observe retrieval-induced forgetting with weak competitors has been used to support the functional view that inhibition should occur for competitors strong enough to interfere with retrieval (M. C. Anderson et al., 1994; Bäuml, 1998). Finally and more broadly, the argument for inhibition in retrieval-induced forgetting resides within a larger argument that advocates have made for the role of inhibition across cognitive and neurophysiological domains (M. C. Anderson, 2003; Levy & Anderson, 2002).

## Conclusion

The idea that retrieval success depends on the relative strength of competing memories has for many years played a central role in theories of recall. However, work in the area of retrieval-induced forgetting has systematically undermined this notion both empirically and theoretically (M. C. Anderson, 2003; M. C. Anderson et al., 1994). First, many studies in the retrieval-induced forgetting literature have failed to observe an effect of strengthening on recall of competitors, seemingly at odds with the principle of relative-strength competition. Second, retrieval inhibition offers alternative explanations for many phenomena previously attributed to relative-strength competition. The present study addressed both of these points. With regard to the first, the relative-strength competition mechanism most commonly portrayed in the literature defines *strength* purely in terms of recallability. The null LSE makes a compelling case against such a mechanism. However, simulations of the SAM-REM model of recall (Malmberg & Shiffrin, 2005) show that a more complex view of strength can accommodate both the presence and absence of an LSE. Regardless of what one thinks about its explanatory usefulness, the model serves to illustrate the fact that the failure to observe retrieval interference as a result of item strengthening does not rule out the role of relative-strength competition in forgetting. The present experiments also provide evidence of an LSE in recall that is not easily attributed to retrieval inhibition. In Experiments 1–3, once-presented (weak) items from mixed-strength study sets (weak condition) were compared with once-presented items from pure weak study sets (baseline condition). Cued recall made it possible to control output position, so that the accumulation of retrieval inhibition during the test phase was held constant for the two conditions. Even so, recall was worse in the weak condition, consistent with the notion that stronger competitors produce greater retrieval interference. In Experiments 2 and 3, this LSE was observed even at the beginning of the test list, before any other retrieval attempts from a set had been made.

It has been suggested that people sometimes engage in covert retrieval even when not explicitly asked to do so. Could differential covert retrieval during the study phase be behind the different rates of forgetting in the two conditions? To address this, in Experiment 3 I controlled for the likelihood and opportunity for covert retrieval. First, the experiment consisted of two study–test blocks. During the first block, subjects were exposed to study pairs with no expectation of an upcoming memory test, so that they would have no reason to engage in preparatory strategies such as covert retrieval. The LSE observed in the first block was not significantly different from that of the second block, when a memory test was expected. Second, the structure of the study list in the previous experiments always placed several strong items at the end of the list, and this could provide additional covert retrieval opportunities for that set. In the repetition–first condition, the additional presentations of the strong items occurred at the beginning of the study list, before any other items were presented. Covert retrieval during these trials could not inhibit the critical competitors; these noun pairings were novel, and the competing words had yet to be associated with the contextual cue. In the remainder of the study phase, there were an equal number of trials involving the study sets associated with the two conditions and thus an equal number of covert retrieval opportunities within each set. Again, this change in procedure had no effect on the size of the LSE. Both the inhibition and retrieval-competition theories considered in this article are of a class that explains forgetting in terms of postencoding factors. In the earlier discussion of Experiment 3, it was noted that an encoding-centered view of forgetting may offer yet another way to interpret this last result.

Although the present study took pains to control for covert retrieval, the use of this concept deserves some scrutiny. The idea that covert retrieval may cause forgetting in situations where subjects are not explicitly asked to retrieve anything certainly seems possible. However, the problem of circularity arises if the covert retrieval argument becomes a way to suggest the hidden influence of retrieval inhibition whenever forgetting is observed. It is not clear that covert retrieval practice is common, and in fact existing evidence suggests the contrary. Bäuml and Aslan (2004) found that subjects did not engage in covert retrieval during additional study opportunities but did so only under specific instructions that related the study to the forthcoming test. The many studies reviewed earlier that failed to connect additional study to forgetting also imply the absence of spontaneous covert retrieval during additional study trials. These studies stand in contrast to the sole study in the retrieval-induced forgetting literature to find direct evidence of covert retrieval in the form of post hoc self-reports (M. C. Anderson & Bell, 2001).

The argument that relative-strength competition has a role to play in forgetting is not meant to detract from the significance of inhibition theory. Inhibition provides a plausible explanation for the output interference consistently observed in these experiments. Moreover, if a competition model predicts a null LSE with additional study trials, the presence of an inhibition mechanism could explain why retrieval trials produce forgetting under the same conditions. It has been argued that many phenomena previously attributed to relative-strength competition may be better explained in part by retrieval inhibition. M. C. Anderson et al. (1994), for example, noted that much of the classic work on interference theory utilized methods that confounded interference manipulations with retrieval practice. Although relative-strength competition has sometimes been used to explain output interference (the decline in recall over the course of a test; e.g., Roediger, 1978), this is also readily explained by retrieval inhibition (M. C. Anderson et al., 1994; Bäuml, 1997). As noted in the discussion of Experiment 2, retrieval inhibition does seem a better explanation than is relative-strength competition for the amount of output interference observed in the present experiments. Finally, Bäuml and colleagues (Bäuml & Aslan, 2004; Bäuml & Kuhbandner, 2003) have made a case for the role of retrieval inhibition in the part-list cueing effect.

However, it has become common in the retrieval inhibition literature to dismiss the detrimental effect of strengthening out of hand rather than test for it explicitly, and the present findings urge more caution in this regard. A converging note of caution comes from work in the area of recognition memory, where retrieval-induced forgetting has also been observed (Hicks & Starns, 2004; Verde, 2004). The frequent citation of such work in the retrieval inhibition literature comes with the implication that retrieval-induced forgetting in recognition and recall depends on a common process. If this is so, then one must consider the fact that the LSE has also been observed in recognition (Norman, 2002; Verde & Rotello, 2004). Like cued recall, recognition controls output order, making it difficult to attribute the LSE to retrieval-induced forgetting.

If relative-strength competition and retrieval inhibition both contribute to forgetting, then there remains the question of how to integrate them within a single theoretical framework. The fact that the two sources of forgetting did not interact (see Experiment 2) allows for a simple possibility. Consider a two-stage model like that described for SAM-REM, in which a memory image is first sampled from a pool of candidate targets and then the sampled image is recovered. The relative strength of memories has been shown to affect response latencies, and this has been attributed to interference that occurs in the sampling stage (Rohrer, 1996; Wixted et al., 1997). Bäuml, Zellner, and Vilimek (2005) failed to observe any effect of retrieval-induced forgetting on response latencies in free recall. They argued that this indicates that the locus of retrieval inhibition is in the recovery stage. Placing relative-strength competition and retrieval inhibition at different stages in the retrieval process makes the two mechanisms compatible in theory. Alternatively, one might look to model architectures that take altogether different approaches to these issues. Norman, Newman, and Detre's (2007) neural network model, for example, explains both retrieval-induced forgetting and the LSE in terms of the weakening of associative connections between representational units, in contrast to the temporary suppression of units described by M. C. Anderson, Green, and McCulloch (2000) or the blocking of units in competitive sampling models like SAM-REM. Interestingly, Norman et al.'s prediction of an LSE was based only on model simulations, which showed an LSE emerging when study items have a high degree of unit overlap in the network. Arguably, the robust LSE observed here under conditions of high interpair similarity provide an empirical confirmation of this prediction. This highlights our dependence on further empirical work to distinguish the various competing models.

## References


Allen, M. M. (1969). Cueing and retrieval in free recall. *Journal of Experimental Psychology, 81,* 29–35.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341–380.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49,* 415–445.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130,* 544–570.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7,* 522–530.

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1141–1159.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68–100.

Bäuml, K.-H. (1996). Revisiting an old issue: Retroactive interference as a function of the degree of original and interpolated learning. *Psychonomic Bulletin & Review, 3,* 380–384.

Bäuml, K.-H. (1997). The list-strength effect: Strength-dependent competition or suppression? *Psychonomic Bulletin & Review, 4,* 260–264.

Bäuml, K.-H. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review, 5,* 459–463.

Bäuml, K.-H., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32,* 610–617.

Bäuml, K.-H., & Kuhbandner, C. (2003). Retrieval-induced forgetting and part-list cueing in associatively structured lists. *Memory & Cognition, 31,* 1188–1197.

Bäuml, K.-H., Zellner, M., & Vilimek, R. (2005). When remembering causes forgetting: Retrieval-induced forgetting as recovery failure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1221–1234.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 95–105.

Basden, D. R., Basden, B. H., & Galloway, B. C. (1977). Inhibition with part-list cuing: Some tests of the item strength hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 100–108.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1403–1414.

Delprato, D. J. (2005). Retroactive interference as a function of degree of interpolated study without overt retrieval practice. *Psychonomic Bulletin & Review, 12,* 345–349.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11,* 125–130.

Houston, J. P. (1967). Proactive inhibition and competition of recall. *Journal of Experimental Psychology, 75,* 118–121.

Keppel, G., & Underwood, B. J. (1968). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161.

Kucera, F., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences, 6,* 299–305.

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322–336.

Mensink, G., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455.

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.

Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition, 12,* 531–552.

Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1083–1094.

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review, 114,* 887–953.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

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.

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology, 92,* 365–372.

Roediger, H. L., III. (1978). Recall as a self-limiting process. *Memory & Cognition, 6,* 54–63.

Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition, 24,* 188–201.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12,* 43–50.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford and N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). New York: Oxford University Press.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology, 92,* 297–304.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition, 32,* 1265–1272.

Verde, M. F., & Rotello, C. M. (2004). Strong memories obscure weak memories in associative recognition. *Psychonomic Bulletin & Review, 11,* 1062–1066.

Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 523–538.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1,* 89–106.

# Appendix A

## SAM-REM Model of Recall

The REM model (retrieving effectively from memory; Shiffrin & Steyvers, 1997) has been implemented in a number of variant forms, all based on a core comparison mechanism that quantifies the degree to which a retrieval cue matches the contents of memory in the form of a Bayesian likelihood ratio. Malmberg and Shiffrin (2005) proposed a variant that integrates this matching mechanism with SAM (search of associative memory; Raaijmakers & Shiffrin, 1980, 1981), a model widely applied to the retrieval process in recall. Their primary goal was to illustrate what they considered to be key principles of encoding and interference in recall. For this reason, they implemented the models in their most simple forms. The goal of the present study is essentially illustrative as well, and the simulations are based on the same model. Some assumptions about context were modified to apply the model to cued recall, but the parameters used in the simulations were identical to those used by Malmberg and Shiffrin wherever possible. I present here a basic description of the model and refer readers to the source articles for a discussion of the theoretical underpinnings of REM and SAM.

REM represents stimulus items and their corresponding memory images as vectors of features, with each feature assigned a value, $V$. It is assumed that in the real world some of these values are more common than others. To simulate this, I drew feature values from a geometric distribution:

$$P[V = v] = (1 - g)^{v-1}g, \quad \text{(A1)}$$

where $v = 1, \ldots, \infty$.

A memory vector consists of $w_i$ features corresponding to the stimulus item as well as $w_c$ features corresponding to the context in which the item is studied. At the initial stage when no information has been stored, these values are set to zero. During study, an attempt is made to copy the feature values of the stimulus item and context onto the memory vector. However, this process is gradual and error-prone and depends on three encoding parameters. For each unit of time, there are $t$ storage attempts during which information may be stored for each feature of the memory vector. On each storage attempt, there is probability $u*$ that something will be stored for a given feature. If something is stored, the correct value (i.e., the true value of the item or context feature) is stored with probability $c$. This means that with probability $1 - c$, an incorrect value, drawn randomly from the geometric distribution, is stored. Once a value has been stored for a given feature, no other attempts are made for that feature. In the present simulations, a study presentation was defined as a single unit of time. The encoding parameters were set to $t = 6$, $c = .80$, and $u* = .02$ ($u*$ was varied in Simulation 2). The memory vector consisted of 40 features such that $w_i = 20$ and $w_c = 20$. The stimuli in the simulations were analogous to word pairs (*A-B*), where the first word (*A*) would later be used to cue the second word (*B*) during the cued-recall test. The second word was defined as the "item" and represented by $w_i$ features of the memory vector. The first word was defined as the "context" and represented by $w_c$ features of the memory vector. Note that in their simulations, Malmberg and Shiffrin (2005) defined *context* as the global experimental context. It was the disparity in stored global context between conditions that drove the LSE in their free-recall design. In the cued-recall design used here, however, it is the disparity in stored specific context (the cue word) that drives the LSE. In this design, all images are equivalent with regard to global context, so it plays no critical role in the LSE. For the sake of simplicity and clarity it was therefore not implemented in the simulations.

During the recall process, a retrieval cue is initially compared with all of the memory images from the active search set. As a standard simplifying assumption, the search set, $k$, is defined to be the items from the study list. The retrieval cue vector is compared with a given memory vector feature by feature, noting not only the matches and mismatches but also the environmental base rates defined by the geometric distribution parameter $g$. The final result yields a likelihood ratio, $\lambda_j$, which represents how well the retrieval cue matches the memory image $I_j$ relative to other images:

$$\lambda_j = (1 - c)^{n_{ij}} \prod_{i=1}^{\infty} \left[ \frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}}. \quad \text{(A2)}$$

(*Appendixes continue*)

In this equation $c$ defines the probability of correct storage, $i$ is a feature value ranging from 1 to infinity, $n_{ij}$ is the number of nonzero features in the memory image that incorrectly match the cue (regardless of value), and $n_{ijm}$ is the number of nonzero features with value $i$ that correctly match the cue.

The recall process, taken from SAM, has two stages: sampling and recovery. Memory images are sampled serially with replacement from the search set. The probability of sampling memory image $I_j$ given retrieval probe $Q$ on a given sampling trial is

$$P(I_j|Q) = \frac{\lambda_j^\gamma}{\sum \lambda_k^\gamma}. \quad \text{(A3)}$$

In this equation $\gamma$ is a scaling parameter (set here to $\gamma = 0.2$) that corrects for the skew of the likelihood ratio distributions (Shiffrin & Steyvers, 1998). The sampling probability depends on how well the cue matches the target image relative to all other images in the search set. Interference from other images occurs when those images also match the cue, decreasing the probability of sampling the target image. When an image is sampled, the information contained in it must then be recovered to a degree sufficient to support a response. This can be assumed to be a monotonically increasing function of the number of correctly stored feature values in the portion of the image not included in the retrieval cue and that must be recovered. As discussed in the text, strength-related interference emerges from the sampling stage. Thus, the recovery stage was not included in the simulations. The outcomes shown in Figures 5 and 6 represent the probability of sampling the target image on a given sampling attempt.

The simulations were set up to be analogous to the design of Experiments 1–3. The simulated-study list consisted of 18 word pairs, and 1,000 simulated subjects were run for each condition. Simulation 1 (see Figure 5) varied pair strength within a study list, and varied pair overlap and number of strong item repetitions between simulated groups. Each study list was created by constructing three sets of six pairs each. Within a set of pairs, half were assigned to the *strong* condition (bold) and half to the *weak* condition. Strong pairs were presented once, twice, or four times in the list, and weak pairs were presented once. In the low overlap condition, all pairs had unique first-word contexts (***A-B***, *C-D*, ***E-F***, *G-H*, ***I-J***, *K-L*). In the medium overlap condition, each pair shared a first-word context with one other pair (***A-B***, *A-D*, ***E-F***, *E-H*, ***I-J***, *I-L*). In the high overlap condition, all pairs shared the same first-word context (***A-B***, *A-D*, ***A-F***, *A-H*, ***A-J***, *A-L*). Each six-pair set used a unique set of words, each word consisting of 20 features with values drawn from the geometric distribution. Analogous to the word-and-letter-stem cued recall test used in the experiments, the retrieval cue vector consisted of the $w_c$ features of the context-word as well as the first two features of the item-word. Simulation 2 (see Figure 6) used the high overlap, four-repetition condition from the previous simulation and varied the $u*$ parameter over three levels ($u* = .02, .15, .25$) between simulated groups. A high value of $u*$ might be thought of as analogous to a situation in which the context is quickly and easily learned, such as when a preexisting association exists between the context and item (for example, with category–exemplar pairs such as *fruit-apple*). Increasing $u*$ has the same effect as increasing $t$. Thus, this simulation might also be thought of as analogous to a situation where the context–item pairs have been learned or trained to a high degree before testing.

## Appendix B

### Results of Analysis of Variance for Experiment 3 With Study Format as a Between-Subjects Variable and Study–Test Block, Test Position, and Condition as Within-Subject Variables

| Variable | $F$[a] | *MSE* | *p* |
|---|---|---|---|
| Study format | 0.67 | 0.17 | .418 |
| Study–test block | 2.50 | 0.13 | .122 |
| Study–Test Block × Study Format | 0.03 | 0.13 | .875 |
| Test position | 6.13 | 0.10 | **.018** |
| Test Position × Study Format | 0.28 | 0.10 | .599 |
| Condition | 10.54 | 0.08 | **.002** |
| Condition × Study Format | 0.04 | 0.08 | .840 |
| Study–Test Block × Test Position | 1.04 | 0.08 | .315 |
| Study–Test Block × Test Position × Study Format | 0.00 | 0.08 | 1.000 |
| Study–Test Block × Condition | 0.03 | 0.12 | .873 |
| Study–Test Block × Condition × Study Format | 0.94 | 0.12 | .340 |
| Test Position × Condition | 0.03 | 0.10 | .860 |
| Test Position × Condition × Study Format | 0.51 | 0.10 | .481 |
| Study–Test Block × Test Position × Condition | 0.38 | 0.13 | .541 |
| Study–Test Block × Test Position × Condition × Study Format | 0.21 | 0.13 | .646 |

*Note.* Boldface indicates significant results ($p < .05$).
[a] $F(1, 38)$.

Received February 18, 2008
Revision received September 4, 2008
Accepted September 13, 2008 ■