

When and why do retrieval attempts enhance subsequent encoding?

Phillip J. Grimaldi · Jeffrey D. Karpicke

Published online: 12 January 2012
© Psychonomic Society, Inc. 2012

Abstract Unsuccessful retrieval attempts can enhance subsequent encoding and learning. In three experiments, subjects either attempted to retrieve word pairs prior to studying them (e.g., attempting to recall *tide-?* before studying *tide-beach*) or did not attempt retrieval and retention of the studied targets was assessed on a subsequent cued recall test. **Experiment 1** showed that attempting retrieval enhanced subsequent encoding and recall relative to not attempting retrieval when the word pairs were semantically related, but not when the pairs were unrelated. In **Experiment 2**, studying a different word pair prior to the correct pair (e.g., studying *tide-wave* prior to *tide-beach*) did not produce the same effect as attempting retrieval prior to studying. Constraining retrieval to a particular candidate word prior to study (e.g., recalling *tide-wa__* before studying *tide-beach*) produced a negative effect on subsequent recall. **Experiment 3** showed that attempting retrieval did not enhance encoding when a brief delay occurred between the retrieval attempt and the subsequent study trial. The results support the idea that a search set of candidates related to the retrieval cue is activated during retrieval and that this retrieval-specific activation can enhance subsequent encoding of those candidates.

Keywords Memory · Semantic memory · Retrieval processes

Retrieving knowledge is not a neutral event: Every act of retrieval modifies the state of memory (Bjork, 1975). Most

research on the mnemonic effects of retrieval has focused on the effects of successful retrieval, and this research has consistently shown that repeated retrieval directly enhances the subsequent retrievability of knowledge (e.g., Karpicke & Blunt, 2011; Karpicke & Roediger, 2007, 2008; Pyc & Rawson, 2009). In addition, other research has suggested that *unsuccessful* retrieval attempts can enhance learning by improving the encoding of unrecalled items during subsequent study trials (e.g., Izawa, 1970; Karpicke, 2009; Karpicke & Roediger, 2007; Kornell, Hays, & Bjork, 2009; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Richland, Kornell, & Kao, 2009; Slamecka & Fevreski, 1983; Wissman, Rawson, & Pyc, 2011). This article explores the circumstances under which unsuccessful retrieval attempts improve subsequent encoding and enhance learning.

Recently, Kornell et al. (2009) revived a paradigm originally developed by Slamecka and Fevreski (1983) to examine the effects of failed retrieval attempts on subsequent encoding. In this procedure, subjects learned weakly associated word pairs (e.g., *tide-beach*). Some items were “pretested” immediately before each study trial, while others were only studied. For pretested items, subjects were given the cue word and told to guess the target word (*tide-?*) immediately prior to studying the target (*tide-beach*). Because the target words were weakly related to the cues, subjects almost always failed to guess the correct target word and typically produced a different related word (e.g., *wave*). Thus, the procedure was designed to ensure high rates of retrieval failure prior to the encoding of the correct target. On a final cued recall test, subjects recalled a greater proportion of the pretested items than of the studied items. Therefore, the failed retrieval attempts seemed to have enhanced the subsequent encoding of the pretested items.

P. J. Grimaldi (✉) · J. D. Karpicke
Department of Psychological Sciences, Purdue University,
703 Third Street,
West Lafayette, IN 47907-2081, USA
e-mail: pgrimald@purdue.edu

Kornell et al. (2009) mentioned three theoretical ideas, which we have elaborated and defined here, that might explain why failed retrieval attempts would enhance subsequent encoding. According to the first theory, which we refer to as a *search set* theory, the presentation of a cue on the pretest (*tide-?*) initiates a search process wherein related candidates become activated. The activation of candidates during retrieval enhances the encoding of those candidates when they are subsequently presented as target words to study (*tide-beach*). The idea that retrieval activates related candidates corresponds with theories of retrieval such as SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) and MINERVA2 (Hintzman, 1986, 1988). According to these models, cues present at retrieval are used to probe memory and to activate relevant traces in parallel based on similarity to the cue information. These active traces form the search set, and any one trace is probabilistically sampled from the set and recovered in order to make a response. Although a strongly associated trace (e.g., *wave*) may be retrieved and produced as a response, other candidates (e.g., *beach*, *surf*, *ocean*, etc.) receive some activation. Thus, according to the search set theory, failed retrieval attempts may enhance learning because activation of the search set facilitates encoding of the target.

The second theory, which we refer to as an *error correction* theory, is that failed retrieval attempts enhance learning through a general error correction process (Carrier & Pashler, 1992; Kang et al., 2011; McClelland & Rumelhart, 1985). When subjects produce a response and then are shown the correct target, any discrepancy between the correct target and a subject's response would produce an error signal, and a general error correction mechanism would guide adjustment of connection weights among nodes in the system in favor of the target answer. In general, the amount of learning that occurs during error correction corresponds to the size of the discrepancy between the response made and the desired response. The idea that learning is driven by the degree of discrepancy between what is expected and what actually occurs is foundational to several learning theories (Rescorla & Wagner, 1972; Rumelhart, Hinton, & McClelland, 1986). Thus, failed retrieval attempts may enhance learning by virtue of general error correction mechanisms that are sensitive to the discrepancy between the outcome of a retrieval attempt and the subsequent study event.

Finally, the third theory, which we will call an *additional-cue* theory, is that the response that a subject produces during an initial failed retrieval attempt is covertly recalled during a future retrieval attempt, functioning as an additional cue that aids in retrieval of the target item (Soraci et al., 1994). For example, if subjects were given *tide-?* and produced *wave* as a response before they studied the pair *tide-beach*, then *wave* would be encoded along with the target pair and could potentially serve as a retrieval cue in

the future. On a final test, when subjects are given *tide-?*, they may implicitly recall *wave* as an additional retrieval cue for *beach*. Therefore, failed initial retrieval attempts may enhance learning because subjects have two cues to aid recovery of target items. Carpenter (2009, 2011) has recently proposed a similar idea, that during retrieval subjects covertly generate many potential words and that those words can serve as “mediators” for the target word (see also Pyc & Rawson, 2010). Here, we focus only on the one word that is explicitly generated.

In the three experiments reported here, we examined the conditions under which retrieval attempts enhance learning, and tested predictions derived from the theories outlined above. The three theoretical ideas need not be mutually exclusive, but the theories lead to specific predictions that were examined in the following experiments.

Experiment 1

Experiment 1 used the procedure developed by Kornell et al. (2009, Exps. 3–6; see also Slamecka & Fevreski, 1983), which was shown to create high rates of failure during initial retrieval attempts. In addition to the pretesting manipulation, the other key manipulation in **Experiment 1** was the relatedness of the cue–target word pairs. Subjects learned a mixed list containing 30 related word pairs (e.g., *tide-beach*) and 30 unrelated word pairs (e.g., *pillow-leaf*). In the no-pretest condition, subjects simply studied each pair one at a time. In the pretest condition, subjects were shown the cue word of each pair (e.g., *tide-?* or *pillow-?*) and attempted to retrieve a target item immediately before they studied the correct cue–target pair. Retention was assessed on a final cued recall test over all items.

The search set theory proposes that semantically related candidates become activated during a retrieval attempt. When one of these related words is presented as a target on the subsequent study trial, the residual activation from the retrieval attempt contributes to the enhanced encoding of the target word. It is unlikely that this enhanced encoding would occur if subjects were to study unrelated words during study trials, because in this case, the initial search set of candidates would not include the target. Thus, attempting retrieval during a pretest should enhance encoding of related cue–target words pairs but should not enhance the encoding of unrelated pairs.

Both the error correction and additional-cue theories lead to predictions that differ from the one derived from the search set theory. According to an error correction theory, an error signal would be produced during the encoding of both related and unrelated word pairs, and thus learning

should be enhanced for both item types. Moreover, it is reasonable to assume that the magnitude of the discrepancy between the subject's response and the correct target would be greater for unrelated than for related pairs. Given that the degree of learning may be related to the degree of this discrepancy (Rescorla & Wagner, 1972), the mnemonic benefit of attempting retrieval prior to study may be greater for unrelated word pairs than it would be for related word pairs.

The idea behind an additional-cue theory is that words produced during a retrieval attempt are encoded as additional retrieval cues. Cues that are generated by subjects are typically more effective than are experimenter-provided cues (Arndt & Jones, 2008; Karpicke & Cross, 2011; Mäntylä, 1986). If subjects can retrieve additional cues at final test, then they could use both experimenter-provided cues and their own self-generated cues to retrieve the target items, and it is reasonable to assume that using both cues would be more effective than using the experimenter-provided cue alone. Thus, like the general error correction theory, an additional-cue theory would predict that because subjects generate an additional cue during a retrieval attempt, performance should be enhanced for both unrelated and related words.

Method

Subjects Thirty-two Purdue University undergraduates participated in exchange for course credit. All subjects indicated that they were fluent in written and spoken English.

Materials A list of 60 word pairs (containing 120 medium-frequency, medium-concreteness words) was used in the experiment. The word pairs were divided into two sets that were matched on frequency and concreteness. One set (30 of the word pairs) included associatively related word pairs (e.g., *tide–beach*, *jelly–bread*, *kite–wind*), and the other set included unrelated pairs (e.g., *stem–candy*, *pillow–leaf*, *spray–bone*). The forward associative strength of the related pairs was between .050 and .054, based on Nelson, McEvoy, and Schreiber's (1998) norms. Forward associative strength represents the likelihood that a word will be the first word generated on a free-association test; therefore, the related target words had about a 5% chance of being correctly guessed by subjects during pretest trials. For unrelated pairs, there was no associative relationship between the words, according to the Nelson et al. norms.

Design The experiment used a 2 (pair relatedness: related or unrelated) \times 2 (learning condition: pretest or no-pretest) mixed factorial design. Pair relatedness was manipulated within subjects, and learning condition was manipulated between subjects. Sixteen subjects were assigned to each condition.

Procedure The subjects were tested in small groups. They were told that they would learn a list of word pairs and take a memory test on the pairs. Subjects in the no-pretest condition studied the list of word pairs (e.g., *tide–beach*) in a study phase. The word pairs were shown on a computer screen in a random order, and each word pair was presented for 5 s with a 500-ms intertrial interval. Subjects in the pretest condition saw each cue word (*tide–?*) on the screen for 7 s prior to each study trial. They were instructed to attempt to guess the target word and to type their response into an input field shown on the screen (following Kornell et al.'s, 2009, instructions). The subjects were encouraged to provide a response for every pair. After each pretest trial, subjects immediately studied the correct cue–target pair (*tide–beach*) for 5 s with a 500-ms intertrial interval.

After the study phase, subjects completed a distractor task that involved solving math problems for 5 min. They then took a cued recall test on the word pairs. On each test trial, subjects were shown a cue word and a blank text input field for 7 s, and they were instructed to recall and type the target word they had studied in the study phase. The cue words were presented in a random order in the test phase. Subjects were told to avoid guessing and to respond only with words they remembered studying, not words they had produced. At the end of the experiment, the subjects were debriefed and thanked for their participation.

Results and discussion

Figure 1 shows the proportions of words recalled on the final cued recall test. For all statistical tests, alpha was set to .05. To ensure that these analyses included only items from initial failed retrieval attempts, items that were correctly guessed during the pretest trials were excluded from the

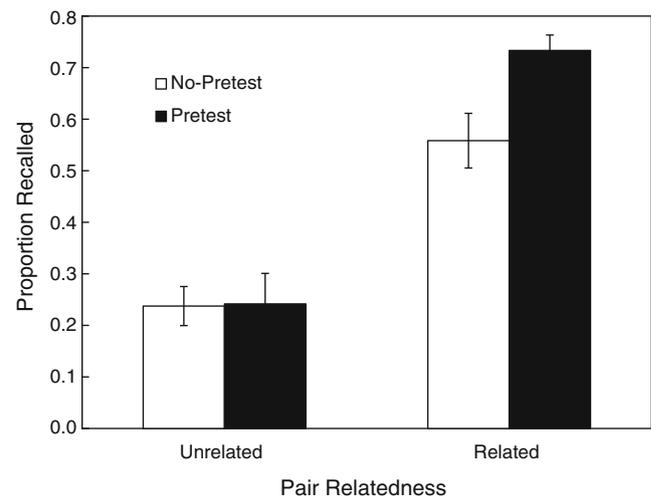


Fig. 1 Proportions of words recalled on the final cued recall test in Experiment 1. Error bars represent standard errors of the means

analyses. On average, subjects correctly guessed the target answer for 6% of the related pairs, a rate that is similar to the rates of correct guessing reported by Kornell et al. (2009). For unrelated pairs, the target responses were never correctly guessed during pretest trials. The removal of correctly guessed pairs from the analyses did not affect the pattern of results.

The final test results were analyzed using a 2 (learning condition) \times 2 (pair relatedness) mixed factorial ANOVA, with learning condition as a between-subjects factor and pair relatedness as a within-subjects factor. There was a main effect of pair relatedness, $F(1, 30) = 180.26$, $\eta_p^2 = .86$, but the main effect of learning condition did not reach significance, $F(1, 30) = 2.14$, $p = .15$. However, there was a significant learning condition \times pair relatedness interaction, $F(1, 30) = 7.67$, $\eta_p^2 = .20$. For unrelated pairs, there was no difference between the pretest and no-pretest conditions (.24 vs. .24), $t(30) < 1$. For related pairs, the pretest condition produced significantly greater recall relative to the no-pretest condition (.73 vs. .56), $t(30) = 2.78$, $d = 0.94$. Thus, attempting retrieval prior to each study trial enhanced subsequent recall of related word pairs but not of unrelated word pairs.

The results of Experiment 1 replicate and extend the results reported by Kornell et al. (2009). Subjects who attempted retrieval prior to each study trial recalled more words on the final test than did subjects who did not attempt retrieval. However, this effect only occurred when the cue and target words were related, and it did not occur when the words were unrelated. The results are consistent with the search set theory described at the outset, which holds that related candidates are activated when the subject establishes a search set during a retrieval attempt. Because semantic relatedness is used to establish the search set, attempting retrieval enhances the subsequent encoding of related but not of unrelated words. The lack of a pretesting effect for unrelated words is more difficult to reconcile with predictions from both an error correction and an additional-cue theory. Both theories would suggest that attempting retrieval would enhance encoding of unrelated word pairs, either due to a general error correction mechanism or by virtue of producing additional retrieval cues.

Experiment 2

The purpose of Experiment 2 was to identify aspects of the retrieval processes that are responsible for enhancing subsequent learning. One question was whether the process of retrieval itself matters for the effects observed in Experiment 1. It may be the case that simply encoding another related word with the cue–target pair facilitates recall, and a retrieval attempt simply provides this additional encoding opportunity.

This idea would be consistent with an additional-cue theory, because the locus of the effect might not be the process of retrieval but rather the product of retrieval (i.e., the presence of an additional cue at encoding). It would also be consistent with associative network models of memory, wherein activation spreads automatically from one concept to related concepts (e.g., Collins & Loftus, 1975). The presence of an additional word (like *wave*) may produce spreading activation to related target words (like *beach*), thereby enhancing encoding of the target words, and this may occur regardless of whether the additional word was studied or retrieved. To examine this possibility, in Experiment 2 we included a *study lure* condition, which was closely matched to the pretest condition but did not involve retrieving a word prior to each study trial. During the study lure trial, subjects studied the cue and a related but incorrect lure word, which was the most frequently produced word based on data from Experiment 1 and from another pilot study (e.g., subjects studied *tide–wave* immediately before studying the target pair *tide–beach*, because *wave* was the word most frequently produced to *tide* in our normative data). If the presence of an additional cue is responsible for the pretesting effect, and if engaging in a retrieval attempt is irrelevant, then the study lure condition should enhance learning in the same way as the pretesting condition.

The second purpose of Experiment 2 was to examine the search set theory by manipulating the nature of the search process that subjects engage in during retrieval attempts. If a set of candidates becomes activated during a retrieval attempt, then constraining the search set to a particular candidate should limit the activation accrued by other candidates and reduce the benefits of attempting retrieval for subsequent encoding. To examine this idea, Experiment 2 included a *constrained-pretest* condition in which subjects were given the stem of a lure word to be generated during pretest trials (e.g., *tide–wa_*). The search set theory predicts that the pretesting effect should be reduced or eliminated in the constrained-pretest condition. Importantly, the error correction and additional-cue theories predict that a constrained retrieval attempt would enhance encoding just as an unconstrained retrieval attempt would, because an error signal or an additional cue would be produced even under constrained-retrieval conditions.

Method

Subjects One hundred twenty Purdue University undergraduates participated in exchange for course credit. None of the students had participated in Experiment 1.

Materials The materials used in Experiment 1 were used again in Experiment 2. Lure words or the stems of lure words were presented in the study lure and constrained-pretest

conditions, respectively. The lure words used in [Experiment 2](#) were the responses produced most frequently to each cue word in the data from [Experiment 1](#) and from an additional pilot study. The lure words selected for [Experiment 2](#) had been produced 22% of the time, on average, in those two previous studies.

Design The experiment used a 2 (pair relatedness: related or unrelated) \times 4 (learning condition: pretest, constrained-pretest, study lure, or no-pretest) mixed factorial design. Pair relatedness was manipulated within subjects, and learning condition was manipulated between subjects. Thirty subjects were assigned to each condition.

Procedure The procedure used in [Experiment 2](#) was similar to that used in [Experiment 1](#). The pretest and no-pretest conditions were identical to those used in [Experiment 1](#). Subjects in the constrained-pretest condition were given the cue and a word stem of the lure word (e.g., *tide-wa_*). The word stems always included the first two letters of the lure and the appropriate number of blanks for the remaining letters in the word. Subjects were instructed to complete the word stem by typing the entire word (e.g., typing *wave*, not *ve*) into an input field. They were also told that the word was often the first word that would come to mind when thinking of the cue word. Subjects were given 7 s to complete the stem prior to each study trial. Subjects in the study lure condition studied the cue and the lure word (*tide-wave*) for 7 s prior to each study trial. Subjects were informed that the first pair was an incorrect answer and that the second pair was the correct target answer. Subjects in both the constrained-pretest and study lure condition were informed at the beginning of the experiment that they would be tested on their memory for the correct second word, not the stem word or lure word. In all other respects, the procedure was identical to the one used in [Experiment 1](#).¹

Results and discussion

Figure 2 shows the proportions of words recalled on the final cued recall test. As in [Experiment 1](#), only items that were not correctly guessed during pretest trials were included in the analyses. For the related pairs, subjects in the pretest condition correctly guessed 8% of the targets on

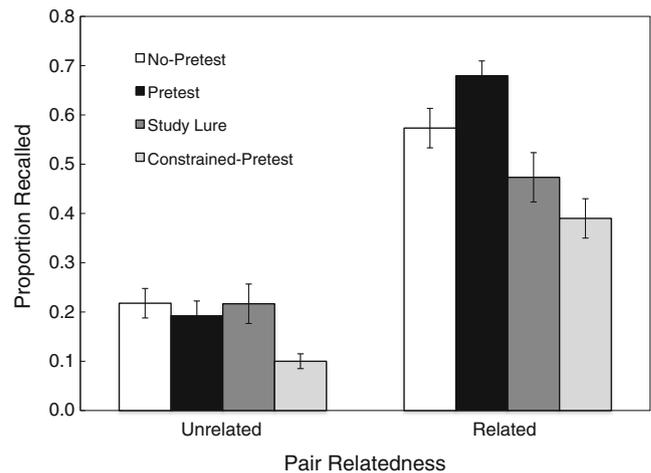


Fig. 2 Proportions of words recalled on the final cued recall test in [Experiment 2](#). Error bars represent standard errors of the means

the pretest, and subjects in the constrained-pretest condition correctly guessed 1% of the targets on the pretest. For the unrelated pairs, target responses were never correctly guessed during pretest trials. The removal of correctly guessed pairs from the analyses did not affect the pattern of results. For completeness, we also computed the probability of producing the lure word during the initial pretest trials. Subjects in the pretest condition produced the lure word on 27% of trials, and subjects in the constrained-pretest condition produced the lure word on 80% of the trials.

The final test results were analyzed with separate one-way ANOVAs on the unrelated and related pairs. For unrelated pairs, subjects in the constrained-pretest condition recalled fewer target words (.10) than did subjects in the other conditions (.22, .19, and .22 for the no-pretest, pretest, and study lure conditions, respectively), $F(3, 116) = 2.99$, $\eta_p^2 = .07$. For related pairs, there was also a significant difference in recall among the four conditions, $F(3, 116) = 10.56$, $\eta_p^2 = .22$. Subjects in the pretest condition recalled significantly more targets than did subjects in the no-pretest condition (.68 vs. .57), $t(58) = 2.42$, $d = 0.58$, replicating the results of [Experiment 1](#). Recall in the study lure condition was worse than recall in the no-pretest condition (.47 vs. .57), although this difference did not reach significance, $t(58) = 1.62$, $p = .11$, $d = 0.41$. Finally, recall in the constrained-pretest condition was significantly worse than recall in the no-pretest condition (.39 vs. .57), $t(58) = 3.49$, $d = 0.81$.

The lack of an effect in the study lure condition suggests that the process of attempting to retrieve a word, rather than the mere presence of a word as a possible additional cue, was responsible for enhancing the encoding of the target word. If spreading activation alone, as it is typically conceived, were responsible for the enhancements to encoding,

¹ A minor change in the instructions was made to the pretest trials in order to match instructions to those of the constrained-pretest condition. Rather than being told to try to guess the target, subjects were instructed to type the first word that came to mind when viewing the cue word. This instructional change did not alter the main result (the pretesting effect) relative to the other experiments.

then there would have been a positive effect in the study lure condition, but there was not. Perhaps the most striking finding of **Experiment 2** is that, whereas engaging in an unconstrained retrieval attempt (in the pretest condition) enhanced performance relative to not attempting retrieval, engaging in a constrained retrieval attempt produced a negative effect, decreasing recall relative to not attempting retrieval. The only difference between the pretest and constrained-pretest conditions was the provision of a stem of the lure word in the constrained-pretest condition. The constrained-pretest condition guaranteed that subjects would produce an error, and the word subjects produced could function as an additional cue. Thus, the results are problematic for both general error correction and additional-cue theories.

It is easier to explain the results in terms of a search set theory. A constrained retrieval attempt likely restricted activation to a particular candidate and limited the activation of other candidates that would occur under unconstrained retrieval conditions. Because other candidates received only limited activation, the retrieval attempt provided no enhancement to the encoding of those candidates during the subsequent study trial. However, the finding that constrained retrieval attempts produced negative effects on subsequent recall, not merely neutral effects, suggests that mechanisms other than the activation of candidates within a search set must be considered to account for the results. We discuss additional possible explanations of these findings in the **General Discussion**.

One might wonder whether the errors from pretest trials and lures from study lure trials produced proactive interference on the final test, and whether or not proactive interference could explain the differences in recall performance. To examine this possibility, we computed the proportions of intrusion errors on the final test. This analysis only included cases in which the word produced on a pretest or constrained pretest, or studied on a study lure trial, was falsely recalled on the final test. These results are shown in Table 1. For unrelated pairs, there were significant differences in intrusions between the conditions, $F(2, 87) = 6.54$, $\eta_p^2 = .13$. The study lure condition produced more intrusions than did the pretest condition, $t(58) = 4.49$, $d = 0.68$, as did the constrained-pretest condition, $t(58) = 3.87$, $d = 0.99$. For related pairs, there were also significant differences in intrusions between the conditions, $F(2, 87) = 10.01$, $\eta_p^2 = .19$. The study lure condition produced more intrusions than the pretest condition, $t(58) = 4.49$, $d = 1.08$, as did the constrained-pretest condition, $t(58) = 4.38$, $d = 1.08$. The intrusion results suggest that more proactive interference occurred in the study lure and constrained-pretest conditions than in the pretest condition. However, subjects in the study lure condition produced more intrusions than did those in the pretest condition for unrelated words, yet the groups' recall

Table 1 Proportions of intrusion errors on the final test as a function of learning condition and pair relatedness in Experiments 2 and 3

	Unrelated	Related
Experiment 2		
Pretest	.04 (.01)	.02 (.00)
Study Lure	.14 (.04)	.16 (.03)
Constrained-Pretest	.20 (.04)	.19 (.04)
Experiment 3		
Immediate-Pretest	.07 (.03)	.03 (.01)
Delayed-Pretest	.17 (.05)	.12 (.04)

These results represent the proportions of final test trials in which an initially recalled word, a lure word, or a stem word was falsely recalled. Standard errors are in parentheses.

performance for unrelated words was equivalent. This suggests that although there was proactive interference, it was not necessarily diagnostic of cued recall performance.

Experiment 3

The purpose of **Experiment 3** was to test a prediction derived from the search set theory. According to the theory, activation of candidates during the initial retrieval attempt should be relatively short lived. Raaijmakers and Shiffrin (1981) proposed that the activation of candidates within a search set was at least long enough to ensure sampling of a candidate. Similarly, research on semantic priming has shown that priming is reduced when one intervening word occurs between a prime and a target word, and some studies have observed no semantic priming when two intervening words occur (McNamara, 1992). In Experiments 1 and 2, there was no lag (no intervening word or trials) between the pretest trial and the subsequent study trial (e.g., *tide-?* was immediately followed by *tide-beach*). In **Experiment 3**, we examined performance in a *delayed-pretest* condition in which subjects experienced an entire block of pretest trials (e.g., *tide-?*, *jelly-?*, *kite-?*) and then studied target word pairs in a separate study block (e.g., *tide-beach*, *jelly-bread*, *kite-wind*). If attempting retrieval results in activation of candidates, and if this activation fades quickly, then increasing the interval between the pretest trials and study trials should eliminate the effect in the delayed-pretest condition. The error correction and additional-cue theories also lead to similar predictions, because presumably a general error correction mechanism would depend on the immediate presence of a correct response, and because an additional cue would need to occur in close temporal proximity to the target in order to establish a link between the two.

Method

Subjects Fifty-four Purdue University undergraduates participated in exchange for course credit. None of the subjects had participated in Experiment 1 or 2.

Materials The materials used in Experiment 1 were also used in Experiment 3.

Design The experiment used a 2 (pair relatedness: related or unrelated) \times 3 (learning condition: immediate pretest, delayed pretest, or no pretest) mixed factorial design. Pair relatedness was manipulated within subjects, and learning condition was manipulated between subjects. Eighteen subjects were assigned to each condition.

Procedure The procedure in Experiment 3 was similar to the one used in Experiment 1. The no-pretest and immediate-pretest conditions were identical to the no-pretest and pretest conditions of Experiment 1. In the delayed-pretest condition, subjects experienced a block of pretest trials for all 60 word pairs followed by a block of study trials for all of the pairs. The order of trials was randomized within each block. Note that total exposure times were matched in the immediate-pretest and delayed-pretest conditions. The only difference between the conditions was the interval between the pretest trial and the study trial for each item.

Results and discussion

Figure 3 shows the proportions of words recalled on the final cued recall test. As in Experiments 1 and 2, the results include only items that were not correctly guessed on the pretest. For related pairs, the proportions of items correctly guessed on the pretests were 6% and 5% in

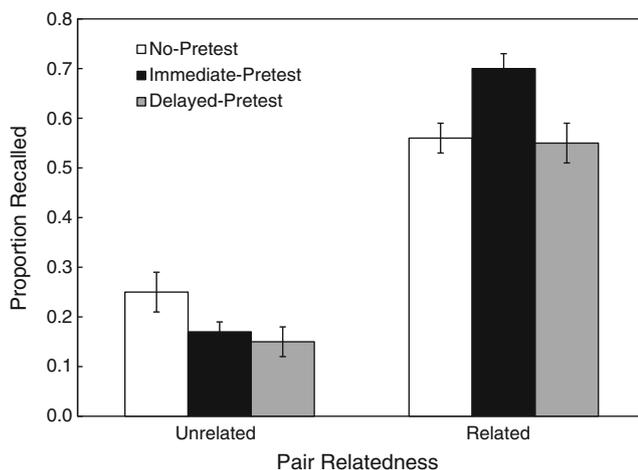


Fig. 3 Proportions of words recalled on the final cued recall test in Experiment 3. Error bars represent standard errors of the means

the immediate-pretest and delayed-pretest conditions, respectively. No unrelated pairs were correctly guessed in either pretest condition. The exclusion of correctly guessed pairs did not affect the pattern of results.

A one-way ANOVA on the unrelated-word-pair results did not indicate a significant difference among the conditions, $F(2, 51) = 2.28, p = .11$. A separate one-way ANOVA on the related-word-pair results did indicate that there was a significant difference among the conditions, $F(2, 51) = 5.13, \eta_p^2 = .17$. Subjects in the immediate-pretest condition recalled more words than did subjects in the no-pretest condition (.70 vs. .56), $t(34) = 3.00, d = 1.08$, replicating the pretest advantage observed in the previous experiments with related word pairs. However, subjects in the delayed-pretest and no-pretest conditions produced similar levels of final recall (.55 vs. .56), $t(34) < 1$. Thus, attempting retrieval only enhanced performance when the cue and target words were related, replicating our previous experiments and those of Kornell et al. (2009), and this enhancement only occurred when the study trial occurred immediately after the pretest trial. This result represents a boundary condition of the pretesting effect, and it also supports the idea that the activation of candidates during retrieval is short lived, and thus, attempting retrieval does not enhance encoding when study trials are delayed relative to the pretest trials.

As in Experiment 2, we conducted an analysis of intrusion errors on the final test in order to examine the role of proactive interference. The results are shown on Table 1. For unrelated words, subjects in the delayed-pretest condition produced more intrusions than did those in the immediate-pretest condition, but this effect did not reach significance, $t(34) = 1.8, p = .08, d = 0.58$. For related pairs, subjects in the delayed-pretest condition again produced more intrusions than did those in the immediate-pretest condition, $t(34) = 2.07, d = 0.72$. Thus, the intrusion rates corresponded to the cued recall results. For unrelated pairs, the two groups had equivalent cued recall performance and equivalent intrusion rates. For related pairs, the immediate-pretest condition had greater cued recall performance and lower intrusion rates for related pairs than did the delayed-pretest group. These results suggest that when a retrieval attempt is followed immediately by encoding of a related item, proactive interference from the initial retrieval attempt is reduced. This is consistent with the search set theory—improving the recall probability of one item in the search set also reduces the recall probability of other items in the search set (Mensink & Raaijmakers, 1988).

General discussion

In three experiments, we identified conditions under which attempting retrieval enhanced learning, had no effect on learning, and hurt learning. Relatively minor changes in

the conditions of initial retrieval attempts were sufficient to bring about sizeable changes in memory performance. In the following section, we briefly review the key findings of these experiments and interpret the results in light of the theoretical explanations discussed throughout this article.

First, attempting retrieval enhanced the subsequent learning of related target words but not of unrelated target words. This pattern of results was observed in all three experiments. The general error correction and additional-cue theories would lead to the prediction that attempting retrieval should enhance subsequent learning of unrelated words, because even when target words were unrelated to the cues, failed retrieval attempts would still generate error signals and still result in the production of an additional cue. Instead, the results are consistent with a search set theory, in which establishment of a search set during retrieval activates related candidates. The encoding of related candidates would be enhanced because of this activation, but the encoding of unrelated words would not be enhanced, because those words were not part of the search set during the retrieval attempt.

Second, the effects of pretesting prior to study trials are retrieval-specific (Exp. 2). Kornell et al. (2009) showed that providing additional study time for target items did not produce the same effects as attempting retrieval prior to studying the items. In [Experiment 2](#), studying the cue paired with a lure word did not produce the same benefit as attempting to retrieve a word (which was typically the lure word) in the presence of the cue. Thus, theoretical explanations of pretesting effects should focus on the nature of the retrieval process, particularly the establishment of a search set.

Third, unconstrained retrieval attempts enhanced learning, but constrained retrieval attempts did not (Exp. 2). In fact, constraining retrieval to a particular word *decreased* subsequent recall, relative to making no retrieval attempt in the no-pretest control condition. Again, a constrained retrieval attempt would yield an error signal and would lead to production of an additional cue, so this result is difficult to accommodate within error correction or additional-cue theories. The result is more consistent with a search set theory. Constraining retrieval likely restricted the search set and limited the activation of other candidates. Therefore, constrained retrieval attempts would not bolster the encoding of candidates during the subsequent study trial.

Finally, attempting retrieval enhanced learning when study trials occurred immediately after the pretest trials but not when they occurred after a delay (Exp. 3). This finding is also consistent with a search set theory. Activation of a search set is assumed to be short lived (Raaijmakers & Shiffrin, 1981). Therefore, a delayed study trial would occur at a time when the search set is presumably no longer activated, so that the studied word would not benefit from a prior retrieval attempt.

Thus, the majority of the present results are consistent with a search set theory of the pretesting effect. The theory suggests that when subjects attempt retrieval, a search set consisting of related candidates becomes activated, and residual activation enhances encoding when a candidate is presented in a subsequent study trial. The theory can account for the findings that attempting retrieval enhanced the learning of related but not of unrelated word pairs, that constraining retrieval to a particular candidate did not enhance learning, and that inserting a delay between the retrieval attempt and the subsequent study trial eliminated the pretesting effect.

One result from the present experiments is not readily explained by the search set theory: Engaging in a constrained retrieval attempt impaired learning relative to making no attempt in the no-pretest control condition. In [Experiment 2](#), while an unconstrained retrieval attempt enhanced recall relative to the control condition (.68 vs. .57), a constrained retrieval attempt impaired recall relative to the control condition (.39 vs. .57). It was likely the case that directing subjects to retrieve a particular candidate introduced competition that was not present under unconstrained retrieval conditions (Barnes & Underwood, 1959). This competition during retrieval may have interfered with the subsequent encoding of candidate words enough to produce a negative effect on recall. In [Experiment 2](#), constraining the initial retrieval attempts reduced the levels of final recall for both related and unrelated word pairs, suggesting that interference was created by general competition during retrieval. At this point, we can only speculate about the possible mechanisms responsible for this negative effect of engaging in initial constrained retrieval attempts. The topic certainly merits further exploration.

Conclusion

The act of attempting retrieval alters the encoding that occurs in a subsequent study episode and thereby affects learning. The effects of attempting retrieval depend not only on the relatedness of the materials and the temporal context of the retrieval and encoding events but also on the nature of the retrieval processes engaged in by subjects. Whereas unconstrained initial retrieval attempts enhanced encoding, constrained retrieval attempts did not. Indeed, constrained retrieval reduced performance, perhaps due to competition afforded by the search for a particular target word. The present experiments provide initial evidence supporting the role of establishing a search set in the positive effects of attempting retrieval. Regardless of the theoretical underpinnings of these effects, it is clear that attempting retrieval has the potential to improve learning by enhancing subsequent encoding.

Author note This research and the writing of this article were supported in part by grants from the National Science Foundation (0941170) and the Institute of Education Sciences in the U.S. Department of Education (R305A110903). The opinions expressed are those of the authors and do not represent the views of the Institute or of the U.S. Department of Education. We thank Lauren Abplanalp, Veronica Crofts, Breanne Lawler, Justin McCloskey, Sohail Mehra, and Shaun Parkins for their help collecting the data. We also thank Althea Bauernschmidt, James Naime, and Megan Smith for several helpful comments on this work.

References

- Arndt, J., & Jones, T. C. (2008). Elaborative processing and conjunction errors in recognition memory. *Memory & Cognition, 36*, 899–912. doi:10.3758/MC.36.5.899
- Barnes, J. M., & Underwood, B. J. (1959). “Fate” of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97–105. doi:10.1037/h0047507
- Bjork, R. A. (1975). Retrieval as a memory modifier. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale: Erlbaum.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1563–1569. doi:10.1037/a0017021
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1547–1552. doi:10.1037/a0024140
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633–642. doi:10.3758/BF03202713
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428. doi:10.1037/0033-295X.82.6.407
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67. doi:10.1037/0033-295X.91.1.1
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review, 93*, 411–428. doi:10.1037/0033-295X.93.4.411
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551. doi:10.1037/0033-295X.95.4.528
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*, 340–344. doi:10.1037/h0028541
- Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103*, 48–59. doi:10.1037/a0021977
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469–486. doi:10.1037/a0017341
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772–775. doi:10.1126/science.1199327
- Karpicke, J. D., & Cross, K. M. (2011). Effects of cue generation and retrieval practice on complex learning. Unpublished manuscript, Purdue University.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151–162. doi:10.1016/j.jml.2006.09.004
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968. doi:10.1126/science.1152408
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 989–998. doi:10.1037/a0015729
- Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 66–71. doi:10.1037/0278-7393.12.1.66
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114*, 159–188. doi:10.1037/0096-3445.114.2.159
- McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1173–1190. doi:10.1037/0278-7393.18.6.1173
- Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95*, 434–455. doi:10.1037/0033-295X.95.4.434
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available at <http://w3.usf.edu/FreeAssociation/>
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 287–297. doi:10.1037/a0021801
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437–447. doi:10.1016/j.jml.2009.01.004
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*, 335. doi:10.1126/science.1191465
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134. doi:10.1037/0033-295X.88.2.93
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243–257. doi:10.1037/a0016496
- Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45–76). Cambridge, MA: MIT Press.
- Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior, 22*, 153–163. doi:10.1016/S0022-5371(83)90112-3
- Soraci, S. A., Franks, J. J., Bransford, J. D., Chechile, R. A., Belli, R. F., Carr, M., & Carlin, M. (1994). Incongruous item generation effects: A multiple-cue perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 67–78. doi:10.1037/0278-7393.20.1.67
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18*, 1140–1147.