Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 1, 283–294

0278-7393/15/$12.00 http://dx.doi.org/10.1037/a0037850

RESEARCH REPORT

Retrieval Attempts Enhance Learning, but Retrieval Success (Versus Failure) Does Not Matter

Nate Kornell
Williams College

Patricia Jacobs Klein
University of Michigan

Katherine A. Rawson
Kent State University

Retrieving information from memory enhances learning. We propose a 2-stage framework to explain the benefits of retrieval. Stage 1 takes place as one attempts to retrieve an answer, which activates knowledge related to the retrieval cue. Stage 2 begins when the answer becomes available, at which point appropriate connections are strengthened and inappropriate connections may be weakened. This framework raises a basic question: Does it matter whether Stage 2 is initiated via successful retrieval or via an external presentation of the answer? To test this question, we asked participants to attempt retrieval and then randomly assigned items (which were equivalent otherwise) to be retrieved successfully or to be copied (i.e., not retrieved). Experiments 1, 2, 4, and 5 tested assumptions necessary for interpreting Experiments 3a, 3b, and 6. Experiments 3a, 3b, and 6 did not support the hypothesis that retrieval success produces more learning than does retrieval failure followed by feedback. It appears that retrieval attempts promote learning but retrieval success per se does not.

Keywords: memory, learning, retrieval, metacognition

There is a great deal of evidence that retrieving information from memory enhances learning (for reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006b). For example, a period of studying followed by a test produces more learning than does a period of studying followed by restudying. This article is about why retrieval enhances learning. We begin by outlining a two-stage framework of retrieval that makes a distinction between the retrieval attempt and the processing that occurs after an answer becomes available. This model motivates the question tested in the experiments presented here: Does retrieving an answer successfully produce more learning than attempting to retrieve, without succeeding, and then being told the answer?

A Two-Stage Framework

Retrieval can be divided into two stages (see Figure 1). Retrieval is initiated when a question is presented. Stage 1 is the retrieval attempt—the period of time when one is searching for the answer, but it is not yet available. Stage 1 ends when the answer is obtained (or the retrieval attempt is terminated). Stage 2, if it occurs, is the period after an answer has become available. There are two ways the answer can become available (thus initiating Stage 2): successful retrieval from memory or a presentation from an external source (see Figure 2). The research presented here investigates which, if either, way of obtaining the answer produces more learning.

This two-stage framework is consistent with what Carpenter (2009) called the "elaborative retrieval hypothesis." When a cue (e.g., *Olympic*) is presented, attempting to retrieve the target (e.g., *Athletic*) activates information related to the cue. Activation spreads through the semantic network associated with the cue (Collins & Quillian, 1972). This spreading activation can activate the target, leading to retrieval success. Stage 2 of the retrieval process begins when the target is identified, either via an external presentation or a successful retrieval. Associations between the nodes that link the cue to the target are strengthened in Stage 2. These include direct associations between the cue and target (e.g., *Olympic–Athletic*). They also include related, mediating connections (e.g., *Olympic–Sport–Athletic*; see Carpenter, 2011). Irrelevant associations (e.g., *Olympic-Mountains*) may be identified and weakened via suppression or inhibition (although this hypothesis has not been tested; Kornell, Hays, & Bjork, 2009). We refer to this two-stage process of activating and then strengthening or weakening related information as *elaborative retrieval* (see also Carpenter, 2009, 2011; Pyc & Rawson, 2010).

This article was published Online First October 20, 2014.

Nate Kornell, Department of Psychology, Williams College; Patricia Jacobs Klein, Department of Mathematics, University of Michigan; Katherine A. Rawson, Department of Psychology, Kent State University.

A James S. McDonnell Foundation Scholar Award to the first author supported this research. Thanks to Hannah Hausman for help conducting the research and to Henry L. Roediger, III, for valuable comments on a draft of this article.

Correspondence concerning this article should be addressed to Nate Kornell, Department of Psychology, Williams College, 18 Hoxsey Street, Williamstown, MA 01267. E-mail: nkornell@gmail.com

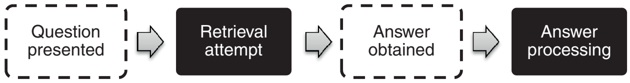

Figure 1. Framework for understanding information processing during retrieval. Stage 1, the retrieval attempt, is initiated when the question is asked. Stage 2, answer processing, begins when the answer becomes available.

Elaborative retrieval explains why retrieval enhances learning more than does re-presentation. When a cue–target pair is presented rather than tested, Stage 1 processing does not occur. Thus, when Stage 2 begins, associative connections emanating from the cue are less active. As a result, correct connections are strengthened less (and perhaps incorrect connections are weakened less). Equally important, the connections emanating from the cue are less varied, so fewer correct paths are strengthened (and perhaps fewer dead-ends are weakened). In short, retrieval causes more cue–target links to be strengthened (or weakened), more powerfully, than does re-presentation.

Does Retrieval Success Matter?

On the basis of the elaborative retrieval hypothesis, Stage 1 and Stage 2 are both crucial. Does it matter, though, whether the transition from Stage 1 to Stage 2 is initiated via successful retrieval from memory or via presentation of the correct answer (see Figure 2)? In other words, retrieval *attempts* clearly matter, but does retrieval *success* matter? This is the question we attempted to answer in the research reported here. As we describe next, prior research has not directly examined this question.

Item Selection Effects in Retrieval Experiments

Table 1 specifies the situations that can occur during the practice phase of a retrieval experiment. Many of the comparisons in Table 1 have been made in past research. For example, retrieval trials that are not followed by feedback (i.e., a mix of C and E trials) produce more learning than do presentation trials (i.e., B trials; e.g., Roediger & Karpicke, 2006a). Retrieval trials followed by feedback (a mix of D and F trials) also produce more learning than presentation trials (i.e., B trials; e.g., Carrier & Pashler, 1992). Feedback following an unsuccessful retrieval attempt enhances learning (i.e., D trials produce more learning than C trials; Pashler, Cepeda, Wixted, & Rohrer, 2005). E trials and F trials are equally effective because feedback has no measurable effect on learning following a successful retrieval (e.g., Hays, Kornell, & Bjork, 2010; Pashler et al., 2005).[1]

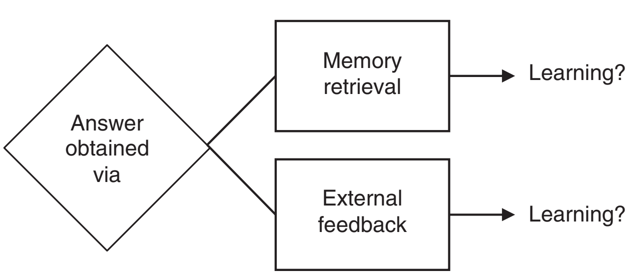

Figure 2. Two ways to arrive at a correct answer following a retrieval attempt. In Experiments 3a, 3b, and 6, we manipulated which of these two routes participants took to reach the answer and measured subsequent learning.

Table 1
Possible Conditions Instantiated in Research on Retrieval Effects

Label	Stage 2 initiated by: Stage 1: Retrieval attempt?	Answer via retrieval success?	Answer via feedback?
A	No		No
B	No		Yes
C	Yes	No	No
D	Yes	No	Yes
E	Yes	Yes	No
F	Yes	Yes	Yes

Our goal was to compare the effectiveness of D and E trials. To our knowledge, they have not been compared in past research without being confounded by item selection effects. Item selection effects are a problem because, by definition, items that are retrieved during practice (E items) are easier for the participant than items that are not retrieved during practice (D items). Differences in final test performance should occur, because of this confound, whether or not retrieval success matters.

Item selection effects can be overcome using random assignment. In the key studies reported herein, Experiments 3a, 3b, and 6, two sets of items went through Stage 1 and Stage 2 processing. The only difference between the sets was how the items made the transition from Stage 1 to Stage 2. Items were randomly assigned either (a) to be retrieved successfully or (b) not to be retrieved, but to be followed by feedback. If retrieval success matters, items that were retrieved should be remembered better on a later test than items that were not retrieved. If retrieval success does not matter, the conditions should not differ.

The hypothesis that retrieval success might not matter relies on the assumption that learning benefits from unsuccessful retrieval attempts that are followed by feedback. A growing body of evidence supports this assumption. For example, Kornell, Hays, and Bjork (2009) found that a test followed by feedback produced more learning than an equal-duration presentation trial, even when retrieval was never successful during the study phase (i.e., D > B; see also Grimaldi & Karpicke, 2012; Hays, Kornell, & Bjork, 2013; Huelser & Metcalfe, 2012; Knight, Ball, Brewer, DeWitt, & Marsh, 2012; Kornell, 2014; Potts & Shanks, 2014; Richland, Kornell, & Kao, 2009; Vaughn & Rawson, 2012). Similarly, increasing the number of unsuccessful retrieval attempts seems to increase the learning that occurs when the answer is presented (Arnold & McDermott, 2013; Izawa, 1970; Karpicke, 2009; Karpicke & Roediger, 2007).

[1] There is evidence that feedback can matter following correct responses on a multiple-choice test, particularly for low-confidence answers (Butler, Karpicke, & Roediger, 2008). This effect may be due, at least in part, to learning that occurs on nominally correct responses made when the participant was guessing and did not know (and was not confident about) the correct answer.

The Present Experiments

The similarities and differences between the six experiments presented in this article are outlined in Table 2. Experiments 1, 2, and 3a–3b were replicated, using different materials and a slightly different procedure, in Experiments 4, 5, and 6, respectively. Experiments 1, 2, 4, and 5 are not especially interesting on their own. They were conducted to test assumptions that were important in interpreting the three key studies, Experiments 3a, 3b, and 6. Experiments 4, 5, and 6 were conducted for two reasons; first, they provided a conceptual replication of the first three studies, and second, the assumption tested in Experiment 2 was not strongly supported by the evidence.

In Experiments 1 and 4, participants studied word pairs, restudied each pair in one of two conditions, and then took a final test (see Figure 3). In the *copy* condition, the correct answer was presented (e.g., *child–mother*), and participants were asked to copy the answer. In the *fragment* condition, the answer was presented with letters missing (e.g., *child–m_th_r* in Experiment 1, *fabric–text__e* in Experiment 4), and participants were asked to type in the answer. These experiments tested the assumption that the fragment condition produces more learning than does the copy condition. Jacoby (1978) first demonstrated the effectiveness of such a condition (see Kornell, Bjork, & Garcia, 2011, for more recent evidence).

In Experiments 2 and 5, participants completed fragment trials identical to those in Experiments 1 and 4 (see Figure 3). Performance on these trials (during the study phase) was compared in two conditions: In one, participants began by completing an initial study phase; in the other, they did not. These experiments tested the assumption that the cognitive processes at work during the fragment trials included episodic retrieval. This assumption was deemed important because our focus in this article is on retrieval of information studied previously, rather than semantic generation of targets based on preexisting knowledge. To the extent that episodic retrieval is at play, fragment completion performance will be greater after initial study versus no initial study.

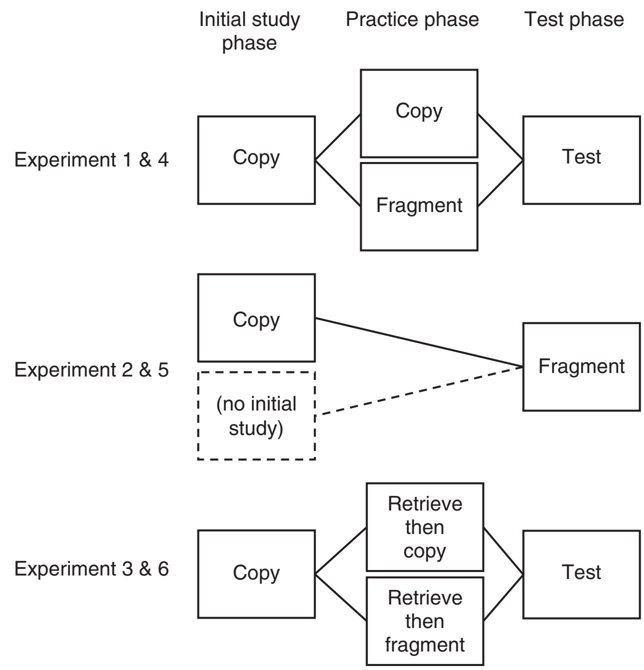

Figure 3. Procedure in Experiments 1–6.

Experiments 3a, 3b, and 6 tested the main question; namely, given that one makes an attempt at retrieval, does it matter whether one retrieves the answer successfully or is provided with the answer via feedback? The procedure was designed such that participants always made retrieval attempts, but their retrieval success was manipulated (see Figure 3). During the practice phase, each trial consisted of two parts. First the participant attempted to retrieve the target given the cue alone. Immediately after this retrieval attempt, they were presented with either a copy trial or a fragment trial. Thus, in the retrieve + copy condition, a retrieval

Table 2
Differences Between Experiments 1–6

Exp.	Retention interval	Materials	Trial timing	Fragment type	Number of items	Phase 1 trials	Experimental question	Answer
1	2 min	Target length 4 + letters	4 s per trial	Consonants only *desk–ph_n_*	60	1	Does the fragment condition produce more learning than the copy condition?	yes
2	same	same	same	same	same	same	Does the fragment condition involve episodic retrieval?	weak yes
3a	2 min or 24 hr	same	same	same	same	same	Does retrieval success matter?	no
3b	2 min	same	6 s per trial	same	same	same	Does retrieval success matter?	weak yes
4	same	Target length 6 + letters	same	First 4–5 and 1 last letter *wine–vine__r*	40	2	Does the fragment condition produce more learning than the copy condition?	yes
5	same	same	same	same	same	same	Does the fragment condition involve episodic retrieval?	yes
6	same	same	same	same	same	same	Does retrieval success matter?	yes, in the wrong direction

Note. Exp. = Experiment.

attempt was followed by a presentation of the correct answer, and in the retrieve + fragment condition, the retrieval attempt was followed by a presentation of the fragment (e.g., *child–m_th_r*). Because participants attempted retrieval in all cases, it was possible to manipulate how they made the transition from Stage 1 to Stage 2. In the retrieve + copy condition, they were shown the answer; in the retrieve + fragment condition, they successfully retrieved the answer. This procedure eliminated item-selection effects.

If retrieval success matters, the retrieve + fragment condition should produce more recall than the retrieve + copy condition does. But both conditions involved a retrieval attempt, so if the retrieval attempt is what matters, then the two conditions should produce equivalent learning.

Experiment 1

The goal of the first experiment was to establish the advantage of the fragment condition over the copy condition.

Method

Participants. Twenty participants took part in the experiment (14 females, six males; $M_{\text{age}} = 28$ years; range = 18–43 years). All reported being fluent in English. Participants were recruited through Amazon's Mechanical Turk, a website where people can sign up to complete jobs online. The experiment took about 20 min, and participants were paid $2.00.

Materials. The materials were 60 paired associates taken from norms established by Nelson, McEvoy, and Schreiber (1998). All items had forward association strength values between .030 and .031. That is, in a free-association task, when people were given the first word in the pair, they produced the second word on 3% of trials. All targets contained two to four vowels, and none began with a vowel. Cues and targets were all at least four letters long. Sample pairs include *butterfly–color*, *soccer–field*, and *sharp–razor*. The full list of pairs appears in Table A1.

Procedure. The experiment took place online. After reading the instructions, participants completed four phases: initial study, practice, delay, and test (see Figure 3).

During the initial study phase, each word pair was presented for 4 s. Assignment of word pairs to conditions, and the order in which items were presented, was randomized anew for each participant. Below the pair, the cue was presented again with an empty box next to it. Participants were told they had 4 s to type the target into the box. They were also told that their goal was to learn the word pairs for a later test.

During the practice phase, 30 items were presented in the copy condition, and the other 30 were presented in the fragment condition. Copy trials were the same as during the initial study phase. In the fragment condition, the target was presented with the vowels missing (e.g., *fence–ch_ _n*). Participants were asked to generate the correct target (*chain*), based on their memory of the initial study phase.

The practice phase was followed by a 2-min distractor task during which participants were instructed to list as many countries as they could. The final test on all 60 pairs followed. Participants were presented with each cue and asked to recall its target. There was no time limit during the final test.

After the final test, participants were asked the following question:

> During the study phase, you encountered two types of trials. Sometimes you were shown the complete answer (e.g., DESK-PHONE), and other times you were shown the answer with the vowels missing (e.g., DESK-PH_N_). If you were to learn another list, and your goal was to learn as much as possible, which study method would you choose?

They were asked to choose one of two responses: "show the answer" and "show the answer without vowels."

Results and Discussion

Response accuracy was determined using a letter-match algorithm that counted responses with minor spelling errors as correct. During the practice phase, participants correctly completed the target on 98% of trials in the copy condition (SD = 3%) and 95% of the trials in the fragment condition (SD = 6%). On the final test, cued-recall accuracy was significantly higher in the fragment condition (M = .61, SD = .24) than in the copy condition (M = .43, SD = .27), $t(19) = 5.64$, $p < .0001$, $d = 1.26$. On the postexperimental questionnaire, 15/20 (75%) of the participants said they would choose to study using the fragment condition rather than the copy condition. In sum, Experiment 1 established the value of the fragment condition.

Experiment 2

In Experiment 2, we tested the assumption that the fragment task involved episodic retrieval of associations learned during the study phase. A competing possibility is that participants were able to use semantic memory to generate answers during fragment trials. In other words, when participants completed fragment trials, did they actually retrieve prior study episodes, or did they generate answers from semantic memory? This distinction is important because retrieval and generation can have different effects. Karpicke and Zaromb (2010) found that asking participants to retrieve produced more powerful effects than did asking them to generate information on the basis of semantic memory.[2] Although participants were instructed to retrieve, not generate, in the studies presented in this article, it remains possible that only generation occurred.

In Experiment 2, we compared two conditions. In the fragment condition, participants did copy trials followed by fragment trials (replicating the fragment condition from Experiment 1). In the fragment-without-copy condition, the initial copy phase was eliminated so that participants began with fragment trials. These two conditions allow equal opportunity for participants to produce answers on the basis of semantic generation, but only the fragment condition allows for episodic retrieval. We reasoned that the two conditions should only produce different results if episodic retrieval plays a role in the fragment condition.

[2] Karpicke and Zaromb's (2010) procedure differed from the ones used in the present studies: Pparticipants studied single words, restudied those words based on related cues, and then took a free-recall test. It is probable that the effects of retrieval versus generation differ depending on the procedure being used.

Method

Sixty-eight participants took part in Experiment 2. Thirty-eight participants were randomly assigned to the fragment condition (22 females, 15 males, one unreported gender; M_{age} = 35 years; SD_{age} = 11.13; range = 19–63 years; one unreported age), and 30 were assigned to the fragment-without-copy condition (18 females, 12 males; M_{age} = 37 years; SD_{age} = 11.13; range = 23–59 years). The mean ages of the participants in the two conditions did not differ significantly, $t(66) = 0.9$, $p = .37$.

In most respects, the procedure was identical to Experiment 1 (see Figure 3). Participants studied 60 items. In the fragment condition, they did an initial study phase followed by a test phase, during which fragments were presented in the same way fragments were presented in the practice phase of Experiment 1 (participants also completed a distractor task and then took a final test, but these phases are not relevant for the present purposes; the findings were consistent with Experiment 1). The fragment-without-copy condition was different in only two respects: (a) there was no initial study phase and (b) during the test phase, participants were asked to input the first answer that came to mind (rather than the word they had studied previously). Study condition was manipulated between participants, not within, so that instructions from one condition would not bleed over to the other condition.

Results

The analysis focused on responses during the test phase, when the cue and fragment were visible. If participants were generating answers on the basis of semantic knowledge, then their ability to produce the targets should not be affected by whether or not they had done copy trials. The results showed that copy trials did affect performance; however, during the practice phase, participants produced the target on 95% of trials in the fragment condition (SD = 6%) and 89% of trials in the fragment-without-copy condition (SD = 9%). This difference was significant, $t(66) = 3.47$, $p < .001$, $d = .83$. Thus, episodic retrieval processes played a role in fragment trials.

Discussion

Experiment 2 suggested that episodic retrieval contributed to performance on fragment trials. The difference was small, however: Participants in the fragment-without-copy condition produced 89% of the items without an episodic trace compared with 95% in the fragment condition. Thus, the assumption that the fragment condition involves episodic retrieval received weak support.

How weak? One might assume that participants in the fragment condition only answered 6% of their questions using episodic retrieval (i.e., 95% − 89% = 6%). This assumption seems unjustified, however, because 6% is the *minimum* percentage of items answered on the basis of episodic retrieval. The remaining 89% of items in the fragment condition could have been answered in multiple ways: based on semantic generation, episodic retrieval, or a combination of both. Given that less than 5 min elapsed between a copy trial and the subsequent consonants trial, it is reasonable to assume that episodic retrieval processes occurred on at least some of those 89% of trials. Thus, we would argue that there is meaningful, if weak, support for the assumption that fragment trials involved episodic memory retrieval (stronger evidence was obtained in Experiment 5). Furthermore, Experiment 3 did not require that every single trial involve episodic retrieval. As long as a meaningful number of trials involve episodic retrieval, one would expect retrieval effects, whatever they may be, to emerge in the data.

Experiment 3a

In Experiments 3a and 3b, participants made retrieval attempts for all items (see Figure 3). The key manipulation concerned the activity that followed immediately after the retrieval attempt. This feedback took two forms: fragment trials, which almost guarantee retrieval success, and copy trials, which provide an external presentation of the answer without retrieval success (see Figure 2). If retrieval success enhances learning, the retrieve + fragment condition should be advantageous. If retrieval success does not matter, and making a retrieval attempt is what matters, then there should be little difference between the retrieve + fragment and retrieve + copy conditions because both involve a retrieval attempt.[3]

In Experiment 3a, we also manipulated the retention interval between the studying and the final test—either 2 min or 24 hr—to allow for the possibility that effects not apparent after a short delay might emerge after a longer delay. Including the delay manipulation had the potential to falsify our prediction that there would not be an interaction between study condition and delay. The other differences between Experiments 3a and 3b are described in the introduction to Experiment 3b.

Method

Thirty-five participants took part in Experiment 3a. Seventeen participants were tested 2 min after the end of the study phase (10 females, seven males; M_{age} = 40 years; SD_{age} = 11.97; range = 25–67 years); the other 18 participants were tested 24 hr after the study phase ended (13 females, five males; M_{age} = 32 years; SD_{age} = 13.98; range = 18–59 years). The mean ages of the participants in the two conditions did not differ significantly, $t(33) = 1.73$, $p = .09$.

The procedure of Experiment 3a was identical to Experiment 1 in most respects (see Figure 3), except that the second phase of the experiment was modified. Each cue was presented alone, and participants had 4 s to attempt to retrieve the answer, immediately after which they were given feedback. The feedback trials, which also lasted 4 s, were identical to the copy and fragment trials from Experiment 1.

To ensure that the groups were equivalent, all participants were required to complete two sessions. They were paid $2.00 after completing the first session, which took about 20 min, and an additional 50¢ after completing the second session, which took about 5 min.

[3] These predictions assume that a full-fledged retrieval attempt can be made during the allotted time; if it cannot, then the additional elaborative processing that would be allowed by a consonants trial *would* be expected to enhance learning even if retrieval success per se is not beneficial.

Results and Discussion

Preliminary analyses showed that participants answered the initial retrieval attempt (during the practice phase) correctly on 31% of trials (SD = 18%). During the subsequent fragment and copy trials, respectively, participants produced the correct answer on 95% (SD = 6%) and 98% (SD = 3%) of trials (exactly matching the previous studies). When conducting the primary analyses, we excluded the 31% of trials answered correctly during the initial retrieval attempt, because it is not possible to manipulate retrieval success after a successful retrieval.

As Figure 4 shows, final recall in the retrieve + fragment condition and the retrieve + copy condition were not significantly different, $F(1, 33) = .02$, $p = .90$, $\eta_p^2 = .001$. There was also no significant interaction between feedback condition and retention interval, $F(1, 33) = .87$, $p = .36$, $\eta_p^2 = .03$. Not surprisingly, recall was significantly lower after the longer retention interval, $F(1, 33) = 20.02$, $p < .0001$, $\eta_p^2 = .38$. These findings suggest that it is the retrieval attempt, not retrieval success per se, that enhances learning. The lack of difference between conditions is particularly striking given the evidence, from Experiment 1, that the fragment condition can be beneficial. (If all items are included in the analysis, the pattern of results is the same: The effect of retention interval was significant, $F[1, 33] = 8.83$, $p < .01$, $\eta_p^2 = .21$, but there was not a significant main effect of practice condition or an interaction, Fs < 1).

One could argue that it is not fair to compare the retrieve + fragment condition with the retrieve + copy condition on the final test because the differential rate of success completing trials during the practice phase (95% vs. 98%, respectively) might have created an advantage for the copy condition. Therefore, in a follow-up analysis, we excluded items that participants did not complete correctly during the practice phase (i.e., 2% of the items in the retrieve + copy condition and 5% of the items in the retrieve + fragment condition). The pattern of results was unchanged: There was a significant effect of test delay, $F(1, 33) = 19.91$, $p < .001$, $\eta_p^2 = .38$, but no significant main effect of condition or interaction effect (Fs < 1).

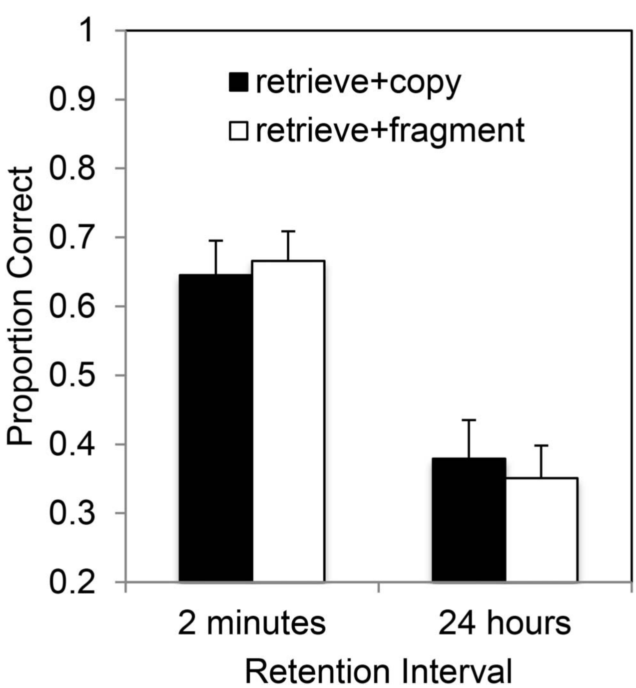

Figure 4. Proportion correct on the final test as a function of study condition and retention interval in Experiment 3a. Error bars represent one standard error of the mean.

On the postexperimental questionnaire, 28/35 (80%) of participants said they would prefer to study in the future using the retrieve + fragment condition rather than the retrieve + copy condition.

Experiment 3b

Experiment 3b was conducted for two reasons. First, because some participants might have failed to answer questions in 4 s that they could have retrieved with a little more time, we increased the time per trial to 6 s. More importantly, because Experiment 3a produced a null result with a relatively small number of participants (35), we sought to increase statistical power by attempting to replicate the result with a larger sample.

Method

Participants. Two hundred fifty-four participants (132 females, 117 males, five did not report gender; M_{age} = 34 years; range = 18–67 years) were recruited online using Amazon's Mechanical Turk and were paid $2.00 for completing Experiment 3b. All the participants were fluent English speakers living in the United States (except four who did not report fluency and three who did not provide a country of residence).

Materials, design, and procedure. Experiment 4 was nearly identical to Experiment 3. Experiment 4 took place in a single session, however, and all participants took the final test after a 2-min delay. The length of all of the trials during the study phase was also increased from 4 to 6 s (and remained 6 s for Experiments 4 through 7). The timing of the final test remained under the participant's control.

Results and Discussion

Similar to Experiment 3a, participants retrieved the correct answer on 29% (SD = 19%) of the initial test trials (during the practice phase), and these trials were excluded from further analyses. During the feedback that followed immediately, participants produced the correct answer at the same rate, 95%, on both fragment trials (SD = 11%) and copy trials (SD = 18%).

As Figure 5 shows, performance on the final test was slightly higher in the retrieve + fragment condition ($M = .62$, $SD = .22$) than in the retrieve + copy condition ($M = .60$, $SD = .24$). This small difference was significant, $t(253) = 2.26$, $p = .02$, $d = 0.09$. Like in Experiment 3a, we conducted a follow-up analysis in which we excluded trials on which participants failed to retrieve the target on the copy or fragment trial (as mentioned earlier, the rate of failure was 5% in both cases). Recall on the final test in the retrieve + fragment condition ($M = .64$, $SD = .21$) was again significantly higher than in the retrieve + copy condition ($M = .61$, $SD = .23$), $t(253) = 3.12$, $p = .002$, $d = 0.13$.

On the postexperimental questionnaire, 179/254 (70%) of participants said they would prefer to study in the future using the

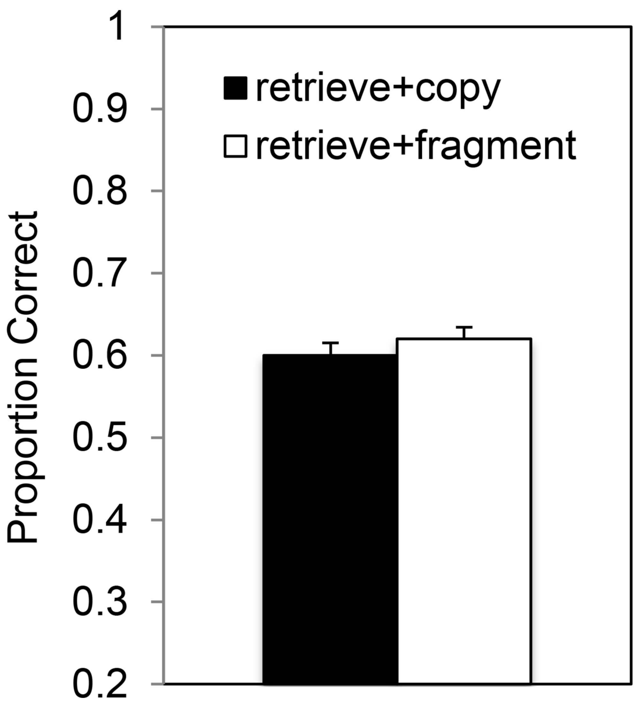

Figure 5. Proportion correct on the final test in Experiment 3b. Error bars represent one standard error of the mean.

retrieve + fragment condition rather than the retrieve + copy condition.

Introduction to Experiments 4–6

Experiment 3a suggested that as long as one makes a retrieval attempt, retrieval success does not affect learning. Using a larger sample ($n = 35$ vs. $n = 254$), Experiment 3b showed that retrieval success might have an effect, although the effect was quite small (on the order of 2 or 3 percentage points).

Experiments 4–6 were close replications of Experiments 1–3, respectively, based on recent emphasis on the importance of replicating novel findings (e.g., Ledgerwood & Sherman, 2012; Pashler & Harris, 2012; Roediger, 2012; Schmidt, 2009). Another key reason for conducting these studies was that Experiment 2 provided only weak support for the hypothesis that our fragment procedure required episodic retrieval. If it did not, it is possible that Experiments 3a and 3b did not actually test the role of episodic retrieval. Although Experiment 2 provided some evidence supporting the episodic retrieval hypothesis, it seemed prudent to run a new set of studies. After extensive pilot testing of eight paradigms and over 300 participants in 15 studies, we devised a new paradigm and set of materials that provided stronger support for the episodic memory assumption (i.e., greater fragment completion after initial copy vs. without initial copy) while also allowing for very high rates of fragment completion success necessary in Experiment 6.[4]

The method in Experiments 4–6 differed from the previous studies in four ways (see Table 2). The main change had to do with the nature of the fragment trials. In Experiments 1–3, the fragments included consonants but no vowels. In Experiments 4–6, the fragments included the first fout letters of the target (or the first five letters if the target was more than eight letters long) as well as the last letter. This change made the fragments more difficult to complete without an initial study phase. The second change was that a different set of word pairs was used because all targets had to be at least six letters long (the targets ranged from six to 12 letters long). However, the pairs came from the same database (Nelson et al., 1998), and forward association strength between the pairs was similar in Experiments 1–3 (range = .030–.031) and Experiments 4–6 (range = .01–0.48). Sample pairs include *wine–vinegar*, *policeman–respect*, and *sentence–compose*. The full list of pairs appears in Table A2. There were three other minor procedural changes, each of which slightly increased overall performance: The trial time was increased from 4 to 6 s (which we also did in Experiment 3b), the number of items was decreased from 60 to 40, and participants cycled through all items twice instead of once during initial study.

Experiment 4

A modified procedure was used in Experiment 4 to test the same assumption as Experiment 1: When not immediately preceded by a retrieval attempt, fragment trials will produce more learning than copy trials.

Method

Participants. Thirty-four participants (20 females, 14 males; $M_{\text{age}} = 34$ years; range = 20–61 years) were recruited using Amazon's Mechanical Turk and were paid \$2.00 for completing Experiment 4. All participants reported being fluent English speakers living in the United States (except for one participant who did not report a country of residence).

Materials, design, and procedure. Experiment 4 replicated Experiment 1 with five modifications to the procedure that were already described (see Table 2): type of fragment, trial timing, items used, number of items, and number of initial study trials.

Results and Discussion

During the practice phase, participants correctly produced the target on 99% of copy trials ($SD = 2\%$) and 95% of the fragment trials ($SD = 5\%$). On the final test, cued-recall accuracy was significantly higher in the retrieve + fragment condition ($M = .76$, $SD = .18$) than in the retrieve + copy condition ($M = .69$, $SD = .23$), $t(33) = 2.72$, $p = .01$, $d = 0.30$. On the postexperimental questionnaire, 24/34 (71%) of the participants said they would choose to study using the retrieve + fragment condition rather than the retrieve + copy condition.

[4] It is important to note that conducting pilot studies until we got the outcome we wanted in Experiment 6 would have constituted "p-hacking" and would have compromised the validity of our findings. Instead, we conducted pilot studies until we found a paradigm that satisfied the conditions necessary to run Experiment 6 and then ran the experiment.

Experiment 5

Method

Participants. Sixty-eight participants who were recruited using Amazon's Mechanical Turk completed Experiment 5. All participants reported being fluent English speakers living in the United States. Thirty-one participants were randomly assigned to the fragments condition (19 females, 12 males; M_{age} = 32 years; SD_{age} = 10.57; range = 19–59 years) and were paid $2.00. Thirty-seven were assigned to the fragments-without-copy condition (43 females, 24 males, one unreported gender; M_{age} = 38 years; SD_{age} = 12.71; range = 20–73 years; one unreported age) and were paid $1.00 because the experiment took less time in this condition. The mean ages of the participants in the two conditions did not differ significantly, $t(65) = 1.39$, $p = .17$.

Materials, design, and procedure. Experiment 5 was nearly identical to Experiment 2, but with the methodological changes outlined earlier (see Table 2). Like Experiment 2, study condition was manipulated between participants.

Results and Discussion

The analysis focused on responses during the test phase, when participants were shown the cue and the target fragment. Copy trials did enhance performance: During the fragment trials, participants produced the target on 95% of trials in the fragments condition (SD = 4%) but only on 77% of trials in the fragments-without-copy condition (SD = 13%). This difference was significant, $t(66) = 7.89$, $p < .0001$, $d = 1.94$. Thus, episodic retrieval processes played a role in fragment trials. This finding suggests that 18% is the minimum percentage of items that were answered correctly because of episodic retrieval. It in no way suggests that 18% is the maximum, however. It seems likely that having studied a pair (e.g., *wine–vinegar*) just minutes earlier and then being presented with the prompt (e.g., *wine–vine__r*), participants frequently did engage in episodic retrieval even if they could have thought of the answer without a study phase, which suggests the amount of episodic retrieval was probably meaningfully greater than 18%. Thus, the findings appear to support the assumption that the fragment condition involved episodic retrieval. Moreover, even if episodic retrieval did not occur on every trial, to the degree that it did occur it should impact learning.

Experiment 6

Method

Participants. Fifty-nine participants (39 females, 20 males; M_{age} = 35 years; range = 19–67 years) were paid $2.00 for completing Experiment 6. All participants reported being fluent English speakers living in the United States (except for one who did not report English fluency and two who did not report country of residence).

Materials, design, and procedure. Experiment 6 was nearly identical to Experiment 3 with the exception of the methodological changes outlined earlier (see Table 2).

Results and Discussion

Preliminary analyses showed that participants correctly retrieved the target 49% of the initial test trials during the practice phase (SD = 26%). As with Experiment 3, we excluded these trials because it is not possible to manipulate retrieval success after a successful retrieval. During the subsequent fragments and copy trials, respectively, participants produced the correct answer on 92% (SD = 9%) and 99% (SD = 2%) of trials.

As Figure 6 shows, final recall in retrieve + copy condition (M = .74, SD = .20) was significantly higher than recall in the retrieve + fragment condition (M = .66, SD = .25), $t(58) = 2.38$, $p = .02$, $d = 0.35$. This difference was in the opposite direction of the difference obtained in Experiment 3b and suggested participants learned *less* when retrieval succeeded than when feedback came from an external source.

In a follow-up analysis, we excluded the 8% of fragment trials and the 2% of copy trials that were not successfully completed. The pattern of data did not change qualitatively, with recall being numerically higher in the retrieve + copy condition (M = .74, SD = .20) than in the retrieve + fragment condition (M = .69, SD = .26). However, this difference was no longer significant, $t(58) = 1.45$, $p = .15$, $d = 0.38$.

On the postexperimental questionnaire, 39/59 (66%) of participants said they would prefer to study in the future using the retrieve + fragment condition rather than the retrieve + copy condition.

General Discussion

In the introduction, we proposed a two-stage framework in which connections related to a cue are activated during retrieval attempts in Stage 1, and then strengthened (or possibly weakened) when the answer becomes available in Stage 2. Prior research has established the benefits of both Stage 1 and Stage 2 processing, but the framework

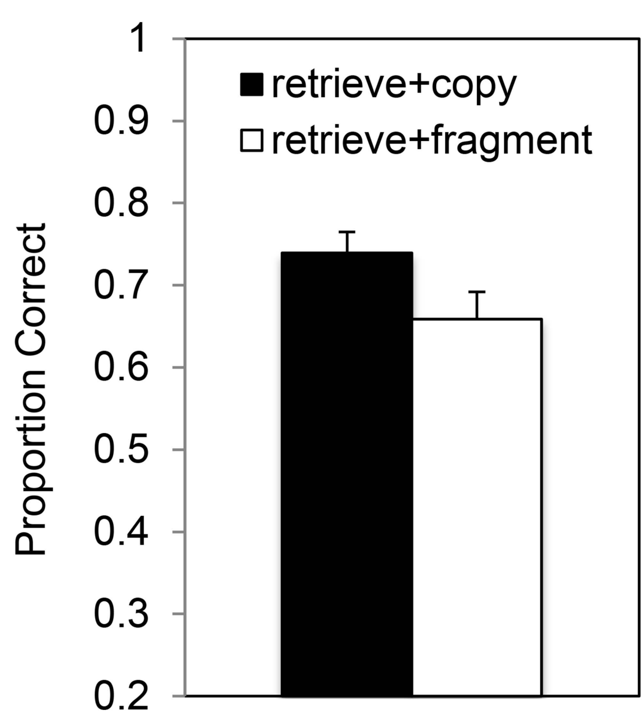

Figure 6. Proportion correct on the final test in Experiment 6. Error bars represent one standard error of the mean.

raised a novel question addressed here: Does it matter whether Stage 2 is initiated via successful retrieval or presentation of the answer?

To summarize the results, Experiments 3a and 3b suggested that retrieval success might lead to slightly more learning than unsuccessful retrieval followed by feedback. However, because of a weak (though significant) effect in Experiment 2, it was not clear whether participants in Experiments 3a and 3b engaged in episodic retrieval. In a second series of studies, this effect was stronger (in Experiment 5) and, thus, provided stronger evidence that episodic retrieval was occurring in Experiment 6. Experiment 6 suggested that retrieval success might lead to slightly *less* learning than unsuccessful retrieval followed by feedback. Together, the results of Experiments 3a, 3b, and 6 do not provide support for the hypothesis that the nature of the transition from Stage 1 to Stage 2 processing affects learning. In other words, the findings suggest that retrieval attempts enhanced learning, but retrieval success per se did not. Like life, retrieval is all about the journey.

Some theoretical explanations of testing effects distinguish between direct and indirect benefits of retrieval (e.g., Arnold & McDermott, 2013; Roediger & Karpicke, 2006b). According to these accounts, successful retrieval has a direct effect on learning, whereas an unsuccessful retrieval attempt works through a different mechanism; it does not have a direct effect, but it does potentiate subsequent learning that occurs when the answer becomes available. In the two-stage framework, by contrast, successful and unsuccessful retrieval attempts do not necessarily involve different mechanisms. During a successful retrieval, like during an unsuccessful retrieval attempt followed by feedback, a Stage 1 retrieval attempt potentiates processing during Stage 2. In other words, according to the two-stage framework, direct and indirect effects might actually be the same effects.

It is possible that future research will demonstrate that retrieval success does matter under some conditions. (Such a finding would not necessarily conflict with the two-stage framework, which is agnostic about whether or not retrieval success is beneficial.) One reason to expect that retrieval success might matter is that the emotional experience associated with retrieval success may boost learning. For example, Finn and Roediger (2011) asked participants to retrieve an answer and then showed an unrelated picture immediately after successful retrievals. They found that participants benefited more from retrieval when shown a negative emotional picture than when the picture was emotionally neutral. This finding suggests that it is possible that the emotions evoked by retrieval success could have similar effects. However, retrieval typically evokes positive emotion, and Finn, Roediger, and Rosenzweig (2012) found that positive emotional pictures did not affect learning. Further research would be necessary to investigate the effect of emotional responding to retrieval success (vs. failure) itself.

Finn and Roediger's (2011) findings have a second implication that is relevant to the present research: Their finding demonstrates that Stage 2 processing, which occurs after the answer becomes available, is important following retrieval success (and it is obviously important when feedback is given following unsuccessful retrieval). This conclusion, in turn, supports the two-stage framework itself. A subsequent study by Finn et al. (2012) provided support for the importance of Stage 1 processing by demonstrating that attempting to retrieve an answer is crucial, but "whether the target is produced by the participant or given by an external source following a commission error does not matter" (p. 1031). Although Finn et al. did not manipulate whether or not retrieval succeeded, their findings are consistent both with the two-stage framework and the hypothesis that retrieval success per se does not affect learning.

Metacognitive Judgments

People's attitudes and beliefs about learning play a crucial role in determining how they study (Bjork, Dunlosky, & Kornell, 2013; Dunlosky & Bjork, 2008). Collapsed across experiments, 295/402 (73%) of participants reported that in the future, they would prefer to study using fragment trials rather than copy trials (75%, 80%, 70%, 71%, and 66% in Experiments 1, 3a, 3b, 4, and 6, respectively; the question was not asked in Experiments 2 or 5 because the manipulation was done between participants). Previous research has similarly shown that people often choose to study via retrieval (even if they do not always think doing so enhances learning; Karpicke, 2009; Kornell & Son, 2009). The fact that consonants trials were effective in Experiments 1 and 4 but ineffective in Experiments 3 and 6 does not appear to have affected participants' judgments.

The fragment task may have been appealing because it was an engaging task that guaranteed success. For students trying to memorize information, particularly using computerized learning tools, fragment trials may be beneficial, and fragment trials following retrieval attempts might improve motivation even if they do not directly enhance learning.

When Is the Best Time to Attempt Retrieval?

We have concluded that retrieval success versus retrieval failure does not matter for items that are *otherwise equivalent*. We are not suggesting that two different sets of items, one of which is easier to retrieve than the other, will benefit equally from retrieval. Nor are we suggesting that educational interventions that differ with respect to how often retrieval succeeds should be equally effective. Neither of these conclusions follows from our claim, because they involve comparing two (or more) different sets of items.

Moreover, the evidence suggests these conclusions are not correct. In fact, learning and knowledge are negatively correlated: People learn relatively more from studying (or retrieving) information they know relatively less well (Bjork & Bjork, 1992). Thus, retrieval success may be a sign that relatively little learning is occurring and that one should have waited longer before attempting to retrieve (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). To be clear, retrieval is clearly far better than re-presentation for well-learned items (Karpicke & Roediger, 2008). But assuming one will be given feedback, the best time to attempt retrieval may be long after an item has ceased to be retrievable.

Conclusion

The experiments reported here suggest retrieval attempts enhance learning, but retrieval success per se does not. This finding suggests that, in the two-stage framework we have outlined (see Figure 1), it is processing that occurs during the retrieval attempt (Stage 1) and after the answer becomes available (Stage 2) that matter, not how one transitions from Stage 1 to Stage 2. The findings also raise a question for future research: Is it fruitful to distinguish between direct effects of retrieval success and indirect effects of retrieval attempts, or are the benefits of retrieval all "indirect" in the sense that they derive from a retrieval attempt followed by Stage 2 processing? Finally, the findings seem to be consistent with the idea that retrieval attempts serve to activate related information (mediators) and that when the answer becomes available, links between the cue, mediators, and the target are strengthened (and, perhaps, unhelpful links are weakened).

References

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 940–945. doi:10.1037/a0029199

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in Honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64,* 417–444. doi:10.1146/annurev-psych-113011-143823

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918–928. doi:10.1037/0278-7393.34.4.918

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

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. doi:10.1037/0033-2909.132.3.354

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory* (pp. 117–137). Oxford, England: Wiley.

Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *A handbook of metamemory and memory*. Hillsdale, NJ: Psychology Press.

Finn, B., & Roediger, H. L. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later recall. *Psychological Science, 22,* 781–786. doi:10.1177/0956797611407932

Finn, B., Roediger, H. L., & Rosenzweig, E. (2012). Reconsolidation from negative emotional pictures: Is successful retrieval required? *Memory & Cognition, 40,* 1031–1045. doi:10.3758/s13421-012-0203-7

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40,* 505–513. doi:10.3758/s13421-011-0174-0

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review, 17,* 797–801. doi:10.3758/PBR.17.6.797

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 290–296.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40,* 514–527. doi:10.3758/s13421-011-0167-z

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

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667. doi:10.1016/S0022-5371(78)90393-6

Karpicke, J. (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., & Roediger, H. L. (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. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62,* 227–239. doi:10.1016/j.jml.2009.11.010

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language, 66,* 731–746. doi:10.1016/j.jml.2011.12.008

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 106–114. doi:10.1037/a0033699

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65,* 85–97. doi:10.1016/j.jml.2011.04.002

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

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17,* 493–501. doi:10.1080/09658210902832915

Ledgerwood, A., & Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science, 7,* 60–66. doi:10.1177/1745691611427304

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from http://w3.usf.edu/FreeAssociation/

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3–8. doi:10.1037/0278-7393.31.1.3

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7,* 531–536. doi:10.1177/1745691612463401

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143,* 644–667. doi:10.1037/a0033194

Pyc, M. A., & Rawson, K. A. (2010) Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335. doi:10.1126/science.1191465

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

Roediger, H. L. (2012). Psychology's woes and a partial cure: The value of replication. *APS Observer, 25,* 27–29.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. doi:10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13,* 90–100. doi:10.1037/a0015108

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review, 19,* 899–905. doi:10.3758/s13423-012-0276-0

Appendix

Word Pairs Used in Experiments 1–6

Table A1
Word Pairs Used in Experiments 1, 2, 3a, and 3b

Cue	Target	Cue	Target
ADVICE	SUGGESTION	MELT	CANDLE
BARREL	WATER	MILK	CHOCOLATE
BUTTERFLY	COLOR	MOOSE	LODGE
CHILD	MOTHER	NORM	REGULAR
CLOTH	COTTON	PATH	SIDEWALK
COIN	QUARTER	PLATE	DINNER
CURVE	CIRCLE	PRONOUN	GRAMMAR
CYLINDER	PISTON	PSYCHOLOGY	SCIENCE
DAWN	SUNSET	RAISIN	CEREAL
DECENCY	RESPECT	REPENTANCE	GUILT
DIFFICULTY	TROUBLE	REPORT	GRADE
DORM	SCHOOL	RIBBON	WINNER
FARMER	VEGETABLES	ROAD	TRAVEL
FENCE	CHAIN	ROUND	TABLE
FINGER	DIGIT	RULER	NUMBER
FORM	LETTER	SCREAM	HORROR
FRESH	ROTTEN	SEEK	SEARCH
GANG	FRIEND	SHARP	RAZOR
GOLD	DIAMOND	SICK	DOCTOR
HAZE	MORNING	SKETCH	PICTURE
HOUSE	FAMILY	SOCCER	FIELD
HURRICANE	DISASTER	SOFT	GENTLE
HURRY	SPEED	SOUP	LIQUID
IDEA	CONCEPT	STARCH	CLOTHES
INSIGHT	WISDOM	STONE	PEBBLE
IRON	CURTAIN	TOSS	SALAD
LAKE	FISHING	WHALE	LARGE
LAUNDRY	DETERGENT	WISE	KNOWLEDGE
LECTURE	LISTEN	WOOD	STOVE
LOYAL	HONEST	WORD	SPEAK

(*Appendix continues*)

Table A2
Word Pairs Used in Experiments 4, 5, and 6

Cue	Target	Cue	Target
ADVENTURE	EXPEDITION	JAR	CONTAINER
BORDER	PATROL	KILOMETER	METRIC
CARDS	BASEBALL	LEADER	COMMANDER
CHURCH	CATHEDRAL	MARBLE	STATUE
CODE	PATTERN	MISSION	ASSIGNMENT
CONTACTS	FRIENDS	MOTEL	VACANCY
DENSE	FOREST	MUSCLE	STEROIDS
DISCOVERY	INVENTION	NERVE	BRAVERY
EDITORIAL	ARTICLE	OATH	PRESIDENT
ELECTRICITY	LIGHTNING	PERCENTAGE	DECIMAL
EXCHANGE	STUDENT	POLE	TELEPHONE
EXECUTIVE	PRODUCER	POLICEMAN	RESPECT
FABRIC	TEXTURE	SENTENCE	COMPOSE
FEAST	FAMINE	SKUNK	STRIPE
FESTIVAL	PARADE	SLOGAN	CANDIDATE
FLOAT	SURFACE	SPEAKER	AUDIENCE
FLOWER	PERFUME	STAIRS	ESCALATOR
GAMBLE	CHANCE	TORCH	LIBERTY
GHOST	SUPERNATURAL	WEDDING	HONEYMOON
GROUNDS	LANDSCAPE	WINE	VINEGAR

Received August 31, 2012
Revision received July 16, 2014
Accepted July 16, 2014 ■