Journal of Memory and Language 67 (2012) 17–29

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Separate mnemonic effects of retrieval practice and elaborative encoding

Jeffrey D. Karpicke *, Megan A. Smith

Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081, United States

ARTICLE INFO

Article history:
Received 24 August 2011
revision received 7 February 2012
Available online 11 March 2012

Keywords:
Retrieval practice
Testing effect
Learning
Elaboration

ABSTRACT

Does retrieval practice produce learning because it is an especially effective way to induce elaborative encoding? Four experiments examined this question. Subjects learned word pairs across alternating study and recall periods, and once an item was recalled it was dropped from further practice, repeatedly studied, or repeatedly retrieved on repeated recall trials. In elaborative study conditions, subjects used an imagery-based keyword method (Experiments 1–2) or a verbal elaboration method (Experiment 3) to encode items during repeated study trials. On a criterial test 1 week after the initial learning phase, repeated retrieval produced better long-term retention than repeated study even under elaborative study conditions. Elaborative studying improved initial encoding when it occurred prior to the first correct recall of an item, but while repeated retrieval enhanced long-term retention, elaboration produced no measurable learning when it occurred after successful retrieval. Experiment 4 used identical item word pairs (e.g., *castle–castle*) to reduce or eliminate verbal elaboration, and robust effects of repeated retrieval were still observed with these materials. Retrieval practice likely produces learning by virtue of mechanisms other than elaboration.

Introduction

When people reconstruct the past, each act of retrieval changes memory in important ways. Practicing retrieval enhances learning and long-term retention more than does spending equivalent time repeatedly studying (Roediger & Karpicke, 2006b). This finding is counterintuitive in light of a number of conventional ideas about how learning happens. Learning is generally thought to occur during study episodes, when people encode new knowledge and experiences, and retrieval provides the opportunity to measure the learning that occurred during study episodes (Karpicke & Roediger, 2007). The fact that repeated retrieval produces learning is surprising because it represents learning that occurs even after people have carried out encoding processes that are sufficient to support successful retrieval (see Karpicke, in press; Karpicke & Grimaldi, 2012).

This article is concerned with the nature of the mnemonic effects of retrieval practice. One idea about the benefits of retrieval practice is that a retrieval event represents an especially effective elaborative encoding opportunity. This idea has been proposed in one form or another by several authors. McDaniel and Masson (1985) wrote that recall testing produced "an elaboration of an existing memory representation that increases the variability of encoded information" (p. 383). Kang (2010) described the idea that "effortful retrieval promotes the activation of more elaborative information, relative to less effortful retrieval or rereading, hence establishing more retrieval routes and increasing later retention" (p. 1009). In a recent review, Roediger and Butler (2011) summarized the general elaboration perspective: "One idea is that retrieval of information from memory leads to elaboration of the memory trace and/or the creation of additional retrieval routes, which makes it more likely that the information will be successfully retrieved again in the future" (p. 24).

Other authors have been more explicit about the meaning of elaboration and its role in retrieval practice effects.

* Corresponding author. Fax: +1 765 496 1264.
E-mail address: karpicke@purdue.edu (J.D. Karpicke).

0749-596X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2012.02.004

Pyc and Rawson (2010) proposed a ''mediator effectiveness hypothesis'' as an explanation for why tests improve learning. Their idea was that tests enhance the processing of mediating words (words that link together cue and target words in paired-associate situations). For example, when subjects study a word pair like *wingu–cloud*, they may produce a word like *bird* as a verbal elaboration or mediator to help form a link between the cue and target. (We use the terms verbal elaboration and mediator synonymously in this article.) Pyc and Rawson's idea seems most relevant to the learning that occurs from failed retrieval attempts. That is, when subjects attempt retrieval and fail, they are likely to create a new, different elaboration during a subsequent study opportunity (Pyc & Rawson, in press; see too Grimaldi & Karpicke, in press). Nevertheless, the mediator effectiveness idea may also mean that the processing of elaborations or mediators is enhanced during the retrieval process itself, and this enhancement could be considered the operative mechanism responsible for the positive effects of repeated retrieval even in the absence of restudy (referred to as direct effects of retrieval on learning; Roediger & Karpicke, 2006a).

Carpenter (2009, 2011) has been explicit about the role of elaboration during the process of retrieval. The idea is that when people attempt to recall a target (again, in paired associate situations), they produce several words related to the cue and the desired target, and the production of related words is what facilitates long-term recall of the targets. The production of related words has been referred to as elaboration, and the idea bears similarity to the one proposed by Pyc and Rawson (2010). The production of elaborating (mediating) words is thought to enhance recall of the target word, and this type of elaboration is thought to be the mechanism responsible for repeated retrieval effects because ''such elaboration seems more likely to occur during retrieval than during restudy'' (Carpenter, 2009, p. 1564).

Considered broadly, elaboration refers to the process of encoding more features or attributes to the representation of an event. Typically, the additional features are conceptualized as semantic or meaning-based, and the encoding of additional features is assumed to aid the ultimate discrimination process that occurs during retrieval. Greater elaboration during encoding is thought to produce detailed and distinctive representations, and these enriched descriptions help distinguish elaborated items from other candidate items at the time of retrieval. Consequently, elaboration may enhance memory because it increases the number of retrieval cues that are potentially effective for recovering elaborated items (see Craik, 2002; Craik & Tulving, 1975; Hunt & McDaniel, 1993; Jacoby & Craik, 1979; Lockhart & Craik, 1990; see too Nairne, 2002).

When any condition is found to enhance learning, it is often assumed that the enhancement must have occurred because that condition induced elaboration. The same type of reasoning has been applied to the effects of repeated retrieval practice. Retrieval may involve deep, elaborative processing, and therefore retrieval practice may operate just like any other elaborative study task. This view would preserve the fundamental idea that elaborative studying is the primary mechanism responsible for producing learning. In other words, practicing retrieval might produce learning not because of processes unique to the act of retrieving knowledge, per se, but because of elaborative encoding processes that happen to occur during repeated tests.

Most research on retrieval practice effects has compared the effects of repeated retrieval conditions to repeated study conditions that are matched on total exposure time (Roediger & Karpicke, 2006b). It is possible that the difference between repeated retrieval and repeated study conditions simply reflects the difference that would occur between any elaborative and nonelaborative learning conditions. This perspective leads to a straightforward prediction: If elaboration is the operative mechanism responsible for the effects of repeated retrieval, then it ought to be possible to induce elaboration directly during repeated study events and produce effects that are the same as or similar to those produced by repeated retrieval.

On the contrary, there are reasons to think that the nature of what happens during repeated retrieval is different from what happens during elaborative encoding. First, consider that the effects of retrieval practice continue to occur post-retrieval, after an item has been successfully recovered (Karpicke, 2009; Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2008). Prior encoding operations would have already established mnemonic features that are sufficient to support successful retrieval. In addition, once items have been recalled, repeated retrieval is largely successful on future recall tests; rates of intertest retention are quite high and rates of intertest forgetting are low, at least with paired-associate materials (Karpicke, 2009). Once a person can successfully retrieve an item, it is not clear that additional elaboration would be necessary to improve the discrimination problem in subsequent retrieval events.

Instead, the mnemonic benefits of retrieval may be due to processes inherent to the act of retrieval itself. Retrieval involves establishing or delimiting a set of retrieval cues (a search set) and then using those cues to discriminate what target events did or did not occur (Raaijmakers & Shiffrin, 1980, 1981; see too Nairne, 2006; Surprenant & Neath, 2009). The processes involved in using retrieval cues to determine the prior occurrence of events by discriminating among candidates and selecting target responses are assumed to be unique to retrieval. That is, these are not the same processes thought to occur when people engage in elaborative encoding. This theoretical perspective suggests that elaborative study methods may not produce the same effects on long-term retention as those produced by engaging in repeated retrieval.

The four experiments reported here examined the effects of retrieval practice and elaborative encoding on learning and long-term retention. In an initial learning phase, subjects studied and recalled word pairs across a series of study and recall periods and continued until they had recalled each item. This method helps control for item-selection differences between repeated retrieval and repeated study conditions by ensuring that subjects recall all items during the learning phase (Karpicke, 2009;

Karpicke & Bauernschmidt, 2011). Once an item was recalled for the first time, it was either dropped from further practice (in drop conditions), repeatedly studied two additional times but no longer recalled (in repeated study conditions), or repeatedly recalled two additional times but no longer studied (in repeated retrieval conditions). Long-term retention was assessed on a final test 1 week after the learning phase.

Prior research has established that repeated retrieval enhances long-term retention more than does repeated studying (e.g., Karpicke & Roediger, 2008). The present experiments examined whether the effects of repeated retrieval practice could be attributed to elaboration. Experiments 1–3 compared repeated retrieval conditions to elaboration conditions in which elaboration was directly induced by having subjects use an imagery-based keyword method (Experiments 1–2) or a verbal elaboration method (Experiment 3). If the effects of repeated retrieval are due to elaboration, then engaging in elaboration during repeated study trials should enhance learning to the same extent as practicing repeated retrieval. On the contrary, if repeated retrieval and repeated elaboration were to produce different effects, this would suggest that the mnemonic effects of retrieval practice might not be attributable to elaboration. Experiment 4 used identical item word pairs (e.g., *castle–castle*) to reduce or eliminate verbal elaboration. If elaboration is responsible for repeated retrieval effects, then the effects should be eliminated when elaboration is prohibited with these materials. Alternatively, if repeated retrieval were to enhance long-term retention of identical item word pairs, then this would suggest that mechanisms other than verbal elaboration likely produced the positive effects of retrieval practice.

Experiment 1

In Experiment 1, subjects used an imagery-based elaborative study method known as the keyword method (Atkinson, 1975; McDaniel & Pressley, 1984). The keyword method involves two steps. First, a familiar English word is extracted from a novel vocabulary word (e.g., the vocabulary word *antiar* sounds like "ant"; the word *loggia* sounds like "log"). This familiar word that is acoustically similar to the vocabulary word is called the keyword. Subjects then form a mental image of a meaningful interaction between the keyword and the definition of the vocabulary word (e.g., *antiar* means *poison*, so subjects might imagine an ant drinking poison; *loggia* means *balcony*, so subjects might imagine a log leaning on a balcony). The keyword mnemonic therefore induces elaboration because it helps subjects make the novel vocabulary words less arbitrary and more meaningful, and it guides them to create enriched descriptions of the vocabulary words and their meanings.

Subjects learned vocabulary words under repeated retrieval, repeated study, or drop conditions. Once items were recalled for the first time in the learning phase, half of the subjects studied the keyword mnemonic and applied it during repetitions. The other half of the subjects did not engage in elaborative processing. If the effects of repeated retrieval are due to elaborative encoding, then inducing elaboration during repeated study trials ought to produce effects very similar if not identical to those produced by repeated retrieval practice. If repeated retrieval and repeated elaborative studying were to produce different effects, however, then this would suggest that the effects of repeated retrieval practice may not be driven by elaborative encoding.

Method

Subjects

Ninety undergraduate students from Purdue University participated for course credit.

Materials

Thirty word pairs were selected from McDaniel and Pressley (1984). Each pair included an uncommon English word and its one-word definition. Each pair also included a keyword mnemonic. For example, for the pair *antiar–poison* the keyword mnemonic was "*antiar* sounds like 'ant' and means *poison*" and for the pair *loggia–balcony* the keyword mnemonic was "loggia sounds like 'log' and means balcony". The students were instructed to form an interactive mental image of the meaning and the sound-alike word (e.g., an image of an ant drinking poison, or an image of a log resting on a balcony; see Bower, 1972). These instructions were patterned after the instructions used in previous research on the keyword method (McDaniel & Pressley, 1984, 1989).

Design

A 3 (learning condition: drop, repeated study, repeated retrieval) $\times$ 2 (elaboration condition: elaboration vs. no elaboration) between-subjects design was used. There were six conditions and 15 subjects were assigned to each condition.

Procedure

Subjects were tested in groups of one to four people. The procedure was similar to the procedure used in previous multitrial retrieval practice experiments (Karpicke, 2009; Karpicke & Roediger, 2008). In Session 1, subjects were told that they would learn a list of English word pairs across a series of study and recall periods. Each study period consisted of several study trials and an individual pair was presented on each trial. During study trials, subjects saw an uncommon English word and its one-word definition below it on a computer screen (e.g., *antiar–poison*, *loggia–balcony*). Each study trial lasted 7 s with a 500-ms intertrial interval. The subjects were told to study the pair so that they could recall the definition when given the uncommon English word. After every study period the subjects performed a 30-s distracter task that involved verifying multiplication problems.

Each recall period consisted of several recall trials and an individual pair was tested in each trial. During recall trials, subjects saw an uncommon English word and a cursor below it. They were told to recall and type the definition for each word (e.g., recall *poison* when given *antiar* and recall *balcony* when given *loggia*). Each recall trial lasted 7 s (with a 500-ms intertrial interval) after which the computer program automatically advanced to the next trial regardless of whether the subject had entered a response.

Following the recall period, the subjects studied the list again in another study period, then were tested on the list again in another recall period, and so on until they had correctly recalled all 30 pairs at least one time in the learning phase. The program terminated after a total of six study/recall periods regardless of whether subjects had recalled the entire list. Most subjects were able to recall all 30 pairs for the first time within the first four recall periods, and therefore they would have experienced two post-retrieval repetitions of all items by the end of six study/recall periods. Two initially-tested subjects in Experiment 1 did not meet this criterion and were excluded and replaced. The order of trials within each period was randomized by the computer.

Subjects in the three learning conditions were treated identically up to the point at which they correctly recalled a pair for the first time. In the drop condition, once a pair was correctly recalled it was removed from further study and recall periods. In the repeated study condition, once a pair was recalled it was removed from further recall periods but presented in two subsequent study periods. In the repeated retrieval condition, once a pair was recalled it was removed from further study periods but presented in two subsequent recall periods. The subjects were informed about the nature of their particular condition at the beginning of the experiment. They were also told they would take a final test over the pairs 1 week after Session 1.

The other critical factor in the experiment was whether subjects were given an elaborative keyword mnemonic following the first correct recall of each pair. Half of the subjects did not study a keyword mnemonic; these subjects were treated identically to subjects in previous multitrial learning experiments (e.g., Karpicke, 2009). The other half studied a keyword mnemonic following the first successful recall of each pair. Subjects in the elaboration conditions were told that after the first time they recalled each definition they would be given an imagery mnemonic to help them remember the definition. They were also told that the mnemonic instructions for each item would be shown on the computer screen. For example, immediately after the first time subjects correctly recalled *poison* as the definition for *antiar*, subjects saw the phrase "*antiar* sounds like 'ant' and means *poison*" on the screen and were instructed to form a mental image of the meaning and the sound-alike word. The keyword mnemonic was shown on the screen for 10 s. The subjects were instructed to think of the keyword mnemonic during all repetitions of that item (all subsequent restudy and recall trials). Of course, in the drop condition the items were not repeated after the first correct recall, so there were no opportunities to repeatedly think of the keyword mnemonic in that condition.

Subjects returned to the laboratory for a final test 1 week after Session 1. On the final test in Session 2, subjects were shown each uncommon English word with a cursor below it and were told to type the definition for each word. Each final test trial lasted 15 s (with a 500-ms intertrial interval). At the end of the experiment the subjects were debriefed and thanked for their participation.

Results and discussion

All results unless otherwise stated were significant at the .05 level.

Learning performance

Fig. 1 shows cumulative learning performance during the learning phase, which is the cumulative proportion of pairs recalled at least once as measured in each recall period. This measure holds the different learning conditions (drop, repeated study, and repeated retrieval) to the same performance criterion (Karpicke, 2009; Karpicke & Roediger, 2007, 2008). The figure shows that there were no differences in initial learning performance across the conditions, which was expected because all conditions were treated identically up to the point when they recalled items for the first time. A 3 (learning condition) × 2 (elaboration condition) × 3 (period) ANOVA was performed on the data from the first three test periods (because performance was essentially at ceiling in periods 4–6). There was a main effect of period ($F(2,168) = 999.18$, $\eta_p^2 = .92$), which reflects the fact that recall increased across periods. There was not a significant main effect of learning condition ($F(2,84) = 2.22$, $p = .12$, n.s.) or elaboration condition ($F(1,84) = 1.02$, n.s.). None of the interactions reached significance (for period × learning condition, $F(4,168) = 1.21$, n.s.; for period × elaboration condition, $F(2, 168) = 1.63$, n.s.; the learning × elaboration and the 3-way interaction terms yielded $Fs < 1$).

It is critical that subjects in the repeated retrieval conditions successfully recall items across repeated tests. If they did not, this would represent a failure to manipulate repeated retrieval. In the repeated retrieval conditions (collapsed across elaboration condition), once an item was correctly recalled on trial *n*, the probability of recalling

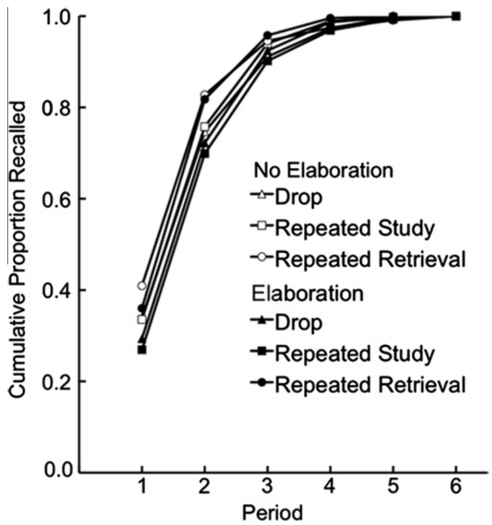

Fig. 1. Cumulative learning in Experiment 1.

it on the next trial ($n + 1$) was .94 and the probability of recall on the trial after that ($n + 2$) was .96. Therefore, once an item was recalled there was very little intertest forgetting on repeated tests (Tulving, 1964). A more detailed picture of intertest retention is presented in Appendix A.

Final recall

The critical results of Experiment 1 are shown in Fig. 2, which shows performance on the final recall test 1 week after initial learning. There was a large effect of repeated retrieval practice on long-term retention, but there was no effect of elaborative studying with the keyword mnemonic. A 3 (learning condition) × 2 (elaboration condition) ANOVA confirmed that there was a main effect of learning condition ($F(2,84) = 27.31$, $\eta_p^2 = .39$), but there was no main effect of elaboration condition and no interaction (both Fs < 1).

To examine differences among the three learning conditions, pairwise comparisons were conducted collapsed across elaboration condition. There was an advantage of the repeated studying ($M = .52$) relative to dropping items ($M = .37$; $F(1,58) = 7.66$, $\eta_p^2 = .12$). The repeated retrieval condition ($M = .73$) outperformed both the repeated study condition ($F(1,58) = 21.94$, $\eta_p^2 = .27$) and the drop condition ($F(1,58) = 62.16$, $\eta_p^2 = .52$). Repeated retrieval enhanced long-term retention relative to repeatedly studying or dropping items.

Additional pairwise comparisons were conducted to examine any possible differences between the no-elaboration and elaboration conditions within each learning condition. No significant differences were found in the drop condition (.38 vs. .35, $F < 1$), the repeated study condition (.51 vs. .52, $F < 1$), or the repeated retrieval condition (.71 vs. .76; $F(1,28) = 1.32$, n.s.). The most important result was that repeated retrieval enhanced long-term retention to a greater extent than did repeatedly studying with the elaborative keyword method.

The results of Experiment 1 cast initial doubt on the idea that the mnemonic effects of retrieval practice originate from elaboration. Once an item had been recalled, repeated retrieval practice enhanced long-term retention more than did repeated studying, and this advantage occurred even when subjects engaged in repeated elaborative studying with the keyword method. In other words, once subjects had encoded the features that were sufficient to support successful retrieval, additional encoding operations—even under enriched, elaborative conditions—did not produce as much learning as repeated retrieval practice.

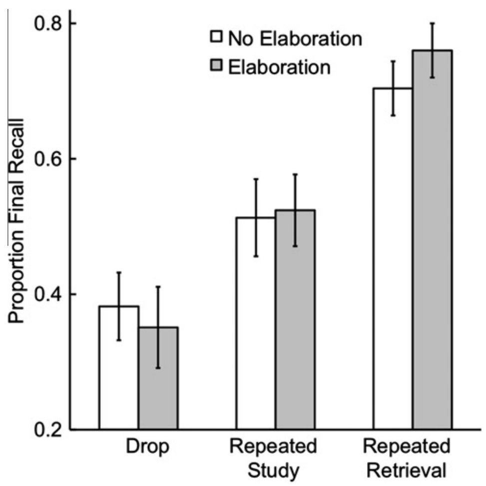

Fig. 2. Final recall in Experiment 1. Error bars represent standard errors.

Experiment 2

Experiment 2 was carried out with two purposes in mind. The results of Experiment 1 might merely reflect a failure of the keyword mnemonic manipulation. Although it is doubtful that this was the case given the large literature demonstrating the effectiveness of the keyword method for improving the initial encoding of vocabulary words (see McDaniel & Callender, 2008), Experiment 2 was conducted to eliminate this explanation. In Experiment 2, subjects in the elaboration conditions were given the keyword mnemonic during every study trial, not just immediately after they recalled each item. Elaborative study with the keyword method should enhance the initial encoding of items, prior to the first time items are successfully recalled, and thus elaborative studying should improve performance during the initial learning phase. But if repeated retrieval were to enhance long-term retention more than elaborative studying, this would replicate the critical result of Experiment 1 and support the idea that the effects of repeated retrieval practice may not be due to elaborative encoding.

A second purpose of Experiment 2 was to examine free vs. forced responding on the final test. It may be the case that subjects who repeatedly studied items were less confident and more likely to withhold responses on the final test than were subjects who repeatedly retrieved items in the initial learning phase. In Experiment 2, subjects were required to produce a response to every item on the final test (forced responding) and then to indicate whether they wished to volunteer or withhold their response (free responding). This procedure allowed us to determine how often subjects could express correct answers when forced to produce responses and how often they withheld correct answers under free report conditions (for related procedures, see Koriat & Goldsmith, 1994, 1996).

Method

Subjects

Ninety undergraduate students from Purdue University participated for course credit. None had participated in Experiment 1.

Materials, design, and procedure

The materials and design were identical to those used in Experiment 1 and the procedure was identical with two exceptions. First, subjects in the elaboration conditions were shown the keyword mnemonic in every study period. Second, subjects took a combined forced/free report final test 1 week after the learning phase. Subjects were required to produce a response for every item, even if they had to guess, and were given an unlimited amount of time to do so. After producing a response for each item, subjects then indicated whether they wanted to keep or omit their response by pressing the 1 key if they wanted to keep their response and the 0 key if they wanted to omit it. They were instructed to keep as many correct responses as possible and to omit as many incorrect responses as possible (Koriat & Goldsmith, 1996).

Results and discussion

Learning performance

Fig. 3 shows cumulative learning in Experiment 2. In the repeated retrieval conditions (collapsed across elaboration condition), once an item was correctly recalled on trial n, the probability of recalling it on the next trial ($n+1$) was .93, and the probability of recall on the trial after that ($n+2$) was .95. A detailed picture of intertest forgetting in Experiment 2 is presented in Appendix A.

The important result depicted in Fig. 3 is a clear effect of elaborative studying on initial learning. The data from the first three periods were entered into a 3 (learning condition) × 2 (elaboration condition) × 3 (period) ANOVA. There was a main effect of period ($F(2,168) = 956.37$, $\eta_p^2 = .92$) and no significant effect of learning condition ($F < 1$). Most importantly, there was a main effect of elaboration condition ($F(1,84) = 14.77$, $\eta_p^2 = .15$). Subjects who were given the keyword mnemonic during study trials performed better during the learning phase than subjects who were not given the mnemonic. There was a period × elaboration condition interaction ($F(2,168) = 4.10$, $\eta_p^2 = .05$) which indicates that the size of the difference between the elaboration and no elaboration conditions differed across trials. Most likely this occurred because performance approached ceiling, and thus the difference necessarily decreased across trials. None of the other interactions reached significance (all $Fs < 1$).

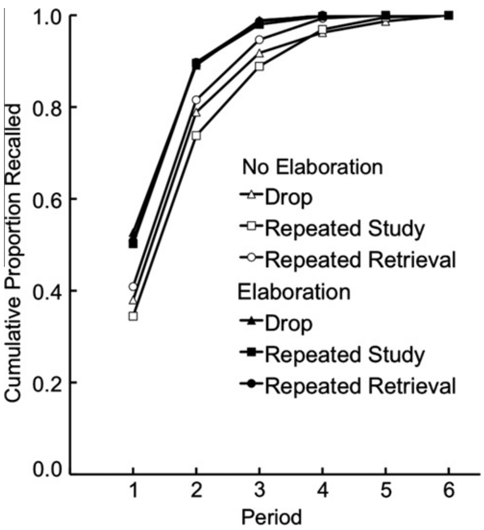

Fig. 3. Cumulative learning in Experiment 2.

Final recall

Fig. 4 shows performance on the forced recall portion of the final recall test. The pattern of results was very similar to the pattern in Experiment 1. A 3 (learning condition) × 2 (elaboration condition) ANOVA indicated that there was a main effect of learning condition ($F(2,84) = 21.52$, $\eta_p^2 = .34$) but no main effect of elaboration condition and no interaction ($Fs < 1$). Pairwise comparisons, collapsed across elaboration conditions, confirmed that there was an advantage of repeatedly studying items relative to dropping items ($Ms = .52$ vs. .34; $F(1,58) = 10.76$, $\eta_p^2 = .16$). There was a significant advantage of repeated retrieval ($M = .67$) relative to repeated studying ($F(1,58) = 8.55$, $\eta_p^2 = .13$) and dropping items ($F(1,58) = 63.49$, $\eta_p^2 = .52$). There was no significant difference between the no elaboration and elaboration conditions in the drop condition (.35 vs. .34, $F < 1$), or in the repeated study condition (.52 vs. .51, $F < 1$), or in the repeated retrieval condition (.66 vs. .68, $F < 1$). Repeated retrieval practice produced greater long-term retention than repeated elaborative studying.

Table 1 shows the analysis of performance under free report conditions. The leftmost column of Table 1 shows the proportion of correct responses that subjects volunteered under free report conditions (cf. to Fig. 2). In general, report option did not impact the outcome of the experiment. An analysis of the data in Table 1 is reported in Appendix B.

The results of Experiment 2 replicated the results of Experiment 1, casting additional doubt on the idea that the mnemonic effects of repeated retrieval stem from elaborative encoding. Once an item was successfully retrieved during a recall trial, further elaborative studying did not improve long-term retention as much as did practicing retrieval. Elaborative studying with the keyword mnemonic improved initial encoding prior to successful recall, indicating that the keyword method did indeed produce its intended effect, but practicing retrieval enhanced long-term retention more than did elaborative encoding with the keyword mnemonic.

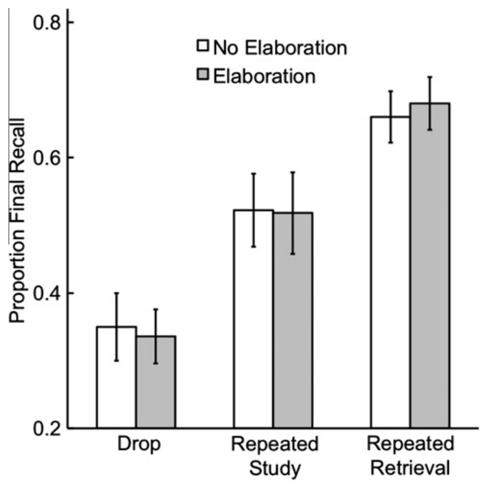

Fig. 4. Final recall in Experiment 2. Note that these data represent recall under forced report conditions. Error bars represent standard errors.

Table 1
Free report performance on the final test in Experiment 2.

	Free report performance			
	Correct volunteered	Correct withheld	Incorrect volunteered	Incorrect withheld
No elaboration				
Drop	.312	.007	.295	.386
Study	.516	.007	.171	.306
Test	.653	.007	.133	.207
Elaboration				
Drop	.331	.004	.187	.478
Study	.516	.002	.142	.340
Test	.664	.016	.116	.204

Note. "Correct" and "Incorrect" refer to whether subjects' forced report responses were correct or incorrect. "Volunteered" and "Withheld" refer to subjects decisions to volunteer or withhold responses under free report conditions. The data are reported to three decimal places because the proportions of correct responses that were withheld were close to zero.

Experiment 3

The purpose of Experiment 3 was to compare the effects of repeated retrieval to the effects of repeated elaboration under conditions intended to directly induce the kind of elaborative encoding thought to produce repeated retrieval effects. As noted earlier, one theory of the mnemonic effects of retrieval practice is that retrieval leads to the production of verbal elaborations or mediators during recall (Carpenter, 2009, 2011; Pyc & Rawson, 2010, in press). For example, when subjects learn foreign language items like *wingu–cloud*, during initial study trials the subjects may encode an elaboration like *bird* to help them relate *wingu* to *cloud*. When subjects are then given *wingu* as a cue to recall *cloud* on a test trial, they may produce the elaboration *bird* as part of the recovery process. The theory is that the production of *bird* as an elaboration during the recall trial is the operative mechanism responsible for the effects of repeated retrieval practice (Carpenter, 2009, 2011; Pyc & Rawson, 2010). If the production of elaborations or mediators during repeated recall trials is responsible for enhancements to long-term retention, then it ought to be possible to directly induce the production of elaborations during repeated study trials and produce effects that are the same as those produced by repeated retrieval practice.

In Experiment 3, subjects studied and recalled Swahili-English word pairs, and once items were recalled in the initial learning phase the words were dropped, repeatedly studied, or repeatedly retrieved (as in Experiments 1 and 2; see too Karpicke, 2009; Karpicke & Bauernschmidt, 2011; Karpicke & Roediger, 2008). In the repeated elaboration condition, subjects studied pairs in repeated study trials and generated words to help them relate the cue and target words (that is, they generated elaborations or mediators). Thus, the repeated elaboration condition directly induced the kind of elaboration thought to occur during repeated retrieval, and based on the idea that this type of elaboration is behind the mnemonic effects of retrieval practice, repeated elaboration should enhance long-term retention to the same extent as repeated retrieval practice.

Experiment 3 also examined an additional idea about the production of elaborations during retrieval, which is that subjects must recover elaborations prior to recall of a target (Carpenter, 2009, 2011). For example, it would be necessary for subjects to recover a mediator like *bird* when they are given *wingu* in order to recall the target word *cloud*. If this were the case, then it stands to reason that it would take less time to produce an elaboration than it would to produce a target word. We examined the response times to produce elaborations in the repeated elaboration condition and to recall targets in the repeated retrieval condition to see whether subjects would be quicker to produce elaborations than they would be to produce target words.

Method

Subjects

Eighty Purdue University undergraduates participated for course credit. None of the subjects had participated in the prior experiments.

Materials

Twenty-four Swahili–English vocabulary word pairs were selected from the norms of Nelson and Dunlosky (1994).

Design

There were four conditions: drop, repeated study, repeated elaboration, and repeated retrieval. Twenty subjects were assigned to each condition.

Procedure

The procedure was similar to the one used in Experiments 1 and 2, except that the initial learning phase was divided into two parts. In the first part, subjects in all conditions learned the list, across a series of alternating study and recall periods, to the criterion of one correct recall of each item. Each study trial lasted 5 s with a 500-ms intertrial interval, and each recall trial lasted 8 s with a 500-ms intertrial interval. In the second part, subjects cycled through the list two additional times under repeated study, repeated elaboration, or repeated retrieval conditions. (There was no second part in the drop condition.) Preliminary pilot testing showed that the level of intertest forgetting was greater in this two-part learning method than it was in the procedure used in the previous experiments. Therefore, to reduce intertest forgetting between the first and second parts, subjects were told whether they were correct or incorrect after each recall trial in the first part of the learning phase.

During the second part of the learning phase, in the repeated study condition, subjects cycled through the list two times in two study periods. In the repeated retrieval condition, subjects cycled through the list two times in two recall periods. In the repeated elaboration condition, subjects cycled through the list two times in two elaboration periods. On each elaboration trial, subjects were shown a word pair and were told to type a word that would help them relate the Swahili word and English word. Subjects were given the example that for the word pair *wingu–cloud*, they might think of a bird flying in the sky, so they might type "bird" or "sky". The repeated study, repeated retrieval, and repeated elaboration trials lasted 8 s.

As was done in the previous experiments, subjects returned to the laboratory for a final criterial test one week after the initial learning phase. The final test procedure was identical to the one used in Experiment 1.

Results and discussion

Learning performance

Fig. 5 shows cumulative learning performance in the first part of the learning phase. Data from the first four periods were entered into a 4 (condition) × 4 (period) ANOVA. This analysis indicated that there was a main effect of period ($F(3,228) = 963.82$, $\eta_p^2 = .93$), but there was no main effect of condition ($F < 1$) and no interaction ($F(9,228) = 1.10$, n.s.). In the second part of the learning phase, in the repeated retrieval condition, the proportions recalled on the first and second repeated tests were .84 and .87, respectively. In the repeated elaboration condition, the proportion of trials on which subjects produced elaborations were .80 and .89 in the first and second elaboration periods, respectively.

During the second part of the learning phase, response times were recorded in the repeated retrieval and repeated elaboration conditions as the time between the onset of the cue and the first keypress of a subject's response. In the repeated retrieval condition, mean response times to produce target words were 2394 ms and 1948 ms in the first and second test periods, respectively. In the repeated elaboration condition, mean response times to produce mediators were 3523 ms and 3055 ms in the first and second elaboration periods, respectively. A 2 (condition) × 2 (period) ANOVA indicated that there was a main effect of condition ($F(1,38) = 24.78$, $\eta_p^2 = .40$), a main effect of period ($F(1,38) = 33.81$, $\eta_p^2 = .47$), and no interaction ($F < 1$). Response times became faster across periods, and response times to produce mediators were substantially longer than the response times associated with producing the correct responses during repeated recall trials. The finding that subjects produced targets more quickly than they produced mediators is puzzling in light of the idea that it is necessary for subjects to produce mediators prior to the recovery of targets during recall trials (Carpenter, 2009, 2011).

Final recall

Fig. 6 shows final recall 1 week after the initial learning phase. An overall ANOVA revealed significant differences among our conditions ($F(3,76) = 8.21$, $\eta_p^2 = .25$). Subjects in both repeated study and repeated elaboration conditions outperformed subjects in the drop condition (the difference between the repeated study and drop conditions was marginally significant: $Ms = .45$ and .36, $F(1,38) = 3.13$, $p = .08$, $\eta_p^2 = .08$; for repeated elaboration vs. drop: $Ms = .48$ and .36, $F(1,38) = 5.43$, $\eta_p^2 = .13$). The difference between the repeated elaboration and repeated study condition was not significant (.48 vs. .45, $F < 1$). Repeated retrieval produced better long-term retention relative to dropping items (.62 vs. .36, $F(1,38) = 43.77$, $\eta_p^2 = .54$)

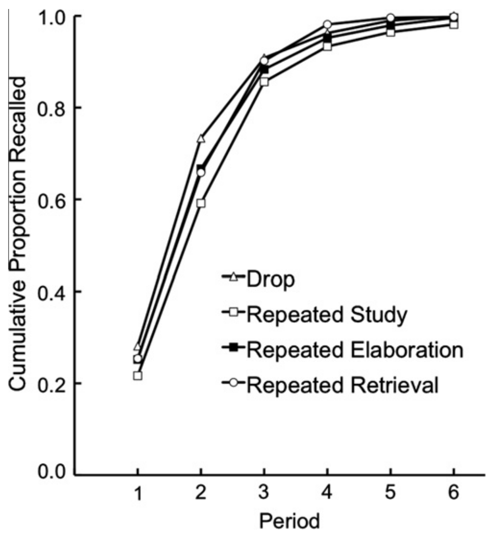

Fig. 5. Cumulative learning in Experiment 3.

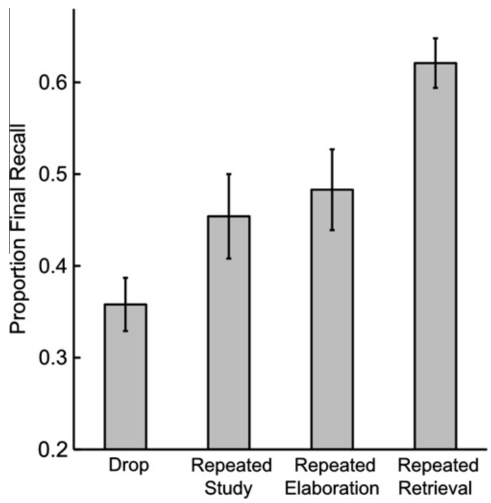

Fig. 6. Final recall in Experiment 3. Error bars represent standard errors.

and relative to repeated studying (.62 vs. .45, $F(1,38) = 9.72$, $\eta_p^2 = .20$). Most importantly, repeated retrieval enhanced long-term retention more than did repeated elaboration (.62 vs. .48, $F(1,38) = 6.76$, $\eta_p^2 = .15$).

The key result of Experiment 3 was that practicing retrieval enhanced long-term learning more than did engaging in repeated verbal elaboration. The repeated elaboration condition was designed to directly induce the type of verbal elaboration thought to underlie repeated retrieval effects. The fact that repeated retrieval was still superior to repeated elaboration under these conditions suggests that the mnemonic effects of retrieval practice may not stem from producing elaborations during recall. Furthermore, the theory that retrieval involves production of elaborations also suggests that the production of elaborations precedes recovery of target words. However, the analysis of response times did not support this theory: access to target words occurred more quickly than the production of elaborations. In agreement with the results of Experiments 1 and 2, the results of Experiment 3 do not support the idea that the mnemonic effects of retrieval practice are due to elaboration.

Experiment 4

Experiment 4 provided an additional test of the idea that the mnemonic effects of repeated retrieval may be due to semantic elaboration, specifically the generation of semantic mediators. The purpose of the experiment was to examine the effects of repeated retrieval under conditions aimed at reducing or prohibiting the production of verbal mediators that would link together cue and target words. Subjects learned a list of unrelated word pairs in which half of the pairs included different words (*mountain–hammer*) and half included identical words (*castle–castle*; see Tulving, 1974). Subjects learned the list under repeated retrieval, repeated study, or drop conditions, following the procedure used in Experiment 3. The reasoning behind this experiment was that the production of verbal elaborations relating cue and target words would be restricted or prevented when the cue and target words were identical. That is, subjects would have no need to produce additional related words to relate a cue word like *castle* to itself as the identical target word. If the mnemonic effects of retrieval depend on this kind of verbal elaboration, as advocated by some authors (Carpenter, 2009, 2011; Pyc & Rawson, 2010), then the effects of repeated retrieval are likely to be eliminated for the identical word pairs. On the contrary, if repeated retrieval enhanced retention of identical word pairs, this would provide additional evidence that the effects of retrieval practice may not depend on the production of verbal elaborations or mediators.

Method

Subjects

Sixty Purdue University undergraduates participated for course credit. None of the subjects had participated in the prior experiments.

Materials

A total of 36 medium frequency words were used. Twenty-four words were used as targets, and the other 12 words were used as cues. Two lists were constructed, and in each list, half of the targets were paired with nominally different cues (e.g., *mountain–hammer* or *dog–chair*), and the other half were paired with the nominally identical word as cue (e.g., *hammer–hammer* or *chair–chair*). In either case, the cue and target words were associatively unrelated according to normative word association data (Castel, McCabe, & Roediger, 2007; Tulving, 1974). The assignment of target words to item condition (different vs. identical cue) was counterbalanced across lists.

Design

A 3 (learning condition: drop, repeated study, repeated retrieval) $\times$ 2 (cue condition: different vs. identical) mixed factorial design was used. Learning condition was manipulated between subjects, and twenty subjects were assigned to each condition. Cue condition (different vs. identical) was varied within-list and thus was a within-subject factor.

Procedure

The procedure was identical to the two-part learning phase used in Experiment 3. Subjects were told that they would learn a list of word pairs in which some pairs contained different cue and target words whereas other pairs contained a cue word paired with itself as the target. Subjects were given examples of each type of word pair. All aspects of the procedures in both experimental sessions were identical to those in Experiment 3.

Results and discussion

Learning performance

Fig. 7 shows cumulative learning performance in the first part of the learning phase. The data were submitted

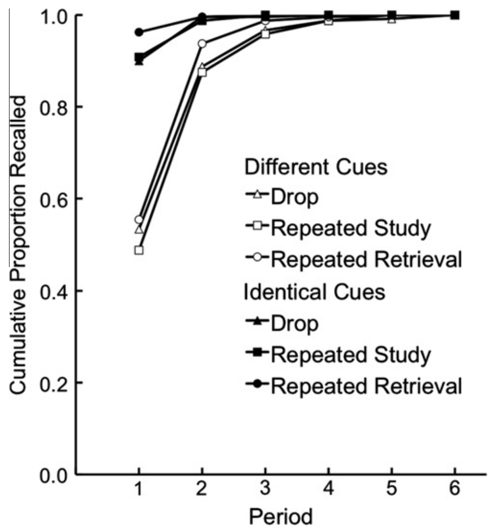

Fig. 7. Cumulative learning in Experiment 4.

to a 3 (learning condition) × 2 (cue condition) × 3 (period) ANOVA. There was no main effect of learning condition ($F < 1$), but there was a main effect of cue condition ($F(1,57) = 110.69$, $\eta_p^2 = .66$) and a main effect of period ($F(2,114) = 174.11$, $\eta_p^2 = .75$). The cue condition × period interaction was significant ($F(2,114) = 115.42$, $\eta_p^2 = .67$), but no other interactions reached significance (all Fs < 1). In the repeated retrieval condition, the proportions of repeated recall of different word pairs on the first and second repeated tests were .93 and .97, respectively. Repeated recall of identical word pairs was perfect (100%) on both repeated tests. In short, identical word pairs were learned much more quickly than were different word pairs, but there were no differences in the rate of learning among the three learning conditions.

Final recall

Fig. 8 shows final recall performance. The data were submitted to a 3 (learning condition) × 2 (cue condition) ANOVA. There was a main effect of learning condition ($F(2,57) = 11.98$, $\eta_p^2 = .30$), a main effect of cue condition ($F(1,57) = 233.50$, $\eta_p^2 = .80$), and a significant interaction ($F(2,57) = 8.00$, $\eta_p^2 = .22$). For different word pairs, repeated studying produced better long-term retention than did dropping items (.45 vs. .27, $F(1,38) = 4.19$, $\eta_p^2 = .10$), and practicing retrieval enhanced retention relative to the other two conditions (.66 vs. .45, $F(1,38) = 5.64$, $\eta_p^2 = .13$; .66 vs .27, $F(1,38) = 35.58$, $\eta_p^2 = .48$). Most importantly, repeated retrieval also enhanced long-term retention of identical items. Final recall in the repeated retrieval condition was almost perfect (99.6% correct) and was significantly better than recall in the repeated study condition (.90, $F(1,38) = 7.02$, $\eta_p^2 = .16$) and the drop condition (.92, $F(1,38) = 10.41$, $\eta_p^2 = .22$). Performance did not differ between the repeated study and drop conditions ($F < 1$).[1]

In Experiment 4, repeated retrieval enhanced long-term retention for different word pairs and for nominally identical word pairs. The identical word pairs were assumed to obviate the generation of elaborations to relate the cue and target words, because in this condition the two words were identical. Therefore, the fact that repeated retrieval enhanced long-term retention of identical word pairs is difficult to explain by recourse to the idea that subjects produce elaborations during repeated recall trials and that the production of elaborations is solely responsible for the mnemonic effects of repeated retrieval practice.

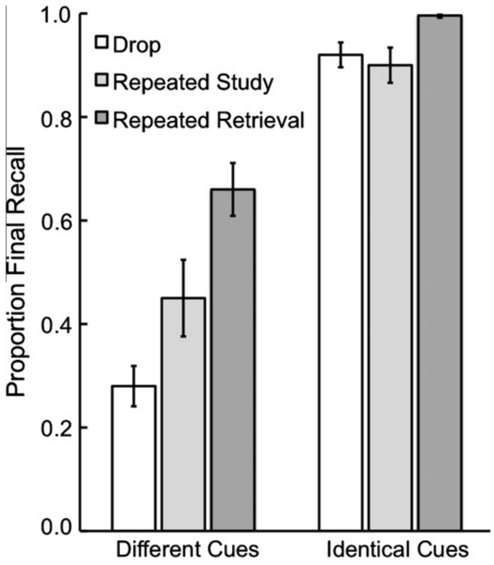

Fig. 8. Final recall in Experiment 4. Error bars represent standard errors.

General discussion

The four experiments reported in this paper examined the idea that the mnemonic effects of retrieval practice occur because of elaborative encoding during repeated tests. The rationale for Experiments 1–3 was that if repeated retrieval produces learning by virtue of elaboration, then it ought to be possible to induce elaborative encoding directly during repeated study opportunities and observe effects on long-term retention that were similar or identical to those produced by practicing retrieval. On the contrary, Experiments 1–3 showed that retrieval practice enhanced long-term retention more than did engaging in imagery-based elaboration (Experiments 1–2) or verbal elaboration (Experiment 3) during repeated study trials. Experiment 4 employed identical item word pairs to reduce or prohibit elaboration, and robust long-term effects of repeated retrieval were still observed with these materials. Taken together, the results cast doubt on the idea that the mnemonic effects of retrieval practice stem from elaborative encoding.

Elaboration is sometimes assumed to be the operative mechanism that underlies the effects of retrieval practice (Carpenter, 2009, 2011; Pyc & Rawson, 2010). It is beyond question that engaging in elaboration enhances initial encoding. Indeed, elaborative encoding produced its expected effects when elaboration occurred prior to the first time words were recalled (Experiment 2). The key issue, however, is why repeated retrieval (in two additional recall trials in the present experiments) enhances long-term retention relative to experiencing materials for the same amount of time during repeated study periods. If the ef-

[1] We also performed non-parametric tests on the identical item data because of differences in variability among the drop, repeated study, and repeated retrieval conditions. The non-parametric analyses showed the same pattern of results as the parametric analyses reported in the text. There was a main effect of learning condition ($H(2) = 12.739$, $p = .002$). Mann–Whitney U tests with a Bonferroni correction for pairwise comparisons indicated that practicing retrieval enhanced retention more than studying ($U = 107.5$, $p = .001$) and dropping ($U = 97.5$, $p = .001$). Performance did not differ between the other two conditions ($U = 194$, n.s.). Because recall of identical items was near ceiling on the final test, we also examined the number of subjects who recalled all items on the final test. In the drop condition, 9 of 20 subjects recalled all of the identical items, and in the repeated study condition, 10 of 20 subjects recalled all items. By comparison, 19 of 20 subjects recalled all identical items in the repeated retrieval condition. One subject in the repeated retrieval condition missed one identical item on the final test.

fects of retrieval practice were due to elaboration during recall, then it would be possible to induce elaboration during repeated study events and produce effects like those produced by repeated retrieval. But in the present experiments, we directly induced elaboration and did not observe effects on long-term retention commensurate with those produced by practicing retrieval. In fact, perhaps the most striking finding of Experiments 1–3 was that repeated elaborative encoding produced no effect relative to repeated studying under neutral, nonelaborative conditions. If elaboration produces no effect when it occurs after successful retrieval, then elaboration cannot be the mechanism responsible for the effects of retrieval practice on long-term retention.

A traditional way of thinking about learning is that the mind is a repository of knowledge, and the central problem facing learners is to get knowledge "in memory". The way to solve this problem is to enhance or enrich the operations that occur during encoding. The finding that the act of retrieving knowledge produces learning has been explained by deferring to this default view and assuming that retrieval must afford elaboration, but the present results suggest that retrieval produces learning by virtue of mechanisms other than elaboration. Indeed, once encoding operations have established the features necessary to support successful retrieval, the encoding of additional features may be unnecessary, and it is possible that additional encoding could harm the discrimination process required during subsequent retrieval.

It is worth wondering why elaboration, as it has been proposed with respect to retrieval practice effects, would have a positive impact on learning at all. If elaboration means the generation of several additional words related to the cue and target (as described by Carpenter (2009, 2011) and Pyc and Rawson (2010)), why would such elaborations not compete for access to the target, thereby disrupting the ability to solve the discrimination problem at retrieval? Some theories of retrieval have proposed that successful recovery of a target depends on inhibiting competitors (see Anderson & Neely, 1996), and regardless of whether that is true, it worth wondering why the generation of several elaborations or mediators would not be expected to flood the search set and ultimately reduce access to the target.

If elaboration is considered more broadly as the addition of features to a memory representation, then elaboration itself does not guarantee an improvement in the likelihood of future recovery. The effects of elaboration will depend entirely on the ultimate retrieval context, and adding features to a memory representation could render it less recoverable in a particular retrieval context. For instance, the added features could change how the search set is composed, and if a dense search set were assembled, then the likelihood of recovering the target would decrease. More importantly, if the features added to a representation were shared with other competitors in the search set, then once again the likelihood of recovering the target would decrease. If elaboration is conceived as the addition of features to the representation of an event, without considering a particular retrieval context, then it is not true that elaboration will always increase the likelihood of remembering information in the future (see Tulving, 1974).

The present results challenge the ideas that retrieval simply represents an effective encoding opportunity, and as an alternative, we suggest that practicing retrieval enhances learning by improving the diagnostic value of retrieval cues (Karpicke & Blunt, 2011; Karpicke & Zaromb, 2010). Retrieval involves a discrimination process in which a set of retrieval cues is established and the cues are used to determine the prior occurrence of a target event. The effectiveness or diagnostic value of retrieval cues for solving this discrimination problem will be a function of how well a cue specifies certain candidates (Tulving, 1974; Tulving & Thomson, 1973) to the exclusion of other competitors (see Raaijmakers & Shiffrin, 1980, 1981; Surprenant & Neath, 2009). Thus, there are a variety of ways by which practicing retrieval may enhance the diagnostic value of retrieval cues. For instance, repeated retrieval might enhance how well a cue specifies a particular candidate, or it might reduce the match between the cue and certain competitors in the search set, or repeated retrieval might constrain the size of the search set, the set of items treated as candidates in the context of a cue. Perhaps one of these components is responsible for the mnemonic effects of retrieval practice or perhaps it is some conjunction of components. This cue diagnosticity perspective identifies potential retrieval mechanisms that may enhance learning and that differ from those mechanisms thought to occur during elaboration.

The present experiments point to a clear conclusion: Repeated retrieval produces learning, but it may not produce learning by the same means thought to occur during elaborative encoding. Once the necessary features have been encoded to support recall, additional elaborative encoding does not produce the same levels of long-term retention as those produced by repeated retrieval. The learning produced by repeated retrieval should not simply be attributed to elaboration. Instead, the mnemonic effects of retrieval practice likely occur because of mechanisms that are unique to retrieval and that differ from those thought to occur during elaboration.

Acknowledgments

This research and the writing of this paper were supported in part by grants from the National Science Foundation (DRL-0941170) and the Institute of Education Sciences in the US Department of Education (R305A110903). The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education. We thank Claire Ballas, Beth Byrer, Kait Cross, and Kalif Vaughn for assistance with data collection and Althea Bauernschmidt, Phil Grimaldi, James Nairne and Roddy Roediger for comments.

Appendix A

Intertest retention in Experiments 1 and 2. The table shows the probability of recalling items in periods $n + 1$ and $n + 2$ after the items were successfully recalled in period n.

Experiment 1

	Probability of recall in periods $n + 1$ and $n + 2$				
First recalled in period n	2	3	4	5	6
1	.92	.94			
2		.96	.98		
3			.96	.99	
4				.90	.90

Experiment 2

	Probability of recall in periods $n + 1$ and $n + 2$				
First recalled in period n	2	3	4	5	6
1	.91	.93			
2		.96	.97		
3			.91	.95	
4				.96	.96

Appendix B

We analyzed both the free and forced report data from the final test in Experiment 2. The first column in Table 1 shows the proportion of correct responses that subjects volunteered under free report conditions. An ANOVA on the free report correct data showed the same pattern as the forced report data. (Due to a programming error, free report data were not collected from one subject in the drop no keyword condition, and therefore there are only 14 subjects in that condition.) A 3 (learning condition) × 2 (elaboration condition) ANOVA showed that there was a main effect of learning condition ($F(2,83) = 22.43$, $\eta_p^2 = .35$) but no main effect of elaboration condition and no interaction (Fs < 1). Pairwise comparisons (collapsed across elaboration condition) showed that there was an advantage of the study condition relative to the drop condition (.52 vs. .32, $F(1,57) = 12.74$, $\eta_p^2 = .18$). The repeated retrieval condition ($M = .66$) outperformed the study condition ($F(1,58) = 7.69$, $\eta_p^2 = .12$) and drop condition ($F(1,57) = 65.81$, $\eta_p^2 = .54$). Thus the main results under free report conditions were the same as the results under forced report conditions (cf. Table 1 to Fig. 4).

The second column in Table 1 shows the proportion of correct responses produced under forced report conditions but withheld under free report conditions. The overall proportion of correct responses produced and then withheld was very low; collapsed across all conditions, the mean was .007. A 3 (learning condition) × 2 (elaboration) ANOVA did not reveal main effects of learning condition ($F(2,83) = 1.29$, n.s.) or elaboration condition ($F < 1$) or an interaction ($F(2,83) = 1.36$, n.s.). The overall proportion of correct but withheld responses was low, and subjects were no more likely to withhold correct responses in the repeated study condition than they were in the other conditions.

The third column of Table 1 shows the proportion of incorrect responses produced under forced report conditions that were also volunteered under free report conditions. Subjects in the drop condition were more likely to volunteer incorrect responses than were subjects in the repeated study or repeated retrieval conditions. A 3 (learning condition) × 2 (elaboration condition) ANOVA showed that there was a main effect of learning condition ($F(2,83) = 3.73$, $\eta_p^2 = .08$). The main effect of elaboration condition did not reach significance ($F(2,83) = 2.09$, $p = .15$) and the interaction was not significant ($F < 1$). Collapsed across elaboration conditions, subjects in the drop condition volunteered a greater proportion of incorrect responses ($M = .24$) than did subjects in the repeated retrieval condition ($M = .12$, $F(1,57) = 6.27$, $\eta_p^2 = .10$). The difference between the drop and repeated study conditions (Ms = .12 vs. .16) did not reach significance ($F(1,57) = 2.85$, $p = .10$, $\eta_p^2 = .05$), and the difference between the repeated study and repeated retrieval conditions was not significant ($F < 1$). Thus an additional benefit of repetition (repeated retrieval and repeated study) was to reduce the proportion of incorrect responses that were mistakenly volunteered as correct.

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 237–313). San Diego, CA: Academic Press.

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30*, 821–828.

Bower, G. H. (1972). Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory* (pp. 51–88). New York: Wiley.

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.

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.

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review, 14*, 107–111.

Craik, F. I. M. (2002). Levels of processing: Past, present…and future? *Memory, 10*, 305–318.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Grimaldi, P. J., & Karpicke, J. D. (in press). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421–445.

Jacoby, L. L., & Craik, F. I. M. (1979). Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 1–21). Hillsdale, NJ: Lawrence Erlbaum Associates Inc..

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition, 38*, 1009–1017.
Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469–486.
Karpicke, J. D. (in press). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*.
Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1250–1257.
Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772–775.
Karpicke, J. D., & Grimaldi, P. J. (2012). *Retrieval-based learning: A perspective for enhancing meaningful learning*. Unpublished manuscript, Purdue University.
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.
Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.
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.
Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297–315.
Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.
Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology, 44*, 87–112.
McDaniel, M. A., & Callender, A. A. (2008). Cognition, memory, and education. In J. Byrne et al. (Eds.), *Learning and memory: A comprehensive reference*. Oxford, UK: Elsevier.
McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371–385.
McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology, 76*, 598–609.
McDaniel, M. A., & Pressley, M. (1989). Keyword and context instruction of new vocabulary meanings: Effects on text comprehension and memory. *Journal of Educational Psychology, 81*, 204–213.
Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*, 389–395.
Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46). New York: Oxford University Press.
Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili–English translation equivalents. *Memory, 2*, 325–335.
Pyc, M. A., & Rawson, K. A. (in press). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*, 335.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associate memory. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York: Academic Press.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20–27.
Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.
Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.
Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York: Psychology Press.
Tulving, E. (1964). Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychonomic Review, 71*, 219–237.
Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*, 74–82.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 359–380.