Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 6, 1787–1794

0278-7393/14/$12.00 http://dx.doi.org/10.1037/xlm0000012

RESEARCH REPORT

Toward an Episodic Context Account of Retrieval-Based Learning: Dissociating Retrieval Practice and Elaboration

Melissa Lehman, Megan A. Smith, and Jeffrey D. Karpicke
Purdue University

We tested the predictions of 2 explanations for retrieval-based learning; while the elaborative retrieval hypothesis assumes that the retrieval of studied information promotes the generation of semantically related information, which aids in later retrieval (Carpenter, 2009), the episodic context account proposed by Karpicke, Lehman, and Aue (in press) assumes that retrieval alters the representation of episodic context and improves one's ability to guide memory search on future tests. Subjects studied multiple word lists and either recalled each list (retrieval practice), did a math task (control), or generated associates for each word (elaboration) after each list. After studying the last list, all subjects recalled the list and, after a 5-min delay, recalled all lists. Analyses of correct recall, intrusions, response times, and temporal clustering dissociate retrieval practice from elaboration, supporting the episodic context account.

Keywords: memory, retrieval, testing, elaboration, proactive interference

Despite decades of research showing that practicing retrieval enhances long-term retention, there is no universally agreed-upon theory of retrieval-based learning. One idea, proposed occasionally throughout the history of retrieval practice research, has been that retrieval represents an especially effective opportunity for elaborative processing. Recent accounts have revived the idea that semantic elaborations are activated during retrieval, producing benefits on delayed criterial tests (Carpenter, 2009, 2011; Pyc & Rawson, 2010). In the present article, we evaluate the elaboration account and propose a new theory, which we refer to as the *episodic context account of retrieval-based learning* (Karpicke, Lehman, & Aue, in press). The present experiment provides a critical test to distinguish the elaboration and episodic context accounts by dissociating the effects of retrieval practice and elaboration on retention.

While some have suggested that retrieval may produce elaboration on existing memory traces (e.g., McDaniel & Masson, 1985), the exact meaning of *elaboration* has not been specified.[1] The *elaborative retrieval* account proposed by Carpenter (2009) is perhaps the only account to clearly suggest a mechanism by which elaboration occurs, which affords testable predictions. Specifically, it says that when people attempt to retrieve a target from memory, they activate several semantically related words while searching for the target, and this semantic elaboration during initial retrieval enhances retention on subsequent tests (Carpenter, 2009). Further, according to this theory, when more difficult retrieval tasks are used, target information is less readily available and a more extensive search of memory is required; this produces more semantic elaborations (e.g., when attempting to recall the target *bread* from the weakly associated cue *basket*, several words that are associated to the cue, such as *eggs* and *wicker*, would be activated). During easier tests (e.g., recalling the target *bread* from the closely associated cue *toast*) or restudy opportunities (e.g., studying the pair *basket–bread*), the target is readily available, and the generation of associated words does not occur (Carpenter, 2009, 2011). Carpenter and DeLosh (2006) argued that retrieval tasks that provide the fewest cues, such as free recall, produce the largest benefit because they allow for the most elaboration.

Support for elaborative retrieval comes from experiments in which the purported difficulty of retrieval is manipulated via the cues used to probe memory. In this context, difficult retrieval tasks are those that provide fewer cues with which to probe memory, such as free recall, or cued recall with weak cues, and these lead to greater performance on a final memory test than tasks that provide more cues, such as recognition, or cued recall with strong

---

Melissa Lehman, Megan A. Smith, and Jeffrey D. Karpicke, Department of Psychological Sciences, Purdue University.

This research and the writing of this article were supported in part by grants from the National Science Foundation (DUE-0941170 and DRL-1149363) and the Institute of Education Sciences in the U. S. Department of Education (R305A110903). The opinions expressed are those of the authors and do not represent views of the Institute or the U. S. Department of Education. We thank Pamela Hanley and Steffanie Francis for help with data collection and Jim Nairne for his helpful comments.

Correspondence concerning this article should be addressed to Melissa Lehman, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907. E-mail: lehman8@purdue.edu

[1] In fact, *elaboration* has been so vaguely defined in memory research that it could refer to adding any type of information to a memory trace or between memory traces or to memory cues. Such a broad definition is of little theoretical value, and it affords no testable predictions.

cues (Carpenter, 2009; Carpenter & DeLosh, 2006). Additional support comes from data suggesting that semantic mediators are more likely to be activated during retrieval than during restudy (Carpenter, 2011). However, the correlational nature of this evidence limits the conclusion that can be drawn; it is not clear whether the activation of semantic mediators during retrieval is a cause of retrieval-based learning, or if the mediators are merely an epiphenomenon of retrieval-based learning (cf. Underwood, 1972). For example, it may be that some mechanism other than elaboration "strengthens" the encoding of targets, and this strengthening activates related mediators. Additionally, although better learning is often attributed to elaboration, elaboration has not been directly manipulated in studies examining the elaborative retrieval hypothesis.

If semantic elaboration is the underlying mechanism responsible for the benefits of retrieval practice, then conditions that directly induce this type of elaboration should produce performance similar to retrieval practice conditions. Recent experiments comparing semantic elaboration to retrieval practice have not confirmed this prediction. Karpicke and Blunt (2011) compared repeated retrieval of texts to creating concept maps, an elaborative, organizational study strategy. Karpicke and Smith (2012) compared retrieval practice to semantic elaboration tasks that involved creating mental images or generating semantic associates. In both sets of experiments, retrieval practice consistently produced better retention than elaborative studying (see also Kang, 2010).

In addition to empirical concerns, another challenge for the elaborative retrieval account is that, according to many prominent views of memory, semantic elaboration during retrieval may create a situation of *cue overload* (Watkins & Watkins, 1975; see Surprenant & Neath, 2009). As the number of items specified by a retrieval cue (referred to as the *search set*) increases, the probability of recalling a particular target decreases and memory performance declines (Raaijmakers & Shiffrin, 1981). The elaborative retrieval account proposes that several semantically related candidate words become activated during retrieval. In principle, this type of elaboration should produce cue overload by increasing the number of candidates in the search set, which would make memory performance worse, not better. It is difficult to see how this type of semantic elaboration could be responsible for the benefits of retrieval practice.[2] On the other hand, the episodic context account of retrieval-based learning proposed here assumes that the mechanism underlying retrieval-based learning is one that alters episodic context representations in such a way that increases the effectiveness of the episodic context cue for eliciting target information.

Over the past few years, emerging results have emphasized the role of context in retrieval practice (e.g., Jang & Huber, 2008; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Sahakyan & Hendricks, 2012). This has led us to develop the episodic context account and was the impetus for the current experiment. The episodic context account was described in detail by Karpicke et al. (in press), and it is built upon a few central ideas borrowed from formal memory models. The first assumptions are that events occur within a slowly changing representation of episodic/temporal context (Mensink & Raaijmakers, 1989), and features of the context become associated with items during encoding (Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981). Next, retrieval involves using cues available in the present to determine what occurred at a particular place and time in the past. Thus, retrieval relies on reinstating a prior episodic context (e.g., Lehman & Malmberg, 2013), and when contextual information stored with items is similar to the reinstated context used as a memory cue, those items become part of the search set. When there are more items in the search set (i.e., more items that match the cue), recall of any given target is less likely (Raaijmakers & Shiffrin, 1981; Watkins & Watkins, 1975; Wixted & Rohrer, 1993). Thus, recall is determined by the ability of the cue to uniquely specify the target, to the exclusion of other items (Nairne, 2002).

The episodic context account of retrieval-based learning assumes that the context reinstatement that occurs during retrieval creates a unique set of context features that become associated with successfully retrieved items; when an item that was studied in a past context is retrieved in the present context, the context representation is updated so that it includes a composite of features from both contexts. On a later test, context is again reinstated to accomplish retrieval, and because items that were previously retrieved are associated with multiple contexts, reinstatement of either context serves as an effective cue for those items, increasing successful retrieval of those items (Karpicke et al., in press). Thus, retrieval drives context change (Jang & Huber, 2008) and produces a set of distinct context features that are encoded with retrieved items, making these items more likely (and extraneous items less likely) to become part of the search set when context is later used to cue memory (this is similar to a suggestion made by Szpunar, McDermott, & Roediger, 2008). In short, the episodic context account suggests that the act of retrieval enhances subsequent memory by altering temporal context representations, allowing people to constrain their search and increase the likelihood of recovering a desired target. To contrast the elaborative retrieval account to the episodic context account, the former attributes the episodic phenomenon of retrieval-based learning to semantic retrieval, whereas the latter attributes this phenomenon to the episodic nature of the initial retrieval task.

A purely semantic explanation of retrieval-based learning seems unlikely, given that merely requiring subjects to think back to the study episode to accomplish the task produces a large mnemonic advantage over relying on semantic memory to complete the task (Karpicke & Zaromb, 2010). Further, there has been little support for a semantic explanation in experiments designed to directly test the elaborative retrieval hypothesis (e.g., Karpicke & Smith, 2012). However, it is possible that in these experiments, retrieval practice simply happened to produce more semantic elaboration, for whatever reason, than elaborative study conditions, even though the latter conditions were designed to induce the kind of elaboration presumed to occur during retrieval. It would be much harder to claim that elaboration is the basis of retrieval practice effects if elaboration and retrieval practice task were shown to produce opposite effects on memory performance. The intent of the present experiment was to dissociate elaboration and retrieval

---

[2] Even if one were to assume that, rather than producing more information associated with a single retrieval cue, elaboration produces more retrieval cues, those cues are not present during recall and must be retrieved using the cues that are available; in the case of free recall, the cue is temporal context present at the time of the test. Thus, the elaboration account still assumes that more information becomes associated with the temporal context cue, and the problem of cue overload remains.

practice. Whereas retrieval practice should restrict the size of the search set and reduce cue overload, as predicted by the episodic context account, elaboration should instead expand the search set and exacerbate the effects of cue overload.

The experiment used a proactive interference procedure developed by Szpunar et al. (2008), depicted in Table 1, to distinguish the effects of retrieval practice and elaboration on retention. Subjects studied a series of five word lists under one of three conditions. In a control condition, subjects studied each list and completed a brief distractor task between lists. In the retrieval practice condition, subjects recalled each list after studying it. In the elaborative study condition, subjects were shown the words and instructed to produce the first two words that came to mind that were associated with the target word. This task directly induces the type of semantic elaboration proposed to occur during retrieval practice, according to the elaborative retrieval account (Carpenter 2009, 2011). After studying the fifth list, subjects in all conditions were instructed to recall List 5. We predicted that practicing retrieval after each list in the learning phase would produce greater levels of correct recall and fewer intrusions from prior lists (Lists 1–4), relative to the control condition, on this List 5 recall test. This result would replicate the key finding from Szpunar et al. (2008) and would support the episodic context account, because retrieval practice affords a restricted search set. If elaboration is the mechanism responsible for retrieval practice effects, then directly inducing elaboration should also increase correct recall and decrease prior list intrusions, just like the retrieval practice condition. In contrast, we suspected that subjects in the elaboration condition would recall fewer correct items and produce more prior-list intrusions, because semantic elaboration would expand the size of the search set.

The episodic context account makes additional predictions about retrieval dynamics that we examined in the present experiment. We measured subjects' response times and examined cumulative recall curves, which display the cumulative number of items recalled throughout a recall period (Karpicke & Roediger, 2007; Roediger & Thorpe, 1978). Cumulative recall curves can be fit by an exponential function, $F(t) = N(1 - e^{-[t-c]/\tau})$ where $F(t)$ represents the cumulative number of items recalled by time t, N represents an estimate of asymptotic recall, the total number of items that would be recalled given unlimited time, τ represents mean response latency to recall those N items, and c represents a delay in onset of recall (Wixted & Rohrer, 1993). A shorter mean response latency (indicated by a smaller τ value) would reflect a smaller, more restricted search set, whereas a longer mean response latency (indicated by a larger τ value) would reflect a larger, expanded search set. The episodic context account predicts that retrieval practice produces a restricted search set, and this would be reflected in the analysis of cumulative recall during the List 5 test (indeed, Bäuml & Kliegl, 2013, recently analyzed cumulative recall in this task, and their data produced parameter values suggesting a restricted search set under retrieval conditions). Finally, at the end of the experiment, subjects freely recalled all words from all lists, and we examined the degree to which subjects clustered items by list during the final test. The episodic context account proposes that practicing retrieval enriches representations of temporal context that are used to guide subsequent retrieval. Thus, retrieval practice should produce greater levels of temporal clustering during final recall.

Table 1
Overview of the Procedure

Condition	Initial phase				Criterial tests	
	List 1	List 2	List 3	List 4	List 5	Recall All
Control	S–	S–	S–	S–	S R	R_{ALL}
Retrieval practice	S R	S R	S R	S R	S R	R_{ALL}
Elaboration	S E	S E	S E	S E	S R	R_{ALL}

Note. S = study period; E = elaboration task; R = free recall test. Subjects in all conditions recalled List 5 and then freely recalled all words (R_{ALL}) at the end of the experiment. Subjects in all conditions completed a math task after each study period.

Method

Subjects and Design

Subjects were 108 Purdue University undergraduates who participated in exchange for course credit. There were three experimental conditions—retrieval practice, elaboration, and a control condition—and 36 subjects were randomly assigned to each condition.

Materials

The five word lists used by Szpunar et al. (2008; Experiment 1B) were used in the experiment. Each list contained 18 unrelated, medium-frequency concrete nouns. List and word order were randomized for each subject.

Procedure

Subjects were tested in groups of four or fewer. They were told they would study multiple lists of words and complete a math task and possibly another task after each list, to be determined randomly by the computer. In reality, subjects completed the same task for the first four lists, depending on experimental condition. Subjects were instructed to study each list because, regardless of the task they completed after each list, they would be asked to recall as many of the words as they could at the end of the experiment.

During each study period, words were presented one at a time in the center of the computer screen for 2 s each, followed by a 500-ms interstimulus interval. After each study period, subjects completed a 1-min math task (solving two-digit addition problems). For Lists 1–4, subjects completed one of three tasks after the math task. In the control condition, subjects completed 1 additional minute of the math distractor task. In the retrieval practice condition, subjects completed a 1-min free-recall test, in which they were told to recall as many of the words as possible from the list they had just studied, in any order. Subjects typed their responses on the computer and pressed the "Enter" key after typing each response. Upon doing so, the response they had typed was added to a list of responses that remained displayed on the computer screen throughout the recall period. In the elaboration condition, subjects were shown the 18 list words simultaneously on the screen, with two blank spaces next to each word. They were told to type the first two words that came to mind for each word on the list and were given 1 min to generate as many elaborations as possible.[3]

For List 5, subjects in all conditions studied the list, completed the 1-min math task, and then took a 1-min free-recall test, following the procedure described previously. Subjects then engaged in another distractor task (playing a video game) for 5 min, and at the end of the experiment, all subjects completed a 5-min final free-recall test, in which they were asked to recall as many words as possible from all of the lists in the experiment.

Results

The key data were correct recall on the List 5 recall test, prior list intrusions on the List 5 test, retrieval dynamics during the List 5 test, and correct recall and temporal clustering on the final free recall test at the end of the experiment. In the retrieval practice condition, the proportion of correct recall on the tests for Lists 1–4 tests averaged .31 (about six of 18 words). Intrusions occurred rarely during initial recall ($M = 0.12$ intrusions per list).

List 5 Recall

The left side of Figure 1 shows the proportion of correct recall on the List 5 test. Retrieval practice increased correct recall relative to both the control condition, $t(70) = 3.99$, $d = 0.95$, 95% confidence interval (CI) [0.46, 1.43], and the elaboration condition, $t(70) = 8.08$, $d = 1.93$, 95% CI [1.36, 2.49]. Indeed, subjects in the elaboration condition recalled fewer correct items relative to subjects in the control condition, $t(70) = 2.52$, $d = 0.60$, 95% CI [0.13, 1.07]. The right side of Figure 1 shows the proportion of prior-list intrusions on the List 5 test.[4] Whereas retrieval practice reduced the recall of intrusions relative to the control condition, $t(70) = 5.34$, $d = 1.28$, 95% CI [0.77, 1.78], elaboration increased the recall of intrusions relative to both the control condition, $t(70) = 1.89$, $d = 0.45$, 95% CI [−0.02, 0.92], and the retrieval practice condition, $t(70) = 7.08$, $d = 1.69$, 95% CI [0.77, 1.78]. These key results support the idea that retrieval practice allowed subjects to restrict their search during recall, whereas semantic elaboration did not. Instead, elaboration expanded the size of the search set and made recall performance worse.

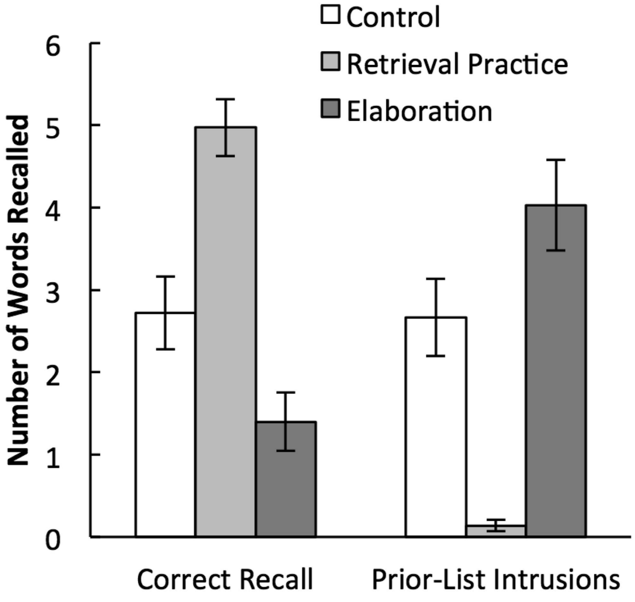

Figure 1. The left side of the figure shows the mean number of List 5 words recalled on the List 5 recall test. The right side shows the mean number of intrusions from prior lists (Lists 1–4) on the test. Error bars represent standard errors.

Table 2 shows the results of an analysis of retrieval dynamics during the List 5 test. The first column in Table 2 shows the average first recall latency, which is the time between the onset of the test and the entry of the first response. Retrieval practice reduced first recall latency relative to the control condition (a 1,275-ms decrease), $t(70) = 0.83$, $d = 0.20$, 95% CI [−0.26, 0.66], while elaboration increased first recall latency relative to both the control condition (a 1,704-ms increase), $t(70) = 2.03$, $d = 0.49$, 95% CI [0.02, 0.96], and the retrieval practice condition (a 2,979-ms increase), $t(70) = 3.37$, $d = 0.81$, 95% CI [0.33, 1.29]. These results could reflect differences in the size of the search set or differences in the time to identify the appropriate search set (Wixted & Rohrer, 1993). The analysis of cumulative recall, reported next, provides more precise estimates of search set sizes across conditions.

For each condition, parameter estimates for τ, N, and c were computed by fitting the exponential function to the cumulative recall data as measured at each 5-s bin in the 60-s recall period. The parameter estimates are presented in Table 2, and Figure 2 shows cumulative recall during the List 5 recall test, with the curves produced by the best-fitting set of parameter values for each condition. Panel A shows cumulative recall of correct items, which shows the clear advantage of retrieval practice depicted in Figure 1. Panel B shows cumulative recall of all items (correct recall and prior-list intrusions), which reflect the contents of the search set. There were small differences in the value of c across conditions, suggesting that differences in first recall latencies were not driven by the time to identify an appropriate search set. There were large differences in τ values; retrieval practice produced smaller response latencies, suggesting a smaller, restricted search set. In contrast, elaboration produced longer response latencies, suggesting a larger, expanded search set.

It is likely that subjects would have output more items if given more recall time. Figure 3 shows the number of correct recalls and number of prior-list intrusions as a function of output position in recall (output quintile, which refer to the first 20% of responses, the second 20%, and so on). The figure shows that number of correct recalls increased across output

---

[3] In the time allotted, subjects were able to enter an average of approximately one word per item.

[4] The analysis of intrusions on the List 5 test is limited to prior-list intrusions, as these are the only type of intrusion that can be compared across conditions. Extra-experimental intrusions were rare ($M < 1$ for all conditions), and semantic associates were only generated in the elaboration condition, though the intrusions rate for those semantic associates was as high as correct recall in that condition ($M = 1.45$).

Table 2
Analysis of Retrieval Dynamics During the List 5 Recall Test

Variable	First recall latency (ms)	τ	*N*	*c*
Correct items				
Control	6572 (789)	20.00 [20.83, 18.87]	2.90 [2.85, 2.95]	2.29 [2.00, 2.59]
Retrieval practice	5297 (420)	15.15 [15.87, 14.49]	5.52 [5.46, 5.58]	2.03 [1.80, 2.27]
Elaboration	8276 (821)	30.30 [33.33, 28.57]	2.50 [2.41, 2.58]	2.19 [1.77, 2.61]
All items				
Control		30.30 [32.26, 28.57]	8.16 [7.96, 8.36]	2.51 [2.21, 2.81]
Retrieval practice		15.38 [16.13, 14.71]	5.91 [5.84, 5.99]	2.21 [1.96, 2.46]
Elaboration		52.63 [55.56, 47.62]	9.74 [9.24, 10.25]	2.90 [2.54, 3.25]

Note. First recall latency refers to the mean time from the onset of the recall period until "Enter" key was pressed to submit the first word recalled. Standard errors are in parentheses. The τ parameter represents mean response latency, *N* represents the total number of items recalled given unlimited time, and *c* represents delay of recall onset. Parameter estimates for τ , *N*, and, *c* were computed based on the cumulative recall data for each condition. 95% confidence intervals for the parameter estimates are presented in brackets. Parameter values for data restricted to correct items are shown on the top; parameter values for all items, including correct items and prior-list intrusions, are shown on the bottom.

position in the retrieval practice condition, whereas number of prior-list intrusions increased in elaboration condition. Thus, if given more time, subjects in the elaboration condition would have produced more words, but those words would likely have been intrusions from prior lists.

Final Free Recall of All Lists

Figure 4 shows the mean number of words recalled on the final free recall test at the end of the experiment. Words from List 5 (the left side of the figure) and Lists 1–4 (the right side of the figure) were analyzed separately, because List 5 was previously tested in all conditions, while Lists 1–4 were treated differently according to the different conditions. Words from Lists 1–4 that were recalled as intrusions on the List 5 test, thus benefiting from retrieval practice, were excluded from the analysis. For List 5 items, subjects in the retrieval practice condition recalled more items than control subjects, $t(70) = 2.51$, $d = 0.60$, 95% CI [0.13, 1.07], and elaboration subjects, $t(70) = 5.94$, $d = 1.42$, 95% CI [0.90, 1.93]. Subjects in the elaboration condition recalled fewer List 5 items than control subjects, $t(70) = 2.53$, $d = 0.60$, 95% CI [0.13, 1.07]. For List 1–4 items, subjects in the retrieval practice condition recalled more items than control subjects, $t(70) = 6.74$, $d = 1.61$, 95% CI [1.07, 2.14], and elaboration subjects, $t(70) = 2.40$, $d = 0.57$, 95% CI [0.10, 1.04]. Subjects in the elaboration condition recalled more List 1–4 items than control subjects, $t(70) = 5.80$, $d = 1.39$, 95% CI [0.87, 1.90].

Finally, Figure 5 shows an analysis of recall transitions. Because subjects could recall the items in any order during the final test, we calculated the proportion of same-list transitions,

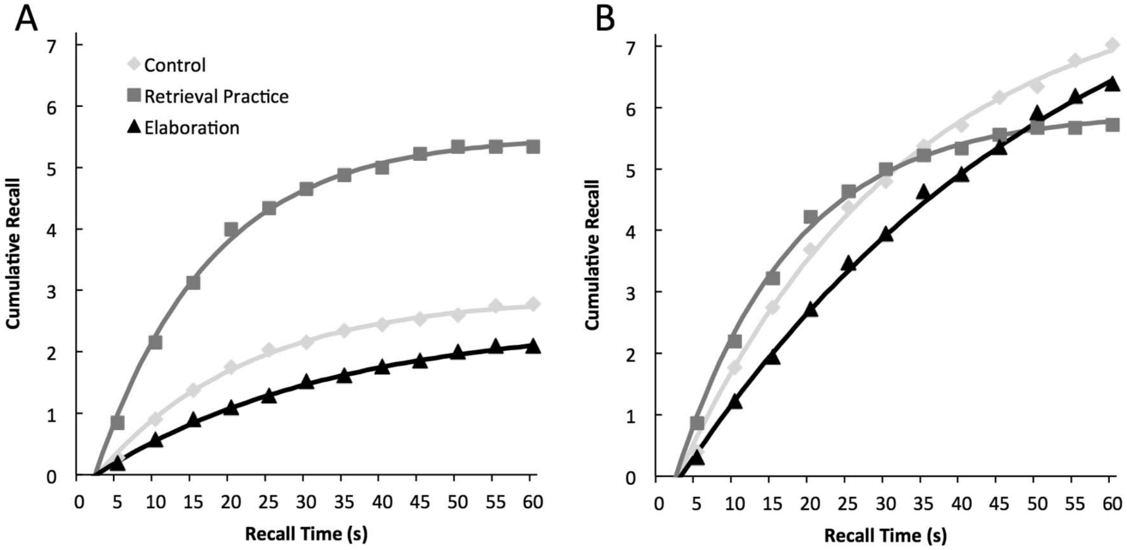

Figure 2. Cumulative recall of words on the List 5 recall test. Panel A shows cumulative recall of words from List 5 (correct recall), and Panel B shows cumulative recall of words from all lists (correct recall and prior-list intrusions). Symbols represent the observed averaged data for each condition, and solid lines represent the best fitting exponential functions (parameters listed in Table 2).

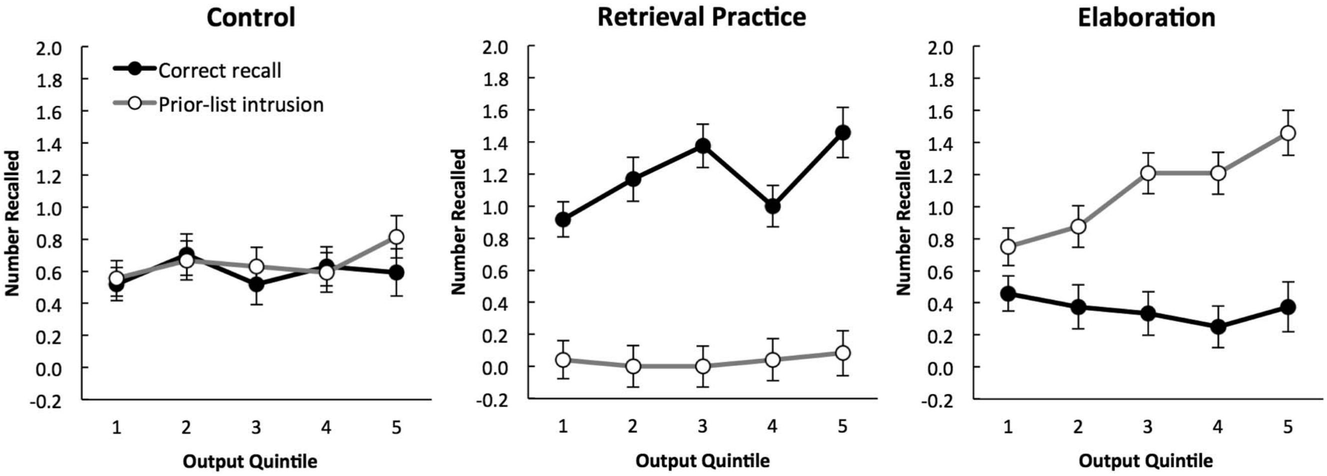

Figure 3. Mean number of correct recalls and prior-list intrusions on the List 5 recall test as a function of output position (output quintile, referring to the first 20% of responses, the second 20%, and so on). Error bars represent standard errors. As the recall period progressed, the number of correct recalls increased in the retrieval practice condition, whereas the number of prior-list intrusions increased in the elaboration condition.

which occurs when Items *n* and *n* + 1 in recall output are from the same list. Higher proportions of same-list transitions in recall output reflect greater reliance on temporal context in final recall. While subjects in the retrieval practice condition were slightly more likely to make transitions within the same list, relative to subjects in the control condition, $t(70) = 0.83$, $d = 0.20$, 95% CI [−0.26, 0.66] (see Figure 3), subjects in the elaboration condition were less likely to make transitions within the same list than subjects in the control, $t(70) = 2.03$, $d = 0.49$, 95% CI [0.02, 0.96], or retrieval practice conditions, $t(70) = 3.16$, $d = 0.75$, 95% CI [0.27, 1.23], referred to as *decreased temporal clustering*.[5]

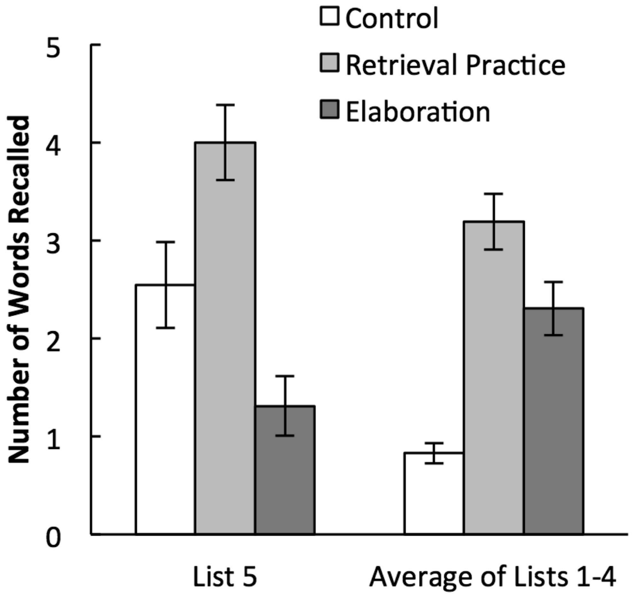

Figure 4. Mean number of words recalled on the final free recall test. Error bars represent standard errors.

Discussion

The increase in correct recall, decrease in prior-list intrusions, and decrease in response latency on the List 5 test in the retrieval practice condition replicate prior work (Bäuml & Kliegl, 2013; Szpunar et al., 2008) and are consistent with the episodic context account of retrieval-based learning. These findings support the notion that the representation of episodic/temporal context is altered during retrieval in such a way that allows people to restrict their search using updated temporal context representations, reducing cue overload. The increase in reliance on temporal context cues produced by context reinstatement during the retrieval task also produces increases in temporal clustering on the final test. In contrast, the same measures suggest an increase in the size of the search set after the elaboration task, revealing dissociation between retrieval practice and elaboration that challenges the elaborative retrieval hypothesis. In opposition to the argument that previous work (e.g., Karpicke & Smith, 2012) failed to support the elaborative retrieval hypothesis because elaboration tasks used in that work do not invoke enough semantic elaboration, we show here that episodic retrieval and semantic retrieval tasks behave quite differently. Combined with the data from Karpicke and Smith (2012), we have failed to find support for the elaborative retrieval hypothesis in both cued recall and free recall designs.

These data support the episodic context account of retrieval-based learning and also add to a growing body of work highlighting the relationship between temporal context and retrieval processes (Jang & Huber, 2008; Lehman & Malmberg, 2013; Sahakyan & Hendricks, 2012). In addition to the role of context change produced by retrieval in segregating lists and reducing interference (Pastötter et al., 2011), we emphasize the role of context reinstatement during retrieval in increasing retention of retrieved information. Indeed, we are not the first to suggest that

---

[5] This was true regardless of whether List 5 was included in the analyses; the data reported here do not include List 5.

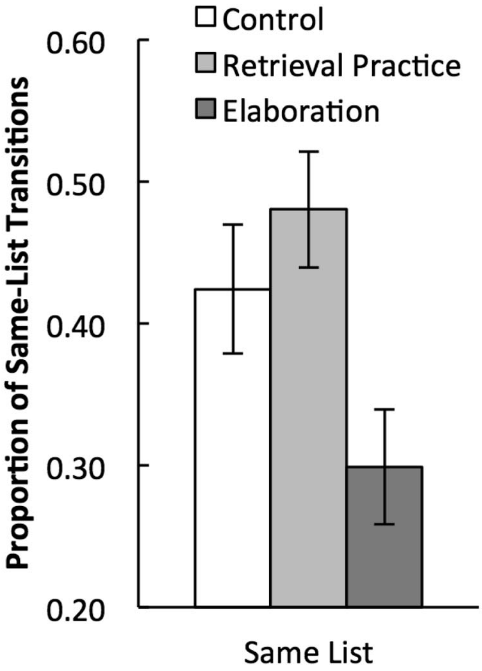

Figure 5. Proportion of same-list transitions during the final free recall test. Error bars represent standard errors. A same-list transition occurs when Items n and $n + 1$ in a subject's recall output come from the same study list. Higher proportions of same-list transitions reflect increased use of temporal context during recall (temporal clustering).

context plays a role in enhancing memory for retrieved information. Delaney, Verkoeijen, and Spirgel (2010) proposed an account of memory that essentially attributes the advantage of spaced study over massed study (e.g., Greene, 1989) to retrieval practice.

Combining study-phase retrieval accounts (Greene, 1989) and contextual encoding accounts (Malmberg & Shiffrin, 2005) of spacing effects, Delaney et al. (2010) proposed that spacing enhances learning when a spaced repetition of a studied item leads to the retrieval of the prior presentation of that item, which produces encoding of additional contextual information, resulting in a stronger link to context. Borrowing from Raaijmaker's (2003) model of spacing effects, Delaney et al. assumed that the amount of contextual strengthening that occurs during study-phase retrieval is inversely related to the contextual strength of the retrieved item; items with low-context strength (i.e., those with a low match to the context used as a cue, such as items spaced with long lags in between presentations) receive greater context strengthening. Thus, the Delaney et al. (2010) account and the Raaijmakers (2003) model upon which it is based attribute the spacing effect to contextual updating associated with retrieval and assume that contextual strengthening is the mechanism by which retrieval produces the benefits of spacing. However, neither is a model of retrieval practice, and thus they do not specify the role of context reinstatement or explicitly describe exactly what context gets strengthened when context reinstatement occurs during retrieval.

Along similar lines, in the retrieval practice literature, researchers have recently suggested that retrieval may add contextual elements to memory traces that help with list discrimination (Szpunar et al., 2008) and that after retrieval practice, subjects may remember previously retrieving an item at a specific point in time (Verkoeijen, Tabbers, & Verhage, 2011), suggesting that retrieval adds distinctive temporal features to the memory trace (see also Bauml & Kleigl, 2013). We extend these assumptions, along with those of Delaney et al. (2010), by proposing that the retrieval process produces the encoding of unique context features that are associated with the retrieval period. The more context reinstatement required to complete the task, the more distinct these features will be, and during a later retrieval attempt, subjects will be better able to utilize the unique context features to probe memory for items that were previously retrieved (Karpicke et al., in press).

The episodic context account explains the finding that retrieval practice benefits are more likely for more difficult retrieval tasks (Carpenter, 2009; Carpenter & DeLosh, 2006), because the more difficult tasks are those that require greater reliance on temporal context cues. This account also explains why benefits of retrieval practice occur for some criterial tests but not others—these effects are most evident on tasks which rely on the use of temporal context, such as free recall, source memory, and exclusion recognition tasks (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Chan & McDermott, 2007). In contrast, it would be difficult for the elaborative retrieval account to explain why if retrieval produces semantic elaboration, it would enhance list discrimination, which is based on contextual information associated with memory traces. In summary, these findings, along with the findings from the current experiment, are troublesome for the elaborative retrieval account. It is possible that other explanations may be consistent with these data; however, at this time, the episodic context account, founded on the core assumptions of contemporary memory models, is the only mechanistic account that explains a variety of findings related to retrieval-based learning, including both those that are challenging for elaborative retrieval and those that elaborative retrieval was developed to explain.

References

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68,* 39–53. doi:10.1016/j.jml.2012.07.006

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory, 18,* 385–393. doi:10.1080/09658211003702163

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

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268–276. doi:10.3758/BF03193405

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 431–437. doi:10.1037/0278-7393.33.2.431

Delaney, P. F., Verkoeijen, P. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 63–147). San Diego, CA: Elsevier Academic Press. doi:10.1016/S0079-7421(10)53003-2

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 371–377. doi:10.1037/0278-7393.15.3.371

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299. doi:10.1006/jmps.2001.1388

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112–127. doi:10.1037/0278-7393.34.1.112

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition, 38,* 1009–1017. doi:10.3758/MC.38.8.1009

Karpicke, J. D., & Blunt, J. R. (2011, February 11). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331,* 772–775. doi:10.1126/science.1199327

Karpicke, J. D., Lehman, M., & Aue, W. R. (in press). Retrieval-based learning: A temporal context account. In B. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 60). San Diego, CA: Elsevier Academic Press.

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., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language, 67,* 17–29. doi:10.1016/j.jml.2012.02.004

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

Lehman, M., & Malmberg, K. J. (2013). A buffer model of encoding and temporal correlations in retrieval. *Psychological Review, 120,* 155–189. doi:10.1037/a0030851

Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322–336. doi:10.1037/0278-7393.31.2.322

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11,* 371–385.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology, 33,* 172–186. doi:10.1016/0022-2496(89)90029-1

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10,* 389–395. doi:10.1080/09658210244000216

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 287–297. doi:10.1037/a0021801

Pyc, M. A., & Rawson, K. A. (2010, October 15). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335. doi:10.1126/science.1191465

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27,* 431–452. doi:10.1207/s15516709cog2703_5

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. doi:10.1037/0033-295X.88.2.93

Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6,* 296–305. doi:10.3758/BF03197459

Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-the-last paradigm. *Memory & Cognition, 40,* 844–860. doi:10.3758/s13421-012-0198-0

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York, NY: Psychology Press.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399. doi:10.1037/a0013082

Underwood, B. J. (1972). Are we overloading memory? In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory* (pp. 1–23). Washington, DC: Winston.

Verkoeijen, P. J. L., Tabbers, H. K., & Verhage, M. L. (2011). Comparing the effects of testing and restudying on recollection in recognition memory. *Experimental Psychology, 58,* 490–498. doi:10.1027/1618-3169/a000117

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 442–452. doi:10.1037/0278-7393.1.4.442

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1024–1039. doi:10.1037/0278-7393.19.5.1024

Received October 26, 2013
Revision received January 22, 2014
Accepted February 18, 2014 ■