CHAPTER SEVEN

Retrieval-Based Learning: An Episodic Context Account

Jeffrey D. Karpicke[*,1], **Melissa Lehman**[*], **William R. Aue**[*,†]
[*]Department of Psychological Sciences, Purdue University, West Lafayette, Indiana, USA
[†]Department of Psychology, Syracuse University, Syracuse, New York, USA
[1]Corresponding author: e-mail address: karpicke@purdue.edu

Contents

1. Introduction 238
2. Current Status of Retrieval Practice Research 239
   2.1 The Effects of Retrieval Practice on Learning 240
   2.2 Delimiting the Key Effects in the Retrieval Practice Literature 245
3. Analysis of Existing Explanations of Retrieval Practice 247
   3.1 Strength and Retrieval Effort 248
   3.2 Storage and Retrieval Strength 249
   3.3 Bifurcation 250
   3.4 Transfer Appropriate Processing 250
   3.5 Encoding Variability 251
   3.6 Elaborative Retrieval 252
4. An Episodic Context Account of Retrieval-Based Learning 258
   4.1 Overview of the Episodic Context Account 258
   4.2 Context Representation and Encoding 259
   4.3 Context Reinstatement During Retrieval 260
   4.4 Context Updating During Successful Retrieval 261
   4.5 Retrieval Practice Restricts the Search Set 262
5. Evidence Supporting an Episodic Context Account 263
   5.1 Effects of Retrieval Practice and Initial Retrieval Difficulty 263
   5.2 Recollection of Context on a Criterial Test 266
   5.3 Manipulations of Initial Context Retrieval 271
6. General Discussion and Final Comments 275
   6.1 Context Variability and Study-Phase Retrieval 275
   6.2 Conclusion 278
Acknowledgments 278
References 278

Abstract

Practicing retrieval is a powerful way to promote learning and long-term retention. This chapter addresses the theoretical underpinnings of retrieval-based learning. We review methodological issues in retrieval practice research, identify key findings to be

Psychology of Learning and Motivation, Volume 61
ISSN 0079-7421
http://dx.doi.org/10.1016/B978-0-12-800283-4.00007-1

accounted for, and evaluate current candidate theories. We propose an episodic context account of retrieval-based learning, which explains retrieval practice in terms of context reinstatement, context updating, and restriction of the search set. Retrieval practice involves attempting to reinstate a prior learning context, and when retrieval is successful, the representation of context is updated to include features of retrieved contexts and the current context. Future retrieval is enhanced because updated context representations can be used to restrict the search set and hone in on a desired target. The context account accommodates a wide variety of phenomena in the retrieval practice literature and provides a comprehensive and cohesive account of retrieval-based learning.

1. INTRODUCTION

We often think of our minds as places that hold copies or records of our past experiences, and perhaps as a consequence, we tend to identify the process of "learning" with the acquisition of new knowledge and experiences—the creation of new information in memory. When learning is viewed as the process of getting information into one's mind, an emphasis naturally falls on the processes involved in encoding knowledge and experiences. Learners may not worry much about how they will retrieve and reconstruct knowledge when they need to use it in the future, but even if they do, they likely view retrieval as the mere expression of knowledge obtained from prior experiences—the evidence that prior learning occurred—but no more. It is in this sense that retrieval is considered "neutral" for learning because the process of accessing knowledge is not thought to change knowledge.

The approach described in this chapter, referred to as *retrieval-based learning*, is based on the finding that accessing knowledge does indeed change one's knowledge. When people practice retrieval, the act of retrieving knowledge in the present enhances one's ability to retrieve and use that knowledge again in the future. Retrieval is not neutral; it does not merely involve accessing static pieces of information held in a storage system. Instead, every time a person retrieves knowledge, that knowledge is changed. Retrieval-based learning is an advantageous feature of our memory systems, one that we might build in if we were designing memory from scratch (Nairne, 2010). Retrieval is typically purposeful and goal directed; when knowledge is successfully reconstructed in the present, it likely happened for a reason, so improving the future retrievability of that knowledge would seem advantageous so that similar problems could be solved again when they occur in the future. Indeed, with every act of retrieval, there is some change that occurs that improves one's ability to retrieve and reconstruct that knowledge in the future.

This chapter is concerned with the nature of the changes that occur as a consequence of retrieval. There has been a surge of research on retrieval practice in the past decade, but this research also has a long history. The fact that active recall improves learning was noted by Francis Bacon in *Novum Organum* (1620) and by William James in the *Principles of Psychology* (1890). Experimental studies of the effects of retrieval on learning date back at least to Abbott (1909), and over 70 years ago, McGeoch (1942, pp. 196–200) summarized the state of research on "Recall During Practice" in his book, *The Psychology of Human Learning*, the gold-standard textbook at the time. Indeed, retrieval practice is not new to *The Psychology of Learning and Motivation* series. Only a few years ago, Delaney, Verkoeijen, and Spirgel (2010) devoted a section of their review of the spacing effect to retrieval practice and Roediger, Putnam, and Smith (2011) surveyed "ten benefits of testing." Nevertheless, as noted by Delaney et al. and Roediger et al., theoretical progress in understanding the nature of retrieval-based learning has been limited.

Section 2 describes methodological issues in retrieval practice research and identifies the key effects that need to be explained by any theory of retrieval practice. Section 3 turns to an overview and analysis of existing accounts of retrieval practice. Several ideas commonly invoked when discussing retrieval practice provide little insight into underlying mechanisms—the deep structure of retrieval practice effects. One exception is the theory that retrieval practice effects stem from semantic elaboration during retrieval. We evaluate the rationale and evidence for this elaboration account. Section 4 sketches an account of retrieval-based learning that we call an *episodic context account*. In Section 5 we review the current evidence in light of the context account, discuss the account in relation to other theoretical ideas, and conclude by offering suggestions for future work based on the predictions of the context account.

2. CURRENT STATUS OF RETRIEVAL PRACTICE RESEARCH

As noted above, there has been a recent surge in research on retrieval practice. In addition to the two reviews mentioned above, recent overviews have been provided by Roediger and Karpicke (2006a), McDaniel, Roediger, and McDermott (2007), Roediger and Butler (2011), Carpenter (2012), Karpicke (2012), and Karpicke and Grimaldi (2012), among others. This review is focused on theoretical explanations of retrieval-based learning, and our aims in this section are, first, to clarify some methodological and conceptual issues surrounding retrieval practice research and, second, to delimit the key effects that must be accounted for by any theory of retrieval practice.

2.1. The Effects of Retrieval Practice on Learning

If a person experiences an event or studies new material and his or her memory is assessed at a later time, the ability to recall the material will decrease as the time between the study and test event increases. The systematic study of forgetting began with Ebbinghaus (1885/1964), and it has long been known that the appropriate way to evaluate forgetting is to test different people or different sets of materials at different points in time. If one were to test the same person or the same material repeatedly, the act of assessing memory at one point in time would influence the measurement of forgetting at a subsequent point in time. The "best practices" for measuring forgetting were emphasized in textbooks by McGeoch (1942) and Deese (1958), for example, during what was perhaps the peak of interest in the study of forgetting curves. The recommended methods for studying forgetting are noteworthy because they acknowledge, tacitly or explicitly, that each act of retrieval influences subsequent retrieval. These effects were traditionally viewed as contaminants that must be removed from experiments to obtain pure measures of forgetting.

The effects of prior retrieval on subsequent retrieval (and the "contaminating" effects on the measurement of forgetting) were demonstrated clearly in an experiment by Hanawalt (1937). He had subjects study simple geometric line drawings and reproduce them at different points in time: immediately, 1 day, 1 week, 1 month, or 2 months after the original study period. Some groups of subjects reproduced the drawings only once at one of the retention intervals (single recall), while another group reproduced the drawings repeatedly at each retention interval (repeated recall). Figure 7.1 shows the proportion of figures correctly reproduced in the single recall and repeated recall conditions. Whereas the typical forgetting curve appears in the single recall condition, there is little or no forgetting over time in the repeated recall condition. Today, we view the effects depicted in Fig. 7.1 as benefits of repeated retrieval rather than as contaminating effects

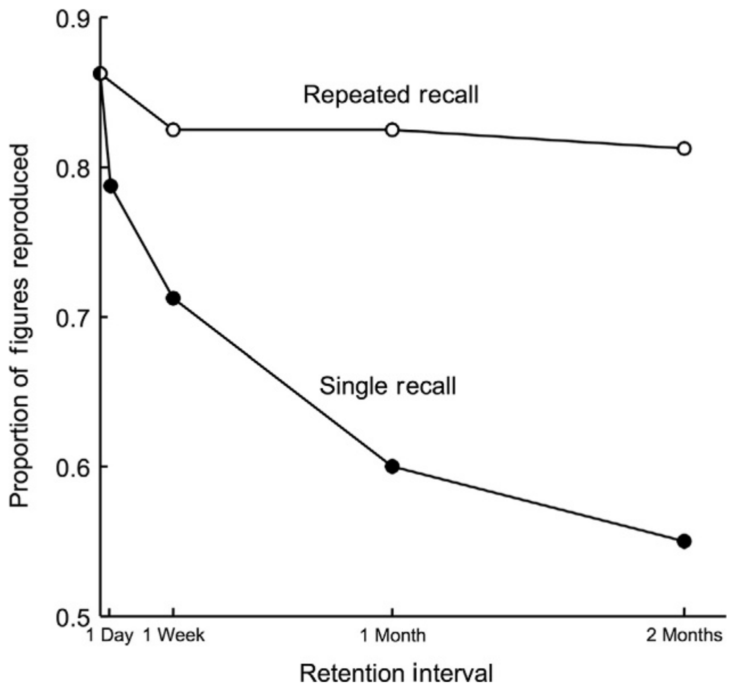

Figure 7.1 Proportion of line-drawing figures correctly redrawn at varying retention intervals under repeated recall or single recall conditions. A typical forgetting trend is observed in the single recall condition, whereas little or no forgetting is seen in the repeated recall condition. *Data are from Hanawalt (1937).*

of retrieval on the measurement of forgetting. The benefit of retrieval practice can be seen at each retention interval. For instance, at the 1-week interval, subjects benefitted from prior recall immediately after learning; at 1 month, subjects benefitted from prior retrieval at the immediate and 1-week intervals; and large benefits of repeated recall were seen 2 months after the original learning episode. Each time subjects practiced retrieving the line drawings, the act of retrieval enhanced the ability to reconstruct the drawings again in the future.

Fast forward to present times and there have been perhaps hundreds of recent demonstrations of retrieval practice effects. One example from Smith, Roediger, and Karpicke (2013) illustrates some important points about the nature of retrieval practice effects. Subjects first studied word pairs that included a category name and an item from the category as a to-be-recalled target (e.g., *vegetable—cucumber*). In the second phase of the experiment, subjects either restudied the word pairs or saw the cue and first two letters of the target (*vegetable—cu*____). For these items, subjects were told to think back the study episode and recall the word that completed the word stem (as we will discuss later, these intentional retrieval instructions are important; Karpicke & Zaromb, 2010). Then in the third and final phase of the experiment, the subjects freely recalled the target words. We refer to this as the *criterial test*, and Fig. 7.2 shows the proportion of words recalled on this test. The results show that practicing retrieval of the target words enhanced recall on the criterial test relative to having restudied the word pairs in phase 2 in the experiment. Smith et al. also varied whether subjects were required to produce a response (overt retrieval) or merely think about the target word in their minds (covert retrieval) in phase 2, and they found little or no effect of overt versus covert retrieval across a series of four experiments. The main point for present purposes is that practicing retrieval of the target words enhanced recall on the criterial test relative to studying the target words.

The Smith et al. experiment highlights some important points about retrieval practice effects. First, subjects did not restudy the target items in the retrieval practice condition. The effect shown in Fig. 7.2 is purely due to the act of retrieval, rather than a combination of retrieval and restudy (or feedback) effects. Second, given that the key effect has to do with

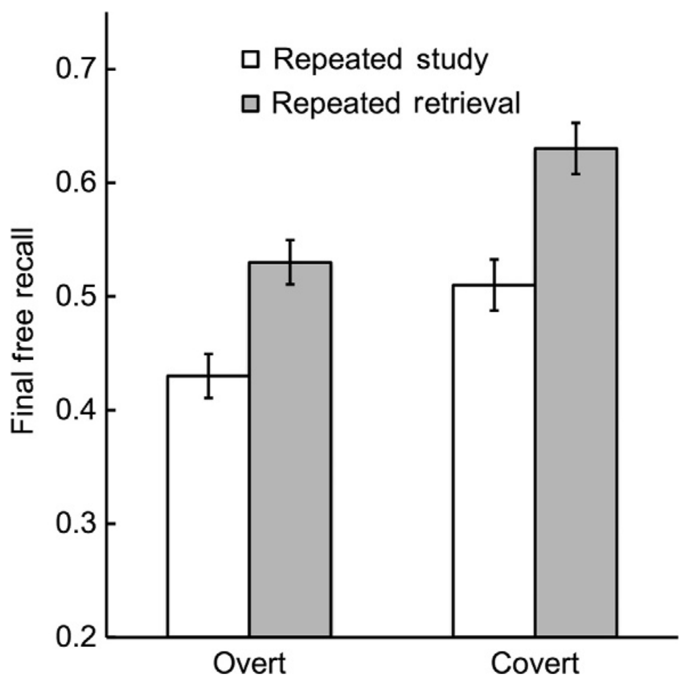

Figure 7.2 Proportion of words recalled on a criterial free recall test as a function of prior repeated study or repeated retrieval practice (where targets were recalled overtly or covertly). Practicing retrieval of the target words enhanced recall on the criterial test relative to restudying the target words. *The figure is redrawn using data from Smith et al. (2013), Experiment 4.*

retrieval in the absence of restudying, successful retrieval is essential. Subjects in the Smith et al. experiment correctly recalled the targets 72% of the time during initial retrieval practice, and the effects in that condition are being compared to the effects of restudying 100% of the items in the repeated study condition. Thus, the comparison is biased in favor of the restudy condition, because all items were reexperienced in that condition, yet there is still a benefit of retrieval practice on the criterial test. Third, the criterial free recall test (phase 3 in the experiment) occurred about 15 min after the initial recall test (phase 2). It is sometimes claimed that the effects of retrieval practice do not appear in the short term and only occur when there is a delay between the learning experience and the final test. Clearly, that is not true; retrieval practice effects are alive and well even in the short term.

The findings of Smith et al. (2013) highlight three key attributes of retrieval practice that are worth emphasizing. First, the best way to examine the effect of retrieval itself is to examine retrieval without restudy or feedback, thereby isolating the effect of retrieval rather than the combined effects of retrieval and subsequent encoding during restudy. Second, the benefits of retrieval practice depend on successful retrieval. When subjects recall relatively little in a retrieval practice condition, one is unlikely to see an effect relative to restudying (for instance, in retrieval practice experiments by Hogan & Kintsch, 1971, subjects recalled about 30% of the items during initial recall, and benefits of retrieval practice over restudying were not seen on immediate criterial tests). When studying the generation effect (Slamecka & Graf, 1978), researchers have been careful to use conditions that ensure successful generation of most if not all target items, yet curiously, when studying testing/retrieval practice effects researchers tend to neglect the importance of initial retrieval success. Third, under the right conditions, benefits of retrieval practice relative to restudying are observable in the short term, a few minutes after the experimental procedure.

An additional noteworthy aspect of retrieval practice is that repeated retrieval and repeated study conditions likely do not represent "pure" manipulations of retrieval practice. Some retrieval of a prior episodic context probably occurs in repeated study conditions. Indeed, "study-phase retrieval" and "reminding" have long been explanations for spaced practice effects (see Greene, 2008). Those ideas essentially attribute spacing effects to retrieval practice and hinge upon the assumption that retrieval practice occurs in repeated study conditions, even when subjects are not instructed to think back to prior events. Interestingly, McGeoch (1942) noted these points in his discussion of "Recall During Practice":

> *There are wide variations in the extent to which a subject depends on the direct stimulus pattern of the material being practiced. Assume that he is memorizing a list of words. At the extreme of complete dependence on the stimulus pattern he repeats each word as he perceives it, but does not attempt to anticipate any word or to recall a word before it is actually presented. At the opposite extreme, beginning with the second presentation he attempts to recall as much as possible of the material without having it presented to his receptors. Between the extremes lie innumerable combinations of presentation and independent recall. (p. 196)*

McGeoch continues:

> *It is probable that the experiments have seldom compared 'pure' reading or presentation with recitation. Instead, the readings may have involved some recitation, so that the comparisons have been between small amounts of recitation and larger ones. The reports of the subjects and the observation of their behavior permit a statement of some of the basic variables underlying the effectiveness of the relatively larger amounts of recall during practice. (p. 199)*

There are a variety of ways to deal with the retrieval success problem and make a fair comparison across restudy and retrieval practice conditions. Two approaches are unsatisfactory. First, one could provide feedback or restudy opportunities after every retrieval attempt. This ensures that subjects reexperience the items even if they cannot recall them, but it clouds inferences about the mnemonic effect of retrieval because the effect is now a combination of direct and mediated effects that cannot be teased apart (Roediger & Karpicke, 2006a). A second possibility is to conditionalize final recall on initial recall. A conditional analysis may provide interesting information, but it raises a host of additional item-selection problems. The items recalled initially are by definition "easier" items, and they are likely to be recalled again on a criterial test by virtue of their easiness alone, rather than because of a retrieval practice manipulation. Conditional analyses are correlational, and inherent item characteristics are tangled with any mnemonic effect of retrieval. Neither the provision of feedback nor a conditional analysis is a satisfactory approach to addressing the retrieval success problem in retrieval practice experiments.

We propose two general solutions to address retrieval success issues. First, experiments can be designed so that levels of initial retrieval success are relatively high, and under such circumstances, one is likely to observe retrieval practice effects even in the short term (i.e., at the end of an experimental session). Karpicke and Zaromb (2010) and Smith et al. (2013) provide examples of how one might accomplish this. Second, methods have been developed to bring subjects to criterion, ensuring that they have successfully recalled each item once prior to introducing the manipulation of repeated study or repeated retrieval practice. There are a variety of ways to implement a criterion procedure (e.g., see Grimaldi & Karpicke, 2014; Karpicke, 2009; Karpicke & Roediger, 2007b, 2008; Pyc & Rawson, 2009). Karpicke and Smith (2012) present and discuss a few possible criterion methods.

We suspect that some researchers may ignore retrieval success problems in retrieval practice experiments because the work is viewed as a "testing effect" rather than as a retrieval practice effect. The locus of the positive effects on learning, however, is in repeated, successful retrieval. Retrieval can occur in a variety of activities that are not "tests." In educational settings, many classroom activities could be modified to incorporate active retrieval (see Blunt & Karpicke, in press). Likewise, tests do not always require people to practice retrieval, as is the case when students take tests with the relevant material available during the test (open-book tests; Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008). The emphasis on "testing" producing learning has sometimes obscured the locus of the effect. It is not testing, *per se*, that produces learning; it is the act of practicing retrieval that produces learning. Retrieval practice will occur with varying degrees of success during a test, it may occur during a condition where students are nominally told to "study," it may not occur on tests where retrieval is not necessary, and it may occur in other activities that do not seem test-like at all.

2.2. Delimiting the Key Effects in the Retrieval Practice Literature

We can now delimit the key effects that would need to be explained by an account of retrieval-based learning and eliminate some effects that are not germane to theorizing about retrieval practice. First, our focus is on the direct effects of retrieval practice, rather than on the mediated effects of retrieval on subsequent encoding or studying (e.g., Grimaldi & Karpicke, 2012; Kornell, Hays, & Bjork, 2009). Second, it is sometimes claimed that the effects of retrieval practice only occur after a delay, and in a similar vein, it has been noted that repeated study and retrieval practice conditions sometimes interact with the timing of the final test, such that repeated study is better in the short term and retrieval practice is better in the long term. Retrieval practice effects do indeed occur at short retention intervals, as described earlier (Karpicke & Zaromb, 2010; Smith et al., 2013). The retention interval interaction is essentially a result of item-selection artifacts favoring the restudy condition, and as such, it is not an important phenomenon in need of a mechanistic explanation. As described earlier, when a repeated study condition reexperiences many more items than a retrieval practice condition, it is no surprise that repeated study would produce better performance than retrieval practice. The "benefit" of repeated study in the short term is illusory; if retrieval success were near perfect (by using one of the methods outlined above), and item reexposure were close to identical across repeated study and retrieval practice conditions, there would be no advantage of repeated study in the short term. Indeed, the benefits of retrieval practice would likely be observed (e.g., Smith et al., 2013).

There are four key effects that, at a minimum, would need to be explained by any theory of retrieval-based learning (see Table 7.1). The first is the main effect of retrieval practice—that practicing retrieval enhances performance on a criterial test relative to an appropriate control condition (either a no-test baseline control condition or a repeated study condition). The next three essentially constitute a set of effects that collectively can be considered effects of "difficult" initial retrievals. This includes the following effects: (1) tests that involve recall tend to produce greater effects than tests that involve recognition (e.g., Glover, 1989; see too Butler & Roediger, 2007); (2) spacing an initial retrieval practice event produces greater effects than massed retrieval (e.g., Jacoby, 1978; Whitten & Bjork, 1977; see too Karpicke & Bauernschmidt, 2011); and (3) recalling with weakly related semantic cues produces greater effects relative to recalling with strong semantic cues (Carpenter, 2009; see too Carpenter & DeLosh, 2006). In these three scenarios, the conditions that produce larger gains in retention (initial recall, spacing an initial retrieval trial, and recalling with weak semantic cues) are thought to afford more difficult retrieval relative to the conditions that produce smaller gains (initial recognition, massed retrieval, and recalling with strong semantic cues).

Table 7.1 Key Effects of Retrieval Practice to Be Explained by Theoretical Accounts of Retrieval-Based Learning

1.	Retrieval practice enhances retention on a criterial test relative to control conditions (no-test or repeated study conditions)
2.	Initial retrieval practice under recall conditions produces greater effects relative to initial retrieval practice under recognition conditions
3.	Spaced initial retrieval practice produces greater effects than massed initial retrieval
4.	Retrieval practice with weak cues produces greater effects than retrieval practice with strong cues

Certainly, there are additional effects that a robust theory would need to handle—for instance, practicing retrieval alleviates the buildup of proactive interference; practicing retrieval enhances recollection on tests that require context retrieval; orienting retrieval toward greater context retrieval produces larger effects; and other phenomena, which we will return to later in this report. Nevertheless, the four central effects described above provide a starting point for any explanation of retrieval practice. In the next section, we examine several existing explanations and evaluate whether the ideas elucidate the underlying mechanisms that may be responsible for retrieval practice effects.

3. ANALYSIS OF EXISTING EXPLANATIONS OF RETRIEVAL PRACTICE

McGeoch (1942) offered four possible explanations for the effect of retrieval practice on learning, some of which align remarkably well with contemporary ideas. He wrote:

> *(1) Recitation furnishes the subject with progressive knowledge of results. This information (a) acts as an incentive condition, (b) brings the law of effect directly to bear, (c) favors early elimination of wrong responses, and (d) by informing the subject which items have been learned, promotes a more effective distribution of effort over the material. (2) Recitation favors articulation of the items and leads to the utilization of accent and rhythm. (3) It likewise promotes grouping of the items, localization in the series, and the search for meaningful connections. (4) In recitation the subject is practicing the material more nearly in the way in which it is to be tested and used – that is, without direct stimulation from the copy. It constitutes, therefore, a more immediately relevant form of practice. (pp. 199–200)*

McGeoch's first explanation falls in the realm of indirect or mediated effects, rather than the direct mnemonic effects of retrieval on learning. His second explanation, which places an emphasis on articulation, is consistent with current ideas about the "production effect" (e.g., MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), but has not been shown to be crucially important for retrieval practice effects (Putnam & Roediger, 2013; Smith et al., 2013). McGeoch's third and fourth explanations bear close relations to two theories described below. The idea that retrieval practice leads students to practice in the way that material will be used in the future is essentially the same as the transfer-appropriate processing idea of retrieval-based learning (see Roediger & Karpicke, 2006a, 2006b). The idea of "searching for meaningful connections" appears similar to the semantic elaboration account, which is given close attention at the end of this section.

This section discusses the most prominent accounts that have been proposed to explain retrieval-based learning, beginning with explanations that attribute retrieval-based learning to modifications of memory trace strength, followed by accounts that propose that retrieval-based learning is due either to practice with retrieval tasks or to encoding variability produced by retrieval tasks. We then turn to an account that attributes retrieval-based learning to semantic elaboration processes.

3.1. Strength and Retrieval Effort

One of the earliest explanations for the mnemonic benefits of retrieval assumed that retrieval processes affect the *strength* of memory traces. This idea assumes that representations of information are stored in memory and those representations can be strengthened in such a way that makes them more retrievable. In a foundational paper, Bjork (1975) proposed that when an item is retrieved from memory, the representation of that item memory is strengthened in some manner. He also proposed that the level of strengthening that takes place is a function of the *effort* required to retrieve the item, an idea termed the retrieval effort hypothesis. Specifically, Bjork suggested that retrieval operated in a way similar to levels of processing during encoding, with retrieval as a deeper level of processing relative to shallow restudying and with more effortful retrieval operations producing even deeper processing and thus greater strengthening (see too Bjork, 1994).

The retrieval effort theory provides an intuitive account of a variety of retrieval phenomena. Initially developed to account for negative recency effects (Craik, 1970), retrieval effort also helped to explain findings that the longer it takes subjects to retrieve words, given their definitions, the more likely those items are to be recalled on a later test (Gardiner, Craik, & Bleasdale, 1973; but see Karpicke & Bauernschmidt, 2011). Retrieval effort has been operationalized and manipulated in a variety of ways such as changing learning criterion (Pyc & Rawson, 2009), providing less informative cues (Carpenter & DeLosh, 2006), or delaying the initial test (Karpicke & Roediger, 2007a, 2007b).

Ultimately, retrieval effort is a redescription of some retrieval practice phenomena and does not delineate mechanisms that would produce the mnemonic effects of retrieval on subsequent retention. The concept of retrieval effort can be problematic, too, because it is not always clear what constitutes effortful retrieval, and the relation between time and effort is ambiguous. Response times are often considered measures of effort, with slower times representing greater effort. It is equally plausible that faster response times reflect greater effort. For instance, it is reasonable to think that running 1 mile requires more effort than walking it, yet running will take less time than walking. These issues aside, retrieval effort is still only a measure that may or may not be correlated with underlying mechanisms that produce the mnemonic effects of retrieval, but it does not specify what those mechanisms might be.

3.2. Storage and Retrieval Strength

An idea related to Bjork's (1975) retrieval effort proposal is the theory of disuse proposed by Bjork and Bjork (1992). A key element of the theory is the differentiation of retrieval strength from storage strength. They suggest that retrieval is a function of both the quality of the item in memory (which they termed storage strength) and the ability of a test cue to elicit the item (retrieval strength; see too Raaijmakers & Shiffrin, 1981; Tulving & Thomson, 1973). To explain the benefit of retrieval, Bjork and Bjork (1992) assumed that both storage and retrieval strength are increased when an item is restudied, and the strengths are increased to a greater degree when an item is recalled. Additionally, they assumed that the increment in strengths is, in part, a function of retrieval difficulty. When an item with low-current retrieval strength is successfully retrieved (i.e., a difficult retrieval), it will receive a greater increment in strengths than an item with high-retrieval strength (i.e., an easily retrieved item).

Bjork and Bjork's (1992) theory of disuse represents, in part, a more concrete version of Bjork's (1975) retrieval effort hypothesis, and the concepts of retrieval strengths and their interplay with storage strengths are consistent with a variety of contemporary models of memory. Moreover, the verbal theory provided by Bjork and Bjork (1992) would seem to account for the key attributes of retrieval-based learning described above. However, the storage/retrieval strength idea simply assumes that strengths increase when an item is recalled; that is, it assumes that retrieval-based learning occurs without proposing a mechanistic explanation for how it occurs. In addition, the primary mechanism for incrementing storage and retrieval strengths is retrieval effort during a test, which is influenced by the retrieval strength of a to-be-recalled item relative to a set of competitors. While Bjork and Bjork describe what a difficult item is in their model (i.e., an item with low-retrieval strength relative to a set of competitors), it is not always clear how to define a "difficult" item in a variety of retrieval situations.

3.3. Bifurcation

Another variant of a strength hypothesis is Kornell, Bjork, and Garcia's (2011) *bifurcation* account, which offers a description of the forgetting rates for retrieved versus restudied items. It is sometimes the case that restudied items are recalled better than tested items on an immediate criterial test, whereas the tested/retrieved items are recalled better than restudied items on a delayed test (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006b). According to the bifurcation model, initial tests produce a bifurcated item distribution, in which items that are successfully retrieved are strengthened, while items that are not successfully retrieved are not. Items that are restudied are strengthened, but to a lesser degree than retrieved items. The model also assumes that items are recalled on a test if they exceed a threshold that varies as a function of test "difficulty" (Halamish & Bjork, 2011). The differential strengthening of retrieved versus restudied items results in advantages for restudied items on an immediate test because more of the items have been strengthened beyond the retrieval threshold. However, after a delay, all items are weakened (i.e., forgotten) at the same rate, and now many restudied items fall below the retrieval threshold, whereas the retrieval practice items, which gained more strength initially, remain above threshold.

In support of the idea, Kornell et al. (2011) showed that when all items are retrieved during initial testing, preventing bifurcation, the aforementioned interaction with delay does not occur; similarly, when repeated study occurs under conditions that presumably produce bifurcation, the interaction is once again present. The bifurcation idea provides a descriptive account of a particular pattern of results, but it is somewhat limited in scope. Kornell et al. (2011, p. 86) acknowledge that their distribution-based framework is merely a descriptive account of this particular pattern of data and is not intended to indicate the underlying mechanisms of retrieval-based learning. Like other strength accounts, the bifurcation model simply relies on the idea that retrieved items are strengthened more than studied items without specifying how or why such strengthening would occur. In addition, as we outlined earlier, the retention interval interaction that the bifurcation model explains is essentially an item-selection artifact that occurs under a particular set of conditions. It is not crucial for understanding the deep structure of retrieval-based learning.

3.4. Transfer Appropriate Processing

Another descriptive account of retrieval practice proposes that intervening retrieval serves as practice that is similar to the conditions of a final criterial test.

Transfer-appropriate processing refers to the idea that test performance will be greatest when the cognitive processes required on a criterial test are similar to the cognitive processes that occurred during original learning (Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977). Some researchers have argued that retrieval practice may be beneficial because the processes necessary for successful initial retrieval are similar to those employed during later retrieval (e.g., Landauer & Bjork, 1978; Roediger & Karpicke, 2006a, 2006b). This idea has been substantiated to some degree by the finding that performance on a criterial test is best when the final test questions are identical to initial test questions (e.g., Butler, 2010; Johnson & Mayer, 2009; McDaniel & Fisher, 1991; see also Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010). However, a strict interpretation of transfer-appropriate processing would predict that performance should be best when the intervening and criterial test formats are identical and thus require exactly the same overlapping mental processes. Some authors have reported such a matching pattern (e.g., Duchastel & Nungester, 1982), but many others have not. Instead, free recall (Carpenter & DeLosh, 2006; Glover, 1989) or short answer test formats (Butler & Roediger, 2007; Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007) have generally been found to produce the best performance regardless of the criterial test format.

It may be the case that, very generally speaking, similarity of processing during original learning and criterial performance is important. However, transfer-appropriate processing still only offers a redescription of the basic retrieval practice effect. The relevant data do not support the idea that retrieval practice effects are greatest when initial and final test formats are matched. The idea of transfer-appropriate processing also does not appear to make clear predictions about the benefit of spaced versus massed retrieval or retrieving with weak versus strong cues, two of the key effects we identified earlier. In sum, transfer-appropriate processing is essentially a statement about the similarity of original learning and later test situations, and it does not specify underlying mechanisms that would produce retrieval practice effects.

3.5. Encoding Variability

Encoding variability has occasionally been proposed as an explanation for retrieval-based learning, although the idea has been discussed less frequently than other ideas reviewed in this section. Encoding variability refers to the idea that when items or materials are experienced multiple times, the materials are encoded in different (variable) ways during each encounter, and this is assumed to increase the number of retrieval routes a person has to access material in the future (Martin, 1968; Melton, 1970). Encoding variability has been explored most extensively as an explanation for the spacing effect, the finding that material is learned better when multiple presentations of an item are spaced over time relative to when presentations are presented back to back with no intervening items (massed practice; see Bower, 1972; Greene, 2008).

Empirical tests of encoding variability as an explanation for retrieval practice have been scarce, and the extant data are mixed. For example, McDaniel and Masson (1985) observed a benefit on a criterial extra-list cued recall test when the intervening test used a different extra-list cue, which may support an encoding variability interpretation. Conversely, Butler (2010) found that varying the conditions of retrieval practice by presenting different questions across repeated tests did not increase retention relative to a condition in which the same questions were presented on repeated tests. However, perhaps the greatest problem for any account based on encoding variability is that attempts to induce variable encoding directly have shown no effect or even decreases in memory performance (e.g., Benjamin & Tullis, 2010; Greene & Stillwell, 1995; Postman & Knecht, 1983; Verkoeijen, Rikers, & Schmidt, 2004).

Our main purpose in mentioning encoding variability is to distinguish that concept from the idea of *contextual* variability, which refers to the specific idea that different temporal/contextual features can be encoded as part of the representation of repeated events. Indeed, contextual variability theories of the spacing effect have received both empirical and theoretical support (Delaney et al., 2010; Greene, 1989; Lohnas & Kahana, in press; Raaijmakers, 2003). In later sections of this chapter, we describe how contextual variability may play an important role in retrieval-based learning.

3.6. Elaborative Retrieval

An explanation that has received considerable attention in recent years attributes the effects of retrieval practice to semantic elaboration that is assumed to occur during the process of retrieval. This theory is known as the "elaborative retrieval hypothesis" (Carpenter, 2009, 2011), though here we refer to it specifically as the *semantic elaboration account* in an attempt to clarify what is meant by "elaboration."

Elaboration generally refers to the process of encoding additional features or attributes in the representation of an event, and this typically refers to semantic or meaning-based aspects of items. When one condition enhances memory performance relative to another, the enhancement can often be attributed to elaboration or "deep processing," perhaps only based on the fact that one condition produced better memory performance than another (Karpicke & Smith, 2012). There is still no universally agreed upon index of elaboration, decades after Craik and Tulving (1975) initiated the search for one. As such, in many circumstances, elaboration essentially remains a "just-so story" when it is invoked to explain memory phenomena.

The semantic elaboration account is clearer than past accounts in describing the type of elaboration assumed to occur during retrieval. The idea is that when subjects use retrieval cues to search for a target response, several items that are semantically related to the cue become activated during the search process (Carpenter, 2009, 2011). For example, when attempting to recall the target word *bread* when given a weakly associated cue such as *basket*, several words that are associated with the cue (like *eggs*, *wicker*, *fruit*, and so on) are thought to become activated, and these semantic associates are assumed to serve as retrieval routes from *basket* to *bread* on a future criterial test. The elaboration account further assumes that there is little if any generation of semantic associates when word pairs like *basket-bread* are repeatedly studied because subjects do not need to search for the target. Similarly, difficult retrieval conditions are assumed to produce more extensive semantic searches relative to less difficult conditions. For example, when a cue is strongly associated to a target (e.g., *toast* as the cue for *bread*), recall of the target is easier, producing a less extensive search of memory and thus less semantic elaboration. "Weak" cues would produce more extensive searches and more elaboration. In experiments related to the semantic elaboration account, weak cues have been defined as ones with weak semantic associations to targets or ones that provide less information about targets (e.g., if the target word were *cabin,* the cue *c* _ _ _ _ would be "weaker" than the cue *c a b* _ _).

The strength of the semantic elaboration account is that it proposes a mechanism for retrieval-based learning: the generation of several semantically/associatively related words during retrieval is assumed to occur and is assumed to produce retrieval practice effects. In addition, the elaboration account attempts to explain key effects in the retrieval practice literature. However, the semantic elaboration account has been challenged on both logical and empirical grounds, as described next.

First, much of the evidence in support of the semantic elaboration account is correlational. For example, Carpenter (2011) showed that practicing retrieval enhanced performance on a criterial cued recall test where the cues were mediators, nonstudied words that were semantically related to the studied pairs. For example, for the word pair *mother–child*, the nonstudied word *father* might be given as a cue to retrieve the target *child*. The idea is that the word *father* came to mind during initial retrieval and mediates the association between *mother* and *child*. Carpenter's (2011) findings were taken as evidence that the activation of semantic mediators occurred during retrieval. However, the generation of semantic mediators, and thus the amount of elaboration, was never directly manipulated. It is perfectly possible that some mechanism other than the generation of semantic associates produced the retrieval practice effect, and the benefits of retrieval practice were seen with extra-list mediators as cues, even though those words never came to mind during initial retrieval. In other words, rather than viewing the activation of mediators as a cause of retrieval-based learning, it may be that some other mechanism produced retrieval-based learning and also produced an effect on final mediator-cued recall.

Second, the idea behind elaborative retrieval seems inconsistent with the principle of cue overload: As more items become associated with a single retrieval cue, the likelihood of recovering a particular target decreases (see Nairne, 2002, 2006; Raaijmakers & Shiffrin, 1981; Surprenant & Neath, 2009; Watkins & Watkins, 1975; Wixted & Rohrer, 1993). The semantic elaboration idea is that subjects generate several semantically related words related to a retrieval cue. This ought to produce massive cue overload, making memory performance worse, yet the generation of semantically related words is proposed to explain the improvement in memory due to retrieval practice. The phenomenon of cue overload is well established, and indeed, the number of words that are implicitly associated with a retrieval cue is negatively associated with recall of target words (Nelson & McEvoy, 1979), which is the opposite of the semantic elaboration idea. Even if we assumed that, rather than becoming part of the search set, the semantically associated information serves as additional retrieval cues, this is still difficult to reconcile with the cue overload principle, as these items are not provided at test and thus would have to be retrieved prior to being used as cues. For example, if a subject were given a cue and generated three associated mediators before reaching the target, the three mediators would still need to be retrieved to access the target on a later test. Thus, four items, rather than one, would be associated with the retrieval cue.

Third, the semantic elaboration account appears at odds with the phenomenon of retrieval-induced forgetting. In retrieval-induced forgetting experiments, subjects study cue–target word pairs where multiple targets share the same cue (e.g., *fruit-orange* and *fruit-banana*) and then practice retrieval of some of targets that were paired with each cue (e.g., subjects might practice *fruit-or*_ _ _ _, but *banana* would not be practiced). The retrieval-induced forgetting effect is that retrieval practice of the targets (*orange* in this example) interferes with subsequent recall of nonpracticed items (*banana*; see Anderson, Bjork, & Bjork, 1994). The semantic elaboration account proposes that several semantically related words are activated during retrieval; presumably, this would mean that nonpracticed items benefit from activation during retrieval practice. If this were generally true, it is difficult to see why retrieval-induced forgetting would occur. Regardless of the particular mechanism proposed to explain retrieval-induced forgetting (see Raaijmakers & Jakab, 2013; Storm, 2011), it is hard to reconcile that effect with the proposal that many semantic associates become activated during retrieval.

The aforementioned concerns are logical and conceptual in nature. As noted above, an advantage of the semantic elaboration account is that it proposes a candidate mechanism for retrieval practice effects that can be induced and experimentally tested. If semantic elaboration is the mechanism responsible for retrieval practice effects, then inducing semantic elaboration directly should produce the same or similar effects as practicing retrieval. Unfortunately, the data from experiments comparing elaboration to retrieval practice have not supported this straightforward prediction.

Some experiments have shown that elaboration tasks do not produce the same results as retrieval practice tasks, which is troubling if retrieval practice effects are presumed to arise from elaboration. For example, Karpicke and Zaromb (2010) had subjects either generate target words from fragments (like *eat*—*di*__) or practice retrieval by recalling the target words from a prior study episode. Active generation is often considered an elaborative task, and both tasks required subjects to produce the target words. Nevertheless, Karpicke and Zaromb consistently found that retrieving the prior occurrence of the target word produced greater effects on a later criterial test than did generating the words. (The Karpicke and Zaromb experiments are described in greater detail in Section 5.3.)

Karpicke and Blunt (2011) also compared the effects of practicing retrieval to the effects of completing an elaborative study task. They had subjects read educational texts and either practice retrieval, by writing down as much as they could remember in the absence of the texts, or create concept maps while viewing the text. Concept mapping is an elaborative study activity where subjects make a node-and-link diagram of the concepts in a set of materials. The task requires subjects to focus on the organizational structure of the material and draw connections among concepts (Novak & Gowin, 1984). Karpicke and Blunt showed that practicing retrieval produced more learning than elaborative concept mapping, a finding that is hard to reconcile with the idea that retrieval practice effects stem from semantic elaboration.

Additional experiments have provided more direct tests of the elaborative retrieval account by inducing exactly the type of semantic elaboration purported to occur during retrieval. In a series of experiments, Karpicke and Smith (2012) had subjects learn lists of word pairs and practice repeated retrieval of the items or engage in various elaborative study tasks. Two experiments examined the effects of forming interactive images of the word pairs. In a critical experiment (Experiment 3), in the repeated elaboration condition, subjects repeatedly generated semantic mediators that connected the cue and target words. The experimental procedure directly induced the type of elaboration proposed to occur during retrieval, and the prediction was that repeated elaboration would produce effects similar or identical to those produced by repeated retrieval practice. Figure 7.3 shows performance on a criterial test 1 week after the original learning phase. Repeated retrieval produced large gains in learning relative to repeatedly studying the items and relative to not reexperiencing the items (in the "drop" condition). Most importantly, repeated elaboration did not produce effects like those seen from practicing retrieval. Again, these data provide a direct challenge to the idea that semantic elaboration is the mechanism responsible for retrieval practice effects. Directly inducing semantic elaboration, in the way proposed by the elaborative retrieval hypothesis, did not produce effects like those produced by practicing repeated retrieval.

The semantic elaboration account faces a number of additional challenges as a general explanation for retrieval-based learning. Retrieval practice effects occur under conditions in which semantic elaboration seems unlikely—for example, when the tasks employ nonverbal materials, such as pictures, symbols, or faces (Carpenter & DeLosh, 2005; Carpenter & Kelly, 2012; Coppens, Verkoeijen, & Rikers, 2011; Kang, 2010), and when cue–target word pairs consist of identical words (e.g., *table–table*; Karpicke & Smith, 2012), which presumably obviate semantic elaboration or the generation of mediators. Even under these circumstances, where it is difficult to imagine how semantic elaboration would occur, the positive effects of repeated retrieval practice are still observed. Additionally, if semantic

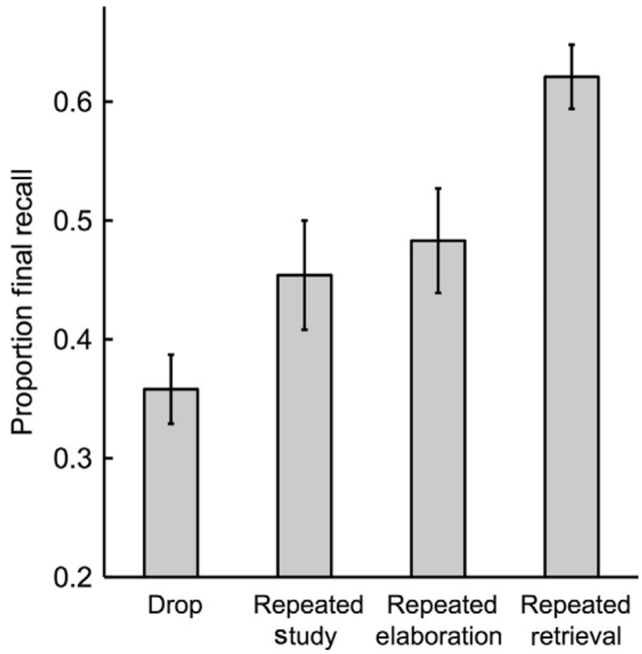

Figure 7.3 Proportion of words recalled on a criterial test in drop, repeated study, repeated elaboration, and repeated retrieval conditions. Repeated retrieval produced large gains in learning relative to repeated studying, whereas repeated elaboration did not. *Figure is reprinted from Karpicke and Smith (2012).*

elaboration occurs during initial cued recall and creates alternate retrieval routes from cues to targets, then the benefits of retrieval practice should be most evident when those cues are provided again on a criterial test. However, initial cued recall produces benefits on subsequent free recall (e.g., Karpicke & Zaromb, 2010; Smith et al., 2013), and in some cases initial cued recall does not improve final cued recall (Carpenter & DeLosh, 2006; Glover, 1989). It is also difficult to see how semantic elaboration from cues to targets would enhance subsequent recall and recognition of the cue words (Carpenter, 2011; Carpenter, Pashler, & Vul, 2006).

In describing encoding variability as an explanatory account of repetition and spacing effects, Greene (2008) wrote, "The notion that some variant of encoding variability underlies the spacing effect has been popular among theorists even in the absence of direct evidence that encoding variability benefits memory at all" (p. 74). A similar statement can be made for semantic elaboration: The idea that semantic elaboration underlies retrieval practice effects remains popular even though attempts to induce elaboration directly have not shown that it benefits memory in the same way as retrieval practice.

Although it may be true that some semantic information becomes activated during retrieval, a broad array of findings suggest that semantic elaboration does not offer a robust explanation of retrieval-based learning. It is important to note that these challenges pertain specifically to the assumption that the retrieval process produces *semantic* elaboration by activating several semantically related words. Perhaps retrieval produces some other form of "elaboration," but if so, such elaboration would need to be explicitly defined. In the next section, we propose an alternative account of retrieval-based learning, the episodic context account, that is based on established ideas in memory theory and that explains a wealth of evidence in the retrieval practice literature.

4. AN EPISODIC CONTEXT ACCOUNT OF RETRIEVAL-BASED LEARNING

4.1. Overview of the Episodic Context Account

Four basic assumptions drawn from memory models underlie the episodic context account. Most of these assumptions are shared by many general theoretical accounts of encoding and retrieval processes, and the context account applies these ideas to retrieval-based learning. We outline the four basic assumptions here and then describe the details of the account in more detail below. First, we assume that events occur within a slowly changing representation of temporal context, and that people encode information about items/events and about the temporal context in which events occur (Howard & Kahana, 2002; Lehman & Malmberg, 2013; Raaijmakers & Shiffrin, 1981). Information about events is stored as an incomplete copy of lexical/semantic item features and temporal context features (Shiffrin & Steyvers, 1997). Second, during retrieval, subjects use cues available in the present to attempt to reconstruct what occurred in the past. When the context during retrieval has changed significantly from the context during study, subjects attempt to *reinstate* the temporal context associated with the study period and use the reinstated temporal context features to guide a search process (Lehman & Malmberg, 2013). Third, when an item that was studied in a past context (context A) is retrieved in the present (context B), the context representation associated with that item is *updated*, such that it includes a composite of A and B context features (e.g., Lohnas & Kahana, in press). Finally, on a later test, subjects again attempt to reinstate context in the service of retrieval, and the updated context representation allows subjects to *restrict their search set*, the set of items considered as candidates for retrieval (Raaijmakers & Shiffrin, 1981). Because items that were studied in context A and retrieved in context B are associated with features of both contexts, the reinstatement of features from context A, context B, or both will serve as effective retrieval cues. When items are repeatedly retrieved in multiple temporal contexts, they become associated with a variety of contextual features that serve as effective retrieval cues on later tests. Ultimately, repeated retrieval may produce a decontextualization process wherein items become more retrievable but are no longer only associated with a specific context (e.g., the original study context).

We propose that these four basic assumptions allow the episodic context account to explain a variety of findings in the retrieval-based learning literature, including those that have been used to support other accounts. As described in more detail below, some of these processes may be differentially affected by the nature of a retrieval practice task, such that more or less contextual reinstatement may occur depending on the demands of the retrieval situation. The context account offers an explanation for the differential effects of retrieval-based learning on different types of tests and retention intervals. The following sections expand on the assumptions of the context account.

4.2. Context Representation and Encoding

The first assumptions of the context account are based in models of memory that assume a role of episodic context. The context account assumes that when material is studied, information is encoded about the items that were encountered in an incomplete and error-prone manner (Shiffrin & Steyvers, 1997). This is assumed to include, but is not necessarily limited to, semantic and phonetic features of the words. Importantly, information about the context in which the event occurred is also stored (Atkinson & Shiffrin, 1968; Howard & Kahana, 2002; Klein, Shiffrin, & Criss, 2007; Lehman & Malmberg, 2009, 2013; Mensink & Raaijmakers, 1989; Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997). Thus, memory representations include both item information and context information.

The term "context" can refer to a variety of aspects of an event, including the external environment (e.g., Smith, 1979) and a person's internal mental state (Klein et al., 2007). The episodic context account focuses on the importance of temporal context, a representation of context that changes with the passage of time (Howard & Kahana, 2002). In general, temporal context features are assumed to change at a slow pace (Bower, 1972; Estes, 1955; Howard & Kahana, 2002; Lehman & Malmberg, 2013; Mensink & Raaijmakers, 1989); however, rapid changes in context can occur when there are significant changes in tasks, goals, or setting

(Lehman & Malmberg, 2013; see too Jang & Huber, 2008). The way context features change from one moment to the next differs across models. Some models assume that context features drift in a random manner (Lehman & Malmberg, 2009, 2013; Mensink & Raaijmakers, 1989), whereas retrieved-context models assume that context change is driven by the retrieval of preexperimental contexts that are associated with items (Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008). The episodic context account offered here is neutral about the exact nature of temporal context change. Regardless of the type of contextual drift one assumes, the key point is that context on trial n is more similar to the context on trial $n+1$ than it is to the context on trial $n+10$. The contexts associated with events that occur in close proximity are likely to be very similar, whereas the contexts associated with events repeated at longer lags will be less similar (Howard & Kahana, 2002; Klein et al., 2007; Mensink & Raaijmakers, 1989).

4.3. Context Reinstatement During Retrieval

During retrieval, a person's goal is to remember what occurred at a particular place and time in the past, using cues available in the present. Consistent with most memory theories, the context account assumes that retrieval is accomplished by comparing available cues to the contents of memory. Memory representations with features that match those of the retrieval cues are assembled into a set of potentially recallable items called a *search set* (Raaijmakers & Shiffrin, 1981). Retrieval is determined by the diagnostic value of retrieval cues, the ability of a cue to uniquely specify a target to the exclusion of competing candidates (Nairne, 2002, 2006; Raaijmakers & Shiffrin, 1981). In other words, the effectiveness of a cue in eliciting a target is positively related to the match between the target and the cue and negatively related to the match between the cue and other candidates stored in memory. Because a search set will likely include both target and nontarget candidates, successful retrieval of a desired target will be most likely when the search set has been restricted (Raaijmakers & Shiffrin, 1981; Wixted & Rohrer, 1993, 1994; Watkins & Watkins, 1975).

While some tasks, such as cued recall, provide subjects with cues with which to probe memory, others, such as free recall, require subjects to generate their own cues. It is worth reiterating that the most robust retrieval practice effects tend to be observed under free recall conditions. When few cues are provided in the retrieval environment, the reinstatement of temporal context plays an especially critical role in retrieval. When temporal context is used as a cue, traces that are most likely to be retrieved are those associated with context features similar to the current retrieval context. Due to contextual drift, the retrieval context will likely be different from the study context. However, because context drifts slowly, the context cue during immediate retrieval will likely still match features of the context associated with traces in memory. As the length of the delay between study and retrieval increases, more contextual drift occurs and the retrieval context is less similar to the context associated with list items. Thus, in order to accomplish a recall task, some of the temporal context information that was present during study must be reinstated to serve as a retrieval cue (Lehman & Malmberg, 2013). The more context has shifted between study and retrieval, and the fewer other cues available, the more contextual reinstatement must be relied on as a retrieval strategy.

4.4. Context Updating During Successful Retrieval

The primary assumption that drives the episodic context account is that retrieval of an item updates the context representation stored with that item, making the retrieved item more recallable in the future. During subsequent retrieval attempts, when temporal context has drifted from the study context, items with updated context representations will be more retrievable via temporal context cues relative to items that have not been updated. Context updating creates a set of items that are more distinctly associated with future temporal contextual cues.

During retrieval in context B, the study context (context A) must be reinstated in order to recall the studied items (Lehman & Malmberg, 2013). When items are successfully retrieved on an initial test (in context B), features from the reinstated context A and the current context B are added to a composite context representation. Thus, the context representations of successfully retrieved items contain features that are associated with both contexts A and B, such that the reinstatement of either context A or context B in the future will serve to evoke the item from memory.

When context has changed very little between study and test (e.g., during massed retrieval), reinstatement of the study context will be less helpful (or unnecessary) to accomplish retrieval, given that the reinstated features will be similar to the current test features. Thus, the retrieved items will enjoy limited benefits from additional encoding of contextual features given that many of these features may be redundant. However, when the study and test contexts are different (e.g., during spaced retrieval), unique context features are added during successful retrieval. During future retrieval, context cues that are similar to A or B will match the context associated with the item. The longer the lag between study and initial retrieval, the more context change occurs, and the more distinct the additional context features will be for successfully retrieved items. As discussed in more detail below, this account of retrieval-based learning is similar to contextual accounts proposed to explain the spacing effect (e.g., Delaney et al., 2010; Lohnas & Kahana, in press; Lohnas, Polyn, & Kahana, 2011; Raaijmakers, 2003). According to contextual variability accounts of spacing effects, spaced repetitions occur in more varied contexts than massed repetitions, producing a larger set of retrieval cues that will be potentially effective for cueing the target information. According to the context account of retrieval-based learning, repeated retrieval leads to the updating of context features each time items are retrieved. The end result is that repeatedly retrieved items are associated with multiple contexts, producing a context representation that will match a variety of context cues.

4.5. Retrieval Practice Restricts the Search Set

The context account proposes that context reinstatement is used to guide a search process and that context is used to restrict the search set, the subset of items treated as candidates during retrieval (Raaijmakers & Shiffrin, 1981). Memory performance will be best when retrieval cues uniquely specify a target to the exclusion of competing candidates; performance can be improved when the cue–target match is increased, when the match between cues and competitors is decreased, and when the size of the search set is restricted to fewer candidates. The episodic context account assumes that the context features associated with successfully retrieved items are effective in uniquely specifying those items because updated context representations help subjects restrict the search set. When the context representation is a composite of A and B features, and subjects attempt to retrieve the items again on a future criterial test, they can restrict their search only to items associated with both A and B contexts. Because only previously retrieved items have representations associated with both context A and context B, the search set can be restricted to those items only. The distinct set of context features associated with retrieved items may aid in list discrimination (e.g., Chan & McDermott, 2007; Szpunar, McDermott, & Roediger, 2008; Verkoeijen, Tabbers, & Verhage, 2011) because items that are strongly associated with contexts A and B (i.e., those that were studied in context A and retrieved in context B) will be likely to be retrieved in response to a cue that includes features of contexts A or B, but unlikely to be retrieved in response to a retrieval cue that includes features of context C (a context in which they did not occur).

The context reinstatement process that occurs during retrieval practice may produce additional effects on the representation of context. First, as predicted by retrieved-context models, the reactivation of context A that occurs during retrieval may cause the current context to become updated with features from the retrieved context A (Howard & Kahana, 2002). In addition, the reinstatement of context A may facilitate future reinstatement of context A. In either case, items that are retrieved in context B via the reinstatement of context A may benefit not only because an updated composite context is created but also because the context features that are reinstated on an initial test are likely to be reinstated and used as cues on future tests. Indeed, recent evidence suggests that practice reinstating a specific environmental context may facilitate later contextual reinstatement necessary to accomplish a recall task, which Masicampo and Sahakyan (in press) have referred to as facilitated reinstatement (see also Brinegar, Lehman, & Malmberg, 2013). If context reinstatement is facilitated on a future test, then the context used as a cue will be more likely to contain features that match those stored with items that were previously retrieved, and the search set will be restricted to those items.

5. EVIDENCE SUPPORTING AN EPISODIC CONTEXT ACCOUNT

5.1. Effects of Retrieval Practice and Initial Retrieval Difficulty

The episodic context account explains the basic retrieval practice effect in terms of context reinstatement, context updating, and restriction of the search set. When people practice retrieval, they attempt to reinstate a prior learning context as they search for and try to recover items, and when retrieval is successful, the representation of context is updated to include features of the retrieved context and the current context. When people attempt retrieval again in the future, they are now better able to restrict the search set and hone in on the desired target by virtue of the updated context representation. In principle, these processes can and sometimes do occur during study events (i.e., study-phase retrieval). If subjects study material and are reminded of a prior occurrence, we assume that processes of context updating and search set restriction may occur while studying. Without intentional retrieval instructions, study-phase retrieval is not obligatory in a repeated study condition. Similarly, we assume that there is a difference in the degree of context updating that occurs with incidental retrieval, which is a person's "mode" during study-phase retrieval, relative to intentional retrieval, which occurs during retrieval practice as people deliberately search memory for information about the prior occurrence of a learning episode. Thus, the context account specifies underlying mechanisms that produce the mnemonic benefits of retrieval practice.

The context account also offers explanations for the effects of "difficult" retrievals outlined earlier (see Table 7.1). First, although retrieval practice effects occur with a variety of initial retrieval formats, the effects are most robust and observed most consistently when initial retrieval involves free recall. This general finding is consistent with the context account: Free recall requires subjects to reinstate a prior context with minimal cues, whereas other retrieval situations (such as recognition tests) may not require as much context reinstatement. Similarly, in cued recall tasks, the nature of the available cues will determine the degree to which subjects must attempt to reinstate context. For instance, practicing retrieval with only the first letter of a target produces a greater gain relative to practicing retrieval with three letters of the target provided (Carpenter & DeLosh, 2006). The context account proposes that subjects must reinstate more episodic context when fewer cues are provided, and greater recollection of the episodic context drives the gains in learning. The context account offers a ready explanation for the general advantage of recall-like retrieval practice conditions.

Second, the context account explains why a spaced initial retrieval produces more learning than massed retrieval. Consider what happens during massed retrieval practice: When an item is studied and successfully retrieved immediately after the occurrence, the temporal context has changed very little between study and retrieval. Context reinstatement may not be necessary, but even if it occurs, context updating under massed retrieval conditions will produce a context representation essentially like one that would exist without retrieval practice. On the contrary, when successful retrieval is spaced relative to a prior study episode, and retrieval occurs in a context that is substantially different from the study context, the context representation is updated to include features of the retrieved and current context. Subjects can then use the distinctive updated context to guide the search process when they attempt retrieval again in the future.

Third, and importantly, the context account readily explains why retrieval with weak semantic cues would produce a greater mnemonic effect relative to retrieval with strong cues, a finding that has been taken as key evidence favoring a semantic elaboration account (Carpenter, 2009). When cues are strongly associated to targets based on preexperimental features such as semantic relatedness, retrieval can be accomplished without much reliance on reinstating episodic context. For instance, if the strong associate *table* is given as a cue to recall *chair*, the target may come to mind easily due to its strong semantic association rather than because of the recollection of episodic context. In contrast, when cues and targets have little or no preexperimental association, reinstatement of prior context is obligatory. To recall *chair* when given a weak associate like *glue* as the retrieval cue, subjects must reinstate episodic occurrence information about when they studied the pair *glue–chair*. The episodic context account explains the advantage of practicing retrieval with weak cues in terms of context reinstatement rather than semantic elaboration purported to occur during retrieval (Carpenter, 2009).

Thus, the context account offers explanations for the general advantage of retrieval practice over repeated studying, the advantage of initial recall versus recognition, the advantage of spaced versus massed initial retrieval, and the advantage of retrieving with weak associates relative to retrieving with strong associates. A strength of the context account is that it specifies mechanisms (degree of context reinstatement and context updating) for the effects of "difficult" initial retrievals. Difficult retrieval conditions are ones that require greater context reinstatement. We hasten to note that similar ideas have been expressed by previous authors. Glover (1989) suggested that benefits of retrieval were dependent on the "completeness" of the retrieval event and Dempster (1996) proposed that "the effectiveness of an intervening test was an inverse function of the availability of retrieval cues" (p. 33). Each statement is essentially about the degree of context reinstatement required during a retrieval opportunity. Importantly, not all "difficult" retrieval situations benefit subsequent retention. For example, dividing attention during retrieval makes retrieval practice more difficult but does not increase the mnemonic benefit of retrieval (Dudukovic, DuBrow, & Wagner, 2009; Gaspelin, Ruthruff, & Pashler, 2013). It is not "difficulty" *per se* that enhances learning; it is the degree to which retrieval practice requires context reinstatement.

The episodic context account offers explanations for several additional effects in the retrieval practice literature, listed in Table 7.2. Specifically,

Table 7.2 Additional Effects of Retrieval Practice that Support the Episodic Context Account of Retrieval-Based Learning

Retrieval practice enhances subsequent context recollection
1. Retrieval practice enhances "Remembering" on a criterial test
2. Retrieval practice enhances recollection on a criterial test, as measured by process dissociation
3. Retrieval practice enhances temporal source memory and list discrimination on a criterial test
4. Retrieval practice reduces the effects of proactive interference
5. Retrieval practice produces a restricted search set and faster response times on a criterial test
Initial context retrieval during retrieval practice enhances subsequent retention
6. Intentional retrieval (being in an episodic retrieval mode) produces greater retrieval practice effects relative to incidental retrieval
7. Reinstating the initial study context during retrieval practice enhances retention
8. Recalling temporal context during initial retrieval practice enhances retention

the context account makes predictions about recollection of context on the criterial test and about the role of retrieving context during initial retrieval practice. We discuss these topics in turn in the next sections.

5.2. Recollection of Context on a Criterial Test

The episodic context account makes specific predictions about what types of final criterial tests will be sensitive to the effects of retrieval-based learning. Because retrieval-based learning occurs when contextual information is reinstated and updated, criterial tests that rely on the use of contextual information should be more sensitive to retrieval practice effects relative to criterial tests that can be accomplished without reliance on temporal context. Indeed, final free recall tests, which involve probing memory with context cues, are more sensitive to the effects of prior retrieval practice relative to final recognition tests, where performance can be accomplished by familiarity or automatic retrieval rather than recollection of context (see Darley & Murdock, 1971; Hogan & Kintsch, 1971; Glover, 1989). However, recognition tests that require subjects to use temporal contextual information ought to be sensitive to prior retrieval practice. This prediction has been confirmed in a variety of studies that we review here.

In an important paper, Chan and McDermott (2007) examined the effects of initial retrieval practice on final recognition memory tests that assessed context recollection, either by estimating subjects' recollection or by directly requiring subjects to recollect prior context. Subjects studied lists of words and then either freely recalled the lists or completed a distracter task. Across a series of experiments, Chan and McDermott examined performance on final recognition tests that assessed recollection with process dissociation (Jacoby, 1991), with source memory judgments (Johnson, Hashtroudi, & Lindsay, 1993), or with remember/know judgments (Tulving, 1985). Chan and McDermott showed that, according to all of these measures, practicing retrieval enhanced context recollection on the criterial test (see too McDermott, 2006; Verkoeijen et al., 2011). These findings are consistent with the episodic context account: Because people reinstate and update context as they practice retrieval, the ability to recollect context is improved on future tests.

In a similar line of research, Karpicke, McCabe, and Roediger (2006) examined the effects of repeated retrieval practice with a variant of process dissociation developed to estimate recollection and automatic retrieval in free recall (McCabe, Roediger, & Karpicke, 2011). Subjects studied four lists, each containing 20 items from a single category (e.g., 20 four-legged animals). One group repeatedly studied the lists four times, while a second group studied once and repeatedly recalled the lists in three consecutive free recall tests (see Roediger & Karpicke, 2006b). The subjects then took final tests either immediately (at the end of the session) or 1 week after the learning phase. In the criterial test phase, the subjects took inclusion tests for two lists (two categories): They were told to recall as many studied words as they could, guessing when necessary in order to produce 20 responses. Subjects took exclusion tests for the other two lists/categories: They were instructed to produce 20 new category members that they had not studied in the original learning phase. The key data from the exclusion tests are the proportions of exclusion errors—words from the studied lists that are mistakenly produced on the test. The results of the inclusion and exclusion tests can be combined to obtain estimates of recollection and automatic retrieval (Jacoby, 1991; see McCabe et al., 2011 for details).

Figure 7.4 shows the key results of the experiment on the 1-week delayed final tests. On the immediate final tests, repeated studying produced better performance than repeated free recall on the inclusion tests (77% vs. 68%; see Roediger & Karpicke, 2006b). Performance under exclusion instructions was near floor, so the process dissociation procedure could

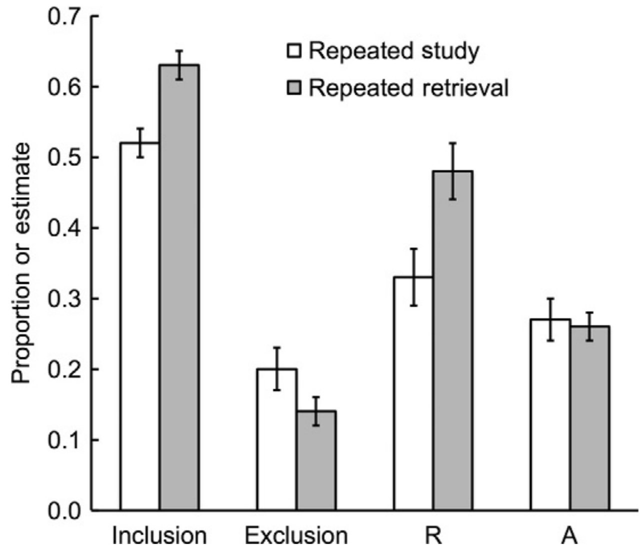

Figure 7.4 The left portion shows the proportion of words correctly produced on an inclusion test and the proportion of words incorrectly produced on an exclusion test. The inclusion and exclusion test data were used to calculate estimates of recollection (R) and automaticity (A) using Jacoby's (1991) process dissociation procedure. Repeated retrieval practice selectively enhanced the recollection estimate and did not affect the automaticity estimate. *Data are from Karpicke et al. (2006).*

not be applied. Thus, the analysis is focused on the 1-week data, which show that subjects in the repeated retrieval condition recalled more than subjects in the repeated study condition on the final inclusion test. In addition, subjects in the repeated retrieval group made fewer exclusion errors than subjects in the repeated study group. Inclusion and exclusion performance can be combined to obtain estimates of recollection and automatic retrieval, and Fig. 7.4 shows repeated retrieval practice selectively enhanced the recollection estimate and did not affect the automaticity estimate. The results suggest that repeated retrieval improved subjects' abilities to recollect which items occurred in the original learning phase so that they could include them on the inclusion test and also correctly exclude them on the exclusion test.

The episodic context account emphasizes that while the recollection of temporal context is essential for retrieval, other surface features of contexts are not as important for remembering occurrence information, and therefore, these features are not likely to be reinstated or updated during retrieval practice. Brewer et al. (2010) provided important evidence about the types of contextual information that are enhanced with retrieval practice. Subjects studied two lists that contained words spoken by male or female speakers. They then either freely recalled each list or completed a distracter task (similar to Chan & McDermott, 2007). On the criterial test, the subjects were shown each word and asked to identify either the list that the word came from (a list discrimination task, which requires temporal judgments) or the gender of the voice that read the word. Initial retrieval practice improved final list discrimination performance but did not enhance the ability to remember the gender of the person who spoke the word. It is noteworthy that similar research on the generation effect has shown that generating words disrupts memory for certain contextual details such as the color or font of a word (see Mulligan, Lozito, & Rosner, 2006). Brewer et al.'s results are consistent with the episodic context account, which proposes that subjects rely on temporal context cues to accomplish free recall, that elements of temporal context were updated during retrieval practice, and that enhanced memory for temporal context is evident on criterial tests.

The research described so far suggests that practicing retrieval enhances the ability to recollect what occurred at a particular place and time. Another way to examine this ability is to have people study several lists of words and instruct them to recall only the last list. The ability to constrain retrieval only to the most recent items, excluding items that occurred on earlier lists, is an index of the degree to which people are able to reinstate and restrict their search to specific prior context. Szpunar et al. (2008) devised a procedure to examine the role of retrieval practice in this way, and we have recently carried out experiments using this method to test predictions of the episodic context account (Lehman, Smith, & Karpicke, in press). Lehman et al. (in press) had subjects study five lists and then freely recall the last (fifth) list. In a control condition, subjects studied and performed a brief distracter task between lists. In the retrieval practice condition, subjects studied and then freely recalled each list after studying it. Finally, in an elaboration condition, after studying each list, the subjects were shown with the words and instructed to generate semantic associates for each word. This task was aimed at inducing the type of elaboration proposed by the semantic elaboration account (Carpenter, 2009; Karpicke & Smith, 2012). Following the fifth list, subjects in all conditions took a criterial recall test on which they were told to recall only words from the most recent (fifth) list. The subjects also completed a final free recall test over all lists at the end of the experiment, but here we focus on the data from the fifth list free recall test.

Figure 7.5 shows the proportion of words recalled from the fifth list (correct recall) and from prior lists (prior-list intrusions). Retrieval practice enhanced correct recall, relative to the control condition, and almost eliminated recall of prior-list intrusions. Importantly, semantic elaboration did

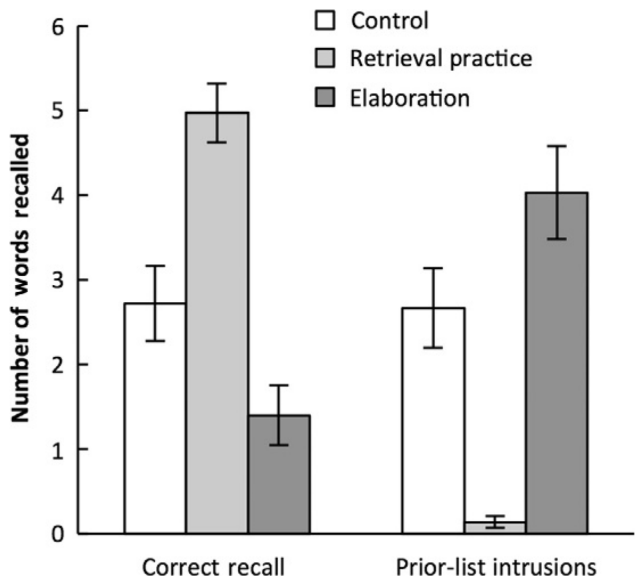

Figure 7.5 The left portion shows the proportion of words correctly recalled, and the right portion shows the proportion of intrusions from prior lists under control, retrieval practice, and elaboration conditions. Retrieval practice enhanced correct recall, relative to the control condition, and all but eliminated recall of prior-list intrusions. Semantic elaboration, on the contrary, reduced correct recall and increased the recall of intrusions. *Data are from Lehman et al. (in press).*

not produce results like those produced by retrieval practice. In fact, elaboration reduced correct recall and increased the production of prior-list intrusions relative to the control condition. Lehman et al. (in press) also examined cumulative recall during the recall period as an indicator of the size of subjects' search sets (Wixted & Rohrer, 1994). When subjects recall from smaller, more restricted search sets, there is an early and rapid approach to asymptote during recall, whereas when subjects recall from larger search sets, there is a slower and more gradual approach to asymptote. Figure 7.6 shows cumulative recall, which includes recall of correct items and intrusions. An analysis of the retrieval dynamics depicted in Fig. 7.6 confirmed that retrieval practice produced a restricted search set, whereas semantic elaboration led to an expanded search set (see too Bäuml & Kliegl, 2013). Overall, Lehman et al.'s results support the predictions of the episodic context account: Practicing retrieval required subjects to reinstate and update context representations, and this improved subjects' abilities to restrict their search to a particular context. Lehman et al.'s results are also important because they cast additional doubt on the idea that semantic elaboration operates in the same way as retrieval practice (Karpicke & Smith, 2012).

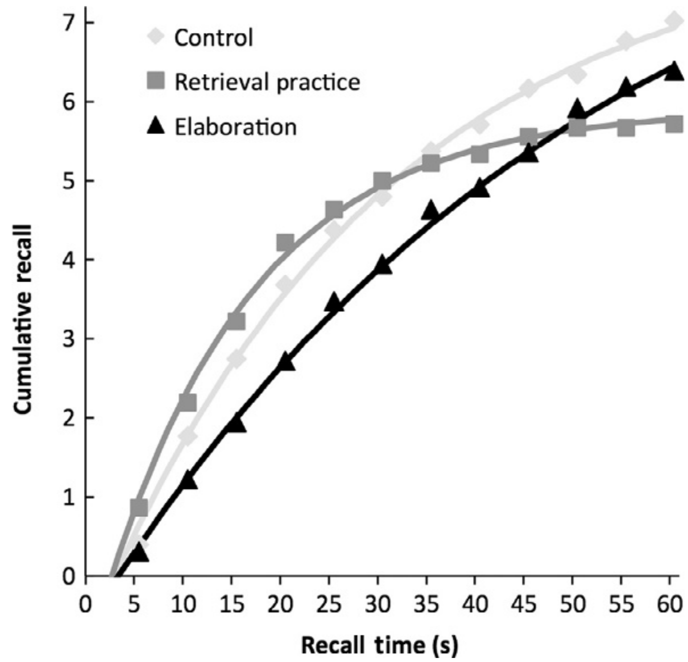

Figure 7.6 Cumulative recall curves showing the cumulative number of words recalled (correct recalls plus intrusions) under control, retrieval practice, and elaboration conditions. Retrieval practice produced a restricted search set, as evidenced by the early and rapid approach to asymptote. Semantic elaboration created an expanded search set, as evidenced by the slower and more gradual approach to asymptote. *Data are from Lehman et al. (in press).*

In sum, several lines of evidence show that practicing retrieval enhances the ability to recollect details of prior episodic contexts, which supports the episodic context account of retrieval-based learning. The next section reviews evidence that manipulating the recollection of episodic context during learning is critically important for retrieval-based learning.

5.3. Manipulations of Initial Context Retrieval

One criticism levied against the elaborative retrieval account is that there have been few attempts to test the theory by directly inducing the kind of semantic elaboration proposed to produce retrieval practice effects (and such attempts have not shown that semantic elaboration produces effects similar to retrieval practice effects; Karpicke & Smith, 2012). The episodic context account must be held to the same standard of evidence. Specifically, the context account leads to the prediction that manipulating the degree to which subjects recollect prior episodic context should be critical for producing retrieval practice effects. Indeed, there have been direct tests of this idea, and the results support the theory that remembering the episodic context matters for retrieval practice.

Karpicke and Zaromb (2010) carried out a series of experiments in which they manipulated whether subjects read (studied) target items, generated the target items, or retrieved the targets from a prior study episode. In their experiments, subjects first studied a list of target words (e.g., *love*). In a second phase, subjects either read the words paired with a related cue word (e.g., *heart—love*) or were given fragments of the target words (e.g., *heart—l_v_*). In a *generate* condition, the subjects were instructed to complete the word fragments with the first word that came to mind, whereas in a *recall* condition, the subjects were told to think back the study episode and recall a word that completed the fragment. This instruction directed subjects to reinstate the prior study context and placed them in what Tulving (1983) referred to as an "episodic retrieval mode." Importantly, there were no differences in the proportion of targets produced under the two conditions (subjects produced 70–75% of the targets in both conditions across a series of four experiments). In the third and final phase of the experiment, subjects freely recalled the target words, and Fig. 7.7 shows the proportion of words recalled on the criterial test. Whereas generating the targets produced no benefit relative to reading the words, retrieving the targets did produce

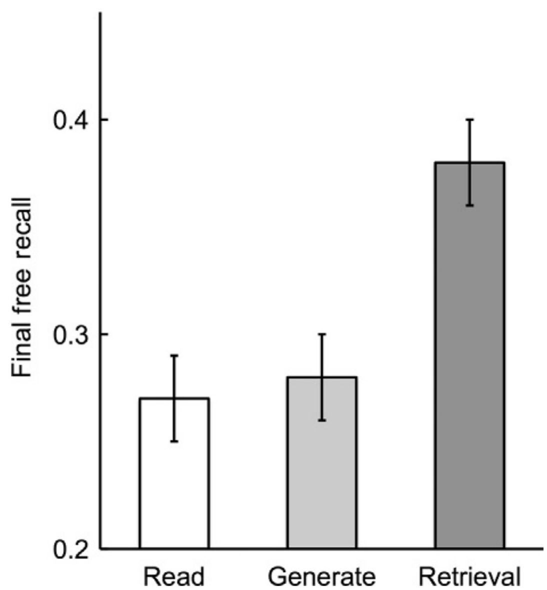

Figure 7.7 Proportion of words recalled on a criterial free recall test under read, generate, and retrieval practice conditions. Whereas generating the targets produced no benefit relative to reading the targets, retrieving the targets did produce an advantage on the final free recall test. *Data are from Karpicke and Zaromb (2010).*

a benefit on the final free recall test. (Karpicke and Zaromb also found similar benefits in other experiments where the criterial test involved item recognition.) The results are consistent with the episodic context account: Intentional retrieval, which involves thinking back to the study context, produced a larger effect on the criterial test than did incidental retrieval (generating the targets without thinking back to the study context).

Karpicke, Lehman, and Gallo (2014) examined the effects of initial retrieval orientation on subsequent recall. The general design was similar to the one used by Karpicke and Zaromb (2010). In the first phase of the experiment, subjects studied a list that contained a mixture of pictures and words presented in red ink (see Gallo, Weiss, & Schacter, 2004). In a second phase, subjects were shown a series of words (this time in black ink), some of which were old studied words and some of which were new, and the subjects were instructed to do one of three things. One group was instructed to form a mental image of each word, a task traditionally considered an elaborative study task. A second group took a standard yes/no recognition test. A third group took a "source constrained" recognition test: They were instructed to say "yes" only if they had studied the word previously as a picture (Gallo et al., 2004; Jacoby, Shimizu, Daniels, & Rhodes, 2005). Thus, both the standard and constrained recognition groups made recognition judgments, but subjects in the constrained condition were required to recollect details about the previous study context. Finally, at the end of the experiment the subjects freely recalled the words. Figure 7.8 shows the proportion of words recalled on the final test (collapsed across whether the items were originally presented as pictures or words). Making a recognition judgment produced greater final recall than did forming a mental image, which represents another demonstration of retrieval practice producing more learning than an elaborative study task (Karpicke & Blunt, 2011; Karpicke & Smith, 2012; Lehman et al., in press). Most importantly, the constrained recognition condition produced better performance than the standard recognition condition. Requiring subjects to recollect details from the study context produced the greatest effects on subsequent retention, a result that supports the episodic context account.

Finally, Whiffen and Karpicke (2013) tested a strong prediction of the episodic context account: If one were to hold all aspects of item presentation constant and manipulate only whether subjects were instructed to think about when an item had previously occurred in time, the act of making that temporal judgment should produce a retrieval practice effect on a later criterial test. Whiffen and Karpicke used a list discrimination task to require

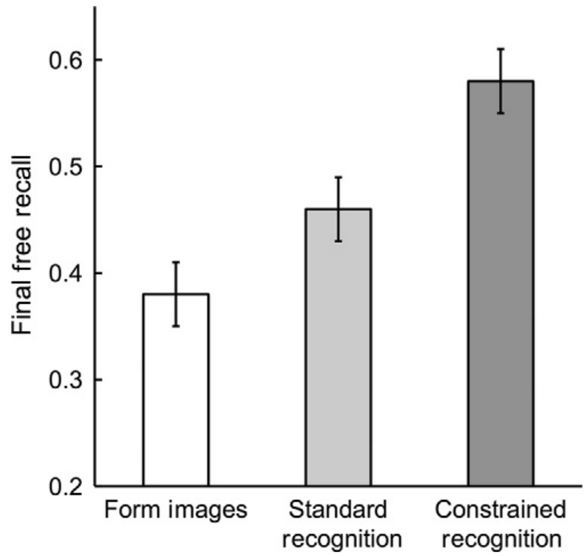

Figure 7.8 Proportion of words recalled on a criterial free recall test after forming mental images of the words, taking a standard yes/no recognition test over the words, or taking a constrained recognition test, in which subjects were asked to remember the original context of the study event. Recognition judgments produced better recall than forming mental images, and constrained recognition condition produced better performance than the standard recognition condition. *Data are from Karpicke et al. (2014).*

subjects to recollect prior temporal occurrence. Subjects studied two short lists of words, separated by a brief distracter task, and were then shown the words again with the two lists mixed together. In the restudy condition, the subjects were simply told to restudy the words in preparation for a final recall test, whereas in the list discrimination condition, the subjects were also told to indicate whether each word came from the first or second list (a list discrimination judgment). The subjects then freely recalled the words, and Fig. 7.9 shows the key results. The left portion of Fig. 7.9 shows the proportion of words recalled on the final test and shows that simply making a list discrimination judgment enhanced final recall. Whiffen and Karpicke also examined temporal clustering on the final test in terms of how often subjects grouped their output based on which list the word came from (this is done by calculating an adjusted ratio of clustering score with "list" as the grouping factor). The temporal clustering scores, shown in the right portion of Fig. 7.9, indicate that subjects relied on temporal order as an output strategy in the retrieval practice condition, clustering their output around list much more than subjects did in the list discrimination condition. The results confirm an important prediction of the context account: Simply thinking about the prior episodic context of an event enhances subsequent memory.

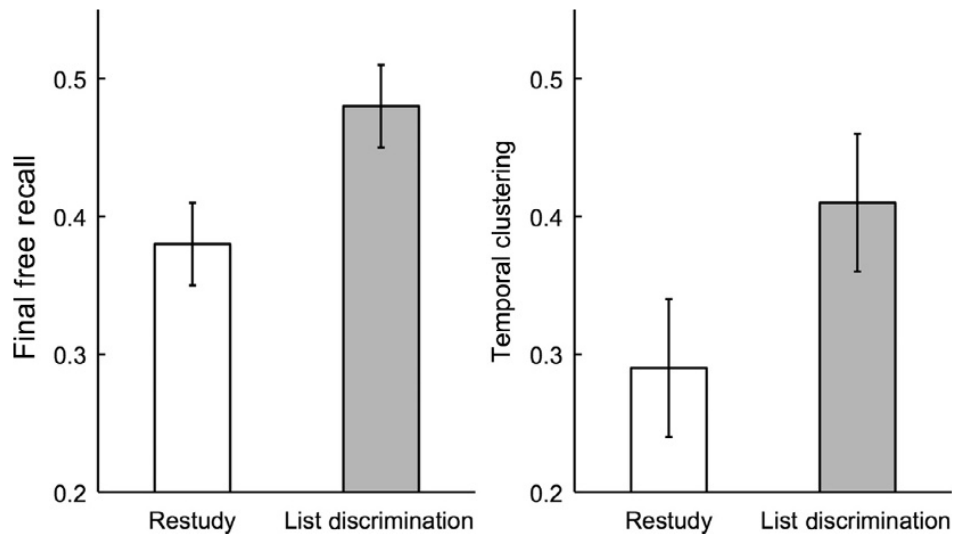

Figure 7.9 The left panel shows the proportion of words recalled on a criterial free recall test after studying the words or making temporal (list discrimination) judgments about the words. Simply making a list discrimination judgment enhanced final recall relative to the restudy condition. The right panel shows clustering scores, which represent the degree to which subjects organized their recall output temporally. Temporal clustering was more prevalent for the list discrimination condition indicating that subjects relied on temporal order as an output strategy relative to the restudy condition. *Data are from Whiffen and Karpicke (2013)*.

The research reviewed in the previous sections supports the episodic context account of retrieval-based learning. Practicing retrieval enhances performance on criterial tests that assess recollection of episodic context, and retrieval practice conditions that emphasize the reinstatement of prior context produce especially large effects on criterial tests. In the remainder of this chapter, we describe how the episodic context account relates to other general theories that emphasize context reinstatement and retrieval.

6. GENERAL DISCUSSION AND FINAL COMMENTS

6.1. Context Variability and Study-Phase Retrieval

The episodic context account presented here is closely related to some accounts that have been proposed to explain the spacing effect (for reviews, see Benjamin & Tullis, 2010; Delaney et al., 2010). The theoretical explanation for the spacing effect that has received the most empirical support attributes spacing effects to a combination of *contextual variability* and *study-phase retrieval*. In this section, we discuss the relation of these ideas to the episodic context account of retrieval-based learning.

According to contextual variability accounts of the spacing effect, the occurrence of a studied item in two different contexts produces a varied set of retrieval cues that are effective for eliciting the item on a later test (Glenberg, 1979; Melton, 1970; see also Lohnas et al., 2011). Because context drifts throughout the study of a list, items that occur in two or more positions on the list experience greater contextual variability than items that are studied in a massed fashion, and the greater the distance between two presentations, the more contextual variability occurs (e.g., Glenberg, 1979; Greene, 1989). Study-phase retrieval accounts propose that spacing effects occur when, upon the second presentation of an item, people retrieve the prior presentation and additional information is added to the memory trace when the prior presentation is retrieved (Greene, 1989; Hintzman & Block, 1973). According to accounts that combine contextual variability and study-phase retrieval, the magnitude of the spacing effect will be greatest when study-phase retrieval is difficult but not impossible, because such conditions produce the greatest contextual variability in encoded traces. For many decades, theoretical explanations of the spacing effect have assumed a role for temporal context (Glenberg, 1979; Greene, 1989; Hintzman & Block, 1973; Kahana & Greene, 1993). These accounts have also assumed that spacing effects will occur only when retrieval tasks require reliance on contextual cues (Greene, 1989; Kahana & Greene, 1993).

Recently, Raaijmakers (2003) proposed an account of spacing effects that implements contextual variability and study-phase retrieval assumptions in a formal model of memory (Raaijmakers & Shiffrin, 1981). In Raaijmakers's model, the likelihood of retrieving an item is a function of the item's context strength, which represents the degree of association between the item and the context cue used to probe memory. The model incorporates the assumption that context drifts over time and that as time passes the current context becomes less similar to the context associated with the studied item, thereby decreasing the trace's context strength (Mensink & Raaijmakers, 1989). Additionally, when a studied item is repeated, subjects may or may not retrieve the trace stored during the prior presentation. If the prior presentation is retrieved, then additional contextual elements are stored with the trace. Importantly, the larger the contextual change between repetitions, the stronger the link to context that is formed (i.e., the greater the increase in context strength).

Delaney et al. (2010) extended Raaijmakers (2003) account, proposing that the amount of additional contextual information that is encoded on spaced presentations during study-phase retrieval is a function of the lag between presentations of the items. When repetitions occur close together on a list, the strength of the item is high during the second presentation, and the increase in context strength will be small. However, when the repetitions are far apart on the list, the strength will be low during the second presentation, and the increase in context strength will be larger. Similarly, various other recent accounts of spacing effects have attributed the benefit of spaced study to some combination of contextual variability and study-phase retrieval (Benjamin & Tullis, 2010; Lohnas & Kahana, in press; Verkoeijen et al., 2004).

Study-phase retrieval/contextual variability accounts of spacing effects share several features with the episodic context account of retrieval-based learning proposed here. The accounts assume that context fluctuates over time, that retrieval produces extra contextual encoding that reflects variation in context, that the amount of variation in encoded context plays a role in the magnitude of the benefit of additional encoding, and that only retrieval tasks that rely on the use of contextual cues will be sensitive to the benefits produced by additional contextual encoding. It is surprising that the most prominent accounts of the spacing effect attribute the advantage of spaced study to contextual encoding driven by retrieval processes, yet none of the prominent accounts of retrieval-based learning (e.g., strengthening, retrieval effort, transfer-appropriate processing, and elaborative retrieval) attributes retrieval practice to contextual factors.

The episodic context account of retrieval-based learning extends the ideas of contextual variability and study-phase retrieval in a few notable ways. The context account emphasizes the importance of intentional rather than incidental retrieval. Whereas study-phase retrieval refers to incidental retrieval that occurs when a person notices or is reminded of a prior occurrence during restudy, retrieval practice refers to conditions where people must intentionally reinstate a study context and remember what occurred during a prior learning episode. The amount of contextual updating is likely greater for intentionally retrieved items than it is for incidentally retrieved items. The context account also assumes that retrieval creates an updated context representation that contains a composite of features from prior contexts and from the current context in which retrieval occurs. The context account further proposes that because context reinstatement is used to guide a search process, people can use distinctive contexts produced by repeated retrieval practice to restrict their search and thereby improve memory performance. In sum, the episodic context account is the natural extension of contextual variability and study-phase retrieval accounts of spacing effects, combined with assumptions from several foundational memory models (e.g., Howard & Kahana, 2002; Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997), to provide a comprehensive and cohesive account of retrieval-based learning.

6.2. Conclusion

The episodic context account is a broad and general account of retrieval-based learning. It provides an account of the key effects in the retrieval practice literature, it accommodates a wide variety of additional retrieval practice effects, and it helps explain some findings that are challenging for other accounts to explain. It is congruent with contextual variability and study-phase retrieval accounts of the spacing effect, and it is founded on assumptions from a variety of fundamental models of memory. Future theoretical work should aim to elucidate specific predictions of the account that can be empirically tested, and the assumptions of the context account remain to be implemented in a formal model. Finally, and perhaps most importantly, the episodic context account identifies the defining feature of retrieval-based learning as the successful remembering of prior learning experiences. The notion that people learn when they practice remembering prior learning experiences can inform future research on retrieval-based learning as well as the design of educational activities to promote student learning.

ACKNOWLEDGMENTS

The writing of this report and some of the research described was supported in part by grants from the National Science Foundation (DRL-1149363 and DUE-1245476) and the Institute of Education Sciences in the U.S. Department of Education (R305A110903). The opinions expressed in this article are those of the authors and do not represent the views of the National Science Foundation, the Institute of Education Sciences, or of the U.S. Department of Education.

REFERENCES

Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs*, *11*, 159–177.

Agarwal, P., Karpicke, J., Kang, S., Roediger, H., III, & McDermott, K. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology*, *22*(7), 861–876.

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89–195). New York, NY: Academic Press.

Bacon, F. (2000). *Novum organum*. L. Jardine & M. Silverthorne, Trans, Cambridge, England: Cambridge University Press (Original work published 1620).
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.
Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247.
Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.
Bjork, R. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition* (pp. 185–205). Cambridge: The MIT Press.
Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn,, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K Estes*: *Vol. 2*. (pp. 35–67). Hillsdale, NJ: Erlbaum.
Blunt, J. R., & Karpicke, J. D. (in press). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*.
Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Meltion, & E. Martin (Eds.), *Coding processes in human memory* (pp. 85–123). Washington, DC: Winston.
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.
Brinegar, K., Lehman, M., & Malmberg, K. J. (2013). Improving memory after environmental context change: A strategy of "preinstatement" *Psychonomic Bulletin & Review*, *20*, 528–533.
Butler, A. C., & Roediger, H. L., III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514–527.
Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *36*, 1118–1133.
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.
Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*, 279–283.
Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology*, *19*, 619–636.
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.
Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review*, *19*, 443–448.
Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, *13*, 826–830.
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.
Coppens, L. C., Verkoeijen, P. P. J. L., & Rikers, R. M. J. P. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology*, *23*(3), 351–357.

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*, 143–148.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology General*, *104*(3), 268–294.

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*, 66–73.

Deese, J. (1958). *The psychology of learning* (2nd ed.). New York, NY: McGraw-Hill.

Delaney, P., Verkoeijen, P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross, & B. H. Ross (Eds.), *The psychology of learning and motivation: Advances in research and theory*: *Vol. 53*. (pp. 63–147). San Diego, CA: Elsevier Academic Press.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork, & R. A. Bjork (Eds.), *Human memory* (pp. 197–236). San Diego, CA: Academic Press.

Duchastel, P., & Nungester, R. (1982). Testing effects measured with alternate test forms. *The Journal of Educational Research*, *75*(5), 309–313.

Dudukovic, N. M., DuBrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory & Cognition*, *37*(7), 953–961.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 145–154.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language*, *51*, 473–493.

Gardiner, F., Craik, F., & Bleasdale, F. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*(3), 213–216.

Gaspelin, N., Ruthruff, E., & Pashler, H. (2013). Divided attention: An undesirable difficulty in memory retention. *Memory & Cognition*, *41*(7), 978–988.

Glenberg, A. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.

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.

Greene, R. L. (2008). Repetition and spacing effects. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 65–78). Oxford, UK: Elsevier.

Greene, R. L., & Stillwell, A. M. (1995). Effects of encoding variability and spacing on frequency discrimination. *Journal of Memory and Language*, *34*, 468–476.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*, 505–513.

Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology*, *106*, 58–68.

Halamish, V., & Bjork, R. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology. Learning, Memory and Cognition*, *37*(4), 801–812.

Hanawalt, N. G. (1937). Memory trace for figures in recall and recognition. *Archives of Psychology (Columbia University)*, *216*, 89.

Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, *99*, 70–74.

Hogan, R., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 562–567.

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

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, *12*, 852–857.

James, W. (1890). *The principles of psychology*. New York: Henry Holt and Company.

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.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*, 621–629.

Kahana, M. J., & Greene, R. L. (1993). The effects of spacing on memory for homogeneous lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 159–162.

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, *38*(8), 1009–1017.

Kang, S., McDermott, K., & Roediger, H. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558.

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. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, *21*, 157–163.

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. *Educational Psychology Review*, *24*, 401–418.

Karpicke, J. D., Lehman, M., & Gallo, D. A. (2014). Retrieval-based learning: The role of initial retrieval orientation. Unpublished manuscript.

Karpicke, J. D., McCabe, D. P., & Roediger, H. L. (2006). Testing enhances recollection: Process dissociations and metamemory judgments. In *Poster presented at the 47th annual meeting of the psychonomic society, Houston, TX.*

Karpicke, J. D., & Roediger, H. L. (2007a). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*, 704–719.

Karpicke, J. D., & Roediger, H. L. (2007b). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162.

Karpicke, J., & Roediger, H. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*, 17–29.

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.

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III*. New York: Psychology Press.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*, 425–449.

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

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 989–998.

Landauer, T. K., & Bjork, R. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*: *Vol. 1*. (pp. 625–632). London: Academic press.

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(4), 970.

Lehman, M., & Malmberg, K. J. (2013). A buffer model of encoding and temporal correlations in retrieval. *Psychological Review*, *120*(1), 155–189.

Lehman, M., Smith, M. A., & Karpicke, J. D. (in press). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Lohnas, L. J., & Kahana, M. J. (in press). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology. Learning, Memory and Cognition*.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of Memory and Language*, *64*(3), 249–255.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*, 671–685.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, *75*, 421–441.

Masicampo, E. J., & Sahakyan, L. (in press). Imagining another context during encoding offsets context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. [Epub ahead of print].

McCabe, D. P., Roediger, H. L., & Karpicke, J. D. (2011). Automatic processing influences free recall: Converging evidence from the process dissociation procedure and remember-know judgments. *Memory & Cognition*, *39*, 389–402.

McDaniel, M., Anderson, J., Derbish, M., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4–5), 494–513.

McDaniel, M., & Fisher, R. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*(2), 192–201.

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., Roediger, H. L. I., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*(2), 200–206.

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, *34*, 261–267.

McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longmans.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, *33*, 172–186.

Morris, C. D., Bransford, J., & Franks, J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533.
Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *32*, 836–846.
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.
Nairne, J. S. (2010). Adaptive memory: Evolutionary constraints on remembering. In B. H. Ross (Ed.), *The psychology of learning and motivation*: *Vol. 53*. (pp. 1–32). Burlington: Academic Press.
Nelson, D. L., & McEvoy, C. L. (1979). Encoding context and set size. *Journal of Experimental Psychology Human Learning and Memory*, *5*(3), 292–314.
Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.
Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.
Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning and Verbal Behavior*, *22*, 133–152.
Putnam, A. L., & Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition*, *41*(1), 36–48.
Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437–447.
Raaijmakers, J. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431–452.
Raaijmakers, J. G. W., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, *68*, 98–122.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.
Roediger, H., & Butler, A. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.
Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.
Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre, & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education*: *Vol. 55*. (pp. 1–36). San Diego, CA: Elsevier Academic Press.
Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology Human Learning and Memory*, *4*(6), 592–604.
Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 460–471.
Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*, 1712–1725.

Storm, B. C. (2011). The benefit of forgetting in thinking and remembering. *Current Directions in Psychological Science*, *20*, 291–295.

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

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 796–800.

Verkoeijen, P. 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*(6), 490–498.

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*(4), 442–452.

Whiffen, J., & Karpicke, J. D. (2013). The role of temporal context in retrieval practice. In *Poster presented at the 54th annual meeting of the psychonomic society, Toronto, ON.*

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 465–478.

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.

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