

University of Massachusetts Amherst  
**ScholarWorks@UMass Amherst**

---

Masters Theses

Dissertations and Theses

---

July 2016

## The Primary and Convergent Retrieval Model of Memory

William J. Hopper  
*University of Massachusetts Amherst*

Follow this and additional works at: [https://scholarworks.umass.edu/masters\\_theses\\_2](https://scholarworks.umass.edu/masters_theses_2)



Part of the [Cognitive Psychology Commons](#)

---

### Recommended Citation

Hopper, William J., "The Primary and Convergent Retrieval Model of Memory" (2016). *Masters Theses*. 350.

[https://scholarworks.umass.edu/masters\\_theses\\_2/350](https://scholarworks.umass.edu/masters_theses_2/350)

This Open Access Thesis is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

**THE PRIMARY AND CONVERGENT RETRIEVAL MODEL OF MEMORY**

A Master's Thesis

by

WILLIAM J. HOPPER

Submitted to the Graduate School of the  
University of Massachusetts Amherst in partial fulfillment  
of the requirements for the degree of

MASTER OF SCIENCE

May 2016

Psychology

© Copyright by William J. Hopper 2016

All Rights Reserved

**THE PRIMARY AND CONVERGENT RETRIEVAL MODEL OF MEMORY**

A Master's Thesis

by

WILLIAM J. HOPPER

Approved as to style and content by:

---

David E. Huber, Chair

---

Jeff Starns, Member

---

Lisa S. Scott, Member

---

Hal Grotevant, Department Head  
Psychological and Brain Sciences

## **ABSTRACT**

### **THE PRIMARY AND CONVERGENT RETRIEVAL MODEL OF MEMORY**

MAY 2016

WILLIAM J. HOPPER, B.S., UNIVERSITY OF CALIFORNIA SAN DIEGO

M.S., UNIVERSITY OF MASSACHUSETTS

Directed by: David E. Huber

Memory models typically assume that recall is a two-stage process with learning affecting both processes to the same degree. This equal learning assumption is difficult to reconcile with studies of the 'testing effect', which reveal different forgetting rates following learning from test practice versus learning from restudy. Here we present a new memory model, termed Primary and Convergent Retrieval (PCR) that assumes successful recall leads to a selective enhancement for the second stage of recall (Convergent Retrieval). We applied this model to existing testing effect data. In two new experiments, we confirmed novel predictions of the PCR model for transfer between retrieval cues and for recall latencies. This is the first formally specified model of the testing effect and it has broad implications for the nature of learning and retrieval.

# TABLE OF CONTENTS

	Page
ABSTRACT .....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTER	
1. THE PRIMARY AND CONVERGENT RETRIEVAL MODEL OF MEMORY .....	1
1.1 Introduction .....	1
1.2 Retrieval and Learning in Prior Models of Memory .....	3
1.3 Study and Recall Produce Unequal Learning .....	7
1.4 An Overview of the Process Theory that Guides PCR .....	10
1.5 A Binomial Instantiation of PCR .....	13
1.6 Application to Data from Roediger & Karpicke 2006a .....	20
2. EXPERIMENT 1.....	25
2.1 Testing the cue-independent learning prediction of PCR .....	25
2.2 Method .....	25
2.2.1 Participants.....	25
2.2.2 Materials.....	25
2.2.3 Procedure.....	26
2.3 Behavioral Results .....	29
2.4 PCR Model.....	32
2.4.1 Implementation .....	32
2.4.2 Results .....	36
2.5 Discussion .....	37
3. EXPERIMENT 2.....	39
3.1 Testing the recall latency prediction of PCR.....	39
3.2 Methods .....	40
3.2.1 Participants.....	40
3.2.2 Materials.....	40
3.2.3 Procedure.....	41

3.3 Behavioral Results .....	42
3.4 PCR Model .....	45
3.4.1 Implementation .....	45
3.4.2 Results .....	47
3.5 Discussion .....	48
4. GENERAL DISCUSSION .....	52
4.1 Overview of Findings .....	52
4.2 Broader Implications .....	53
APPENDICES	
A. BETA APPROXIMATION TO THE BINOMIAL .....	75
B. APPROXIMATING THE REACTION TIME DISTRIBUTION .....	77
REFERENCES .....	80

## LIST OF TABLES

Table	Page
Table 1: Best fitting PCR model parameters as applied to Roediger and Karipicke, 2006, Experiment 1 .....	58
Table 2: Average best fitting parameters for the PCR model used in Experiment 1.....	59
Table 3: Average best fitting parameters for the PCR model used in Experiment 2.....	60



## LIST OF FIGURES

Figure	Page
Figure 1: Recall accuracy in Experiment 1 of Roediger and Karpicke 2006a. ....	61
Figure 2: Schematic diagram of primary retrieval .....	62
Figure 3: Schematic diagram of primary retrieval process at encoding .....	63
Figure 4: Schematic diagram of successful (left panel) and unsuccessful (right panel) convergent retrieval process. ....	64
Figure 5: Reaction time function in the PCR model .....	65
Figure 6: Recall accuracy performance predicted by the PCR model from Experiment 1 of Roediger and Karpicke 2006. ....	66
Figure 7: Outcomes of subsequent retrieval attempts following successful convergent retrieval and convergent retrieval learning. ....	67
Figure 8: Recall accuracy on the final cued recall test of Experiment 1. ....	68
Figure 9: Conditional probability of recall on final test in Experiment 1. ....	69
Figure 10: Joint distribution of recall outcomes on the final and practice tests in Experiment 1. ....	70
Figure 11: Free recall accuracy observed in Experiment 2. ....	71
Figure 12: Elapsed time (in seconds) between recall outputs in each condition of Experiment 2. ....	72
Figure 13: Free recall accuracy observed predicted by the PCR model. ....	73
Figure 14: The observed and predicted time (in seconds) between recall outputs in each condition of Experiment 2. ....	74

## CHAPTER 1

### THE PRIMARY AND CONVERGENT RETRIEVAL MODEL OF MEMORY

#### 1.1 Introduction

The efforts of psychologists to investigate explicit memory can be roughly divided into studies using two types of tasks: recognition tasks, which require a participant to judge a stimulus as previously encountered or novel, and recall tasks, which require a participant to generate detailed remembrances of a previous event's content or meaning. It is generally agreed upon that there is a process dissociation between recognition and recall such that the need to dredge up the details of the past must require an additional process compared to when the task is to judge something as old or new (Hintzman & Curran, 1994; Mandler, 1980). This need to successfully complete a two-stage process to recall information is thought to underlie the "tip of the tongue" (TOT) phenomenon, where the complete recall of a particular piece of information fails, but certainty about knowing the information is achieved (Brown & McNeill, 1966). Computational models of memory have long incorporated a two-stage recall process, but all of these models have assumed that learning must produce effects in both processes. Here we present a new memory model, the Primary and Convergent Retrieval model of memory, which includes the assumption of a two-stage recall process, but also posits separate learning in each stage of retrieval. This assumption allows us to make sense of findings from retrieval practice paradigms, such as studies of the testing effect (see Roediger &

Karpicke, 2006a), where different types of practice seem to have differential effects on retention of practiced material.

Before introducing PCR more formally, we broadly define the concepts of primary and convergent retrieval as they relate to recall, and then relate these concepts to prior models of recall and recognition. When faced with the need to recall information, a necessary first step is to isolate the relevant memories from the vast store of information accumulated over a lifetime of experience. In short, there is a need to filter out the unwanted memories (or conversely highlight the sought-after memories). This process is guided by the retrieval cues at hand, which assuredly include temporal context (i.e., the stuff that is currently on your mind), but may also include item information, such as when given an explicit retrieval cue (e.g., a word, face, or picture). Primary retrieval describes this initial process of isolating relevant memories, based on their associations with the retrieval cues. However, this process may result in a partial retrieval for the desired memory, and, furthermore, it may result in the retrieval of many other competing memories. Thus, before an overt recall response can be emitted, it is necessary to 'clean up' the content of the memory response, such that the response systematically converges on complete retrieval of one, and only one response. Convergent retrieval describes this second stage, which recovers the full content of a single memory from the noisy output of primary retrieval through a pattern-completion process.

## 1.2 Retrieval and Learning in Prior Models of Memory

Several previously proposed models of recall possess mechanisms analogous to Primary Retrieval (PR) and Convergent Retrieval (CR). The Search of Associative Memory (SAM, Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) model describes recall with two discrete stages, both dependent on the strength of the associative connections between the current retrieval cues and the memories in a long-term store. When a memory search is initiated, the current retrieval cues activate many stored traces simultaneously. To successfully output information from this long-term store (i.e., recall), a specific item must first be selected from the set of active traces (called the sampling process). The strength of the associative connections between a set of retrieval cues and a single memory trace, relative to other memory traces, determines if an item is likely to be sampled. When an item is sampled, the information it holds must be extracted (called the recovery process). The probability of successful recovery depends only the absolute value of the connection strength, not its value relative to other competing memories. Learning is modeled as an increase in the strength of the associative connections between cues and memory traces, which can occur whenever both are concurrently active in a short-term memory store (e.g., when given an opportunity to restudy a pair of words together, or following successful recall of a word in response to a word cue). Thus, any learning affects both the sampling stage and recovery stage alike, because the strength of the associative connections determines both the relative likelihood of sampling an item as well as the probability of recovering its details. The

Retrieving Effectively from Memory model (REM, Shiffrin & Steyvers, 1997)

extended the core ideas of cue-dependent sampling and recovery process for recall, but modeled the traces stored in long-term memory using vectors of many individual features instead of with a single abstract strength value.

Like the REM model, MINERVA 2 (Hintzman, 1984, 1988) represents items in the long-term store as integrated collections of features, and assumes that retrieval cues (called a “probe” in MINERVA) activate items stored in memory based on number of feature values that match between each stored item and the probe. Comparing the probe to the contents of memory returns a composite response from all items called the *echo*, which has two components: *intensity*, representing the overall level of activation in the entire memory system, and *content*, representing the amount of each unique feature's signal in the echo (Hintzman, 1988). Echo content is critical for a recall response, but when many items with different features are all activated by the probe, the resulting echo content may not be a good match to any one item in memory. The state of ambiguity can be resolved by iteratively redeploying the echo content as a new probe and measuring the new echo, until the echo content is “sharpened” to most closely resemble a single item. In MINERVA, learning increases the probability of correctly encoding a feature into the long-term store, which applies independently to each feature for each item trace. Because both the echo's intensity and content is determined by the degree of correspondence between probe and item features, learning affects both the process of determining that there *is* a relevant memory trace, as well as recovering what that trace's content is.

Norman and O'Reilly (2003) have developed a dual-process model of recognition memory behavior, based on the Complementary Learning Systems framework (McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & Rudy, 2001). This framework posits two specialized and anatomically distinct learning mechanisms: a hippocampal network which rapidly encodes the details of individual events and items, emphasizing their distinctiveness, and a slower learning cortical network (representing the structures of the medial temporal lobe) designed for encoding the similarities of items across episodes. The implementation by Norman and O'Reilly (2003) uses the amount of activation signal from the cortical network as an index of "familiarity" to a recognition probe, assuming frequently encountered stimuli (such as those presented in a memory experiment) will have strong representations in this signal. But due to the cortical network's emphasis on encoding similarities, it lacks the ability to represent details of a single item or event and thus cannot support recall; this responsibility falls upon the hippocampal network. In the hippocampal network, the recognition probe serves as a retrieval cue that partially activates the stored representations (the activation is only partial because changes in temporal context between the encoding episode and the retrieval attempt over the course of the retention interval mean that retrieval cues are not an exact match to the stored representation). Though only partially activated, an item still may be able to be retrieved thanks to positive associative links exist between the units that define individual items in the network. With sufficiently strong associations between the units defining an individual item, a partially activated item may be able

to “bootstrap” its way into complete activation via a pattern completion process between these units, therefore boosting its contribution to the content signal.

Learning from any situation makes the magnitude of memory system response larger and the content clearer under the Complementary Learning Systems model. Learning results in a sharpened signal from the cortical network, meaning a probe item seems more familiar, as well as a new episodic encoding in the hippocampal network. Multiple encodings in the hippocampal network means there will be more matching units between a probe and a stored item, so there will be stronger evidence to select that item for output relative to competitors. However, learning does not increase the ability of the hippocampal network to complete specific patterns from their partial activation.

This brief overview of previous memory models of recall serves to highlight their common assumptions. Each of these models assumes that the response from the memory system is two-fold, with one response component related to the overall magnitude of the match between retrieval cues and stored memory traces, and another content component holding detailed but often incomplete information about the stored trace. Additionally, each of these models assumes that the process of learning both increases the magnitude of the memory response as well as our ability to recover the specific content details in a memory trace. This second assumption, that any learning produced through experimental manipulations affects all aspects of the recall process, is difficult to reconcile with findings from studies using retrieval practice paradigms, where the primary focus is on how recalling information from memory affects subsequent memory performance. Below, we

summarize key findings from studies comparing the effects of study and test practice on memory retention, and explain why they are problematic for an assumption of uniform learning across retrieval processes.

### **1.3 Study and Recall Produce Unequal Learning**

Laboratory studies comparing the effects of study and test practice proceed by first giving participants some new material (e.g. a list of words) to learn by studying it. Then either the studied material or the participants are subdivided into different groups for further practice, either by restudying the material, or by taking a practice test. A variable delay period (the retention interval) follows this second practice session, after which all participants are given a final test of their memory. These studies consistently show a distinct advantage on the final test for material practiced with a test relative to material practiced by restudying, and this advantage grows with duration of the retention interval. This phenomenon is commonly referred to as the “testing effect” (see Roediger & Karpicke, 2006b for a recent review).

When the practice tests are administered without feedback, a surprising paradox is revealed: information studied by re-reading and by taking a practice test appear to be forgotten at different rates (Carpenter, Pashler, & Vul, 2006; Kuo & Hirshman, 1996; Roediger & Karpicke, 2006b; Toppino & Cohen, 2009; Wheeler, Ewers, & Buonanno, 2003). More specifically, when the retention interval between the practice phase and the final test is short (i.e. 5 minutes), there is a small overall advantage on the final test for restudying over taking a practice test. But, if the



retention interval is longer (i.e. 24 hours or one week), this relationship is reversed, and a much higher proportion of information practiced with a test can be remembered on the final test than information practiced by restudying. Thus, it appears that more can be learned through restudying practice, but that this information is quickly forgotten. Conversely, it seems less is learned overall by taking a practice test, but what is learned becomes extremely robust against forgetting. This result is problematic for the extant memory models because they do not differentiate between learning from testing versus restudy; thus, whichever type of practice produces the better performance for an immediate test should also produce the better performance in the long run. In short, this crossover interaction suggests that these two types of practice involve different encoding operations, with the type of encoding that occurs with test practice resulting in a memory trace that is more resistant to forgetting.

An important caveat to this conclusion arises when separately considering items that are correctly recalled during test practice versus those that are not. More specifically, the items that are correctly recalled during test practice reveal even higher performance on a final immediate test as compared to restudied items. Thus, when considering that only some items that might experience learning from test practice (i.e. correctly recalled items, with no accuracy feedback on incorrect trials), the rank ordering of conditions is the same across all time points: items recalled on a practice are remembered better than restudied items, which are remembered better than items that were not successfully recalled on the practice test.

This observation was made by Kornel, Bjork and Garcia (2011), and has been labeled 'the bifurcation model' of the testing effect. However, this account is somewhat unsatisfactory because it does not tell us *why* successful test practice should result in substantially more learning than restudy. Furthermore, this account ignores item differences and item selection effects; more specifically, items that were successfully recalled during test practice are likely to be items that are more easily recalled in general. The issue of item effects was investigated by Jang, Wixted, Pecher, Zeelenberg, and Huber (2012) by including a pre-test of all of the items before assigning items to either test practice or restudy. When only considering items that were retrieved on this pre-test (i.e., the 'easy' items), there was still a crossover interaction between type of practice and retention interval. Thus, it appears that this crossover interaction reflects a qualitative rather than a quantitative difference in the learning that occurs during test practice as compared to restudy. To explain this difference, and more generally to explain in detail the different stages of recall, we developed a process model in which the kind of learning from test practice is qualitatively different than the kind of learning from restudy.

Next we describe the PCR model in two parts. First, we describe the theoretical processes that explain how engaging in recall will result in different encoding operations as compared to passive study. This theory assumes that the ordering of events matters, such that the partial retrieval that occurs with Primary Retrieval, followed by full retrieval with Convergent Retrieval, results in new intra-item associations between the content retrieved in each stage of recall. In contrast,

this intra-item learning does not occur with passive study because a full appreciation of the item occurs all at once rather than in stages of retrieval. Second, following presentation of this theory, we describe a specific mathematical instantiation of this theory that captures the essential elements of the theory in an abstract manner.

#### **1.4 An Overview of the Process Theory that Guides PCR**

PCR assumes that all memory traces and retrieval cues are composed of a finite number of individual features. When features are simultaneously active, unidirectional associative connections between the features are formed, according to the temporal order in which the features became active. For instance, if feature A becomes active at some point in time  $T$ , and feature B becomes active at time step  $T+1$ , an excitatory connection is formed from feature A to feature B. Thus the next time feature A is activated, it will activate feature B. However, activating feature B will *not* activate feature A, as the connection between them is unidirectional. This feature is consistent with findings showing that the precise temporal ordering of pre and post synaptic action potentials determine the strength and direction of the modifications to synaptic weights (i.e. Spike Timing Dependent Plasticity (STDP), for examples see G. Bi & Poo, 1998; G.-Q. Bi, 2002; Dan & Poo, 2006). Importantly, these associative connections are formed both across perceptual groupings (i.e., across distinct items) and within groupings (i.e., intra-item).

As mentioned previously, the first stage of retrieving an item from memory is Primary Retrieval (PR). When information is to be retrieved from memory, all currently active features serve as retrieval cues to activate features in memory traces, with the features serving as cues being sourced from the current temporal context, or via cueing with an associated item. The pattern of activation in the memory system in response to the features in the retrieval cues follows a one-to-many relationship, where a single active feature may activate many other features within stored memory traces. Thus, PR can be said to be cue-dependent, as the active features serving as retrieval cue completely determine the magnitude and the content of the memory system's response. Figure 2 shows a schematic outline of the PR process in a retrieval episode.

This process of currently active features activating other features in stored memory traces is not restricted to occurring during a retrieval episode; it also occurs during opportunities to study or restudy items. By definition, studying presents the to-be-remembered information along with any associated retrieval cues in a specific temporal context, and thus positive associative connections are formed between all currently active features (a schematic diagram is shown in Figure 3). The main difference in between a restudy and a retrieval episode is that during a retrieval episode, not all features of the target memory are activated by the retrieval cues. For an accurate, overt recall response to be given, all features of a memory trace must be simultaneously active, and thus a pattern-completion processes is necessary to “fill in the blanks” of the inactive features.

We use the term Convergent Retrieval (CR) describes this pattern completion process, where features activated by the PR process subsequently activate initially dormant features “missed” by the PR process, via excitatory associative connections between the individual features within a memory trace. If the CR processes succeeds, and all features are successfully pattern-completed, learning takes place: new excitatory associations are formed between the initially activated features, and the subsequently activated features. A schematic diagram of the CR process is shown below in Figure 4. The convergent retrieval learning mechanism is the major theoretical departure of PCR from previously proposed memory models. Because CR learning strengthens the internal associations within the set of features defining a single item, that item will become easier to recall in the future, regardless of the route by which the initially active features became active. Put another way, successful CR depends on getting enough strongly interconnected item features activated by the PR process, not on how those features became active. Thus, the CR process and CR learning can be said to be cue independent.

So far, we have described the theoretical assumptions that underlie the PCR model. We now turn towards developing a specific model implementation of this theory, to examine whether it provides a sufficient explanation of data. Here we introduce an abstract implementation of the PCR model, with specific application to data from free recall experiments in “testing effect” paradigms, without implementing a full neural network model. Nevertheless, the constraints on this abstract model implementation are dictated by a more full-fledged process model involving features, weight changes, and the timing of events while learning.

## 1.5 A Binomial Instantiation of PCR

We assume each item encoded into memory consists of a finite number of discrete features, set here to 100 for convenience. We assume each item requires a specific number of features to be activated by retrieval cues in the PR process in order to support pattern completion in CR<sup>1</sup>. The value of each item's CR “threshold”,  $\theta$ , is binomially distributed with probability  $t$  and 100 trials (one trial for each feature). Throughout this section, we will use the symbol  $B(\cdot)$  in equations to refer to a binomial distribution.

$$\theta \sim B(t, 100) \tag{1}$$

At the point of initial study, when all cues and targets are presented together, we assume positive associative connections are formed from cue items and context features to the target items as part of the PR process. Re-presenting the cue items within the same temporal context without the target item will result in reactivation of a subset of the target item's 100 features. This amount of feature reactivation during the retrieval episode is also binomially distributed, with probability parameter  $e$  and 100 trials.

$$pr \sim B(e, 100) \tag{2}$$

---

<sup>1</sup> Using a fixed number of activated features to determine convergent retrieval success is a simplifying assumption made for the binomial instantiation of PCR. In general, the absolute number of features activated in the PR process is not strictly as important as which specific features become activated. There is likely to be an asymmetry amongst the intra-item feature associative connections, such that some features have more associations and are thus better suited to support CR, and we appeal to such an asymmetric throughout.

The parameter  $e$  may be thought of as an encoding rate parameter, as it represents how strongly the retrieval cues are linked to the target item during the encoding episode. It is important to note that we will use the abbreviation PR to refer to the primary retrieval *process* (i.e. the process of features from retrieval cues activating features in target memories and forming excitatory associations), and the lowercase italic  $pr$  in a mathematical context to represent the outcome of the PR process, i.e. the *number* of features in the target memory activated by the retrieval cues.

Both the number of features activated by retrieval cues and the pattern completion threshold for each item can be modulated over the course of time and with experimental manipulations. In general, the number of features of a target memory that are activated by retrieval cues will decrease over time, as changes in the temporal context mean that the target memory will not be strongly matched by context cues and as weights on the positive associative connections between cue items paired with the target item decrease. We chose to model forgetting from all sources through reducing  $pr$  by a binomially distributed random quantity  $F$ , with probability parameter  $f$  and  $pr$  trials.

$$F \sim B(f, pr) \tag{3}$$

Having the number of binomial trials for  $F$  set to  $pr$  reflects the assumption that only features that would have been activated initially can be lost due to forgetting. After forgetting, from any source,  $pr$  becomes equal to  $pr - F$ .

The value of  $pr$  is increased whenever the features of the retrieval cues and the features of the target memory become co-active, in that order. Thus, additional study practice or successful test practice increases  $pr$ . We model this additional

learning by increasing  $pr$  according to a binomially distributed random quantity,  $L$ , with a probability parameter  $l$ , and  $100-pr$  trials

$$L \sim B(l, 100 - pr) \quad (4)$$

Setting the number of binomial trials for  $L$  to  $100-pr$  reflects the assumption that only features that have *not* already been associated with the retrieval cues may become associated with the retrieval cues upon restudy or successful recall. The value of  $pr$  after successful test practice and  $pr$  after restudy practice is equal to  $pr + L$ . When an item is not successfully recalled, no PR learning takes place, and the value of  $pr$  is unchanged as a result of the retrieval attempt.

However, successful retrieval of an item does not only increase future  $pr$ , it also reduces the items pattern completion threshold  $\theta$ . This threshold reduction reflects the assumption that intra-item learning takes place following pattern completion in CR. Again, in this implementation the magnitude of the threshold reduction for each item is a binomially distributed random quantity  $R$ , with probability parameter  $t$  and  $\theta$  trials.

$$R \sim B(t, \theta) \quad (5)$$

Having the number of trials for CR threshold reduction set to  $\theta$  reflects the constraint that an items CR threshold cannot go below zero. Thus, after recall, an items threshold  $\theta$  becomes  $\theta - R$ . Restudy practice does not produce this extra learning because we assume that the order of feature activation matters for forming associative connections: Since all the features of a target item become activated simultaneously when it is re-presented, no intra-item associations are learned and CR is not strengthened.



Recall accuracy performance is determined by the probability of successful convergent retrieval, *cr*. Again, note that we will use the uppercase CR to refer to the convergent retrieval *process*, and the lowercase italic *cr* in a mathematical context to refer to the probability of successful CR. Since  $p(cr)$  depends on the difference between the number of features activated by the retrieval cues,  $pr$ , and the items CR threshold  $\theta$ , then  $p(cr) = p(pr > \theta)$ . In other words, if an item's  $pr$  exceeds  $\theta$ , the item will be recalled.

We also assume that the CR process takes time to complete, and that each item's time to successful CR depends on the difference between the number of features activated and the items threshold. The distance to CR threshold is related to the amount of elapsed time between initiating a retrieval attempt and outputting the target item (i.e. reaction time) through a negative exponential function.

$$RT = T_{min} + (T_{max} - T_{min})(e^{-\lambda|pr-\theta|}) \quad (6)$$

$T_{max}$  and  $T_{min}$  are parameters that set the upper and lower bounds on the possible elapsed time to output a recall response. The  $\lambda$  parameter controls the rate of decrease in reaction time as the distance from the CR threshold grows; large values result in a gradual decrease in reaction time with increasing threshold distances, while small values result in a rapid decrease of reaction time with increasing threshold distances. The shape of this function means that items with a small difference between  $pr$  values and  $\theta$  values will take a much longer time to recall than items with large differences between  $pr$  values and  $\theta$  values. This follows from the idea that if most features are initially activated by retrieval cues, the CR process should proceed much more quickly than if the CR process must pattern complete a

relatively large number of features. Figure 5 below shows the predicted RT as function of distance to CR threshold across several values of  $\lambda$  with a fixed  $T_{\max}$  and  $T_{\min}$ . In general, the  $\lambda$ ,  $T_{\max}$  and  $T_{\min}$  parameters are constrained to positive values. Thus far, all the described mechanics of the binomial implementation of the PCR model are equally applicable to cued recall and free recall experimental designs. However, applying the model to free recall paradigms requires an additional specification of the order in which target items are recalled. Here, the serial order of recalled items is controlled by the order in which the memory search progresses. We have chosen to determine the order of the memory search by rank ordering items by their  $pr$  values, from highest to lowest. A retrieval attempt is initiated for each item according to this rank ordering, which can be thought of as ordering items by the overall strength of the memories. Once an item has been successfully retrieved, or the retrieval attempt is abandoned as unsuccessful, no other attempts to retrieve it are made, and the item does not compete for retrieval or interfere with retrieval processes during later attempts to retrieve other items. Note that searching in order of  $pr$  values is not the same as searching in order of retrievability. An item's retrievability is determined by the difference between its  $pr$  value and its CR threshold  $\theta$ , so items with high values of  $pr$ , and thus the first items attempted to be recalled, may also have high CR thresholds which make them unable to be recalled. A memory search order based on information about an item's retrievability is not logically possible, as the item would necessarily have to be retrieved to know about its ability to be retrieved.

Importantly, PCR assumes that items with  $pr$  below the CR threshold  $\theta$  also contribute to observed recall latencies. Even if an item is ultimately unable to be recalled, some amount of time is still spent attempting to recall the item before moving on to initiating a retrieval attempt for another item. We relate the amount of time spent attempting (unsuccessfully) to recall a below threshold item to the difference between  $pr$  and  $\theta$  in the same fashion as above threshold items. When there is a small difference between  $pr$  and  $\theta$ , it takes longer to reject as “unrecallable”, while items with large differences are rejected relatively quickly. Practically, this predicts that large “gaps” will sometimes appear between recall outputs during free recall experiments. Participants may attempt and fail to retrieve several items in between two successively output items, rather than giving a relatively uniformly spaced series of correct outputs and abruptly terminating recall.

Another consequence of relating reaction time to the distance to CR threshold is that learning which increases this distance should also produce faster recalls on later tests. Both restudy and test practice increase this distance, however, successful test practice will increase this distance by a larger magnitude because it results in both an increase in  $pr$  and a reduction in  $\theta$ . This reduction in recall latencies can also have an effect on the observed recall accuracy, depending on the design used in a specific experiment. For example, participants in a free recall experiment are often only given a relatively short amount of time in which to recall items from a studied list, meaning that some potentially recallable items may not be output because the time spent recalling earlier takes up the entire allotted response

window. If given a sufficiently unlimited period to respond, participants can often recall additional items late in the response window, even minutes past the point in which the response window would close in an average free recall experiment (Roediger & Tulving, 1979). If the earliest items recalled were able to be recalled more rapidly, this leaves open a longer portion of the response period to try and retrieve items that have not yet been output. Considering that the PR and CR process may be lengthy, especially for items near the CR threshold, this means that it is possible to increase the observed recall accuracy by speeding up the retrieval for easily output items and leaving more time free to attempt retrieving the more difficult items.

The model we have presented has an analytic solution for predicted memory accuracy, utilizing joint binomial distributions of  $pr$  and  $\theta$ . However, this form is both complex and computationally expensive when considering the need to account for the effects of test practice, which make values of  $pr$  and  $\theta$  dependent. Thus, we will utilize Monte Carlo simulation methods in applying the model where the quantities of interest (e.g.  $pr$ ,  $\theta$ ,  $L$ ,  $F$ ,  $R$ ) will be determined by random binomial sampling. The simulation method has the additional advantage of producing a distribution of reaction times when plugging the distance to threshold into Equation 6. Now that we have presented a basic outline of the mathematical mechanics of the PCR model, the next task is to fit a specific implementation of the model to extant data. We chose to model the free recall data from Experiment 1 reported in Roediger and Karpicke 2006a. In the next section, we show that the PCR model can capture the

key cross-over interaction between test practice and retention interval, and explain what facets of the model allow it to do so.

### **1.6 Application to Data from Roediger & Karpicke 2006a**

Experiment 1 from Roediger and Karpicke 2006a employed a 2 x 3 mixed factorial design, which manipulated which practice method (test practice vs. restudy) within subjects and retention interval (5 minutes, 2 days, and 1 week) were manipulated between-subjects. All participants were initially given two short prose passages to study. Following this initial study, participants took a practice free recall test on one of the passages, and restudied the other passage. Exposure time was equated between the study and practice test conditions, with seven minutes allotted for each. On average, participants recalled 70% of the information from the passage on the practice test. After either a 5 minute, 2 day or 7 day retention interval, participants took a final test in which they were asked to recall as much information from each passage as possible. Performance for test practice and restudy practice across these three intervals is shown above in Figure 1.

We fit the PCR model to the reported recall probabilities from each of conditions in this experiments, including the average practice test performance. In describing the model implementation and parameters, we will use the nomenclature and symbols defined in the previous section. The model included 6 free parameters: an encoding rate parameter  $e$ , a learning rate parameter  $l$ , a threshold reduction parameter  $r$ , two forgetting rate parameters  $f_2$  and  $f_7$  ( applied to the two day and seven day retention interval conditions, respectively), and a parameter governing

the minimum time between recall outputs,  $T_{\min}$ . The maximum time between recall outputs was fixed at 60 seconds. The distribution of item CR thresholds,  $\theta$ , was determined by binomial random sampling, with the probability parameter  $t$  fixed at .5, and the number of features for each item fixed at 100. These parameters we used to estimate six key quantities of interest for each item in the studied passages (30 items per passage) via Monte Carlo simulation, with 1,000 simulations per item<sup>2</sup>:

- $pr$ , the number of features able to be activated by retrieval cues following the initial encoding episode.
- $L$ , the increase in the number of features activated by retrieval cues following PR practice, relative to initial encoding.
- $\theta$ , the number of features that must be activated by retrieval cues to support convergent retrieval.
- $R$ , the reduction in CR threshold ( $\theta$ ) following successful retrieval practiced.
- $F_2$ , the reduction in  $pr$  due to forgetting over a two-day retention interval.
- $F_7$ , the reduction in  $pr$  due to forgetting over a seven-day retention interval.

The probability of recalling an item on the practice test is the same the probability of successful convergent retrieval, thus performance is given by calculating  $p(pr > \theta)$ . As previously outlined, PCR assumes that recall strengthens the associations between the cue features and target features, as well as the

---

<sup>2</sup> A continuous approximation to the binomial was used for these simulations. In general, the state activation of features composing memory traces is not binary, and is better described as having a continuous range of activation. See the Mathematical Appendix A for details of the implementation.

associations between the within-target features, while only cue-target associations are strengthened with restudy. Thus, performance on the five minute delayed practice test following restudy is given by finding  $p(pr + L > \theta)$ . Calculating performance on the five minute delayed final test requires summing over two quantities: the probability of recalling given a prior successful recall,  $p(pr + L > \theta - R)$ , and the probability of recalling given a failure to previously recall,  $p(pr > \theta)$ . Without any forgetting, the sum of these probabilities would always come out be the same as the probability of recall on the initial practice test (as the increase in  $pr$  and decrease in  $\theta$  are irrelevant for items already above the threshold, and items below the threshold are not strengthened by virtue of the fact that they are below threshold and unable to be recalled) if it were not for the speed up of reaction time following successful retrieval. This speed up in recall and output of items already recalled previously allows there to be enough time left in the test period to retrieve items with a  $pr > \theta$  after initial encoding, but were not able to be output on the practice test because of the response deadline. Thus, it is possible to observe a slight increase in the number of recalled items on the final test relative to the practice test, given forgetting of a sufficiently small magnitude.

We assumed that forgetting occurred between the practice phase and final test phase in the two day and seven day retention interval conditions, but not the five minute retention interval. To model this forgetting, the value of  $pr$  following the practice phase was decreased by  $F_i$ , where the subscript  $i$  is a categorical index referring to a specific retention interval. Thus, delayed final test performance was given by calculating  $p(pr + L - F_i > \theta)$  for items receiving restudy practice,  $p(pr + L -$

$F_i > \theta - R$ ) for items recalled on the practice test, and  $p(pr - F_i > \theta)$  for items not recalled on the practice test. Since the five minute retention interval condition was relatively short, we assumed that any decrease in the number of features activated due to changes in current temporal context were small enough to ignore. Another practical reason to not apply forgetting for this condition would mean that the model would be over parameterized, with as many free parameters as observed values.

The likelihood of the data was maximized using the binomial likelihood ratio test, which provides a chi-square goodness-of-fit statistic. The low value of this statistic demonstrates the model was an extremely good fit to the observed data ( $\chi^2(1) = .042, p = .83$ ). The best fitting model parameters are shown in Table 1, and the model predictions as compared to the observed data are shown in Figure 6. The assumption of learning in both the PR and CR stages of retrieval allows the PCR model to readily explain the cross over interaction between retention intervals and practice method commonly observed in studies contrasting restudy and test practice.

While showing that the PCR model is a good fit to extant data, to which (to the best of our knowledge) no computational models have been applied to, is satisfying and supports our theoretical claims, we now turn to our own experiments which test direct predictions of the PCR model. As a result of positing that retrieval practice strengthens or builds new intra-item feature associations, PCR predicts recall practice has hidden benefits on an immediate final test in the testing effect paradigm. Experiment 1 tests the prediction of cue generalization produced CR



learning: if successful retrieval practice results in learning a pattern completion process between the features of a stored memory trace, then this learning should generalize across retrieval cues. Specifically, we test the predication that recall accuracy on a final test will be elevated as a result of successful retrieval practice, even if the retrieval cues used on the final test are not the same as those used during practice. Experiment 2 directly tests the prediction that CR learning increases the speed of later recall outputs: if the CR process requires active features of a memory trace to successively activate other dormant features, and successful retrieval practice strengthens or builds new intra-item associations, then subsequent iterations of the CR pattern completion process for that item will proceed more rapidly.

## **CHAPTER 2**

### **EXPERIMENT 1**

#### **2.1 Testing the cue-independent learning prediction of PCR**

#### **2.2 Method**

##### **2.2.1 Participants**

64 individuals from the University of Massachusetts were recruited from the undergraduate subject pool. Participants were given one unit of credit that could be applied either toward class participation requirements or extra credit opportunities in undergraduate psychology classes. Participants were randomly assigned to either the immediate (n=30) or delayed (n=34) final test conditions. The data from one subject who participated in the delayed condition was discarded from analysis and model fitting because of extremely low performance on the cued recall task (< 5% correct recalls in all conditions), leaving a grand total of 63 participants in the experiment.

##### **2.2.2 Materials**

100 cue-target pairs consisting of English nouns were randomly selected from a word pool to serve as test materials. The word pool was constructed using the English Lexicon Project database (Balota et al., 2007). All words in the pool were

moderate frequency<sup>3</sup> English nouns, have between four and 10 characters, and have concreteness and imageability ratings of over 500 (Brysbaert & New, 2009).

### **2.2.3 Procedure**

Both groups of participants in the experiment were tasked with learning four lists of 25 cue-target word pairs. The first list that all participants studied presented each cue-target word pair alone in the center of a computer monitor for five seconds, and was immediately followed by a 30 second math distractor task. After the distractor task, participants immediately took a cued-recall test on all 25 of the pairs presented in the first list. The presentation order of the cue words was randomized, and for each pair, the first item in the pair (the cue word) was presented in the same location on the monitor as during the initial study episode. This word served as a cue to recall the second word in the pair, which was replaced by a “?” in the visual presentation. Participants had 10 seconds to enter their response using the computer’s keyboard, but could advance to the next test item at any time by pressing the “Enter” key.

Memory performance on this test was scored using an automated procedure, based on calculating the Levenshtein edit distance between the target word and the string entered by the participant in the response to the cue for that target word. Input responses with a Levenshtein edit distance of less than or equal to 1 from the

---

<sup>3</sup> Moderate frequency was defined as having between five and 200 usages per million words, as measured by the SUBTLEX<sub>US</sub> database

target word were considered to be correct responses<sup>4</sup>. Memory accuracy on this test was used to modulate the presentation duration of cue-target pairs in subsequent lists for each subject: if less than 25% of responses were correct, presentation duration was increased to six seconds, and if more than 75% of responses were correct, presentation duration was reduced to four seconds. This adjustment was made with the aim of modulating performance on subsequent practice tests towards 50%, as performance of this level will allow for an analysis of conditional recall probabilities (i.e. the probability of recalling a target word on the final test given that it was recalled on the practice test). Following this calibration, 12% of participants had their presentation time increased to six seconds, and 35% had their presentation time decreased to four seconds.

The structure of the cue-target word lists was changed following the test on the performance modulation list. Subsequent lists still presented 25 cue-target word pairs, but within each list, five target words occurred twice in the list and were paired with a different cue word each time (making 10 total pairs). The remaining 15 targets in the list appeared a single time, and cue words were always unique. Put another way, each cue-target *pair* was unique, but 40% of targets had two cues. Cue-target pairs from within each list were randomly selected to receive restudy practice, test practice, or no practice (a control condition). The 15 non-repeating target words were equally subdivided into these three conditions (5 restudied, 5 tested with cued recall, and 5 not practiced). The 10 repeating target words were

---

<sup>4</sup> A Levenshtein edit distance of 1 meant that substituting, adding, or deleting 1 character from the input string would transform it into the target string. This weak string matching scoring procedure was used to deal with typos in the input.

subdivided between the no practice and test practice condition. The two types of practice (study and test) were blocked within the practice phase for each list, and the order in which the items were restudied/tested was randomized. The order of the test and study blocks was counter balanced across subjects. Responses for the cued recall practice test were given using the computer keyboard, with a response deadline of 10 seconds. No feedback was given about performance on the practice test.

At the critical final test, participants took a cued recall test on the restudied, tested and unpracticed non-repeating target words, but were only tested on the repeating target words using the unpracticed cue word. Thus, the final tests reflect memory for four distinct classes of target items: 1) targets cued with word cues that were studied as a pair and the pair was not practiced again (control condition), 2) targets cued with word cues that were studied and then restudied as a pair (restudy condition), 3) targets cued with word cues that were studied as a pair and then given as retrieval cues on a cued recall practice test (same-cue test condition), and 4) targets cued with word cues that were studied as a pair, and the pair was not practiced again but the target word itself had retrieval practice using a different cue word during the practice phase (other-cue test condition). For example, if the word pair “Orchid – Light” was learned initially and “Orchid - ?” was used as a cue to retrieve “Light” during the practice test, then the responses to the cue “Orchid - ?” during the final test belong to the “same cue tested” target condition. On the other hand, if the word pairs “King – Table” and “Foot – Table” were learned initially, and “King - ?” was used as a cue to retrieve “Table” during the practice test, then the responses to the

cue “Foot - ?” during the final test belong to the “other-cue tested” condition. Just as during the practice test, the order the cues were given in was randomized and responses were given using the computer keyboard with a response deadline of 10 seconds for each cue.

The temporal structure of the initial learning, practice, and final test phases differed between the immediate and delayed final test groups. Participants in the immediate final test condition completed the three phases in order for each list. Participants in the delayed condition completed the study and practice phase for *each list* of pairs before moving into the final test phase. Thus, participants had to retrieve items from the second and third lists before taking the final cued recall test on items from the first list. This intervening cued recall test causes a shift in temporal context (Jang & Huber, 2008), and participant’s performance will be negatively impacted as a result of the context of the final test not strongly matching the context at encoding. This manipulation mimics the effect of a longer retention interval on memory performance, allowing us to assess the interaction between practice type and retention interval without requiring participants to attend two separate experimental sessions.

### **2.3 Behavioral Results**

Recalled items were checked for accuracy first by an automated routine performing strict string comparison between responses and list items. Recall responses that were scored incorrect by the automated procedure were then double checked and scored by hand. Recall performance on the practice tests was

comparable across both the immediate and delayed groups (57% of targets correctly recalled in the immediate group, 54% percent of targets correctly recalled in the delayed group). The percent of targets correctly recalled from all four classes of target items in both groups of subjects are shown in Figure 8. Performance on the final cued recall test was analyzed using a 2 x 4 Mixed ANOVA. There was a significant main effect of retention interval group ( $F_{1,61} = 15.07, p < .001$ ), with the delayed final test group performing worse than the immediate final test group. There was also a significant main effect of practice type ( $F_{3,183} = 72.95, p < .001$ ) and a significant interaction between practice type and group ( $F_{3,183} = 6.87, p < .001$ ). In the immediate final test condition, memory accuracy for restudied targets was 25% higher than targets cued with the same cue that was given on the practice test (81% recalled versus 56% recalled), and 36% higher than for targets that were never practiced with any cue (45% recalled). Memory accuracy for targets that received test practice with the same word cue used on the final test was nearly identical to accuracy on the practice test (only a 2% decline in accuracy). The critical data point for assessing the predictions of the PCR model is final test performance for targets cued using a word that was never itself practiced with the target after the initial study phase, but received cued recall test practice using another word cue (the other-cue target condition). Accuracy for these other-cue target items was nearly identical to accuracy for the same-cue test practice target items (56.6% recalled for same-cue recall practice vs 56.2% recalled in other-cue recall practice).

In the delayed final test condition, memory accuracy was lower overall than in the immediate final test condition, but the same qualitative pattern of results

between the four target practice types emerged. Memory accuracy was highest for study practiced targets, followed by same-cue tested targets, other-cue tested targets, and finally unpracticed targets. However, the performance gap between the restudy and same-cue test practice conditions narrowed from 25% to 6%. This was driven by a precipitous drop in accuracy for restudied targets (54% recalled, down from 81% in the immediate condition), while accuracy for same-cue test practice targets remained relatively stable (48% recalled, down from 56% in the immediate condition). Memory accuracy for other-cue test practiced targets was reduced by a greater amount than same-cue tested targets (falling by 23% relative to the immediate final test condition, down to 33% recalled) but was still greater than the control condition of no practice with any cue (only 29% recalled).

Figure 9 shows the breakdown of final test memory accuracy for target items that received test practice, conditional on the outcome (correct or incorrect recall) of the practice test. For both the immediate and delayed final test conditions, recall for same-cue test practice targets mirrored the outcome of the practice test. Targets that were recalled on the practice test were almost always recalled on the final test when the same cue was given ( $p(\text{correct} | \text{correct}) = .95$  in the immediate condition,  $p(\text{correct} | \text{correct}) = .86$  in the delayed condition). Targets that were *not* recalled on the practice test were almost never recalled when the same cue was given ( $p(\text{correct} | \text{incorrect}) = .02$  in the immediate condition,  $p(\text{correct} | \text{incorrect}) = .03$  in the delayed condition). Similarly, a successful retrieval on the practice test resulted in better retrieval of that target item on the final test, even when the practice test was taken with different word cue. In both the immediate and delayed



final test conditions, targets were more likely to be recalled in response to the unpracticed cue word when they had successfully recalled the target on the practice test using a different cue word ( $p(\text{correct} | \text{correct}) = .63$  in the immediate condition,  $p(\text{correct} | \text{correct}) = .37$  in the delayed condition) relative to when the retrieval attempt was unsuccessful on the practice test ( $p(\text{correct} | \text{incorrect}) = .48$  in the immediate condition,  $p(\text{correct} | \text{incorrect}) = .23$  in the delayed condition). However, this relative advantage was much narrower than in the same-cue test practice condition. A statistical analysis of this interaction was prevented because of missing cells introduced by grouping final test outcomes based on practice test outcomes for that item (3 subjects in the immediate condition made no errors on the practice tests for same-cue targets, and 1 subject in the immediate condition made no errors on the practice tests for other-cue targets items, resulting in missing cells for these participants in the  $p(\text{correct} | \text{incorrect})$  condition).

## **2.4 PCR Model**

### **2.4.1 Implementation**

We fit the PCR model to the proportion of correctly recalled target items on the final test for the no practice and restudied targets, as well as the joint proportion of items recalled/not recalled on the practice and final tests for targets that received test practice. We fit the model to these proportions for each subject that participated in the experiment. Goodness of fit was assessed by minimizing the negative binomial log likelihood of the model's accuracy predictions. It was not possible to calculate the likelihood ratio to perform statistical test of the model's

goodness of fit due to missing cells in the observed data of some participants<sup>5</sup>. To generate model predictions in each of the target item practice type conditions, we simulated six quantities of interest:

- $pr_1$ , the number of features able to be activated by retrieval cues following the initial encoding episode using the first cue word.
- $pr_2$ , the number of features able to be activated by retrieval cues following the initial encoding episode using the *other* cue word.
- $L$ , the increase in the number of features activated by retrieval cues following PR practice, relative to initial encoding.
- $\theta$ , the number of features that must be activated by retrieval cues to support convergent retrieval.
- $R$ , the reduction in CR threshold ( $\theta$ ) following successful retrieval practiced.
- $F$ , the reduction in  $pr$  due to forgetting during the retention interval.

The values of each of these quantities for each item were determined using Monte Carlo simulation of binomial random samples. We used five free parameters to fit the data from each subject: an encoding rate probability parameter  $e$ , which controlled the value of  $pr_1$  and  $pr_2$  for each item, a learning rate probability parameter  $l$ , which controlled the value of  $L$  for each item, a threshold reduction probability parameter  $r$ , which controlled the value of  $R$  for each item, and a

---

<sup>5</sup> Some participants has a joint probability of recalling on the practice test and not recalling on the final test equal to zero. This zero in the observed data ends up in the denominator of the likelihood ratio statistic, causing its value to be undefined.

forgetting rate probability parameter  $f$ , which controlled the value of  $F$  for each item. The distribution of item CR thresholds,  $\theta$ , was determined by binomial random sampling, with the probability parameter  $t$  fixed at .5, and the number of features for each item fixed at 100. No parameters related to retrieval time were included in model for these data. We also included a ‘space out’ parameter  $s$  that was allowed to vary for each subject, representing the probability that on any given trial, a participant would not attempt to retrieve the target item in response to the cue. This was necessary since the outcomes of the successive retrieval attempts using identical memory cues is deterministic in the PCR model, and some randomness was needed in order to have a non-zero probability of recalling a target item on the final test following a failure to recall that item on the practice test with the same cue word<sup>6</sup>.

The probability of recalling an item on the practice test is given by calculating  $p(pr_1 > \theta)$ . When the associations from cue features to target features are strengthened through restudy or successful retrieval practice, PCR assumes that amount of features activated increases by the quantity  $L$ , but that changes in the current temporal context reduce the amount of features activated by the quantity  $F$ . Thus, performance on final test following restudy is given by  $p(pr_1 + L - F > \theta)$ . Calculating performance on the final test for targets cued with the same cue that received test practice requires summing over two quantities: the probability of

---

<sup>6</sup> We chose to use the “space out” parameter as a matter of mathematical simplicity. Other sources of randomness within the memory system could be proposed to handle this fluctuation in outcomes between practice and final tests, such as random changes in the number of features activated by retrieval cues.

recalling given a prior successful recall,  $p(pr_1 + L - F > \theta - R)$ , and the probability of recalling given a failure to previously recall,  $p(pr_1 - F > \theta)$ , because items which are not recalled receive no boost to their  $pr$  values and no reduction to their CR threshold  $\theta$ . Calculating performance on the final test for targets that were practiced with a cued recall test but cued with a different, unpracticed cue word on the final test also requires summing over two quantities: the probability of recalling given successful practice test retrieval,  $p(pr_2 - F > \theta - R)$ , and the probability of recalling given a failure to recall on the practice test,  $p(pr_2 - F > \theta)$ . Note that there is no need for two separate forgetting rates between the immediate and delayed final test group because the immediate/delayed manipulation was done between subjects and the model was applied to each individual subject.

Importantly,  $pr_1$  and  $pr_2$  were sampled independently of one another, so the amount of feature activation from one cue word did not affect the amount of feature activation from the other cue word. Following from this, any increase to the number of features activated from the retrieval cues (via restudy or successful test practice) applies *only* to the cue utilized during practice. But while the amount of feature activation from each different cue is independent, the target item itself has the same CR threshold throughout the experiment, regardless of retrieval cues. In other words, the CR threshold value is assumed to be a property of the target item itself, while the amount of feature activation in PR is property of the associations learned between the cue and target item. Thus, while the amount of feature activation from one of the learned cues may be below the CR threshold, the amount of feature activation from the other cue may be above the CR threshold. This distinction is

especially important to the other-cue test practice condition. After receiving successfully retrieving an item on the practice test in response to cue one, the items CR threshold is assumed to be lowered, meaning it becomes easier to retrieve regardless of the cue used to activate its features. Thus when the unpracticed second cue associated with the target word is presented on the final test, the target item will be more readily retrieved, even though it received no explicit practice with that cue.

### **2.4.2 Results**

After fitting the model to each subject's data, the model's predicted recall accuracies and best fitting parameters were averaged across the fits to individual subjects. The average final test recall accuracy predictions for the four target item practice types, and the conditional recall accuracy predictions on the final test for the two types of tested target items are shown Figure 8 and Figure 9 respectively. The model's fit to the joint probability of recall on the practice test and final test can be seen in Figure 10, along with observed joint averages from both groups of participants. The model's best fitting parameters (which are shown in Table 2) produce predictions which closely approximate the observed data. Predicted performance is higher for the immediate final test than for the delayed test, and the pattern recall accuracies for the four target practice types are qualitatively matched in both groups (study practice > same-cue test practice > other-cue test practice > no practice). The only significant discrepancy between the fitted predictions and the observed data is found in the conditional probabilities of other-cue tested target

items, where the model under predicts the observed accuracy for items not recalled on the practice test, and over predict the accuracy for items that were correctly recalled on the practice test. However, this discrepancy results primarily from some subject's low number of observations in some of these conditions (i.e. nearly perfect recall on the practice test, or nearly all errors on the practice tests).

## **2.5 Discussion**

Experiment 1 tested and confirmed the prediction of the PCR model that successful retrieval practice should result in learning that generalizes to memory tests using other retrieval cues. We observed the surprising result that retrieving a target on a practice test is beneficial to memory on a later test regardless of whether the cues used on the final test match those used at the practice test. This can be seen in the data for the other-cue test practice condition, such that targets that received test practice with another cue are always better recalled than the control condition, and are recalled just as frequently as same-cue test practice targets on an immediate final test. Thus, it appears that test practice can enhance the retrievability of an item itself. This benefit was primarily observed for an immediate final test however, meaning that the retrieval cues themselves are not rendered unimportant to later retrieval outcomes by testing.

This prediction of cue generalization stems from the convergent retrieval learning hypothesis as proposed by the PCR model. Specifically, we argue that that the act of retrieving from memory requires a type of pattern completion of the partially activated memory trace, and successful retrieval results in increases to the

associations that support this process. The observation that practicing retrieval of a target word with one cue word results in a higher probability of retrieving that target word in response to an independent cue which received no additional practice supports the convergent retrieval learning hypothesis. Importantly, the PCR model was able to capture this cue generalization effect in the model fitting exercise, while also fitting the classic pattern of resulting found in testing effect paradigms. Extant theories of memory which model learning as strengthening of the associations between retrieval cues and memory traces cannot readily explain the generalization of memory benefits from retrieval practice with one experimentally associated cue word to another cue word. We believe this result instead points to a form of learning about the target memory itself independently of retrieval cues. We now turn to testing a second prediction of the convergent retrieval learning hypothesis, specifically that if memory retrieval strengthens intra-item feature associations, that these strengthened associations will lead to reduced retrieval latency on subsequent memory tests.

## CHAPTER 3

### EXPERIMENT 2

#### 3.1 Testing the recall latency prediction of PCR

As outlined previously, we assume that the primary retrieval process partially activates stored memory traces using retrieval cues, and that it takes time to completely activate and “fill in” the missing features during the CR process before a recall response may be given. After successful convergent retrieval forms new associative connections between items, attempting to recall again with the same memory cues will not require as many time steps to completely activate the remaining features in the item. Such a situation is schematically outlined in Figure 7. The left panel of Figure 7 shows the pattern of intra-item feature associations after a successful convergent retrieval attempt. If the same cues are used to guide a later retrieval attempt (i.e. the same two features are activated, as in the top right panel), then the remaining three features can simultaneously become active via intra-item feature associations immediately afterwards. Without the increase in intra-item associations learned from successful convergent retrieval, this process would take much longer, as it requires each remaining feature to become activate one at a time to support pattern completion (such a situation is shown in the left panel of Figure 7.) Thus, this predicts that even when no difference in accuracy is observed between a practice test and a final test (i.e. when convergent retrieval is supported by retrieval cues), the retrieval time will be faster on a the final test than on the practice test



Such a pattern of results has been recently demonstrated in a cued recall paradigm (van den Broek, Segers, Takashima, & Verhoeven, 2014). In this experiment, we utilized a free recall paradigm, as allowing participants more flexibility in structuring the retrieval process provides better window into the dynamics of the retrieval process, and a more constraining data set for modeling. Participants were instructed to learn simple word lists in an initial study phase, after which the list was either restudied, or a free recall test was given. Immediately after this practice phase, a final free recall test was given. The results showed that while restudy practice produced the highest memory accuracy for the lists, test practice produced the fastest responses despite having little effect on overall accuracy, in line with our predictions. Below, further details of the experimental method, behavioral results and modeling results with the PCR model are given.

## **3.2 Methods**

### **3.2.1 Participants**

34 individuals from the University of Massachusetts were recruited from the undergraduate subject pool. Participants were given one unit of credit that could be applied either toward class participation requirements or extra credit opportunities in undergraduate psychology classes.

### **3.2.2 Materials**

List items consisted of 180 English nouns randomly selected from a pool of words. The word pool was gathered using the English Lexicon Project database

(Balota et al., 2007). All words in the pool were moderate frequency English nouns<sup>7</sup> with lengths between four and 10 characters and concreteness and imageability ratings of over 500 (Brysbaert & New, 2009).

### **3.2.3 Procedure**

The experiment used a single factor (test practice vs. study practice) within-subjects blocked design. Each block consisted of three phases: an initial learning phase, a practice phase, and a final test phase. For the initial learning phase, participants studied a list of 15 serially presented words, where each word was presented alone for three seconds in the center of a computer monitor. During study practice blocks, word lists were re-presented to participants in identical serial order. During test practice blocks, participants took a 90 second free recall test where they were instructed to recall as many words as possible from the list of words they had just studied. Practice test responses were given with the computer keyboard. No feedback about performance on the practice test was given. During the final test phase for all blocks, participants took a 90 second free recall test (an identical format as the practice test).

Between initial study and practice, as well as the practice and final test phases, participants completed a 20 second math distractor task. The distractor task required participants to calculate a running cumulative sum of 5 consecutively presented single digit integers (5 seconds for each sequence presentation, 5 seconds to enter a response, with 2 repetitions of the sum task between each phase). Four

---

<sup>7</sup> As in Experiment 1, moderate frequency was defined as having between five and 200 usages per million words, as measured by the SUBTLEX<sub>US</sub> database.

study practice and four test practice blocks were presented in alternating order for each participant, with the method of practice for the initial block counter balanced across participants. This design yielded 12 total test observations per subject: four baseline practice tests, four final tests following test practice, and four final tests following restudy practice. The entire experiment lasted approximately 45 minutes.

### **3.3 Behavioral Results**

Recalled items were checked for accuracy first by an automated routine performing strict string comparison between responses and list items. Recall responses that were scored incorrect by the automated procedure were then double checked and scored by hand. Because there are no explicit instructions or temporal cues given by the experimenter telling participants to give each response in a free recall paradigm, the reaction time for a specific response in this data set is considered to be the elapsed time between responses (also known as the inter-retrieval time). This elapsed time between each response given was measured by the time between confirming the last correct response<sup>8</sup> (confirmation was given by hitting the “Enter” key after typing in the response) and the first keystroke of the next entered response, except for the first item output, where it was measured as the elapsed time between the onset of the response window and the first keystroke of the first response.

---

<sup>8</sup> If an item from the word list was output more than once, it was not treated as a correct response after the first time.

Free recall accuracy and reaction time across the three conditions (baseline or initial practice test, final test after restudy practice, and final test after free recall test practice) were analyzed with a one-way repeated measures ANOVA. There was a significant effect of practice on accuracy ( $F_{2,66} = 174.9, p < .001$ ), and Bonferroni-corrected pairwise t-tests showed significant differences between recall accuracy on the final test between the restudy and test practice conditions (81% correct vs 59% correct,  $t_{33} = 14.2, p < .001$ , Cohen's  $d_z = 1.83$ ), a significant improvement in accuracy on the final test after restudy practice from the practice test (81% correct vs 61% correct,  $t_{33} = 13.1, p < .001$ , Cohen's  $d_z = 1.64$ ) and a small but significant decrease in accuracy on the final test after test practice from the practice test (61% correct vs 59%,  $t_{33} = -3.9, p < .01$ , Cohen's  $d_z = .12$ ).

The average reaction times in the three conditions were not normally distributed, each failing the Shapiro-Wilk test for normality, so the data were transformed using the natural logarithm for statistical analysis. After the natural logarithm transform, reaction times from each condition no longer failed the Shapiro-Wilk test for normality. There was a significant effect of practice on the log reaction times ( $F_{2,66} = 20.72, p < .001$ ). Bonferroni corrected pairwise t-tests showed no significant decrease in the log transformed reaction times on the final test following test practice relative to restudy practice (.289 seconds versus .327 seconds,  $t_{33} = -1.01, p = .95$ , Cohen's  $d_z = .13$ ), a significant decrease in log reaction time after restudy practice from the practice test (.327 seconds versus .51 seconds,  $t_{33} = -5.66, p < .001$ , Cohen's  $d_z = .69$ ) and a significant decrease in log transformed reaction time on the final test after test practice from in log transformed reaction

time the practice test (.51 seconds vs .289 seconds,  $t_{33} = 5.3$ ,  $p < .001$ , Cohen's  $d_z = .82$ ).

A breakdown of accuracy and response time by each items serial position in the output responses is shown for the three conditions in Figure 11 and Figure 12, respectively. Figure 11 shows the probability of recalling at least one items from the list, the probability of recalling at least two items from the list, etc., all the way up through the probability of recalling all 15 items from the list. Thus, perfect performance would be shown in the figure as a horizontal line at  $y = 1$ . This figure demonstrates that restudy practiced increased the probability of recalling more items from the list relative to the amount recalled on the practice test, while the probability of recalling more items from the list is slightly smaller on the final test than on the practice test following test practice. When breaking the reaction times down by the items serial position in recall output, it can be see that reaction times are relatively stable across output positions on the final test following test practice, as evidence by the relatively straight line connecting the response times at each output position. On the other hand, reaction times increased as a function of output position on both the practice test and on the final test following restudy practice, as shown by the upward slope in the connecting the reaction times observed for items output later in the list. This pattern of results (an accuracy advantage for restudy practice, but a speed advantage for test practice) is in line with the predictions of the PCR model.

### 3.4 PCR Model

#### 3.4.1 Implementation

The PCR model was simultaneously fit to both the accuracy and reaction times of each participant. Monte Carlo simulation methods were used to estimate four quantities of interest for each of the 15 items on the test list (1000 simulations per item):

- $pr$ , the number of features able to be activated by retrieval cues following the initial encoding episode
- $\theta$ , the number of features that must be activated by retrieval cues to support pattern completion.
- $L$ , the increase in the number of features activated by retrieval cues following PR practice, relative to initial encoding.
- $R$ , the reduction in CR threshold ( $\theta$ ) following successful retrieval practiced.

Seven free parameters were allowed to vary in the simulations of these quantities. As in the previous implementations, a probability parameter  $e$  was used to control the rate of feature activation after the initial encoding opportunity, a probability parameter  $l$  was used to control the amount of increase in feature activation after restudy or successful retrieval on the practice test, and a probability parameter  $r$  was used to control the amount of reduction to the convergent retrieval threshold  $\theta$  after successful retrieval on the practice test. The parameters controlling the minimum and maximum time spent retrieving (or attempting to retrieve) an item from memory,  $T_{min}$  and  $T_{max}$  were allowed to vary for each subject,

as well the  $\lambda$  parameter which controlled the rate of decrease in reaction time as the distance to the CR threshold increased. The variability of the distribution of convergent retrieval threshold was allowed to vary, and was controlled by the  $t$  parameter<sup>9</sup>.

As in each previous implementation, the probability of recalling an item on the practice test is given by calculating  $p(pr_1 > \theta)$ . However, we also model the cumulative amount of time spent retrieving items from the list, and any items which would only be able to be recalled outside the 90 second response window are considered a failure to recall. The amount of time spent time spent retrieving (or attempting to retrieve) an item from memory was controlled by the negative exponential function show in Equation 6. Restudy or successful retrieval practice increases the amount of PR feature activation by the quantity  $L$ . Thus, performance on final test following restudy is given by finding  $p(pr + L > \theta)$ , while performance on the final test following test practice requires summing over two quantities: the probability of recalling given a prior successful recall,  $p(pr + L > \theta - R)$ , and the probability of recalling given a failure to previously recall,  $p(pr - F > \theta)$ . Since increases in the value of  $pr$  and reductions in the value of  $\theta$  both change the distance to the convergent threshold ( $pr - \theta$ ), both restudy practice and successful test practice can reduce the response time of retrieval attempts subsequent to

---

<sup>9</sup> As mentioned in the section outlining the implementation of the PCR model used to fit the data from Roediger and Karpicke's Experiment 1, a continuous approximation to the binomial distribution using the beta distribution was utilized in the simulations. Here, the distribution of item CR threshold was a symmetric beta distribution with one free parameter (i.e. beta shape parameter  $a =$  shape parameter  $b =$  PCR model parameter  $t$ ). For more details of the beta approximation, see Mathematical Appendix A.

practice. In contrast to the implementations used to model other experiments in this paper, forgetting between the practice phase and the final test phase was not included in the model, to reduce the number of free parameters and because the delay between the practice and final test was minimal.

The model simulation method produced output detailing both the predicted probability of recalling at least  $x$  items from the word list (where  $x$  ranges from 1 to the total number of words on the list, 15) in each of the three conditions, as well as distribution of reaction times for items recalled at each output position (output item number 1 through output item number 15). Best fitting model parameters were identified by maximizing the likelihood of each observed data point from each subject under a joint reaction time and accuracy probability distribution approximated by smoothing and normalizing the distribution of reaction times produced by the PCR model<sup>10</sup>. Best fitting model parameters are shown in Table 3.

### **3.4.2 Results**

The accuracy predictions of the best fitting PCR model as applied to each subject data were averaged at each output position, and these averages are shown as compared to the observed data in Figure 13. The median reaction times of the best fitting PCR model as applied to each subject data was calculated at each output position, and these medians are shown in Figure 14. As these plots show, the PCR model is able to capture the structure of the accuracy and reaction times observed

---

<sup>10</sup> Details of the smoothing and normalization procedure used to create the joint accuracy and reaction time distributions are given in Mathematical Appendix B.



in the data. In the accuracy plots showing the probability of recalling an item at each output position, predicted performance on the practice test and final test after practice testing pattern together, with more items recalled from the lists following restudy practice, just as in the observed data. In the plots showing the median reaction times at each output position, the model captures the trend of increasing reaction times as a function of output position in the baseline practice test and final test after restudy practice conditions, as well as the stable reaction times across output positions on the final test practice test, just as in the observed data.

### **3.5 Discussion**

The results of Experiment 2 confirmed the prediction of a second hidden benefit following test practice. The development of the PCR model predicted that successful test practice would increase the strength of intra-item feature associations, resulting in the ability to more quickly recall items on a subsequent final test. In a free recall paradigm, we observed that response times decreased on an immediate final test following both restudy and test practice relative to response times on the baseline practice test, but that responses times were decreased by the highest magnitude following test practice. This decrease in reaction time following test practice is consistent with previous experimental findings (van den Broek et al., 2014). This speed up in responding was in contrast to the effects of restudy and test practice on accuracy: restudy strongly boosted recall accuracy, whereas accuracy accurately decreased slightly on the final test where the list of words was practiced with a free recall test. A negative correlation has been observed previously between

the number of items output in a free recall test and the rate at which the items are output (see Wixted & Rohrer, 1994 for a review of these free recall latency findings) but we argue that that such a small difference in the number of items recalled alone (an average of 9.15 recalled per list on the practice test and an average of 8.85 items recalled per list) is unlikely to explain the sharp decrease in response times. Additionally, response times also decreased while the number of items output *increased* in the restudy condition, pointing towards other mechanisms working to produce the observed changes in reaction time apart from simple changes in the number of items recalled.

When fit to the data observed from each subject, the PCR model was able to capture the pattern of accuracy and reaction time changes in each condition, further supporting the convergent retrieval learning hypothesis. The key feature of the model which allows it to fit the observed data is the assumption that it takes time to “fill in the blanks” of features in a memory trace not activated by retrieval cues, and that successful convergent retrieval speeds up subsequent retrieval attempts by strengthening the associative connections between intra-item features. The model fitting exercise provided a strong challenge for the PCR, as fitting each observed data point from each subject places significant constraint on the model. Despite this constraint, the only discrepancy between the pattern of observed results and the predictions of the PCR model was that the observed data showed a small decrease in accuracy for final tests following test practice relative to accuracy on the practice test, while the PCR model predicted a slight increase in performance. This stems from that fact that the model predicts that some failures to recall an item from the

list on the practice test are due to running out of time on the test, and that the increased response speed for items that *were* recalled can mean that on later tests, there is just enough time to retrieve an item that was left out on previous tests due to the time limit.

We can consider two possible explanations for this small discrepancy. The first is that there was forgetting between the practice and final test, a process which was not included in the model of Experiment 2. A small degree of forgetting between the two tests might be able to lower the weakest items in memory (those which are just above the CR threshold, and likely the last to be attempted in the list) below the CR threshold. Alternatively, this discrepancy may be due to the fact that this implementation of the PCR model has a relatively simple memory search process, in which items are attempted to be retrieved in the order of their *pr* values, and items do not interfere with the retrieval of other items. In reality, there is good evidence that list items do interfere with one another at retrieval (Gillund & Shiffrin, 1984; Ratcliff, Clark, & Shiffrin, 1990). A possible explanation the small decrease in accuracy is that the retrieval of items target items on the practice and final test made them so strong in memory that they repeatedly outcompeted weaker items for retrieval (i.e. a list-strength effect). Some evidence for this possibility can be found in an analysis of repeats (the number of times a target word was output more than once) during the practice and final test phases. On the practice test, there were an average of .26 repeated words per list, while repeats on the final test increased to an average of .43 words per list, suggesting that some words became more difficult to *not* repeatedly retrieve after test practice. Despite this discrepancy, we argue that

the model's predictions are consistent with the observed data, and provide strong support for the PCR model's novel theoretical component of convergent retrieval learning.

## **CHAPTER 4**

### **GENERAL DISCUSSION**

#### **4.1 Overview of Findings**

This paper introduces a new model of memory retrieval, dubbed the Primary and Convergent Retrieval model. This model's chief contribution to the field of existing memory models is the assumption that associative learning can take place between the individual features that compose a memory trace, an assumption we argue is warranted by evidence from experimental paradigms showing that not all forms of practice have equivalent effects on long term memory. We provide support for the CR learning hypothesis by applying the PCR model to data from experiments previously reported in the literature, and running two novel experiments which confirmed key predictions of the PCR model of hidden benefits for retrieval practice on immediate final tests. In the first experiment, we tested and confirmed the prediction that because convergent retrieval learning from retrieval practice strengthens intra-item feature associations, benefits from successful retrieval guided with one cue should generalize to another unpracticed cue. In the second experiment, we tested and confirmed the prediction that because convergent retrieval learning from retrieval practice strengthens intra-item feature associations, subsequent retrieval attempts would proceed more quickly thanks to the increase in connectivity between features. Here, the implications of the success of the PCR model in explaining past and present data are discussed first for the field of research surrounding the testing effect and then for the field of memory models.

## 4.2 Broader Implications

Despite being first documented nearly 100 years ago, there have been no well specified (i.e., mathematical) models of the cognitive processes that underlie the testing effect. Prior verbal theories have explained the benefit of retrieval practice in various ways, appealing to mainly the desirable difficulty induced through the effort required to retrieve items from memory (Bjork & Bjork, 1992; Jacoby, 1978; Roediger & Karpicke, 2006b), the opportunity for elaborative processing strategies (Carpenter, 2009; McDaniel & Masson, 1985; Pyc & Rawson, 2010), and a transfer appropriate processing framework (Roediger & Karpicke, 2006a). A serious problem in debates over different theoretical accounts of test practice benefits stems from the fact that the explanations offered by various authors are not mutually exclusive from one another, making strong tests of their hypothesis difficult. The current paper offers a more concrete and testable theory about the source of retrieval practice benefits by connecting the literature on the testing effect with the broader literature describing computational models of memory.

The PCR model explains the long term benefits of test practice by hypothesizing that two independent learning processes take place when information is recalled (PR learning and CR learning), while only a single learning process takes place when information is restudied (PR learning). Since retrieval practice engages the additional CR learning mechanism, it provides additional protection against interference/forgetting. This protection comes in the form of strengthened intra-item feature associations, where this strengthening is driven by the need to activate

features of memory traces left inactive by retrieval cues. As shown by applying the PCR model to the data from Roediger and Karpicke 2006a, The CR learning mechanism provides a process based explanation of the cross-over interaction with practice method and retention interval, which has not been present in previous literature.

The CR learning mechanism also provides an explanation of why test practice benefits fail to appear when the practice test involves recognition. Carpenter and DeLosh (2006), as well as Glover (1989) tested participants using all possible combinations of free recall, cued recall, and recognition memory tests for the practice test and the final test. Both studies found the highest performance for items that received a free-recall practice test, regardless of what type of memory test was taken for the final test. Their findings show that the testing effect seems related to the degree to which the test practice requires the act of retrieval. This is consistent with the explanation of retrieval practice benefits offered by the PCR models. Since CR learning is hypothesized to underlie the benefits of retrieval practice, and CR learning requires that intra-item features associations activate features of the memory trace left dormant by the retrieval cues to occur, then recognition practice will not induce this pattern competition process since presentation of the recognition probe necessarily activates the features of that item, and the pattern completion process is irrelevant.

It is also important to situate the PCR model within the context of extant models of memory. Despite differences between existing models of memory and the functional form of the implementation of PCR used in this paper, PCR incorporates

many ideas from these existing models of memory. PCR still includes the key assumptions from the memory models previously discussed (SAM, REM, MINERVA, and the Complementary Learning Systems model). Specifically, we assume that retrieval from long term memory is a two stage process requiring the use of information from memory system about the overall strength of memories, as well as information about the specific content of the memory traces. Indeed, direct comparisons between the processes of the PCR model and the retrieval processes in these previous model can be drawn. The amount of features activated by retrieval cues during the Primary Retrieval process in PCR is analogous to the sampling strength of a memory in SAM, or the echo intensity in MINERVA, or the magnitude of the signal from the cortical network in the Complementary Learning Systems model. Each of these provides information about the absolute strength of the memory for a particular item, or about the magnitude of the memory system's response across many items. An items Convergent Retrieval threshold in PCR is analogous to the recovery strength of a memory in SAM, or the echo content in MINERVA, or the magnitude of the signal from the hippocampal network in the Complementary Learning Systems. Each of these quantities provides a measure of how easily the content of a specific episodic memory can be restored and thus recalled. The Convergent Retrieval process in PCR is most similar to pattern completion process in the hippocampal network from the Complementary Learning Systems model, and the echo content sharpening process in the MINERVA model. Each of these processes uses the currently active features of a memory trace to try and help bring



the currently active features “online” so that the memory system settles into a stable state to recall an item.

However, PCR differs from these previous models in one significant way. Prior models assumed that any learning process which took place would only strengthen the association between the retrieval cues and the features of the target memory trace. Then, these associations would make the magnitude of the memory response larger and its content clearer on subsequent attempts to retrieve the target memory with those retrieval cues. PCR retains this assumption, contained in the Primary Retrieval process, but posits an additional form of associative learning between the individual features of an item themselves, contained in the Convergent Retrieval Process. While this learning is indeed a separate *process*, it does not require the invocation of any additional learning *mechanisms*. Convergent Retrieval learning is still based on learning to associate one feature with another, based on their concurrent activation, just as in the Primary Retrieval process. The difference between the two is that in PR, the features associations cross discretized items, while in the CR learning process the associations are simply learned within the features that define a discretized items. Thus, the PCR model implies that when the convergent retrieval process is invoked (which is any time recall is necessary), something is learned about the internal consistency of the target memory itself. This greater internal consistency allows it to be more easily retrieved in the future, regardless of context or cue, and helps the memory to “spring to mind” more readily. This final point is a significant departure from the structure of extant memory

models, and will hopefully spark further investigation and research of item-specific learning in the future.

**Table 1: Best fitting PCR model parameters as applied to Roediger and Karipicke, 2006, Experiment 1**

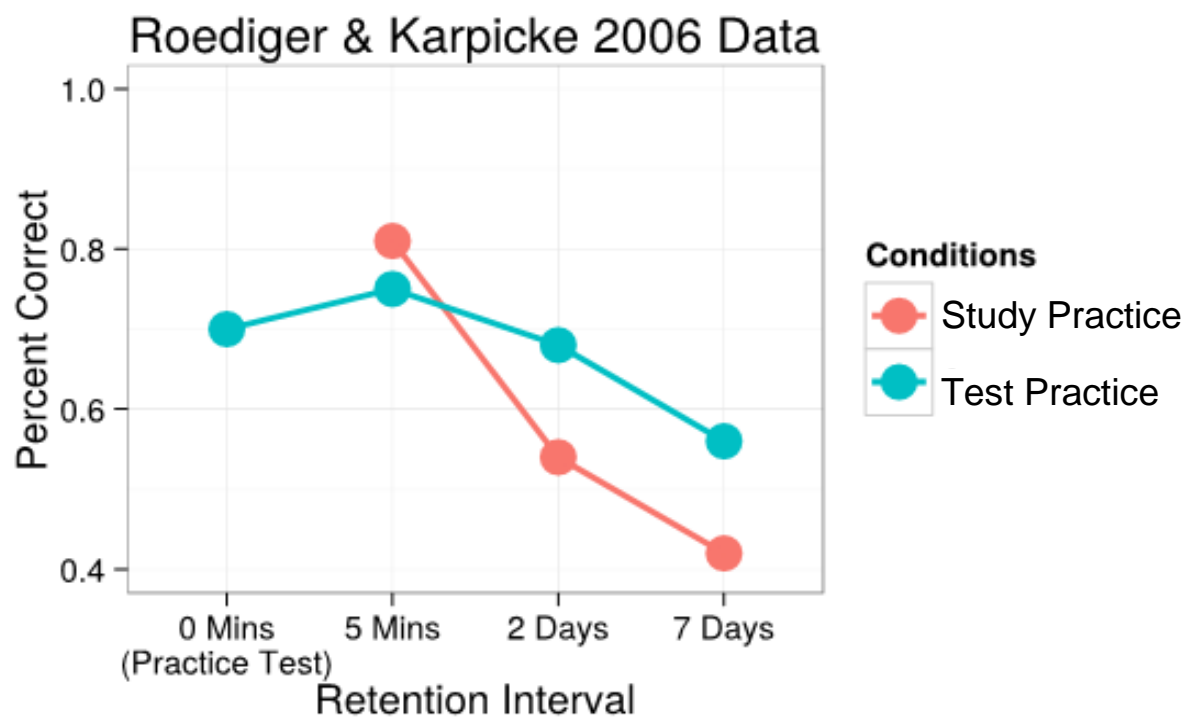
<i>Parameter</i>	<i>e</i>	<i>l</i>	<i>r</i>	<i>f<sub>2</sub></i>	<i>f<sub>2</sub></i>	<i>T<sub>min</sub></i>	<i>G<sup>2</sup></i>
<i>Value</i>	.585	.102	.090	.137	.064	15.725	.042

**Table 2: Average best fitting parameters for the PCR model used in Experiment 1**

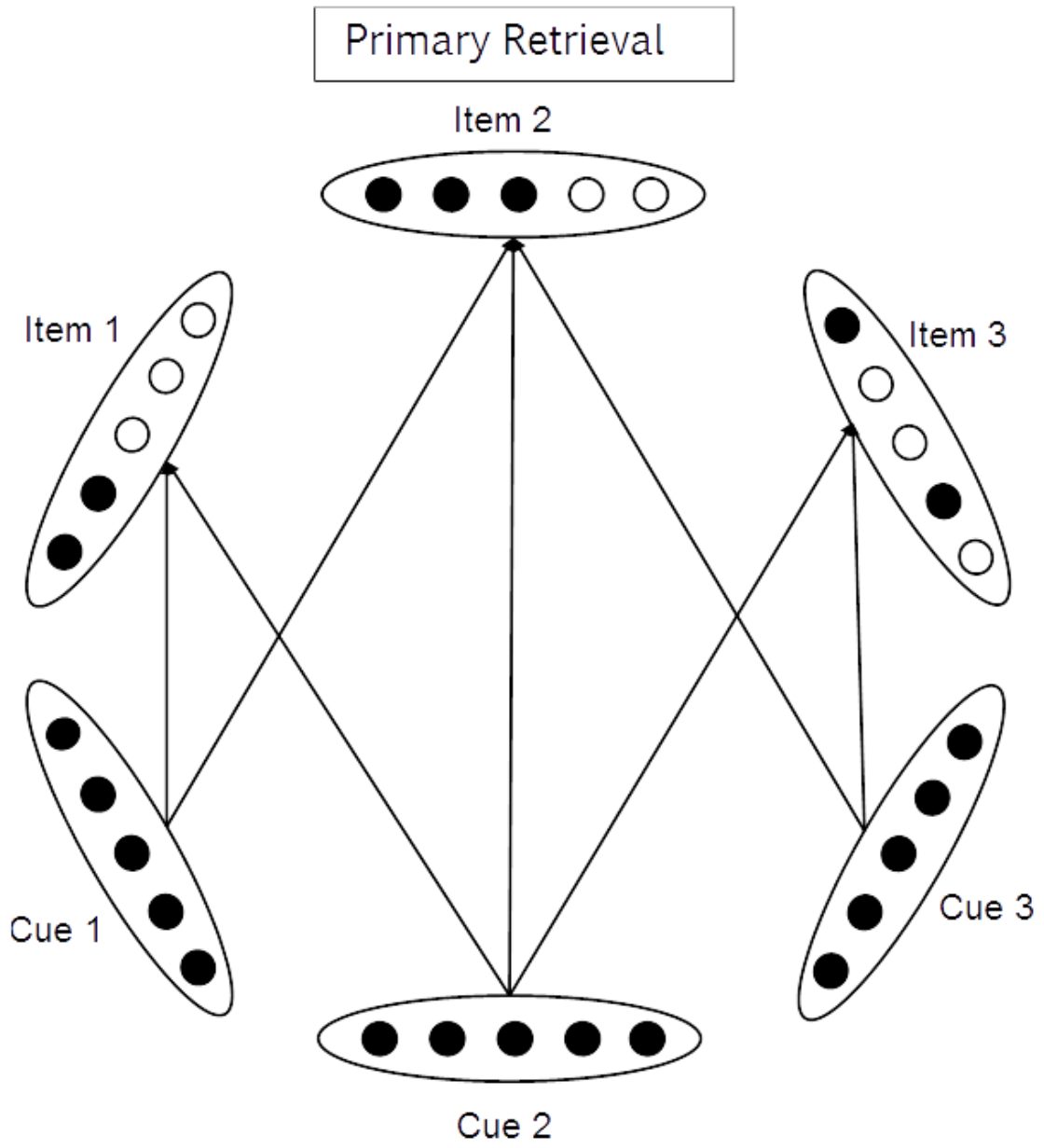
	<i>Parameter</i>	<i>e</i>	<i>l</i>	<i>r</i>	<i>f</i>	<i>s</i>	<b>-log likelihood</b>
<i>Immediate Final Test</i>	<i>Value</i>	.528	.185	.064	.043	.045	43.89
<i>Delayed Final Test</i>		.521	.137	.035	.109	.08	45.4

**Table 3: Average best fitting parameters for the PCR model used in Experiment 2**

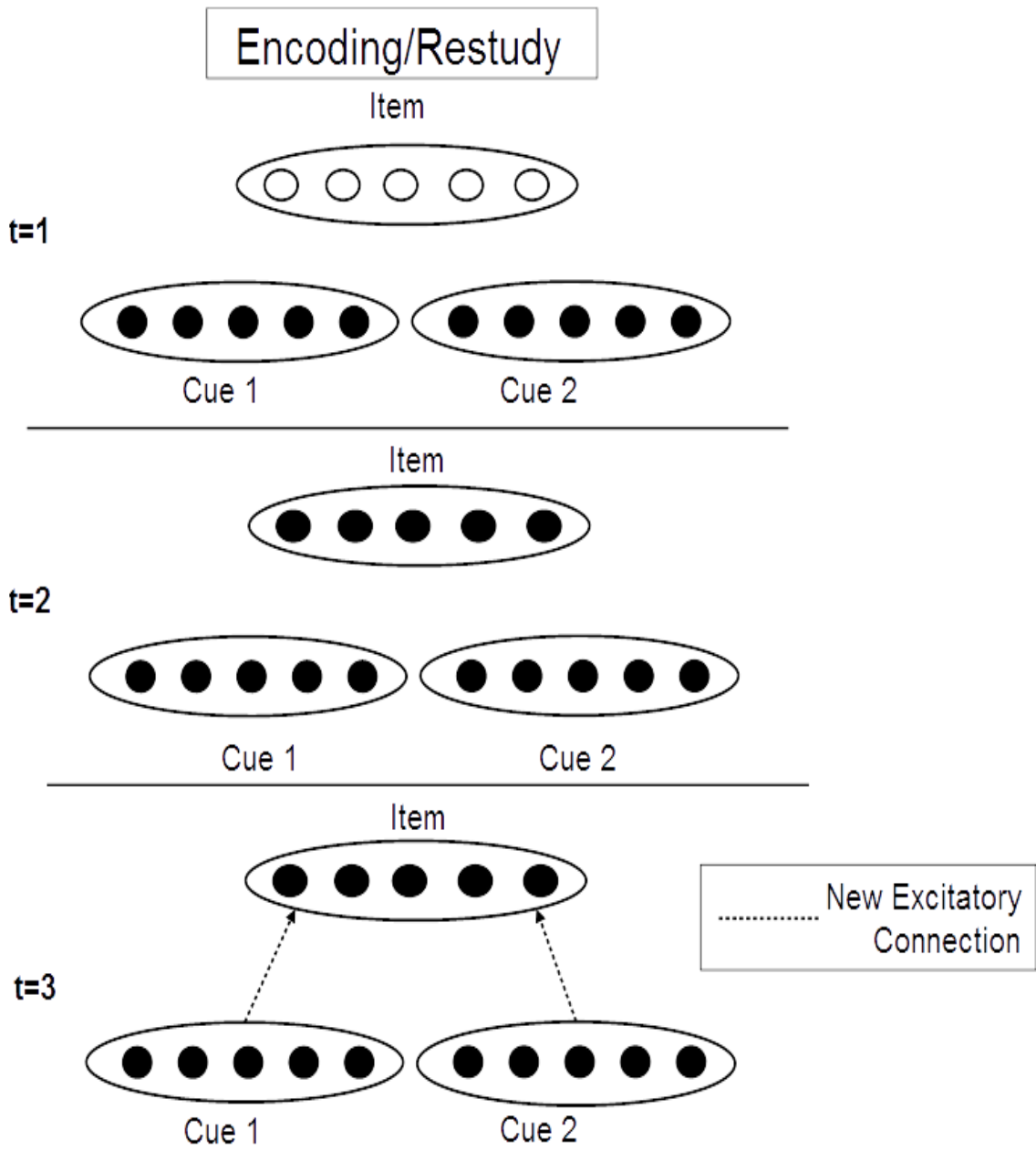
<i>Parameter</i>	<i>e</i>	<i>l</i>	<i>t</i>	<i>r</i>	$T_{min}$	$T_{max}$	$\lambda$	<b>-log likelihood</b>
<i>Value</i>	0.535	0.070	146.932	0.039	0.908	53.132	0.808	509.688



**Figure 1: Recall accuracy in Experiment 1 of Roediger and Karpicke 2006a.**

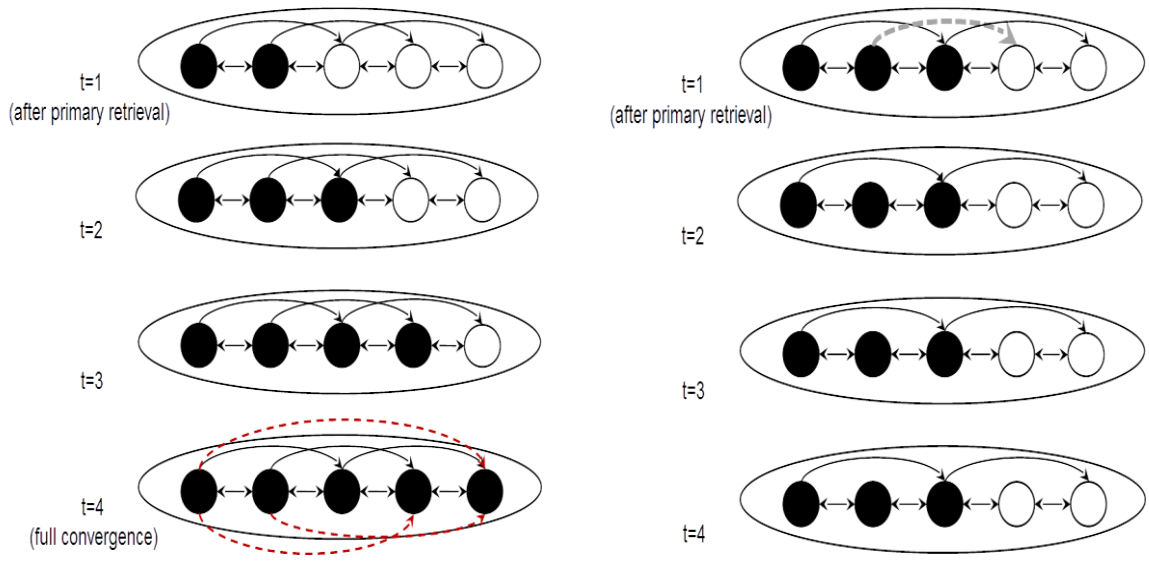


**Figure 2: Schematic diagram of primary retrieval**

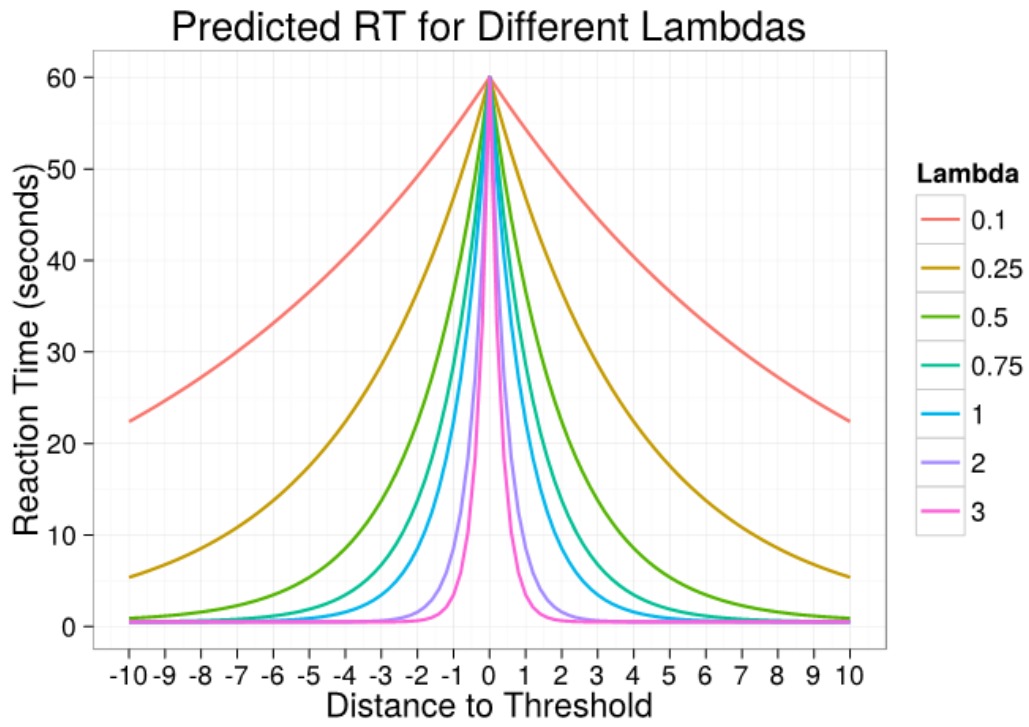


**Figure 3: Schematic diagram of primary retrieval process at encoding**

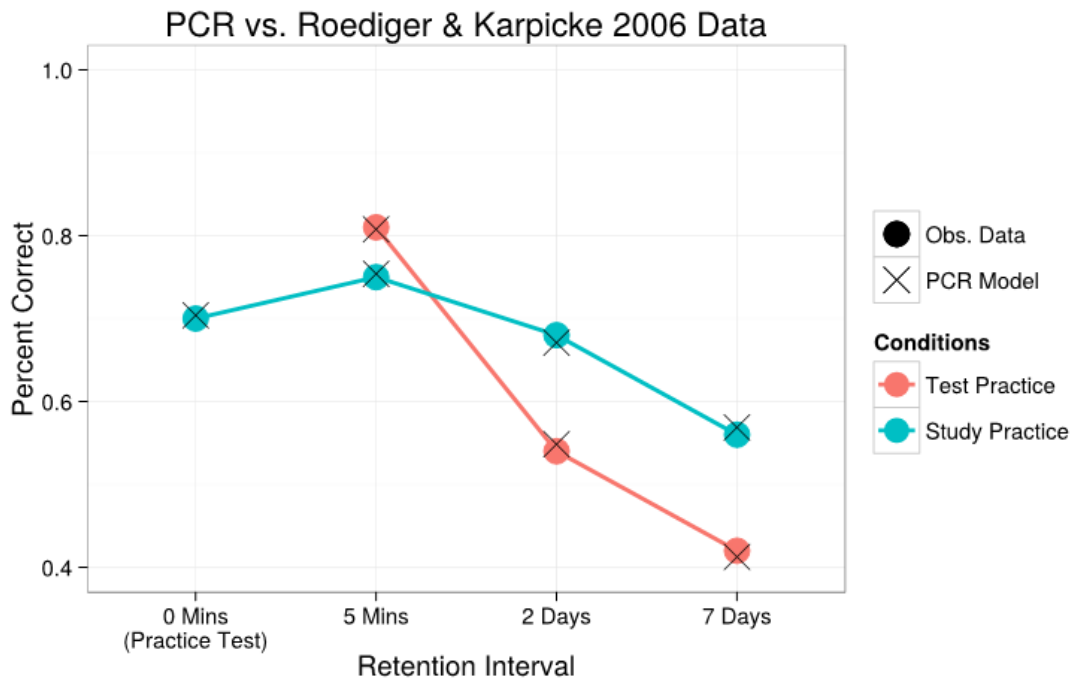




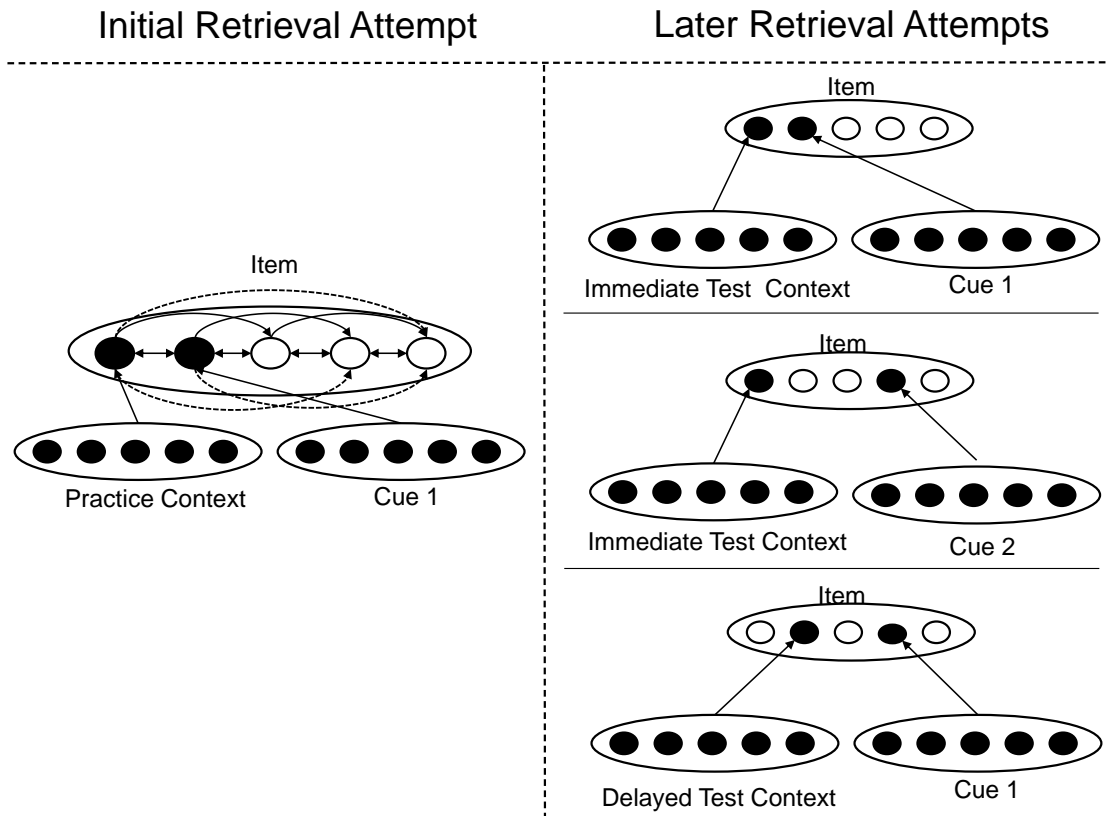
**Figure 4: Schematic diagram of successful (left panel) and unsuccessful (right panel) convergent retrieval process.**



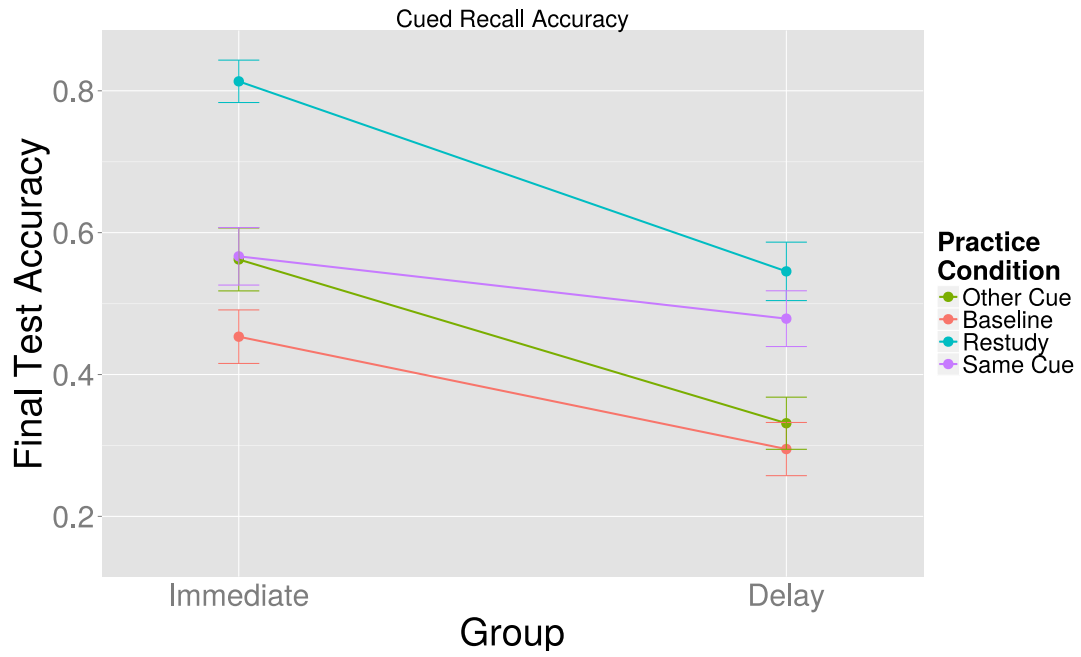
**Figure 5: Reaction time function in the PCR model**



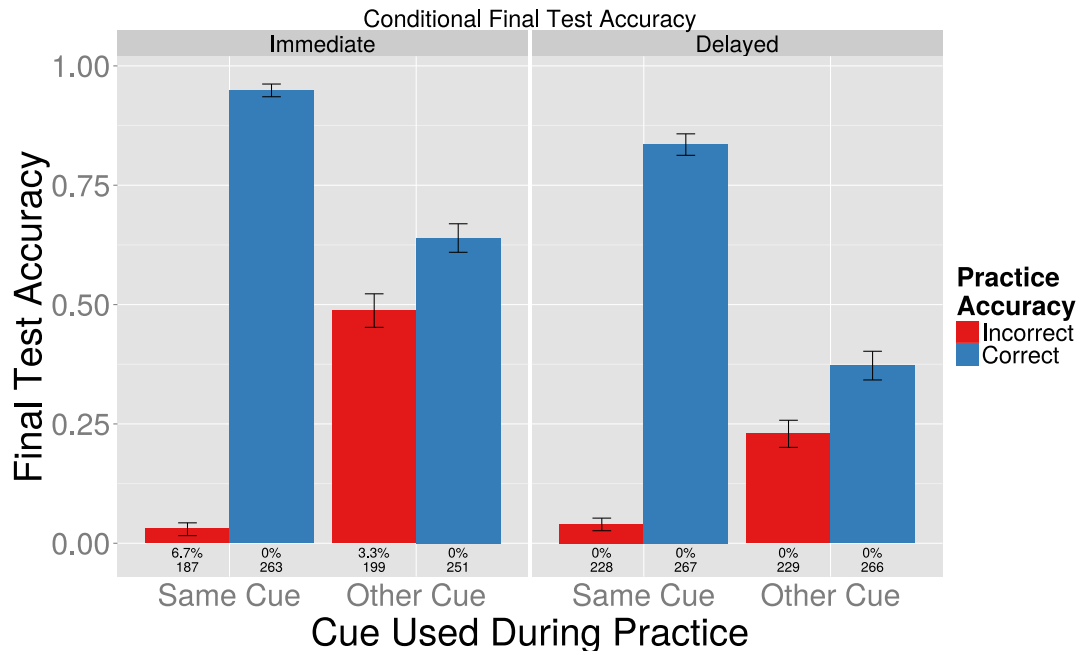
**Figure 6: Recall accuracy performance predicted by the PCR model from Experiment 1 of Roediger and Karpicke 2006**



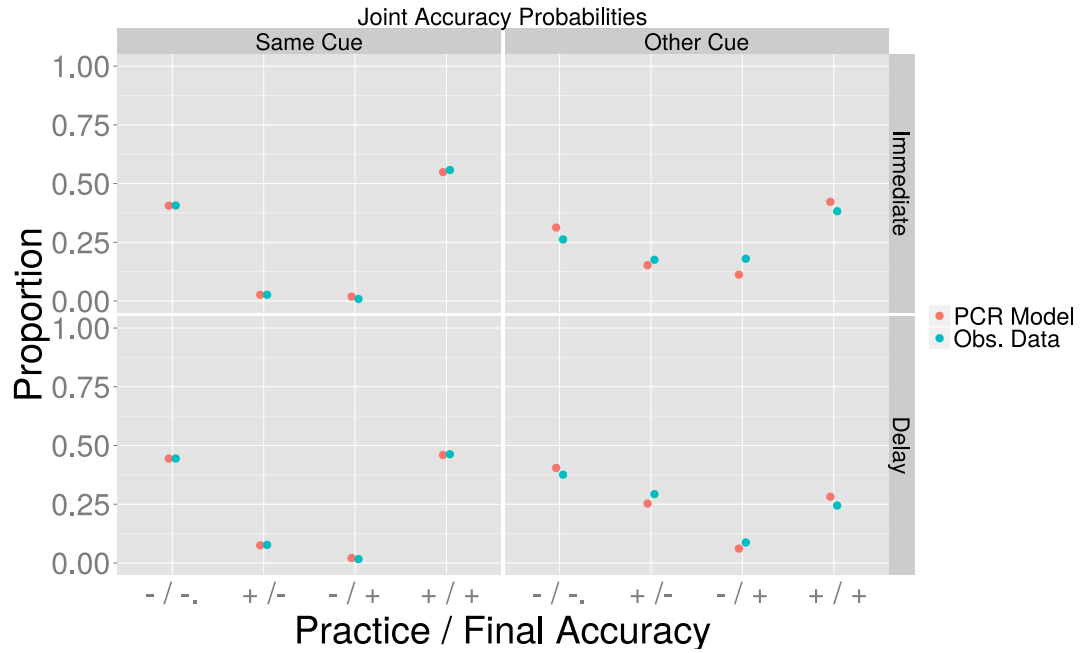
**Figure 7: Outcomes of subsequent retrieval attempts following successful convergent retrieval and convergent retrieval learning.**



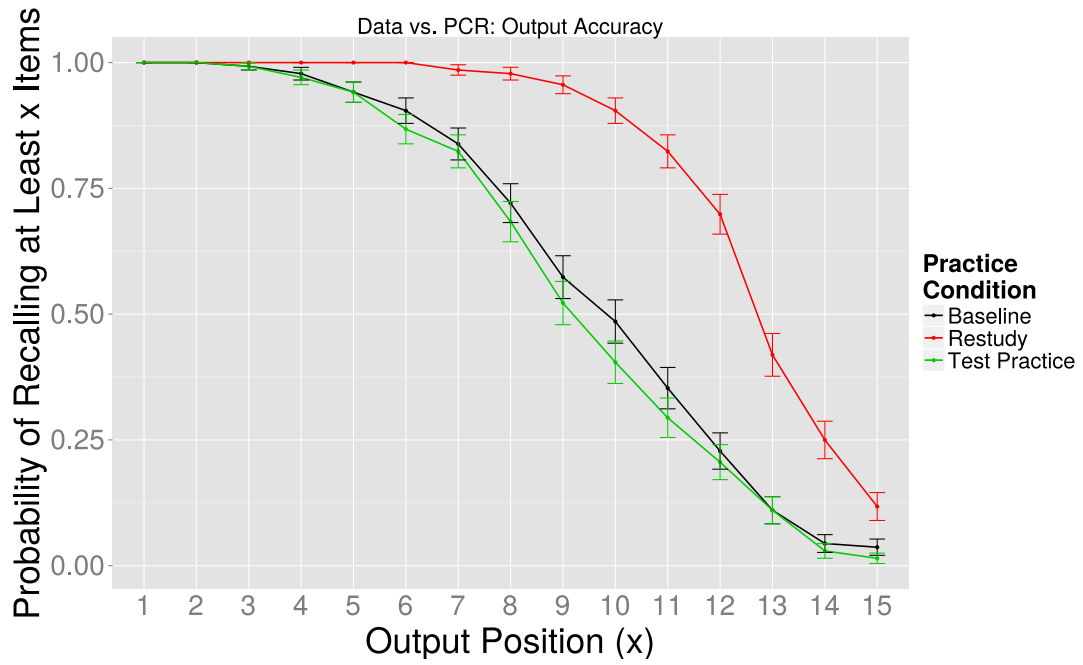
**Figure 8: Recall accuracy on the final cued recall test of Experiment 1.**



**Figure 9: Conditional probability of recall on final test in Experiment 1.**

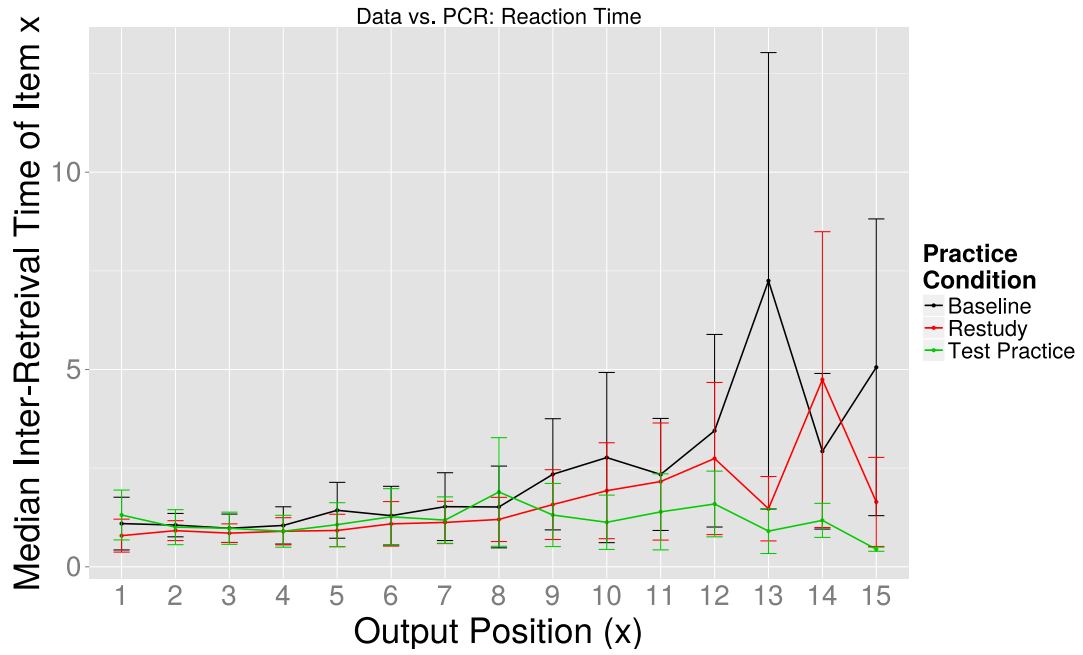


**Figure 10: Joint distribution of recall outcomes on the final and practice tests in Experiment 1.**

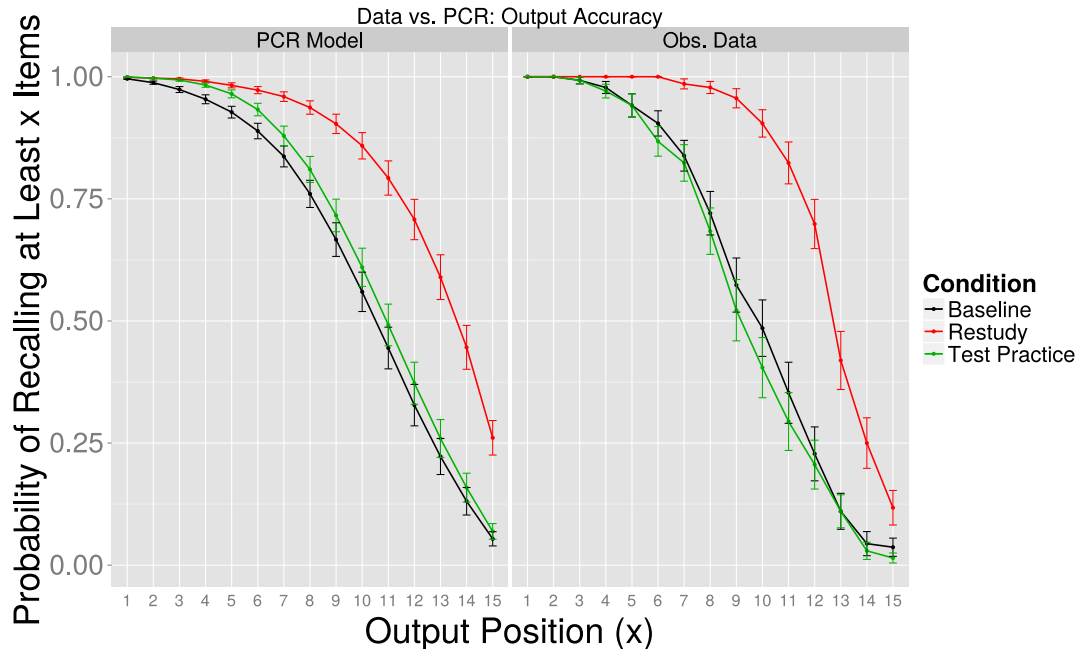


**Figure 11: Free recall accuracy observed in Experiment 2.**

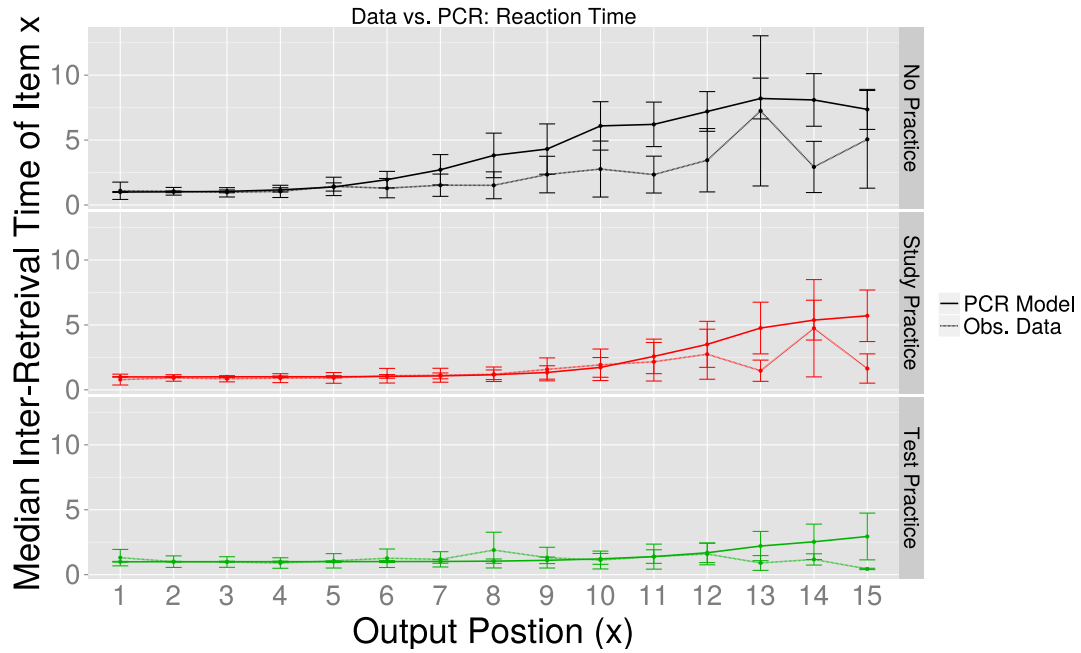




**Figure 12: Elapsed time (in seconds) between recall outputs in each condition of Experiment 2.**



**Figure 13: Free recall accuracy observed predicted by the PCR model.**



**Figure 14: The observed and predicted time (in seconds) between recall outputs in each condition of Experiment 2.**

## APPENDIX A

### BETA APPROXIMATION TO THE BINOMIAL

The beta distribution was used to model the amount of feature activation in the PR process and the CR threshold because of the assumption that features of an item stored in memory are not discretely active or inactive. However, the parameters of the Binomial distribution (probability of success, and the number of trials) remained more interpretable than the shape parameters of the beta distribution and were used throughout as free parameters. So, the probability of success and total number of trials were used to find the mean and variance of the binomial distribution they parameterized, and this mean and variance were used to solve for the shape parameters of the beta distribution with the identical mean and variance.

For any given binomial distribution with parameters  $p$  (the probability of success) and  $N$  (the total number of trials), the mean,  $\mu$ , of this binomial distribution is given by  $\mu = Np$  and the variance,  $\sigma^2$ , is given by  $\sigma^2 = Np(1 - p)$ . In our simulations,  $p$  is generally a free parameter, and  $N$  is fixed at 100 (the number of features for each item).

After finding  $\mu$  and  $\sigma^2$ , these values were divided by  $N$  to put them on the same scale as the mean and variance of the beta distribution. Then, these values were used to solve for the two shape parameters of a beta distribution with the same mean and variance as the binomial distribution parameterized with  $p$  and  $N$ . The first shape parameter  $a$  was found with the equation:

$$a = \mu \left[ \frac{\mu(1 - \mu)}{\sigma^2} - 1 \right]$$

And the second shape parameter  $b$  was found with the equation:

$$b = (1 - \mu) \left[ \frac{\mu(1 - \mu)}{\sigma^2} - 1 \right]$$

After simulating random samples from a beta distribution with shape parameters  $a$  and  $b$  100 times for each item, the resulting samples (bounded in the range  $(0, 1)$ ) were multiplied by  $N$  to put them back on the original scale of the binomial distribution.

## APPENDIX B

### APPROXIMATING THE REACTION TIME DISTRIBUTION

As described in the “PCR implementation and results” section of Experiment 2, the PCR model simulates the memory accuracy and reaction time (time spent attempting convergent retrieval before actually recalling or rejecting the item as unrecallable) 1000 times for each item in the word list. Thus, in some proportion of the 1000 list simulations in each of the 3 conditions (baseline practice test, final test after restudy, and final test after test practice) only 1 item will be output, or only two items will be output, etc., up to all 15 items output. Each of the simulated outputs is associated with a specific reaction time, and thus an “empirical” distribution of reaction times associated with the items recalled at each output position can be formed from the model’s output.

To assess goodness of fit to the observed data, this distribution of reaction times at each output position was smoothed using Gaussian kernel density estimation with a kernel bandwidth of 1. The kernel density was estimated at 900 evenly spaced points between .1 seconds and the maximum observable reaction time, 90 seconds. We will denote the estimated density at point  $i$  for output position  $j$  as  $D_{ij}$ . Next, the density at each of these 900 points was divided by the grand sum of the density across all 900 points, thus making the new grand sum of the density at all 900 points equal to 1. Mathematically, the new density at point  $i$  in output position  $j$ ,  $D'_{ij}$ , is given by:

$$D'_{ij} = \frac{D_{ij}}{\sum_{i=1}^{900} D_{ij}}$$

This was done for the reaction times in each of the observed output positions from the model's simulations (i.e. one up to a maximum of 15). A final step of normalization was done in order to take into account the models predicted accuracy for each condition. The density each of the 900 points in each of the output positions was multiplied by the proportion of simulations (out of 1000) in which the model produced an output at that positions. Mathematically, the new density at point  $i$  in output position  $j$  is given by

$$D''_{ij} = D'_{ij} * \frac{s_j}{1000}$$

Where  $s_j$  gives the number of simulated lists in which the model produced an output at position  $j$ . In other words, the densities were normalized by the probability of recalling at least  $j$  items. Thus, the grand sum of the densities over all the 900 points, spaced evenly between .1 to 90 seconds, across all output positions one through 15 was one (in each of the conditions), creating a joint reaction time and accuracy probability distribution.

Finally, these doubly normalized density values were used to assess the likelihood of each observation from the subject. For example, the height of the density curve for output position one,  $D''_1$  at each of the observed reaction for the first item output in the practice test baseline condition would determine the likelihood of that data. Since the density was calculated for responses times at tenths of a second intervals, observed reaction times were rounded to the nearest

tenth of a second before calculating the density at that point. Then these likelihoods were log transformed and summed to compute the overall likelihood of all the observations.



## REFERENCES

- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–459.
- Bi, G., & Poo, M. (1998). Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *The Journal of Neuroscience*, 18(24), 10464–10472.
- Bi, G.-Q. (2002). Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms. *Biological Cybernetics*, 87(5/6), 319.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, R. M. Shiffrin, A. F. (Ed) Healy, S. M. (Ed) Kosslyn, & R. M. (Ed) Shiffrin (Eds.), *Essays in honor of William K. Estes, Vol. 1: From learning theory to connectionist theory; Vol. 2: From learning processes to cognitive processes*. (pp. 35–67). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning & Verbal Behavior*, 5(4), 325–337.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990.
- 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(6), 1563–1569.
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13(5), 826–830.
- Dan, Y., & Poo, M.-M. (2006). Spike Timing-Dependent Plasticity: From Synapse to Perception. *Physiological Reviews*, 86(3), 1033–1048.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91(1), 1–67.
- Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers*, 16(2), 96–101.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95(4), 528–551.

- Hintzman, D. L., & Curran, T. (1994). Retrieval Dynamics of Recognition and Frequency Judgments: Evidence for Separate Processes of Familiarity and Recall. *Journal of Memory and Language*, 33(1), 1–18.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, 17(6),
- 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(1), 112–127.
- Jang, Y., Wixted, J. T., Pecher, D., Zeelenberg, R., & Huber, D. E. (2012). Decomposing the interaction between retention interval and study/test practice: The role of retrievability. *The Quarterly Journal of Experimental Psychology*, 65(5), 962–975.
- 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(2), 85–97.
- Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *The American Journal of Psychology*, 109(3), 451–464.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87(3), 252–271.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 371–385.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110(4), 611–646.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, 108(2), 311–345.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335–335.

- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134.
- Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 163–178.
- Roediger, H. L. I., & Karpicke, J. D. (2006b). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249 – 255.
- Roediger, H. L. I., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210.
- Roediger, H. L. I., & Tulving, E. (1979). Exclusion of learned material from recall as a postretrieval operation. *Journal of Verbal Learning & Verbal Behavior*, *18*(5), 601–615.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252–257.
- van den Broek, G. S. E., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, *22*(7), 803–812.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571 – 580.
- Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*(1), 89–106.