The Primary and Convergent Retrieval Model of Recall

William J. Hopper (whopper@psych.umass.edu)
Department of Psychological and Brain Sciences, 135 Hicks Way
Amherst, MA 01003 USA

David E. Huber (dehuber@psych.umass.edu)
Department of Psychological and Brain Sciences, 135 Hicks Way
Amherst, MA 01003 USA

Abstract

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.

Keywords: Memory Modeling; The Testing Effect, Retrieval Practice

Two-Stage Models of Recall

As briefly reviewed here, memory models typically assume that recall is a two-stage process. For instance, the Search of Associative Memory model (SAM, Raaijmakers & Shiffrin, 1981) contains a *sampling* stage that selects a specific memory trace from a pool of active traces, followed by a *recovery* stage that extracts the details of the sampled memory. Similarly, the MINERVA 2 model (Hintzman, 1984) differentiates between an *intensity* response (measuring overall activation) that is used to weight the contribution of memories to the *echo content*, which is subsequently 'cleaned up' through a recursive process to produce the desired content. Norman and O'Reilly (2003) assumed these separate processes reflect the operations of different brain regions, with parahippocampal cortex providing a *familiarity* response, such as used in recognition, whereas recall requires *pattern completion* that depends on the actions of the hippocampus.

These models, and others, include a stage that isolates relevant memories based on a scalar value for retrieval strength followed by a stage in which sufficient detail is extracted to produce on overt recall response. However, these models also assume that learning is passive and any opportunity to encode a memory will affect both stages in a similar manner. Additionally, the time course of this second stage is not specified by extant memory models. Here, we present a new memory model which specifies the learning processes and time course of this second stage, motivated in large part by studies of retrieval practice effects.

Restudy, Test Practice, and Forgetting Rates

The assumption of passive learning appears at odds with testing effect studies that indicate greater learning from retrieval practice (see Roediger & Karpicke, 2006a for a review). In these studies, participants learn some new material, after which the material is practiced by restudying, or by taking a practice test (either with or without feedback). A retention interval follows this practice, after which participants take a final test. The final test often reveals an advantage for material practiced with a test relative to material practiced by restudying, and this advantage grows with retention interval. In other words, these two types of practice produce different forgetting rates. A striking example of different forgetting rates is found when test practice occurs without feedback. In this case, if the retention interval is short (e.g., 5 minutes), restudying produces higher accuracy than a practice test. However, if the retention interval is longer (e.g., 24 hours), this relationship is reversed and test practice produces better accuracy than restudying (Roediger & Karpicke, 2006b; Toppino & Cohen, 2009; Wheeler, Ewers, & Buonanno, 2003).

This crossover interaction comparing restudy to test practice without feedback is partially explained by realizing that there is no opportunity for additional learning for the items that fail to be retrieved during test practice. Thus, the test practice reflects a bifurcated distribution (Kornell, Bjork, & Garcia, 2011). However, this account still assumes greater learning for the items that were recalled during test practice as compared to the learning from restudy and yet no explanation is provided as to *why* this is the case. Furthermore, the crossover interaction occurs even when considering recallable items, as determined by an initial test prior to subsequent test practice or restudy (Jang, Wixted, Pecher, Zeelenberg, & Huber, 2012). To explain why the act of recall produces qualitatively different learning than passive restudy, we developed the Primary and Convergent Retrieval (PCR) model of recall.

The Primary and Convergent Retrieval Model

The PCR model retains the two-stage recall architecture of previous memory models, with the first stage termed *Primary Retrieval* (PR) while the second is termed *Convergent Retrieval* (CR). Primary Retrieval describes the initial process of activating all relevant memories based on their associations with the current retrieval cues (e.g., the current temporal context and any item information such as a word or picture provided as a cue). However, even the most active memory in PR may be incomplete (e.g., some of the features of the memory remain inactive). Convergent Retrieval describes a second stage in which the memory system focuses on the most active memory and attempts to activate any inactive features through intra-item associations between the features of that memory.

General Assumptions of PCR

PCR assumes that all memory traces and retrieval cues are composed of a finite number of features. New unidirectional associations between features are formed according to the temporal sequence of events. For instance, if feature A becomes active at time *t*, and feature B becomes active at time *t*+1, an excitatory connection is formed from feature A to feature B. This form of learning is consistent with spike time dependent plasticity in which potentiation of a synapse only occurs if the pre-synaptic cell fires before the post-synaptic cell (for examples see Dan & Poo, 2006). We assume this learning rule applies in all situations. Thus, context features become associated with item features during study because context is active prior to presentation of the item. More importantly, in terms of explaining the testing effect, initially active features of an item become associated with initially inactive features of the same item, provided that those features are subsequently retrieved. This intra-item learning explains the extra benefit of retrieval practice.

Primary Retrieval

When a retrieval attempt is initiated, all currently active features (e.g., the current temporal context and any retrieval cues) serve as cues to activate features of memory traces. Thus, PR is cue-dependent, meaning that the content and magnitude of the memory system's response depends completely on the features of the retrieval cues. During initial study, the retrieval cues (context and an item presented as a cue) are typically active first, followed by the target item, allowing associations between these retrieval cues and the features of the target item. However, encoding is likely to be incomplete and error prone. Furthermore, the temporal context will naturally change between study and test (e.g., Howard & Kahana, 2002). Thus, the activation of the target features that occurs with PR will be incomplete. Because a naming response requires full retrieval of the item, a pattern completion process is needed for recall success. We term this process 'convergent retrieval'.

Convergent Retrieval

Convergent Retrieval is the process by which initially dormant features that were missed by PR become active, via intra-item excitatory connections between the individual features that define the target item. Even for readily known items (e.g., high frequency words), the retrieval cues may fail to activate enough of the item, such as occurs with 'tip-of-the-tongue'. If CR succeeds in activating all the remaining features, two things occur: 1) the item can then be recalled; and 2) new associations between the features activated by PR and the features subsequently activated by CR can be formed (note that this also holds true when the target item is provided after CR failure, such as occurs in test practice with feedback; this explains why the testing effect is more powerful with feedback).

This second outcome, called intra-item learning, represents a theoretical departure from most memory models, which do not explicitly model the learning between features of an item. This intra-item learning makes it easier to recall the item *regardless of the initial state of activation that occurs with PR*; because the associations between the individual features that compose an item are a property of the item, rather than the association between the item and retrieval cues, the intra-item learning that takes place following successful test practice benefits recall in a cue-independent manner. This intra-item learning reduces the numbers of steps required for CR, resulting in faster recall. Thus, even in situations where intra-item learning fails to increase the probability of recall, it will decrease retrieval latency for the items that are recalled.

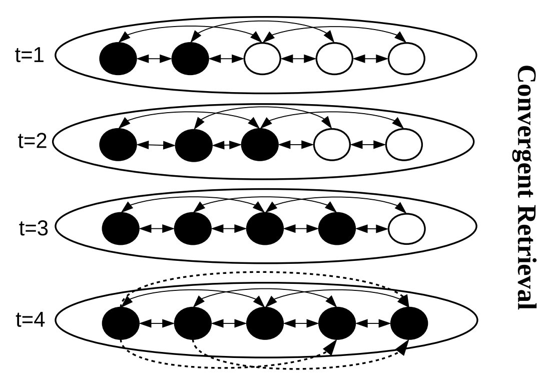

Figure 1: An example of successful Convergent Retrieval. The features initially activated by retrieval cues during Primary Retrieval activate the remaining inactive features according to the associative connections between each feature. Following convergence, additional associations are learned between the initially active features, and the subsequently activated features.

Figure 1 shows an example of the CR process, beginning where PR ends. In this example, retrieval cue features only activate two of the five features of an item in memory (features of an item are shown as circles inside an oval, with currently active features represented by the filled circles; the existing intra-tem associations are indicated by the solid arrows). In this example, if a feature requires two excitatory inputs to become active, all of the initially dormant features will eventually become active across three time steps. New intra-item associations are now formed according to this progression of events (represented by the dashed arrows).

This discussion outlines the guiding principles behind the PCR model. However, a full-scale neural network implementation would require many auxiliary assumptions (e.g., the exact time function underlying learning, the exact rule for feature activation during CR, the nature of temporal context change, etc.). Next we present a simple abstract mathematical model that approximates these guiding principles, noting that a fully mechanistic instantiation of these principles might deviate somewhat from this simple model.

A Binomial Instantiation of PCR

We assume each item consists of a finite number of features (set to 100 for convenience) and that each item requires a specific number of active features for CR to be successful (features are discretely active or inactive). This captures pre-experimental item differences in which some items are more easily recalled even if encoding is incomplete or context has changed greatly (i.e., even with weak PR). The value of each item's CR 'threshold', θ_i, is sampled from a binomial distribution with probability *t* and 100 counts (one count for each feature). We use the symbol $B(\,)$ to indicate a binomial distribution.

$$\theta_i \sim B(t, 100)$$

We assume that initial encoding is incomplete, or prone to errors such that even an immediate final test (i.e., one for which context has not changed) fails to activate all of the target item's features. As with item differences in threshold, we assume some items are better encoded than others, and the number of features that are encoded for each item, α_i, is also sampled from a binomial distribution, but in this case the probability parameter is *e*.

$$\alpha_i \sim B(e, 100)$$

The parameter *e* may be thought of as an encoding rate parameter, and more time spent studying will result in higher values of *e*.

Restudy and *successful* recall provide another opportunity for encoding item features by associating them with retrieval cues and both forms of practice increase the value of α_i (there is no learning for unsuccessful recall without feedback). This learning during the practice phase of an experiment is again captured with a binomial sample, but in this case the probability parameter is *l*. However, unlike initial study, this learning only applies to features that were not originally encoded and so the number of counts for this binomial distribution is 100 - α_i, resulting in the following expression for change in the number of encoded features.

$$\Delta\alpha_i \sim B(l, 100 - \alpha_i)$$

Forgetting is implemented in the model by reducing the value of α_i. Modeling forgetting by reducing the number of features activated by retrieval cues corresponds to the assumption that temporal context shifts during the retention interval such that PR only activates a subset of the item features that were previously encoded in relation to the context at the time of initial study. The reduction in features activated during PR is captured by a binomial sample with probability *f* as follows.

$$\Delta\alpha_i = -B(f, \alpha_i)$$

As outlined previously, CR success produces new intra-item learning, as the features of an item become associated with each other because they were activated in a progressive manner. We capture intra-item through a reduction in the CR threshold θ_i. This change in threshold is again a binomial sample, but with probability parameter *r* and the number of counts equal to the current threshold.

$$\Delta\theta_i = -B(r, \theta_i)$$

This reduction makes items more easily recalled regardless of how they are cued (i.e., even when PR is weak). In simulations with the model, recall success for each item is discretely determined according to a comparison between α_i and θ_i. If $\alpha_i > \theta_i$, then PR has activated more features than are required for CR success, and thus the entire content of the item is retrieved and available to be named in a recall response. However, as described next, a key component of the recall process is the order in which items are considered for CR and the time that it takes to attempt CR.

In a cued recall test, it is likely that the target memory is the only memory that has any appreciable activation (although this is not true in cued recall experiments that pair the same cue with multiple targets). In the case where the target is the only active memory, the important question is whether that memory is sufficiently active to support CR success and whether CR can be achieved in the allowed time. As outlined in Figure 1, we assume that CR takes time as dormant features progressively activate. We implement this by assuming that the time needed for CR relates to the distance from threshold, α_i - θ_i. More specifically, the Reaction Time (RT) to recall (or fail to recall) target item *i* is a negative exponential function.

$$RT_i = T_{min} + (T_{max} - T_{min})\left(e^{-\lambda|\alpha_i - \theta_i|}\right)$$

Here, T_{max} and T_{min} are upper and lower bounds on possible response times, respectively. Importantly, the term in the exponent uses the absolute magnitude of the difference. This captures the intuition that an item that is on the 'tip-of-the-tongue' is one that will take a long time before it is recalled or before the memory system admits defeat. In contrast, items that are far above threshold are recalled very quickly. Similarly, items that are far below threshold fail to progress in the CR process, and the retrieval attempt is quickly abandoned.

In a free recall test, the order in which memories are considered for CR plays a crucial role. More specifically, a considered item may fail to be recalled owing to CR failure but another possibility is that the item was not recalled because CR was never attempted. We assume that the memory system does not directly know whether CR success is possible for each memory; this knowledge requires the actual engagement of the CR process. However, the amount of feature activation (i.e., the amount of PR) is a good proxy for CR success. In other words, it is more likely that $\alpha_i > \theta_i$ when selecting for items with high α_i. Thus, rather than random sampling as in the SAM model, we assume that CR is attempted for each item in descending rank order of α_i. In this way, interference occurs in the construction of the rank ordered list; strong memories will be near the top of the list, making it more likely that a test taker runs out of time before considering items farther down the list.

The predicted probability of recall in a free recall experiment is determined as follows. Monte Carlo simulations generate 1,000 hypothetical lists of items in which each item has an initial threshold θ_i, and α_i value owing to initial study, followed by changes to these values with restudy or test practice. Simulating final recall, CR is attempted for each item in the rank ordered list of α_i values for each list, while keeping track of the total time elapsed during the recall session until the allowed time is exceeded or the entire list has been considered for CR.

Applying PCR to Roediger & Karpicke 2006b

Experiment 1 from Roediger and Karpicke 2006b manipulated practice method (test practice without feedback vs. restudy) and retention interval (5 minutes, 2 days, and 1 week), using free recall for both test practice and the final test. Participants were given two prose passages that each contained 30 'idea units'. Participants then took a practice test on one passage, and restudied the other. After one of the 3 retention intervals, participants took a final test on both passages. Figure 2 shows recall accuracy in each condition, as well as mean accuracy on the practice test.

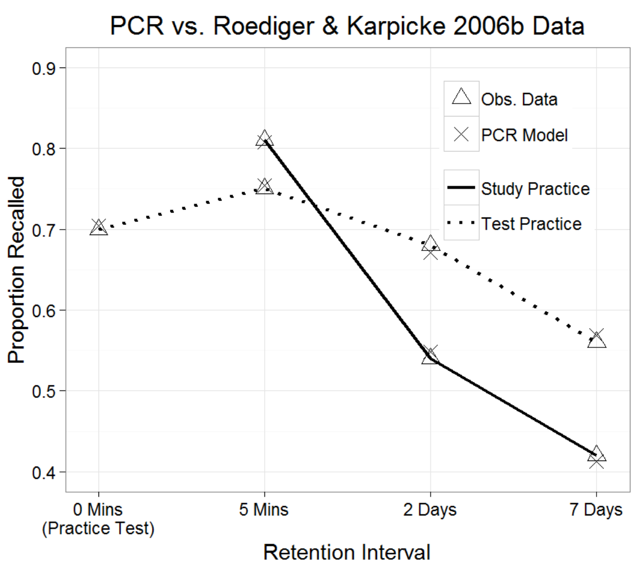

Figure 2: Recall accuracy from Roediger and Karpicke 2006b shown with the accuracy predicted by PCR.

We fit the PCR model to the average free recall accuracies reported by Roediger and Karpicke 2006b (shown in Figure 2). The e, l, r, T_{min}, f_2, and f_7 parameters were allowed to freely vary, while T_{max} was fixed at 60 seconds. Each item's CR threshold θ_i was drawn from a Binomial distribution with probability parameter t fixed at .5. We assumed that no forgetting occurred in the immediate final test condition (i.e., f = 0). 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 indicates that the model is not rejected ($\chi^2(1) = .042$, $p = .83$). The predicted accuracy of the best fitting model is shown along with the observed data in Figure 1.

One the one hand, the accuracy of this fit is not surprising considering that 6 free parameters were used to fit 7 conditions. On the other hand, this crossover interaction between type of practice and retention interval is theoretically challenging and these results are problematic for any memory model, regardless of the number of free parameters. Ultimately, the fit to these data can be considered a "proof of concept" that the PCR model is able to explain the pattern of recall accuracy results observed in studies of the testing effect. In the following sections, we test novel predictions of the PCR model in a data set which imposes a much greater amount of constraint.

Hidden Benefits of Retrieval

The PCR model includes the notion of a bifurcated distribution (e.g., learning only applies to items that were recalled during test practice). The notion of a bifurcated distribution suggests that a great deal of learning has occurred for the items that were recalled on the practice test, but does not specify the nature of this extra learning for tested items. The PCR model attributes this extra learning to increased associations between the individual features of a target item. The incremental nature of CR during test practice promotes this type of learning whereas this does not occur with restudy considering that all of the item features are presented simultaneously with restudy. However, in the absence of feedback during test practice, this extra learning is not apparent *when examining accuracy* on an immediate final test because it only occurs for the items that would have been recalled *even if* they hadn't received test practice. Thus, the effects of this additional intra-item learning are masked in the short term and only emerge after a delay period, where this extra learning serves as additional protection against interference and forgetting.

However, it should be possible to reveal the benefit of retrieval practice on an immediate final test by measuring how long it takes for each item to be retrieved, instead of just how many items are ultimately retrieved. The PCR model predicts faster retrieval following test practice even if overall accuracy does not improve on an immediate final tests.

Practice Tests Reduce Retrieval Latency

To see why intra-item learning reduces the latency of subsequent retrievals of the same item, consider the example of CR shown in Figure 1. Even if successful test practice failed to strengthen associations between the retrieval cues and the item features (i.e., even if PR still only activated 2 of the 5 features), CR would be achieved in fewer time steps during subsequent retrieval attempts. The newly formed intra-item associations (shown by the dashed arrows), allow CR to occur in 2 time steps rather than 4 time steps. This is because the 2 features initially activated during PR are now associated with *all 3* of the initially dormant features, instead of just 1.

Restudy also reduces retrieval latency, but for a different reason. Following restudy, the reduction in latency occurs because more features are active in the first time step (i.e., after PR). In other words, learning from restudy increases the number of features activated by the retrieval cues such that fewer features need to be filled in during CR. It is important to note that this type of learning also occurs with successful test practice. In other words, both restudy and test practice result in more features activated by the retrieval cues. However, successful test practice also boosts intra-item learning, and thus the CR process operates more efficiently (i.e., fewer time steps) to fill in the remaining features. In summary, for an immediate final test, accuracy is predicted to be higher following restudy because restudy boosts PR for all of the items. In contrast, retrieval latency is predicted to be faster following test practice because recalled items will have benefited both from better PR and from more efficient CR.

We tested these predictions in a free recall experiment in which 34 participants studied lists of 15 words, followed by either restudy or a practice test, and then a final test (this procedure was repeated 8 times). On the memory tests, participants were given 90 seconds to recall the 15 words in any order. Participants completed a 30 second math distractor task between both the initial study and the practice phase, and between the practice phase and the final test. This experimental design yields 3 conditions to measure: performance on the practice test itself (i.e., the baseline or "no prior practice" condition), performance on the final test following restudy practice, and performance on the final test following a practice test. In addition to recall accuracy, we analyzed the elapsed time between each item recalled (the Inter-Response Time, or IRT). The accuracy and latency results were in line with the PCR models predictions: restudy produced the highest accuracy (average of 81% correct) and faster IRTs than baseline, while IRTs were the fastest on a final test following test practice, despite no substantial change in accuracy (61% correct on the practice test and 59% correct on the final test). The median IRTs for each possible output position in this experiment are shown in Figure 3. Note that the lack of a "testing effect" in terms of recall accuracy is expected in this situation, as the learning benefits of test practice only apply to items *already* able to be recalled, and thus should only be expected to emerge with a longer retention interval. Demonstrating the generality of these latency effects, a similar speed-up has been found with a final cued recall test following cued recall practice (Broek, Segers, Takashima, & Verhoeven, 2014).

Applying PCR to Free Recall Latencies

The PCR model was simultaneously fit to the observed recall accuracy and IRTs from each item on each list of the free recall experiment, separately for each subject. More specifically, the PCR model predicts the shape of the IRT

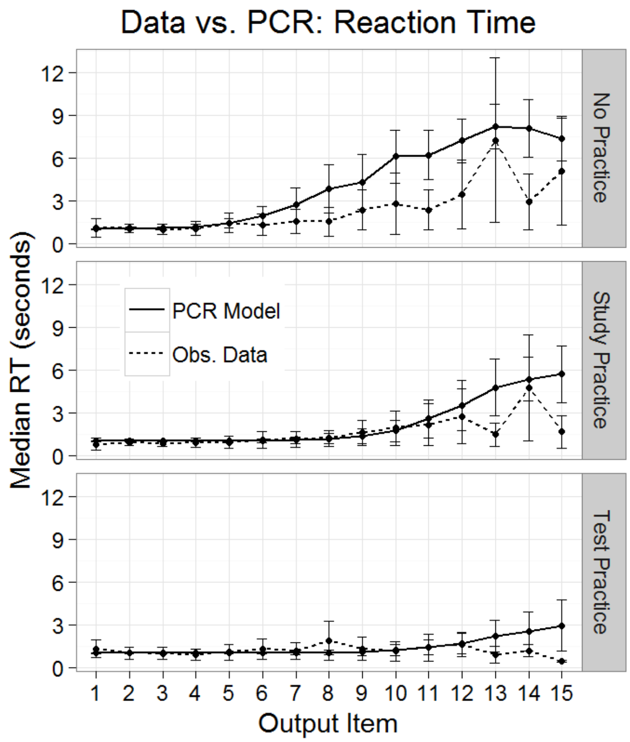

Figure 3: Predicted and observed inter-retrieval times in the no practice, study practice, and test practice conditions. Inter-retrieval times increased as function of output position in the no practice and study practice conditions, but not in the test practice condition.

distributions, and these were used to produce a maximum likelihood fit of the joint probability of producing each observed latency at the observed output order position within the test list (e.g., taking 4.3 seconds after recall of a 4th item to then recall a 5th item). In this manner, the model explained the shape of the separate IRT distributions as a function of output position within the test list, and as a function of prior restudy or prior test practice. The predicted IRTs using the best fitting model parameters for the baseline, restudy, and test practice conditions are shown along with the observed data in Figure 3.

To allow the model to capture the data, the T_{min}, T_{max}, and λ parameters were allowed to freely vary in addition to the *e*, *l*, *r*, and *t* parameters. When using a binomial distribution, the model generated IRT distributions are discrete because there is a finite number of possible RTs representing a finite number of possible values for the difference from threshold (however, this finite number is large when considering that an observed IRT may reflect some number failed CR attempts before a successful CR). A continuous IRT distribution was produced owing to two factors. First, rather than using the discrete Binomial distribution, we used a continuous Beta distribution that approximates the Binomial by having the same mean and variance. Theoretically, this corresponds to consideration of partial feature activation. Second, rather than assuming that each CR attempt was initiated precisely at the offset of the last CR, we imposed a standard normal (0 mean and 1 second standard deviation) for the start time of each CR. This was implemented with Gaussian kernel estimation of the Monte Carlo data. This procedure produced a family of IRT distributions across the 15 possible output positions (i.e., these 15 distributions integrate to value 1.0), and in this way the model simultaneously explained both the accuracy data and trial-by-trial IRTs.

General Discussion

The PCR model has important theoretical implications for the testing effect literature as well as the broader field of memory research. To this date, there have been no well specified (i.e., mathematical) models of the cognitive processes that underlie the testing effect. Beyond serving this need, the PCR model provides a conceptually novel account of the benefits from retrieval practice. In the testing effect literature, the leading theories include overlearning (i.e., testing simply produces more learning), transfer appropriate processing (i.e., the best way to learn to recall at a later date is to practice recall), and desirable difficulties (i.e., testing requires more effort). Intra-item learning is conceptually different than any of these accounts. It is closest to transfer appropriate processing, but it supposes that the act of recall introduces a king of learning that will benefit future recall attempts for that item under a variety of circumstances (i.e., beyond situations that are identical to those of test practice). This theory provides an explanation of why test practice benefits are largest following recall practice (Carpenter & DeLosh, 2006); because CR is not reliably engaged by tests that do not require recall (e.g., a recognition or some forms of multiple choice tests), these tests do not produce as much intra-item learning.

While PCR incorporates many ideas from existing models of memory, the assumption of different kinds of learning for the two stages that underlie recall is an important departure from prior models. These models concern the association strength between retrieval cues and memories, and then use these associations in different ways to explain the difference between familiarity responses (the first stage) and recall responses (the combination of two stages). By additionally considering intra-item learning, PCR goes beyond passive theories of memory formation to explain why the act of recalling something from memory results in a qualitatively different kind of learning; a kind of learning that is unique to the item, allowing better/faster recall of the item even when retrieval cues change. Nevertheless, in the PCR model, these two types of learning follow from the same learning mechanism that builds associations between features according to the temporal order in which features become active. Thus, the key distinction between passive study and active recall is that passive study is an all-at-once event in terms of items and their features, whereas recall is a gradual unveiling of an item's features.

References

Broek, G. S. E. van den, 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. http://doi.org/10.1080/09658211.2013.831455

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268 – 276.

Dan, Y., & Poo, M.-M. (2006). Spike Timing-Dependent Plasticity: From Synapse to Perception. *Physiological Reviews*, *86*(3), 1033–1048. http://doi.org/10.1152/physrev.00030.2005

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers*, *16*(2), 96–101. http://doi.org/10.3758/BF03202365

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. http://doi.org/10.1006/jmps.2001.1388

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. http://doi.org/10.1080/17470218.2011.638079

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. http://doi.org/10.1016/j.jml.2011.04.002

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.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. http://doi.org/10.1037/0033-295X.88.2.93

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. http://doi.org/10.1111/j.1745-6916.2006.00012.x

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252–257. http://doi.org/10.1027/1618-3169.56.4.252

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571 – 580.