Using Testing to Enhance Learning: A Comparison of Two Hypotheses

Michael C. Mozer
Department of Computer Science & Institute of Cognitive Science
University of Colorado
Boulder, CO 80309 USA

Michael Howe
Department of Computer Science & Institute of Cognitive Science
University of Colorado
Boulder, CO 80309 USA

Harold Pashler
Department of Psychology
University of California at San Diego
La Jolla, CA 92093 USA

Abstract

Students learning facts such as foreign language vocabulary often rely on a self-testing procedure in which they cue themselves with the English word and try to recall the foreign language target, instead of simply memorizing cue-target pairs. The value of this strategy has been empirically verified by a long history of research, yet existing computational models of human learning do not address the enhancing-learning-through-testing phenomenon. Using a simple, well studied model—a feedforward neural net with no hidden units—we propose two different hypotheses for characterizing the phenomenon. Hypothesis 1 is that self-testing generates a target which is used for additional training. Hypothesis 2 is that self-testing produces a more reliable error signal for training than rote memorization. Through simulation studies, we find that hypothesis 2 readily explains the phenomenon whereas hypothesis 1 does not. Further, hypothesis 2 makes predictions worthy of further empirical study, and can be viewed as a natural consequence of temporal difference learning.

When learning foreign language vocabulary and other facts, students often study using index cards that have an English vocabulary word (or *cue*) on one side and a foreign language vocabulary word (or *target*) on the other. The intuition is that by testing oneself, the associations are better learned and retained.

This intuition has been supported by a long history of empirical demonstrations (e.g., Izawa, 1966; Young, 1971). For example, Bartlett and Tulving (1974) asked participants to learn a list of paired associates (the *study phase*), and later tested retention of the pairs using free recall or recognition (the *final test*). Before the final test, subjects were given a cued-recall test (a *self test*) of some of the paired associates. Retention was better on the final test for those items that received the self test.

In this paradigm, it is unclear whether the benefit of the self test is attributable to attempting retrieval per se, or to the fact that successful retrieval of an associate also results in a re-presentation of the pair—an additional training trial.

An obvious strategy for examining the effect of retrieval is to conduct and experiment with, in addition to the initial study phase and the final test, an intervening phase in which participants are given either a self test or an experiment-provided re-presentation of the paired associate (which we'll refer to as *study only*). In this paradigm, the outcome is ambiguous (Carrier & Pashler, 1992): self testing outperforms study in some experiments (e.g. Hogan & Kintsch, 1971), but not others (e.g., McDaniel & Masson, 1985). One explanation for the inconsistency is that the rate of retrieval success on self test trials varies among experiments, and the mechanisms of learning are likely to be dependent on retrieval success. The experiments have other problems, including different amounts of time for study-only and self-test conditions, and failure to control the time spent on individual items (Carrier & Pashler, 1992).

To overcome these metholological difficulties, Carrier and Pashler (1992) compared a *study-only* or *SO* condition in which each cue-target pair was presented for ten seconds to a *test/study* or *TS* condition in which the cue was presented alone for five seconds and then the target appeared for the final five seconds. In TS trials, participants were supposed to use the cue to retrieve the target, but even if retrieval failed, the trial still had value due to the presentation of cue and target together for five seconds. Consequently, the dependence on retrieval success rate is minimized. Also, the paradigm matches the total time per item in SO and TS conditions. If anything, self testing is at a disadvantage because the total viewing time for cue plus target was lower.

In Experiments 1 and 2 of Carrier and Pashler, 40 cue-target pairs were used, half each assigned to the SO and TS conditions. The experiment began with a study only phase in which participants viewed each of the 40 pairs once for ten seconds. Then two more passes were made through the pairs, presented in the manner designated for that pair—SO or TS. For both conditions, participants were instructed to say aloud the target. In the TS condition, this instruction required that participants recall the target, or if they failed to recall, to wait until the target appeared. Following the three presentations of each pair, a final test phase evaluated performance in the two conditions via cued recall.

In Experiment 1, the cues were consonant-vowel-consonant trigrams and the targets were two digit numbers. For the sake of ecological validity, Experiment 2 used a language learning task with English language word cues and the corresponding Siberian Eskimo Yupik language translation targets. Table 1 shows the percentage of error responses. In both Experiments 1 and 2, performance was better in the TS condition than

TABLE 1. Performance in Enhancing-Learning-Through-Testing Experiments

	Human Data (% Error)		Simulation (Mean Squared Error)	
	Study Only	**Test Then Study**	**Study Only**	**Test Then Study**
Carrier & Pashler, Expt. 1	42.0%	36.0%	.389	.321
Carrier & Pashler, Expt. 2	43.0%	36.0%		
Carrier & Pashler, Expt. 3	40.0%	32.7%	.356	.258

the SO condition. These results indicate roughly 10% fewer errors with testing, therefore, having to retrieve the target is more effective than simply studying the target, when all else is controlled for.

Carrier and Pashler conducted a further experiment to rule out an alternative explanation of their results. In Experiments 1 and 2, participants may have used the first retrieval attempt in the TS condition to determine which items were difficult, and then increased their encoding effort for the difficult items on the second retrieval attempt, thereby learning the items better. Experiment 3 ruled out this explanation by giving participants only a single pass through the items in either the TS or SO conditions, following two passes through the items as study-only trials. The total number of items was reduced to 30. The results were similar to Experiments 1 and 2 (see Table 1), suggesting that the effect of attempting retrieval on later retention does not depend on strategic allocation of encoding effort.

Mechanism Underlying Enhanced Learning Through Testing

Why does testing oneself—i.e., attempting to retrieve a target from memory—have beneficial effects for later retention, above and beyond the effects due to mere study? A variety of explanations have been proposed for the self-testing benefit.

- Landauer and Bjork (1978) considered that retrieval attempts provide a general sort of practice or context that boosts performance at a future time. However, this account predicts that the benefits would not be item specific, i.e., SO and TS items would benefit equally in an experiment where the item types were mixed within subject.
- Mandler (1979) suggested that cued recall might strengthen the structural, integrative information about a cue-target pair. Cooper and Monk (1976) proposed that retrieval requires neural activity that consolidates the representation of the target in memory. However, both of these accounts do not provide a strong explanation for why TS should be better than SO, because both conditions involve simultaneous activation of cue and target.
- Bjork (1975) hypothesized that the act of retrieval may strengthen existing retrieval routes to the target representation, or may create new routes. Although interesting and consistent with the data, it is unclear what this hypothesis corresponds to in computational terms, and seems as if it might require novel, custom learning mechanisms.

This paper explores two alternative hypotheses concerning the enhancement of learning through testing, and we evaluate their plausibility via simulation studies. In proposing hypotheses, our aim was to determine whether an existing, well-accepted model could explain the basic phenomenon without requiring additional assumptions. A model is not convincing if two novel assumptions are needed to explain two data points. Further, an existing model is already constrained and therefore has the power to make strong predictions, which can guide the design of behavioral experiments.

Our hypotheses lie within the framework of neural network models. We explore the simplest architecture that might be capable of explaining the phenomenon: an associative network consisting of a pool of n_I input units fully connected to a pool of n_O output units. The activity of output unit j, y_j, is simply a weighted sum of the inputs, x_i, passed through a sigmoid squashing function that limits the output in the range [–1, +1]:

$$y_j = \tanh\left(\sum_{i=1}^{n_I} w_{ji} x_i\right).$$

A training set consists of n_L paired associates to be learned, $\{(\boldsymbol{x}^1, \boldsymbol{d}^1), \ldots, (\boldsymbol{x}^{n_L}, \boldsymbol{d}^{n_L})\}$, where the superscript is the index over pairs in the training set, and $\boldsymbol{x}$ and $\boldsymbol{d}$ are the activity vectors of the cue and target of the pair, respectively. To reflect the fact that items to be learned in the behavioral studies are arbitrary, make little contact with existing knowledge, and have no systematic similarity to one another, we assume that the cue and target activity vectors are random. (Further details in the methodology section that follows.)

In neural net models of cognition, the training of the model is often viewed as an abstract procedure for loading knowledge into a network, and as having no direct correspondence to the sequence of episodes a human learner experiences. In contrast, we commit to a one-to-one correspondence: An SO trial in a behavioral experiment is modeled as one weight update in the neural network. For many neural net architectures and learning procedures, this correspondence is implausible; training the network requires dozens if not hundreds of passes through the training examples, and training on one example can result in catastrophic interference with other examples. We avoid these problems in two ways. First, our architecture has direct connections from input

units to output units, in contrast to strictly layered architectures with hidden units. Second, we endow our architecture with as many inputs as training examples, i.e., $n_L = n_I$; consequently, cues are approximately orthogonal to one another, and interference among examples is minimal. Due to the architecture, the model can learn associations with roughly the same number of exposures as a human participant in a paired-associate experiment.

We use the standard supervised learning procedure for associative networks, a generalization of the Widrow-Hoff or LMS learning algorithm (Widrow & Hoff, 1960) to nonlinear outputs. Following presentation of a cue $\boldsymbol{x}$ to be paired with target $\boldsymbol{d}$, a weight update is performed:

$$\Delta w_{ji} = \varepsilon(d_j - y_j)x_i(1 + y_j)(1 - y_j)$$

where ε is a step size (learning rate).

Having described the general class of models we consider, we turn to two specific hypotheses concerning the nature of learning via self testing.

Hypothesis 1: Self-generated training

One hypothesis is based on the notion of Guthrie (1952) that one learns what one does. That is, when individuals test themselves, they generate a candidate response, and then learn the association between the cue and the candidate response, whether it is correct or incorrect. If the candidate is correct, existing connections are strengthened and are therefore more resilient to decay or interference; if the candidate is incorrect, the wrong association is reinforced, making it more difficult to unlearn.

This interpretation of self testing suggests that testing should benefit an individual only if the material is already somewhat familiar. Some evidence indeed suggests that testing on novel paired associates—when individuals cannot possibly know the correct response—is detrimental to learning (Cunningham & Anderson, 1965).

By this hypothesis, a TS trial involves the following steps: (1) The cue is presented and a candidate response is generated. (2) The LMS weight update is computed for the candidate response. (3) When the target is eventually presented, the LMS weight update is computed for the experiment-provided target. In contrast, an SO trial involves only the third step. In a TS trial, two weight updates are generated; the weight updates are added together and performed at the end of the trial.

How does the model generate a candidate response? It might produce an output and then deterministically select the nearest *well formed state*, defined as a state which has a meaning in the domain (e.g., the set of all targets used for training, plus some distractor alternatives, plus a null or "no response" state). However, individuals are essentially guessing at early stages of learning, and the deterministic rule implies an ability to find the best response among a set of barely-known alternatives. Instead, one might wish to introduce a stochastic selection rule. A standard stochastic procedure for reading out from a neural network is to use a Luce choice or Boltzmann rule (Luce, 1959). By this rule, the distance between each possible response, $\boldsymbol{r}^i$, and the network output, $\boldsymbol{y}$, is computed, $v_i = \|\boldsymbol{r}^i - \boldsymbol{y}\|^2$, and the probability of choosing response i is

$$p_i = \exp(-\beta v_i) / \sum_j \exp(-\beta v_j),$$

where large β achieves a more deterministic selection.

Rather than treating β as a free parameter, we chose β such that the mean correct-response probability is 0.95 if the network produces the correct response on each trial. The model has other free parameters, though, including learning rates for supervised and self-generated targets, the number of distractor vectors considered as candidates for response selection, and the possibility of memory (weight) decay that introduces forgetting.

Hypothesis 2: Complete processing of cue

Carrier and Pashler (1992) speculated on an intriguing basis for the self-testing benefit. They reasoned that in neural net models that learn by error correction, which includes the LMS algorithm, learning requires a comparison between the desired output and the *actual* output—the output that the network produces given its current state of knowledge. If presentation of the target simultaneously with the cue "contaminates" (to use their term) production of the actual output, learning would be less efficient. Essentially, presentation of the target terminates ongoing processing and interferes with the estimation of error needed for learning.

An elegant instantiation of this hypothesis in the context of neural net models is via the incorporation of time into the neural net, specifically, the notion that units in a neural net are slow integrators of information and therefore require many time steps for information to propagate from the input layer to the output layer (McClelland, 1979). We can do this in the network by indexing its output by the (discrete) time step t, i.e., $y_i(t)$, and adding a time constant to the activation dynamics:

$$y_j(t) = (1 - \tau)y_j(t - 1) + \tau\tanh(\sum w_{ji}x_i),$$

where $y_i(0) = 0$ Asymptotically, the output is independent of the time constant τ, for $0 < \tau \le 1$, but τ determines the rate at which convergence is achieved, i.e., how rapidly information is transmitted from the input to the output.

If we assume that activation dynamics freeze—equivalent to setting $\tau = 0$—when the target is presented, the model will produce a more accurate estimate of its output in the TS condition than in the SO condition, and the learning procedure will have a better estimate of the error. From another perspective, note that if activation dynamics freeze at $t = 1$, the actual output y will be zero, and the training procedure reduces to a form of Hebbian learning; at the other extreme, if the activation dynamics do not freeze and the asymptotic

value $\boldsymbol{y}(\infty)$ is used for training, the learning procedure is exactly the LMS gradient descent step. Because LMS is a more powerful procedure than Hebb, it should yield better performance.

For our simulations, we established a relatively coarse-grained correspondence between time steps in the neural net and real-world time by designating the duration of each time step to be 250 msec. Rather than leaving the time constant τ as a free parameter, we chose τ in advance such activation would reach halfway to asymptote by 2000 msec. To match the TS condition in the behavioral studies, 20 time steps (= 5 seconds) of processing was allowed before the onset of the target. For the SO condition, we had some freedom to determine the time step at which activation dynamics freeze. Although the cue and target appeared simultaneously in the behavioral experiments, participants may nonetheless have done some amount of processing of the cue before the target is processed. We experimented with 0, 1, and 2 time steps of processing in the SO condition, and all yielded similar results; we chose 1 time step in modeling Carrier and Pashler, because that was a sufficient amount of processing to ensure that with enough practice, the model could learn the items in the SO condition. For evaluation of the model during the final test, the activity at time step 80 was used.

Simulation Studies

General Methodology

In all simulations, we used networks with $n_L = n_I = n_O$. Cue and targets were random binary vectors in $\{-1, 1\}^{n_I}$. To model Experiments 1-3 of Carrier and Pashler, we used the same number of items as in their experiment, $n_L = 40$ for Experiments 1 and 2, and $n_L = 30$ for Experiment 3. Because Experiment 2 is essentially a replication of Experiment 1 with different stimulus materials, and the materials in both experiments were intended to be unfamiliar to participants, we capture both experiments with one simulation. Half of the items were assigned to the SO condition and half to the TS condition.

Weights in the neural network were initialized to zero. We also conducted stimulations in which the initial weights were chosen from a normal distribution with mean zero and standard deviation 0.001. However, because the variability of the weights had no systematic effect on the results, we simplified by eliminating this source of noise from the simulation.

The experiments of Carrier and Pashler each involved three *epochs* of training, followed by a final test. An epoch is a presentation of all items in the training set. Within an epoch, order of presentation was randomized, with the constraints that Carrier and Pashler imposed to ensure an intermixing of SO and TS items.

Epochs were of two sorts: in a *pure study* epoch, all items were studied without testing, regardless of whether they were SO or TS items; in an *experimental* epoch, presentation of an item depended on whether it was assigned to the SO or TS condition. In Experiments 1 and 2, the first epoch was pure study (denoted *S*), and epochs 2 and 3 were experimental (denoted *E*); we use the shorthand notation *SEE* for this design. In Experiment 3, the first two epochs were pure study and the third was experimental, i.e., an *SSE* design.

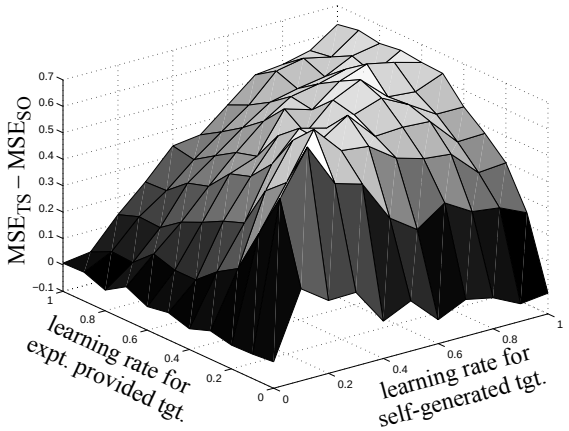

FIGURE 1. Testing error for TS minus SO as a function of the learning rates for self-generated and experiment-provided targets. The difference is nonnegative everywhere, indicating no enhancement through testing. For this simulation, no additional distractor states or weight decay are included.

All results reported are a mean computed from 1,000 independent simulations, where the simulations differ from one another in the choice of random training vectors and the randomization of items within an epoch.

We use mean squared-error (MSE) as a measure of performance of the model. With additional assumptions, we could classify a response as correct or incorrect (e.g., using the stochastic read out procedure that is built into Hypothesis 1), but there is little value in transforming a qualitative fit to a quantitative fit if several new assumptions are required. Consequently, we focus on obtaining qualitative measures of recall, and determining how manipulations of the model affect relative recall.

Hypothesis 1: Self-generated training

After a systematic exploration of the model parameter space, we failed to find any parameter settings that yielded an enhancement of learning by testing. Error was consistently higher in the TS condition than in the SO condition. The two conditions converge as the learning rate for the self-generated target approaches zero, where at the limit SO and TS become identical. Figure 1 illustrates one exploration of the parameter space.

In retrospect, the negative result should not have been surprising. After one or two epochs, the model—like people—is about as likely to make an error as to guess correctly. Consequently, the model will receive as much training from self-generated targets that steers it away from veridical recall as training that steers it toward veridical recall.

Hypothesis 2: Complete processing of cue

Fortunately, our second hypothesis yields more encouraging results. Consistent with Carrier and Pashler, the model produces an enhancement of learning by testing—a lower error for TS than SO—in simulations of Experiment 1/2 (one simulation for both experiments, since they are identical except for the stimulus materials) and Experiment 3 (right side of Table 1). In these simulations, we chose a learning rate that yielded the best possible performance, averaged over TS and SO items. However, the testing benefit was robust over the choice of learning rate.

Figure 2 facilitates a better understanding of the phenomenon in terms of the model. The Figure shows mean-squared error for TS and SO items for four different experimental designs. All designs involve three epochs of training, but they differ in how many epochs of pure study (*S*) precede the experimental (*E*) epochs. The designs range from all study (*SSS*) to all experimental (*EEE*). *SEE* and *SSE* correspond to Experiment 1/2 and Experiment 3, respectively.

The Figure shows that two testing trials helps more than one (*SEE* versus *SSE*). Interestingly, three testing trials shows little benefit over two (*EEE* versus *SEE*). This latter result was at first surprising to us, because it would seem that the more accurate error estimate that is obtained via testing would benefit epoch 1 as well as epochs 2 and 3. However, with an untrained net whose weights are all zero or close to zero, the initial output of the net is close to zero regardless of the number of time steps of activation dynamics.

Comparing SO items in *SSS* versus *EEE* designs, it appears that the SO items benefit from being in a context where testing is occurring; this is a bit surprising considering that the training of these items is identical and learning rates are identical across designs.The result also cannot be explained by virtue of generalization from the better-learned TS items to the SO items, because the items were generated with no systematic similarity structure. Instead, we suggest that the transfer from TS to SO is due to the TS items generating a more meaningful error signal—an error signal that reflects the sort of outputs the network is likely to produce if it is allowed to run to asymptote. Although the precise outputs will differ from one cue to another, the TS items provide information about the distribution of activity values for each output unit across items. This information can certainly be used to determine characteristics of the weight vector (e.g., its overall magnitude, and the sign of biases).

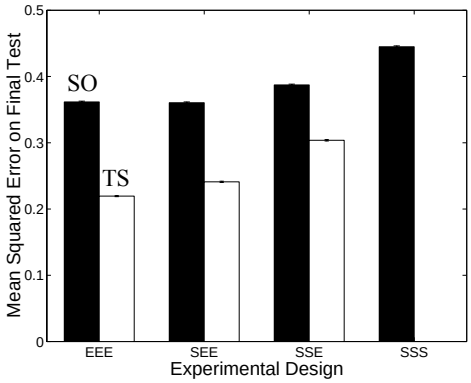

FIGURE 2. Mean-squared error in TS (white bar) and SO (black bar) conditions for experimental designs with three training epochs. +/–1 standard error of the mean is indicated.

We discuss the implication of these results next.

Discussion

In simulations of two models, we found that one hypothesis for the enhancement-of-learning-through testing effect—the hypothesis that self-generated responses are used as targets for further training—is not supported. Another hypothesis—that presentation of the target terminates processing of the cue—is consistent with the experimental data. In the remainder of the paper, we discuss predictions, extensions, and implications of the second hypothesis.

Predictions

- Our model predicts little difference between an *EEE* design and the *SEE* design used in Experiment 1/2. That there is no cost to testing on the first epoch runs against at least one experimental study (Cunningham & Anderson, 1965), but that study used a quite different methodology, and the finding of an initial-epoch-testing cost has not been widely reported in the literature.
- Our model predicts that an SO item should benefit from being embedded among TS items. If it is observed experimentally, a natural interpretation of this effect is that the greater effort on TS items spills over to the SO items. However, the model achieves this spillover without any notion of generalized "effort."
- Our model predicts the relative magnitude of the testing enhancement as a function of the cue-target asymmetry (CTA), i.e., the difference in time between the onset of the cue and the onset of the target. The Carrier and Pashler experiment used a CTA of 5 seconds (20 time steps in the model). One could conduct an experiment in which the CTA was longer or shorter. (Because the time scale of retrieval in the model was set arbitrarily, there is a degree of indeterminacy in the model's predictions. Nonetheless, with one free parameter tied down, the model can characterize the effect of shorter or longer CTAs) Figure 3 shows the model's performance as the CTA is varied. For small CTAs, there is little difference between SO and TS conditions; for large CTAs, the conditions are similar to those studied in the present simulations. Clearly, increasing the CTA has diminishing returns.

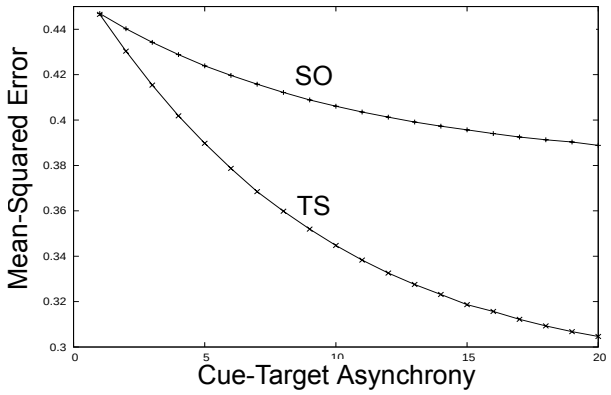

FIGURE 3. Test performance as the cue-target asynchrony in training is varied in the TS condition.

Extensions to the Model

In both SO and TS conditions, each simulation trial began by presenting the cue for T time steps—T being different for SO and TS conditions—at which point a weight update was performed to reduce the difference between the actual response, $\boldsymbol{y}(T)$, and the target. An alternative procedure involves updating the weights to reduce the difference between *each* of $\boldsymbol{y}(1)$, $\boldsymbol{y}(2)$, ..., $\boldsymbol{y}(T)$ and the target. This alternative procedure encourages the net to produce the target as rapidly as possible, and is equivalent to a form of *temporal difference* (TD) learning known as TD(1) (Sutton, 1988). Temporal difference learning is concerned with learning to predict the future given successively better information over time—exactly the situation experienced by the network with time constants, because the propagation of information occurs gradually. However, TD(1) is often not useful in practice because the earliest predictions, e.g., $\boldsymbol{y}(1)$, are treated as important as later, better predictions, e.g., $\boldsymbol{y}(T)$. To remedy this problem, Sutton proposed a family of algorithms, denoted TD(λ), for $0 \leq \lambda \leq 1$, where λ is roughly the emphasis on achieving correct early predictions. The λ that yields optimal performance depends on the domain. Although it would be interesting to discover how the self-testing benefit depends on λ, the deeper contribution of casting the learning procedure in the TD framework is that it offers a rationale for the termination of processing when the target is presented.

The TD framework is based on the notion that learning mechanisms are fundamentally concerned with predicting eventual outcomes at the earliest possible moment. The adaptive value of prediction is clear; accurate prediction can avoid danger and missteps. Considering the associative learning task in this manner, the cue is a predictor of the target, and TD learning aims to get from the cue to the target as rapidly as possible. However, once a target has been presented, nothing remains to be predicted. The TD framework has been valuable for explaining a broad range of data, from the animal conditioning literature (Sutton & Barto, 1981) to the neural basis of reward (Schultz, Dayan, & Montague, 1997). It seems a natural extension to the mechanisms of associative learning, although one must confront the finding that associative learning appears symmetric (Kahana, 2002).

REFERENCES

Bartlett, JC, & Tulving, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *JVLVB*, *13*, 297-309.

Bjork, R.A. (1975). Retrieval as a memory modified: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition,* 20*,* 632-642.

Cunningham, DJ, & Anderson, RC (1968). Effect of practice time within prompting and confirmation presentation procedures on paired associate learning. *JVLVB,* 7, 613-616.

Guthrie, E. (1952). *The Psychology of Learning (Rev. Edition)*. New York: Harper.

Hogan, RM, & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *JVLVB*, *10*, 562-567.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, *18*, 879-919.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*, 823–840.

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

Luce, R.D. (1959) *Individual choice behavior: A theoretical analysis*. New York: John Wiley & Sons.

Mandler, G. (1979). Organization and repetition: Organizational principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th Anniversary* (pp. 293-327). Hillsdale, NJ: Erlbaum.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287-330.

McDaniel, MA, & Masson, MEJ. (1985). Altering memory representations through retrieval. *JEP:LMC*, *11*, 371-385.

Schultz, W., Dayan, P. & Montague, R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Sutton, R. (1988). Learning by the method of temporal differences. *Machine Learning, 3,* 9–44.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*, 135–170.

Widrow, B., & Hoff, M.E. (1960). Adaptive switching circuits. In *IRE WESCON Convention Record*, *pt. 4*, 96-104.

Young, JL (1971). Reinforcement-test intervals in paired associate learning. *J ournal. of Math. Psych.*, *8*, 58-81.