Adaptation of Cue-Specific Learning Rates in Network Models of Human Category Learning

Mark A. Gluck **Paul T. Glauthier**
Center for Molecular and Behavioral Neuroscience
Rutgers University

Richard S. Sutton
GTE Laboratories Incorporated

Abstract

Recent engineering considerations have prompted an improvement to the least mean squares (LMS) learning rule for training one-layer adaptive networks; incorporating a dynamically modifiable learning rate for each associative weight accellerates overall learning and provides a mechanism for adjusting the salience of individual cues (Sutton, 1992a,b). Prior research has established that the standard LMS rule can characterize aspects of animal learning (Rescorla & Wagner, 1972) and human category learning (Gluck & Bower, 1988a,b). We illustrate here how this enhanced LMS rule is analogous to adding a cue-salience or attentional component to the psychological model, giving the network model a means for discriminating between relevant and irrelevant cues. We then demonstrate the effectiveness of this enhanced LMS rule for modeling human performance in two non-stationary learning tasks for which the standard LMS network model fails to adequately account for the data (Hurwitz, 1990; Gluck, Glauthier, & Sutton, in preparation).

Introduction

In earlier papers, we have explored a simple adaptive network as a model of human learning (Gluck & Bower, 1988a,b; Gluck, Bower, & Hee, 1989; Gluck, 1991). This network model is based on Rescorla & Wagner's (1972) description of classical conditioning; the learning rule is the same as the least mean squares (LMS) learning rule for training one-layer networks (proposed by Widrow & Hoff, 1960), where the goal of learning is to minimize the discrepancy between the expected and the actual outcome.

Correspondence should be addressed to Mark A. Gluck, Center for Molecular and Behavioral Neuroscience, Rutgers University, 197 University Avenue, Newark, NJ 07102. E-Mail to gluck@pavlov.rutgers.edu

There are, however, many different engineering algorithms for minimizing classification error; the LMS rule is only the simplest of them. Recent engineering considerations have prompted an improvement to the least mean squares (LMS) learning rule for training one-layer adaptive networks; incorporating a dynamically modifiable learning rate for each associative weight accellerates overall learning and provides a mechanism for adjusting the salience of individual cues (Sutton, 1992a,b). From a psychological perspective, this enhanced LMS rule is analogous to adding a cue-salience or attentional component to the psychological model, giving the network model a means for discriminating between relevant and irrelevant cues. Thus, it is similar to psychological ideas of learning cue-specific saliences, associabilities, and attentional parameters (Pearce & Hall, 1980; Mackintosh, 1975; Frey & Sears, 1978). We call the class of learning methods that dynamically adjust cue-specific learning rates *dynamic-learning-rate* (DLR) methods.

Dynamic-Learning-Rate Methods

Dynamic-learning-rate (DLR) methods are meta-learning algorithms for adapting step-size parameters (i.e. learning rates) during a base-level learning process, which in this paper is the Rescorla-Wagner (1972) or LMS rule (Widrow & Hoff, 1960). The step-size parameters are incrementally adjusted by a gradient descent process to optimize convergence and tracking performance. Such methods have been of interest within the neural network community as a way of speeding the relatively slow convergence of learning methods such as back-propagation (e.g., Jacobs, 1988; Silva & Almeida, 1990; Lee & Lippman, 1990; Sutton, 1986; Barto & Sutton, 1981; Tollenaere, 1990) and have also been proposed as relevant to a key problem in machine learning: finding good individualized learning rates to speed and direct learning (Sutton, 1992a). Recently, Sutton (1992b) has argued that these dynamic-learning-rate

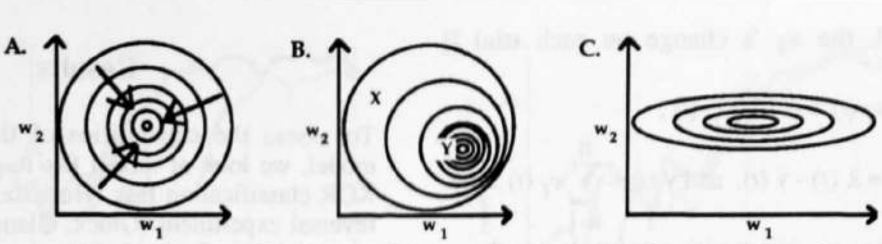

Figure 1. **A.** Expected squared error surface in the weight space of a network with two weights of equal diagnosticity (relevance). The two axes shown represent the possible values of weights 1 and 2, respectively. The third axis (shown as a contour plot) represents the expected squared error, and the global minimum of the surface specifies the ideal weight solution. The LMS network learns by changing the weights in the direction of steepest descent. **B.** A surface with different slopes in different places. In the standard model, the largest changes are made to those weights where the surface drops most sharply, as in part Y. A better strategy would be to take larger steps in the shallow, gently curving part X to traverse it efficiently, and smaller steps in part Y to prevent instability and oscillation in the network. **C.** A ravine: a surface with different slopes in different directions. Here, the network should take larger steps along the horizontal axis where the slope is gentle (along the ravine) and smaller steps along the vertical axis where it is steep (across the ravine) (After Sutton, 1986).

methods may improve classical engineering methods for estimation such as least-squares methods and the Kalman filter. The idea behind these DLR methods is a generalization of Kesten's (1958) method for accellerating stochastic approximation. Consider one of the base-level modifiable parameters--one of the weights in a connectionist network, for example--and how it changes over time. If the weight changes are all in the same direction--e.g., all increases--this signifies that the step-size parameter is too small. The weight could reach its asymptotic value faster if it took larger steps. On the other hand, if the weight changes are in opposite directions--e.g., first up and then down--this signifies that the step-size parameter is too large. For example, opposite-signed weight changes will occur when the weight is overshooting its optimal value. The basic idea behind current DLR methods is to adjust the step size according to the correlation between successive weight changes, with the goal of obtaining zero correlation. Jacobs (1988) proposed correlating the current weight change with a recency-weighted average of previous weight changes. This update rule was written $\overline{\Delta}(t-1)\Delta(t)$ and was called the Delta-Bar-Delta algorithm. The extension of this method to the incremental case is called the Incremental Delta-Bar-Delta (IDBD) method (Sutton, 1992a). This is the method we use to form the extended psychological model explored in this paper.

DLR methods have advantages for both static problems, in which the correct solution does not change, and for non-static problems, in which the correct solution does change over time and must continually be tracked.

In static problems, DLR methods help overcome well-known limitations of *steepest descent* methods such as LMS and backpropagation. In the weight space of a network (i.e. the space formed by assigning each connection weight its own dimension), the expected squared error forms a surface. The minimum of this surface is the point at which the error is smallest; this identifies the ideal asymptotic weights for a particular learning task. In the standard model, a step is taken in weight space at each trial in the direction in which performance is expected to improve most rapidly (Figure 1A). Steepest descent methods are well known to perform poorly for surfaces with different slopes in different places (Figure 1B) and for those containing *ravines*--places which curve more sharply in some directions than others (Figure 1C). In both of these cases, following the direction of steepest descent does *not* take you directly to the minimum. Jacobs (1988) and others have shown that DLR methods can significantly increase the speed of convergence on static problems.

Another advantage of DLR methods is on non-static "tracking tasks", in which the correct solution is not fixed, but continues to change. For example, suppose a subject is faced with a sequence of categorization tasks. Even if the correct solution differs from task to task, the same subset of cues may always be relevant. If cue-relevance can be learned on the early tasks, learning performance on later tasks can be greatly improved. Advantages of this sort have been shown in an engineering context for DLR methods (Sutton, 1992a,b).

The DLR Model

In this section we present the specifics of the standard LMS model and of its extension with a DLR method. We will refer to the extended model as *the DLR network model*. In the standard LMS model, the network operates in a training environment in which feedback (the US or the correct classification) is given after each stimulus pattern. At each time step, or trial, t, the learner receives a set of inputs, $x_1(t)$, $x_2(t)$, ... , $x_n(t)$, computes its output, $y(t)$, and compares this to the desired output, $\lambda(t)$. In the standard model, the w_i 's change on each trial according to:

$$w_i(t+1) = w_i(t) + \alpha\,\delta(t)\,x_i(t),$$

$$\text{where } \delta(t) = \lambda(t) - y(t), \text{ and } y(t) = \sum_{i=1}^{n} w_i(t)\,x_i(t)$$

The learning rate, α, is a positive constant (on the order of .01 in most simulations) that determines how much all the weights change when the output differs from the training signal.

With the DLR algorithm, however, there is a different learning rate, α_i, for each input, x_i, and these change according to a meta-learning process. The base-level learning rule is:

$$w_i(t+1) = w_i(t) + \alpha_i(t+1)\,\delta(t)\,x_i(t)$$

(The α_i are indexed by $t+1$ rather than t to indicate that their update, by a process descibed below, occurs *before* the w_i update.) To insure that the learning rates remain positive, they are expressed and stored in the form: $\alpha_i(t) = e^{\beta_i(t)}$. The IDBD algorithm we used updates the β_i by:

$$\beta_i(t+1) = \beta_i(t) + \frac{\theta\,\delta(t)\,x_i(t)\,h_i(t)}{\sqrt{\alpha_i(t)}}$$

where θ is a positive constant, the *meta-learning rate*, and h is an additional per-input memory variable initialized at zero and updated by:

$$h_i(t+1) = h_i(t)\left[1 - \alpha_i(t+1)\,x_i^2(t)\right]^+ + \alpha_i(t+1)\,\delta(t)\,x_i(t)$$

where $[x]^+$ is x, if $x > 0$, else 0. The first term in the above equation is a decay term; the product $\alpha_i(t+1)\,x_i^2(t)$ is normally zero or a positive fraction and this causes a decay of h_i towards zero. The second term increments h_i by the previous error. The memory, h_i, is thus a decaying trace of the cumulative sum of recent errors (Sutton, 1992).

Results

To assess the capabilities of the extended DLR model, we look at model fits to data from both an XOR classification task (Hurwitz, 1990) and a new reversal experiment (Gluck, Glauthier, & Sutton, in preparation). Both of these experiments involved relevant and irrelevant dimensions. The Gluck et al study is a non-stationary task in the sense that the correct response to the stimuli changes over time. Non-stationary tasks are especially appropriate here because they test the DLR model's ability to learn biases during early learning and then use these biases to improve later learning.

XOR Experiment (Hurwitz, 1990)

Hurwitz (1990) describes an experiment in which subjects learned to classify words from a new language into one of two categories. The design involved 16 patterns, each defined on 4 binary dimensions. The assignment of patterns to categories A and B was determined by two relevant dimensions, related to the categories by the XOR rule: the patterns 11•• and 22•• were assigned to category A, and 12•• and 21•• to category B; the irrelevant dimensions are indicated by bullets (•). The trials were broken into four divisions of 80 trials each with only a subset of the patterns presented during each division (Figure 2). Each subset was designed so that both the first and second pairs of dimensions could produce XOR relationships to the categories. However, combinations of the subsets preserved this relationship only for the relevant pair of dimensions. Subjects who used the XOR relationship defined on the irrelevant pair of dimensions in making their categorization responses would therefore show reduced performance when a new subset was introduced. Generalization to new patterns in a new subset could only occur if the relevant features had been discovered during training on previous subsets.

A.

Set	Patterns A	Patterns B	XOR Classification ●●--	XOR Classification --●●
1	●●○○ ○○●●	●○●○ ○●○●	A	A
2	●●●● ○○○○	●○○● ○●●○	A	A
3	●●●○ ○○○●	●○○○ ○●●●	A	B
4	●●○● ○○●○	●○●● ○●○○	A	B

B.

Block	Pattern Set 1	2	3	4
1	10	10	0	0
2	5	5	10	0
3	0	0	5	15
4	5	5	5	5

Figure 2. Hurwitz' XOR experimental design. **A.** Patterns were divided into 4 subsets of 4 patterns each so that within each subset both the first and second pairs of dimensions could produce XOR relationships to the categories. Combinations of subsets preserved this relationship only for the first pair of dimensions. **B.** Subsets were presented with varying frequencies in each block of 80 trials.

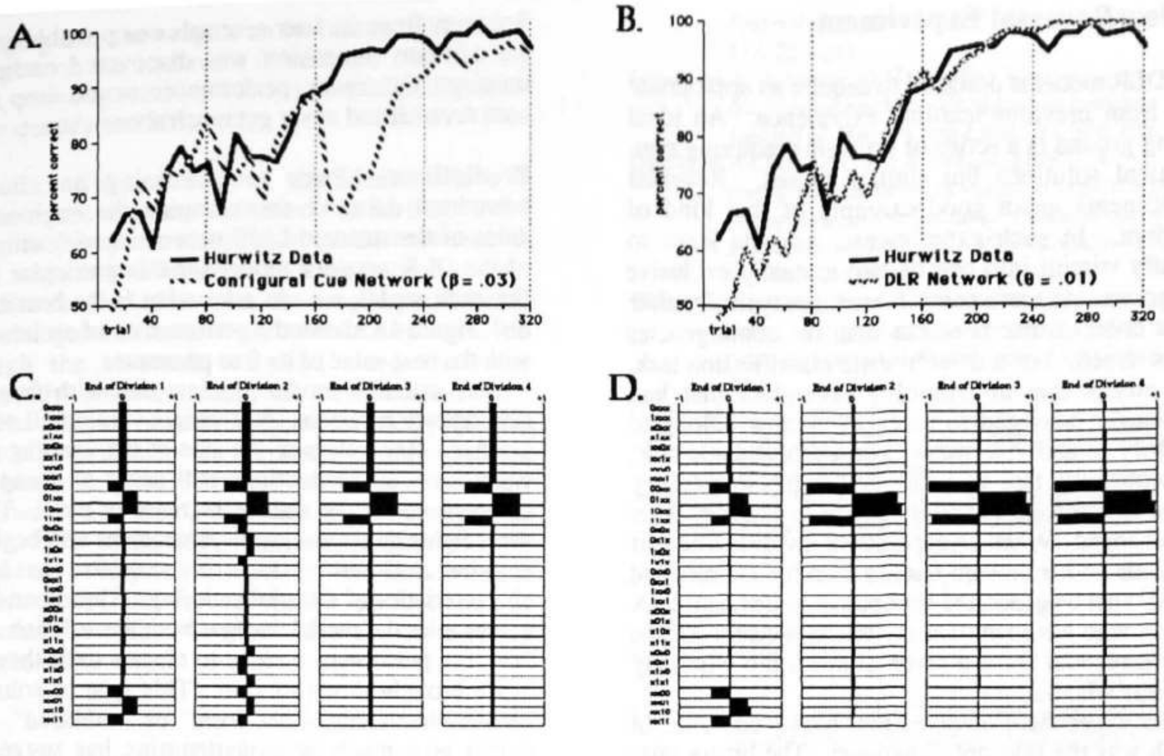

Figure 3. Fits to performance on Hurwitz' XOR experiment. A and B show average percent correct (y axis) throughout training in blocks of 10 trials each (x axis). The vertical lines mark the divisions where a new subset of patterns was introduced. A Standard Configural-Cue LMS Network (β=.03). B. DLR Network (θ=.01). C and D show network weights at the end of each division. Each of the 32 rows represents the connection weight for a feature node (both component and configural) indicated at the left. The middle of a row represents a weight of 0, the left represents -1, and the right represents +1. C. Standard Configural-Cue LMS Network Model which reaches a stable solution by the end of Division 3. D. DLR Network Model which stabilizes by the end of Division 2.

The results of Hurwitz's experiment suggest that many subjects were able to distinguish the relevant dimensions from the irrelevant ones. Subjects' acquisition curve was relatively smooth, showing a downward trend only at the beginning of the second trial division. In the third division, their curve remained smooth, despite the fact that the exemplar distribution was changing again. By the fourth division, they were at virtually perfect performance.

The model fits from Hurwitz's simulations show that the LMS configural-cue network (Gluck & Bower, 1988b; Gluck, Bower, & Hee, 1989) and exemplar models are unable to account for the subjects' ability to generalize to the new patterns introduced at the beginning of the third trial division. As shown in Figure 3A, performance of the standard configural-cue LMS model drops significantly at this point, whereas the subjects' performance continued to improve.

Hurwitz argued that the configural-cue LMS model has no mechanism to allow it to differentiate relevant from irrelevant cues. Thus, it does not predict the solvers' generalization to new patterns in the third division of trials.

The extended model presented here, however, does have the ability to differentiate between relevant and irrelevant cues. As with the standard network model, the DLR network does not use hidden layers or backpropagation. The individual, dynamic learning rates on each connection allow the network to differentiate relevant from irrelevant features. As shown in Figure 3B, performance of the configural-cue DLR model continues to improve at the beginning of the third trial division, despite the fact that the exemplar distribution changes again. This is consistent with the empirical data and suggests how the model successfully distinguishes relevant from irrelevant dimensions.

To illuminate the difference between the standard configural-cue model and the DLR model, we compared their solutions at the end of each trial division. The major difference between the models' solutions was apparent at the end of Division 2. Whereas the standard model still attributed significant weight to the local XOR relationships defined in pattern subsets 2 and 3 (Figure 3C), the DLR model reduced weights on all irrelevant dimensions to virtually zero (Figure 3D). Only the XOR solution that remained consistent across subsets 1 and 2 was reflected in the weights. For the remainder of the trials, this solution remained essentially unchanged.

A New Reversal Experiment

The DLR model is designed to acquire an appropriate bias from previous learning experience. An ideal testing ground is a series of problems requiring non-identical solutions but similar biases. Reversal experiments are a good example of this kind of problem. In such experiments, subjects learn to classify stimuli into one of two mutually exclusive categories. At some point in learning, typically after some criterion has been reached, the contingencies are reversed. For a deterministic classification task, this means that all stimulus exemplars that had previously belonged to category A now belong to category B, and vice versa. For a probabilistic task, the probability that an exemplar belongs to category A and the probability that it belongs to category B are interchanged. When an experiment involves frequent reversals and irrelevant cues, a bias for the relevant cue(s) must be generated for optimal performance. A subject who has identified the relevant cue(s) will be able to recover from a reversal much more quickly than one who has not.

Our experiment involved four binary cues, one of which was the relevant dimension. The binary cues for the relevant dimension determined the category assignment of exemplars with a probability of 0.9. The cues for each of the three irrelevant dimensions were assigned to each category equally often. The contingencies were reversed 11 times over the course of the experiment; the cue that had been diagnostic of category A became diagnostic of category B, and vice versa. The reversals occurred more frequently as the experiment progressed. Of the 420 trials presented, reversals occurred after trials 80, 140, 200, 240, 280, 300, 320, 340, 360, 380, and 400. Recovery from the later reversals was possible only if the relevant dimension was discovered earlier in training. Otherwise, performance would drop after each reversal and never get much above chance.

Predictions. Prior to examining any human behavioral data, we can compare the engineering value of the standard LMS network model with that of the DLR network model for this particular task. For each model, we can ask: what is the best it can do? Figure 4A shows the performance of each model with the best value of its free parameter.

The standard model predicts that with frequent contingency reversals, ideal weight values will not be reached. Regardless of the size of the learning rate, weights on all connections will begin to change to compensate for the sudden increase in error. Thus, the weights on the irrelevant dimensions will begin to acquire non-zero values in response to local characteristics of the trial ordering. These non-zero values take the model farther from the solution, and keep the performance close to chance until they are again brought down to zero. This type of solution can be characterized as "local" or "unbiased". No matter how much previous learning has suggested that the weights on the irrelevant dimensions be kept at zero, weights for all present features still change.

The DLR model predicts that the ideal weights for a block of trials between reversals will be approached more and more rapidly as the experiment progresses. With each reversal, the bias toward the relevant dimension will become stronger. In other words, the learning rates on the connections from the node(s) for the relevant dimension will become larger, while all other learning rates will drop toward zero. When a reversal occurs, the model is "biased" toward

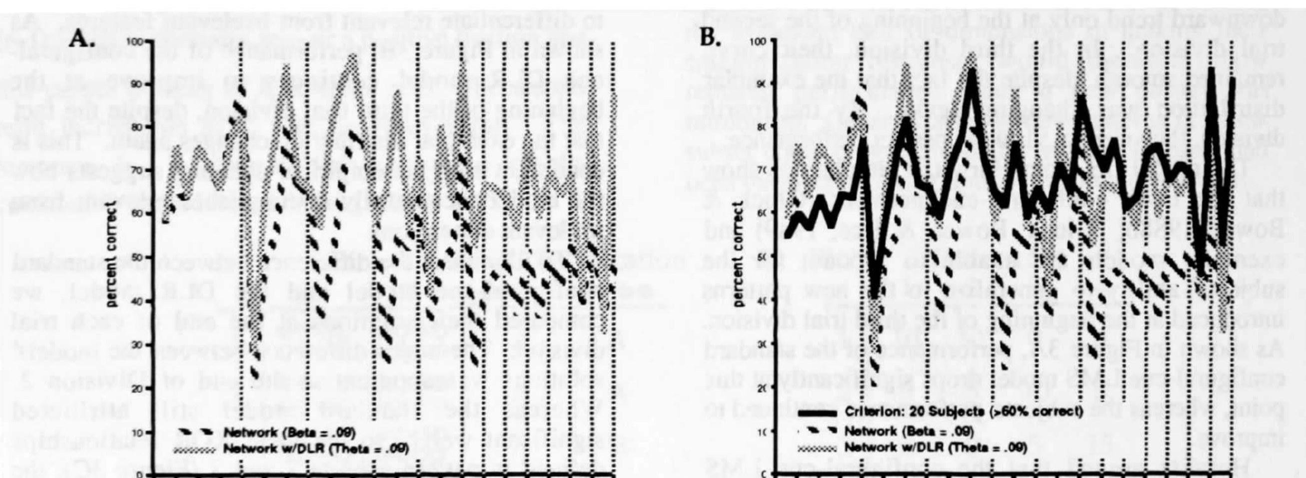

Figure 4. **A.** Optimal parameters for the standard and DLR network models. The graph shows average percent correct (y axis) throughout training in blocks of 10 trials each (x axis). The vertical lines mark the divisions where a reversal occurred. **B.** Fits to performance on Reversal Experiment. Subjects who met criteria were able to recover from reversals late in training. Whereas the standard model exhibited only chance performance, the DLR network was able to form a selective bias to the relevant dimension and recover from the reversals.

changing only the weights on the relevant dimension. Weights on the irrelevant dimensions stay close to zero, allowing the model to improve its performance more quickly and efficiently.

Results. Subjects showed a gradual increase in performance over the first block of 80 trials and a sharp drop in performance after the first reversal. It took subjects an average of 35 trials to recover from this first reversal. However, the drop in performance on subsequent reversals was considerably less, even though the reversals occurred more and more frequently. Recovery rates were likewise improved.

The data we obtained suggest that many subjects were able to identify the relevant dimension and use this bias to help them recover performance after later reversals. The last 7 reversals, occuring every 20 trials, were of particular interest because of their frequency. Whereas the standard LMS network is unable to account for the solvers' performance on these later reversals, the DLR model fits the data well (Figure 4B).

Conclusion

The new psychological model presented here emerges from engineering considerations; it comes from a search for a better way to minimize the expected error (Sutton, 1992a,b). Giving each input its own dynamic learning rate is analogous to adding a salience or attentional component to the learning mechanism. Learning is significantly accellerated on static learning tasks. On series of related tasks with common relevant cues and on selected non-stationary tasks, the extended model shows superior learning capacity and provides a better account of human learning behavior. We have shown that for two such tasks, in which the standard LMS configural-cue model (Gluck & Bower, 1988b) fails to account for the behavioral data, the extended DLR model succeeds far better.

References

Barto, A.G. & Sutton, R.S. (1981) Adaptation of learning rate parameters. Appendix C of *Goal Seeking Components for Adaptive Intelligence: An initial assessment*. Air Force Wright Aeronautical Laboratories/Avionics Tech. Report.

Frey, P.W. & Sears, R.J. (1978) Model of conditioning incorporating the Rescorla-Wagner associative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review, 85*, 321-340.

Gluck, M.A. & Bower, G.H. (1988a) From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General, 117*, 227-247.

Gluck, M.A. & Bower, G.H. (1988b) Evaluating an adaptive network model of human learning. *Journal of Memory and Lang., 27*, 166-195.

Gluck, M.A., Bower, G.H., & Hee, M. (1989) A configural-cue network model of animal and human associative learning. *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, Ann Arbor, MI.

Gluck, M.A. (1991) Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science, 2(1)*, 50-55.

Gluck, M.A., Glauthier, P.T., & Sutton, R.S. (in preparation) A Theory of dynamic variations in stimulus associability.

Jacobs, R.A. (1988) Increased rates of convergence through learning rate adaptation. *Neural Networks 1*, 295-307.

Hurwitz, J.B. (1990) A hidden-pattern unit network model of category learning. Ph.D. thesis, Harvard Psychology Dept.

Kesten, H. (1958) Accelerated stochastic approximation. *Annals of Mathematical Statistics 29*, 41-59.

Lee, Y. & Lippmann, R.P. (1990) Practical characteristics of neural network and conventional pattern classifiers on artificial and speech problems. In *Advances in Neural Information Processing Systems 2*, D.S. Touretzky, Ed., 168-177.

Mackintosh, N.J. (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Pearce, J.M. & Hall, A.G. (1980) A model for Pavlovian conditioning: Variation in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*, 532-552.

Rescorla, R. & Wagner, A. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. Black & W. Prokasy (Eds.) *Classical Conditioning II: Current Research and Theory*. New York, Appleton-Century-Crofts.

Sutton, R.S. (1982) A theory of salience change dependent on the relationship between discrepancies on successive trials on which the stimulus is present. Unpublished working paper.

Sutton, R.S. (1986) Two problems with backpropagation and other steepest-descent learning procedures for networks. *Proceedingd of the Eighth Annual Conference of the Cognitive Science Society*, 823-831.

Sutton, R.S. (1992a) Adapting bias by gradient descent: An incremental version of Delta-Bar-Delta. AAAI-92.

Sutton, R.S. (1992b) Gain adaptation beats least squares?. *Proceedings of the Seventh Yale Workshop on Adaptive and Learning Systems.*

Tollenaere, T. (1990) Super SAB: Fast adaptive back propagation with good scaling properties. *Neural Networks 3*, 561-574.

Widrow, B. & Hoff,, M. (1960) Adaptive switching circuits. *Inst. of Radio Engineers, WESCON Convention Record, 4: 96-194.*