Some biological implications of a differential-Hebbian learning rule

MARK A. GLUCK, DAVID B. PARKER, and ERIC REIFSNIDER
Stanford University, Stanford, California

Klopf (1988) presents a formal real-time model of classical conditioning which generates a wide range of behavioral Pavlovian phenomena. We describe a replication of his simulation results and summarize some of the strengths and shortcomings of the drive-reinforcement model as a real-time behavioral model of classical conditioning. To facilitate further comparison of Klopf's model with neuronal capabilities, we present a pulse-coded reformulation of the model that is more stable and easier to compute than the original, frequency-based model. We then review three ancillary assumptions to the model's learning algorithm, noting that each can be seen as dually motivated by both behavioral and biological considerations.

Klopf (1988) presents a formal real-time model of classical conditioning that predicts the magnitude of conditioned responses (CRs), given the temporal relationships between conditioned stimuli (CSs) and an unconditioned stimulus (US). When compared with alternative accounts in the animal learning literature (e.g., Donegan & Wagner, 1987; Sutton & Barto, 1981, 1987; Wagner, 1981), Klopf's drive-reinforcement (D-R) model ranks as a simple, elegant, and powerful account of many real-time conditioning phenomena. The possibility that this model might be implemented with a single neuron is intriguing but largely speculative.

In this short note we review Klopf's model, describe our replication of his simulations, and summarize some of the strengths and shortcomings of the D-R theory as a real-time behavioral model of classical conditioning. To facilitate a comparison between Klopf's model and neuronal capabilties, we develop and evaluate a pulse-coded reformulation of Klopf's model.

Summary of Klopf's Model

Hebb (1949) proposed that learning involved changes in the efficacy of plastic synapses at the neuronal level. He suggested that these changes occurred through correlations between approximately simultaneous pre- and postsynaptic levels of neuronal acitivity. In an attempt to create a model that more accurately simulated the behavioral properties of animal learning, Klopf (1988) proposed a variation on Hebb's model that incorporated and extended some of the ideas presented in Sutton and Barto's (1981) real-time generalization of Rescorla and Wagner's (1972) model of classical conditioning.

First, in lieu of correlating levels of pre- and postsynaptic activity, Klopf proposed that changes in presynaptic levels of activity should be correlated with changes in postsynaptic levels of activity. Second, instead of correlating approximately simultaneous pre- and postsynaptic levels of activity, Klopf suggested that earlier changes in presynaptic signal levels should be correlated with later changes in postsynaptic signal levels. Third, in order to produce S-shaped acquisition curves (similar to those observed in animal learning), he proposed that changes in synapse efficacy be proportional to current synapse efficacy. The mathematical specification of Klopf's D-R model consists of two equations: one that calculates output signals on the basis of a weighted sum of input signals (drives), and one that determines changes in synapse efficacy due to changes in signal levels.

The specification of signal output level is defined as

$$y(t) = \sum_{i=1}^{n} w_i(t)x_i(t) - \theta, \qquad (1)$$

where $y(t)$ is the measure of postsynaptic frequency of firing at time t, $w_i(t)$ is the efficacy (positive or negative) of the ith synapse, $x_i(t)$ is the frequency of action potentials at the ith synapse, θ is the threshold of firing, and n is the number of synapses on the "neuron." This equation expresses the idea that the postsynaptic firing frequency depends on the summation of the weighted presynaptic firing frequencies, $w_i(t)x_i(t)$, relative to some threshold, θ. Note that the US is conceived of as a nonplastic, strongly weighted input to the neuron.

The learning mechanism is defined as

$$\Delta w_i(t) = \Delta y(t) \sum_{j=1}^{\tau} c_j |w_i(t-j)| \Delta x_i(t-j), \qquad (2)$$

where $\Delta w_i(t)$ is the change in efficacy of the ith synapse at time t, $\Delta y(t)$ is the change in postsynaptic firing at time t, and τ is the longest interstimulus interval over which delayed conditioning is effective. The c_j are empirically

established learning-rate constants, each corresponding to a different interstimulus interval.

The learning mechanism is thus defined in terms of four factors: (1) the learning rate constants, c_j, (2) the absolute value $|w_i(t-j)|$ of the efficacy of the synapse at time $t-j$, when the change in presynaptic level of activity occurred, (3) the change in presynaptic activity level $\Delta x_i(t-j)$, and (4) the change in postsynaptic level of activity $\Delta y(t)$. Through the summation of changes in presynaptic signal levels, the learning mechanism correlates earlier changes in presynaptic signal levels with later changes in postsynaptic signal levels. Note that the calculation uses the absolute value of the synapse efficacy, $|w_i(t-j)|$, allowing for both inhibitory and excitatory synapses in the simulation of classical conditioning.

To accurately simulate various behavioral phenomena observed in classical conditioning, Klopf added three ancillary assumptions to his model. First, he placed a lower bound of 0 on the activation of the node. Second, he proposed that changes in synaptic weight, $\Delta w_i(t)$, be calculated only when the change in presynaptic signal level is positive—that is, when $\Delta x_i(t-j) > 0$. Third, he proposed separate excitatory and inhibitory weights in contrast to the single real-valued associative weights in other conditioning models (e.g., Rescorla & Wagner, 1972; Sutton & Barto, 1981).

Replication and Evaluation of Behavioral Properties

We have reimplemented and successfully replicated all of the D-R model simulations presented in Klopf (1988). Like the Sutton and Barto (1981, 1987) model, Klopf's D-R model is a real-time extension of Rescorla and Wagner's (1972) trial-level model of classical conditioning; because of this, the D-R model accounts for all the behavioral properties associated with the Rescorla-Wagner model, including acquisition, extinction, conditioned inhibition, partial reinforcement, overshadowing, and blocking. In addition, Klopf's model is able to account for some aspects of the following real-time phenomena: second-order conditioning, enhanced reacquisition, interstimulus interval effects, delay and trace conditioning, as well as the effects of interstimulus intervals and conditioned and unconditioned stimulus duration and amplitude.

The strength of Klopf's model as a simple formal behavioral model of classical conditioning is quite evident. Although the model has not yielded any new behavioral predictions, it has demonstrated an impressive abilty to reproduce a wide, though not necessarily complete, range of Pavlovian behavioral phenomena with a minimum of assumptions. Many conditioning phenomena are presumably the result of complex circuit-level interactions among many neurons; thus it might seem unlikely that many conditioning phenomena can be traced to the properties of synaptic learning rules. It is possible, however, that gross characteristics of system-level behavior may, in certain situations, mimic the properties of component mechanisms. For example, the learning behavior of large mul-

tilayer adaptive networks will, in some situations, but not all, mimic the properties of their component neuron-like elements (Gluck & Bower, 1988a, 1988b).

In addition to its strengths as a behavioral model, Klopf's D-R theory contains a number of intriguing similarities to the computational characteristics of single neurons. Some of these similarities are discussed in Klopf (1988).

In the succeeding sections of this paper, we highlight some of the nontrivial similarities between neuronal capabilities and the computations necessary for implementing the D-R theory. One way of laying the groundwork for developing more complete integrated theories of the biological bases of associative learning is to note the similarities and differences between behavioral theories and biological capabilities (Donegan, Gluck, & Thompson, in press).

To evaluate the strengths and weaknesses of the D-R theory as a model of neuronal functioning requires that one distinguish between two aspects of the model. First, one key idea behind Klopf's model is the differential-Hebbian learning rule, by which earlier changes in input signals are correlated with later changes in output signals to produce changes in synaptic weights. In addition to this learning rule, Klopf made a number of other ancillary assumptions. In evaluating the biological plausibility of the model, we have found it useful to consider the learning algorithm and the ancillary assumptions separately.

Klopf (1988) specified the D-R learning algorithm in terms of activation or frequency levels. Because neuronal systems communicate through the transmission of discrete pulses, it is difficult to evaluate the biological plausibility of an algorithm when so formulated. For this reason, we present and evaluate a pulse-coded reformulation of Klopf's D-R model. We then review the ancillary assumptions Klopf made to the learning algorithm, noting that each can be seen as dually motivated, by both behavioral and biological considerations.

Pulse-Coding in Neuronal Systems

Neurons, it is well known, use pulses as input and output. Frequency (or activation) is a useful abstraction of pulse trains, especially for bridging the gap beween whole-animal and single-neuron behavior. To evaluate the implications of a neuronal model more precisely, it is preferable to develop a model that transmits discrete pulse-coded information. We describe here a reformulation of the D-R model that uses pulses in place of frequencies. Our motivation is that such an extension will enable the D-R model to be compared in a more detailed manner with the biological properties of neuronal functioning.

We will begin by outlining the general theory and the engineering advantages of pulse-coding and then describe a pulse-coded reformulation of differential-Hebbian learning. The key idea, which is quite simple, can be summarized as follows: Frequency can be seen, loosely speaking, as an integral of pulses; conversely, therefore, pulses can be thought of as carrying information about the derivatives of frequency. Thus, computation with the deriva-

tives of frequency, as in Klopf's model, is analogous to computation with pulses. As described below, our basic conclusion is that differential-Hebbian learning, when reformulated for pulse-coded system, is both more stable and easier to compute than is apparent when the rule is formulated in terms of frequencies. These results have important implications for any learning model that is based on computation with time-derivatives, such as the Sutton and Barto (1987) temporal-difference model.

There are many ways to transmit analog information from point to point. Perhaps the most obvious way is to transmit the information as a signal level. In electronic systems, for example, data that vary between 0 and 1 can be transmitted as a voltage level that varies between 0 and 1 V. This method can be unreliable, however, because the receiver of the information can't tell whether a constant dc voltage offset has been added to the information, or whether crosstalk has occurred with a nearby signal path. To the exact degree that the signal is interfered with, the transferred data will be erroneously altered. Furthermore, if computations based on derivatives of a signal are required (as in Klopf's model), then even a small, but sudden, change in signal level can drastically alter its derivative.

A more reliable way to transmit analog information is to encode it as the frequency of a series of pulses. Even in the face of dc voltage offsets or moderate crosstalk, a receiver can reliably determine that it has received a pulse. Because most errors will not be large enough to constitute a pulse, they will have no effect on the transmitted information. The receiver can then count the number of pulses received in a given time window to determine the frequency of the pulses. Further information on encoding analog information as the frequency of a series of pulses can be found in many electrical engineering textbooks (e.g., Horowitz & Hill, 1980).

As noted by Parker (1987), a further advantage of coding an analog signal as the frequency of a series of pulses is that the time derivative of the signal can be calculated easily and stably. If $x(t)$ represents a series of pulses ($x = 1$ if a pulse is occurring at time t; otherwise $x = 0$), then we can estimate the frequency $f(t)$ of the series of pulses by using an exponentially weighted time average:

$$f(t) = \mu \int_{-\infty}^{t} x(\tau)e^{-\mu(t-\tau)}d\tau, \qquad (3)$$

where μ is the decay constant. The well-known formula for the derivative of $f(t)$ is

$$\frac{df(t)}{dt} = \mu[x(t) - f(t)]. \qquad (4)$$

Thus, the time derivative of pulse-coded information can be calculated without using any unstable differencing methods. It is simply a function of the presence or absence of a pulse, relative to the current expectation (frequency) of pulses. The calculation of time derivatives is a critical component of both Klopf's (1988) and Sutton and Barto's (1981, 1987) real-time models of associative

learning. They are also an important aspect of second-order (pseudo-newtonian) extensions of the backpropagation learning rule for multilayer adaptive "connectionist" networks (Parker, 1987). Further details on the computational modeling of pulse-coding in neuronal systems can be found in Colbert and Levy (1988). The details of a pulse-coded reformulation of Klopf's (1988) D-R model are presented in our Appendix.

Ancillary Assumptions to the D-R Model

The fundamental idea behind Klopf's D-R model is the differential-Hebbian learning rule described above. In addition, Klopf made a number of ancillary assumptions. It is intriguing that all of these assumptions are not only sufficiently justified by constraints from behavioral data but also motivated by neuronal constraints. Three assumptions in particular are doubly (and thus convergingly) motivated by both behavioral and neuronal constraints.

The first such assumption is the lower bound of 0 for frequency calculations. Turning first to the behavioral evidence, we note that despite the many successes of the Rescorla–Wagner trial-level model of classical conditioning, this model does have several well-known limitations. One shortcoming is that the Rescorla–Wagner model erroneously predicts that one can drive a conditioned inhibitor (i.e., one with net negative associative strength) to zero strength by presenting it without the US. The behavioral data indicate that this is not the case (Zimmer-Hart & Rescorla, 1974). In the Rescorla–Wagner model, the net activation from a stimulus (analogous to the output signal in Klopf's model) can range over both negative and positive values. If, in Klopf's model, the output frequencies were allowed to be both negative and positive, then, as in the Rescorla–Wagner model, Klopf's model would mistakenly predict the extinction of conditioned inhibitors. That is, if the activity at the inhibitory synapse were allowed to push y below zero (corresponding to a positive x and negative y), then, when the output activity ceased, a positive Δy would ensue. The result would be the pairing of positive Δx and a later positive Δy, thus leading to the extinction of the associative weight. When frequencies in the D-R model are limited to the nonnegative, however, a conditioned inhibitor will not become extinguished, because the positive Δx that occurs at the time of the conditioned inhibitor's onset is not followed by a positive Δy.

Thus, there is strong behavioral justification for allowing only positively valued output signals. This assumption can be justified from a biological standpoint, because negative output signals would have no meaning if they were meant to correspond to negative output firing frequencies. There is no biological meaning for a negative "firing frequency," as the term is used in this paper.

Klopf's second ancillary assumption is the use of only positive-valued presynaptic changes. The behavioral justification for this assumption is clear: the performance of the model deviates radically from experimental evidence when this restriction is relaxed; in particular, it be-

comes possible for negative Δxs to be paired with negative Δys, with a net positive change in synaptic weights. The offset of a CS followed by a decrement in the output of the model would reinforce that CS; this would not be consistent with the behavioral data.

The biological justification for the use of only positive-valued presynaptic changes can be most easily appreciated with reference to Equation 4 of the pulse-coded reformulation of the D-R model. In the pulse-coded model, incoming pulses carry information about positive changes in presynaptic frequency. Thus, a system that mediates changes on the basis of incoming pulses can be seen to be equivalent to a system in which synaptic changes are mediated by positive-valued presynaptic changes. Negative-valued changes in presynaptic frequency reflect the internal decay of the frequency counter in Equations 3 and 4.

The third ancillary assumption in the D-R model is the use of separate excitatory and inhibitory weights (cf. the single associative values of the Rescorla–Wagner model). One behavioral motivation for these separate weights is that reacquisition is quicker than initial acquisition when the model is implemented in this way. With separate excitatory and inhibitory weights, the reinforcement of the excitatory weight is not entirely lost during extinction. Extinction ends when the inhibitory output to the model is strong enough to cancel the excitatory output. When this is the case, there is no CR, and thus no conditioning occurs. Since more excitatory reinforcement remains after extinction than before the initial training, the excitatory weight rises more quickly to its asymptotic value than it would if it had been at its starting value. This is both because it is closer to asymptote than the minimum excitatory weight is, and also because the nature of the algorithm allows larger weights to grow faster.

The CR also depends on the inhibitory input, which adds to the rapidity of reacquisition. While the excitatory weight rises, the inhibitory weight decreases in magnitude. Thus, the CR grows rapidly as the inhibition disappears. We point out that the D-R model's account of reacquisition is novel in the conditioning literature in that other accounts of this phenomenon (e.g., Kehoe, in press) have necessitated postulating two-stage processes. The biological motivation for separate inhibitory and excitatory weights is clear. Synapses in real neurons are generally presumed to be either exclusively inhibitory or exclusively excitatory, regardless of the level of plasticity they demonstrate.

To summarize, each of the three ancillary assumptions that Klopf incorportes in the D-R model is justifiable solely with reference to the behavioral phenomena of conditioning. Still, it is intriguing that if one were designing a physiologically realistic neuronal model, one would similarly be drawn to these same three assumptions.

General Discussion

Klopf (1988) has presented a formal real-time model of classical conditioning that generates a wide range of Pavlovian behavioral phenomena. In this paper we first reviewed Klopf's model, described our replication of his results, and summarized some of the strengths and shortcomings of the D-R theory as a real-time behavioral model of classical conditioning. We then considered the possibility that this model might be implemented at the level of a single neuron. To facilitate comparison of Klopf's model with neuronal capabilities, we formulated a pulse-coded variation of Klopf's model. Our basic conclusion is that differential-Hebbian learning, when reformulated for a more neuronally realistic pulse-coded system, is more stable and easier to compute than is apparent when the rule is formulated in terms of frequencies.

In addition to a differential-Hebbian learning rule, the D-R model includes three ancillary assumptions, each of which can be motivated from both a biological and a psychological (behavioral) perspective. First, a lower bound of 0 is placed on the output activation of the node. Second, changes in synaptic weights, $\Delta w_i(t)$, are calculated only when the change in presynaptic signal levels is positive. Third, separate excitatory and inhibitory weights are used.

In reviewing Klopf's D-R model, we have noted that the model stands on its own as a real-time behavioral model of classical conditioning; the possibility that the model might be implemented within a single neuron is intriguing but largely speculative. The pulse-coded D-R model is intended to simulate more closely the physical processes that occur in neurons than the original formulation of the D-R model can. In doing this, our motivation has been to enable us to draw further parallels and connections between real-time behavioral models of learning and biological circuit models of the substrates that underlie classical conditioning (e.g, Donegan, Gluck, & Thompson, in press; Gluck & Thompson, 1987; Thompson, 1986).

REFERENCES

COLBERT, C. M., & LEVY, W. B. (1988). What is the code? In *Proceedings of the International Neural Network Society* (pp. 246). Boston: Pergamon Press.

DONEGAN, N. H., GLUCK, M. A., & THOMPSON, R. F. (in press). Integrating behavioral and biological models of classical conditioning. In R. D. Hawkins & G. H. Bower (Eds.), *The psychology of learning and motivation: Vol. 22. Computational models of learning in simple neural systems.* New York: Academic Press.

DONEGAN, N. H., & WAGNER, A. R. (1987). Conditioned diminution and facilitation of the UCR: A sometimes-opponent-process interpretaton. In I. Gormezano, W. Prokasy, & R. Thompson (Eds.), *Classical conditoning II: Behavioral, neurophysiological, and neurochemical studies in the rabbit* (pp. 339-369). Hillsdale, NJ: Erlbaum.

GLUCK, M. A., & BOWER, G. H. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory & Language,* 27, 166-195.

GLUCK, M. A., & BOWER, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General,* 117, 225-244.

GLUCK, M. A., & THOMPSON, R. F. (1987). Modeling the neural substrates of associative learning and memory: A computational approach. *Psychological Review,* 94, 176-191.

HEBB, D. O. (1949). *Organization of behavior.* New York: Wiley.

HOROWITZ, P., & HILL, W. (1980). *The art of electronics*. Cambridge, England: Cambridge University Press.

KEHOE, E. J. (in press). A layered network model of associative learning: Learning-to-learn and configuration. *Psychological Review*.

KLOPF, A. H. (1988). A neuronal model of classical conditoning. *Psychobiology*, **16**, 85-125.

PARKER, D. B. (1987). Optimal algorithms for adaptive networks: Second order back propagation, second order direct propagation, and second order Hebbian learning. In *Proceedings of the IEEE First Annual Conference on Neural Networks* (pp. 593-600). San Diego.

RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

SUTTON, R. S., & BARTO, A. G. (1981). Toward a theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135-170.

SUTTON, R. S., & BARTO, A. G. (1987). *A temporal-difference model of classical conditioning* (Tech. Rep. No. 87-509.2). Waltham, MA: GTE Laboratories.

THOMPSON, R. F. (1986). The neurobiology of learning and memory. *Science*, **233**, 941-947.

WAGNER, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. Spear & G. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

ZIMMER-HART, C. L., & RESCORLA, R. (1974). Extinction of Pavlovian conditioned inhibition. *Journal of Comparative & Physiological Psychology*, **86**, 837-845.

APPENDIX

A Pulse-Coded Reformulation of the D-R Model

We illustrate here a pulse-coded reformulation of Klopf's D-R model (Klopf, 1988). The equations that make up the model are fairly simple. A neuron is said to have fired an output pulse at time t if $v(t) > e$, where e is a threshold value and $v(t)$ is defined as follows:

$$v(t) = (1-d)v(t-1) + \sum w_i(t)x_i(t), \tag{A1}$$

where $v(t)$ is an auxiliary variable, d is a small positive constant representing the leakage or decay rate, $w_i(t)$ is the efficacy of synapse i at time t, and $x_i(t)$ is the frequency of presynaptic pulses at time t at synapse i. The input to the decision of whether the neuron will fire consists of the weights and efficacies of the synapses, as well as information about previous activation levels at the neuronal output. Note that the leakage rate, d, causes older information about activation levels to have less impact on current values of $v(t)$ than does recent information of the same type.

The output of the neuron, $p(t)$, is as follows:

If $v(t) > e$ then $p(t) = 1$ (pulse generated)

If $v(t) \le e$ then $p(t) = 0$ (no pulse generated)

It is important that once $p(t)$ has been determined, $v(t)$ will need to be adjusted if $p(t) = 1$. To reflect the fact that the neuron has fired [i.e., $p(t)=1$], $v(t) = v(v) - 1$. This decrement occurs after $p(t)$ has been determined for the current t. Frequencies of pulses at the output node and at the synapses are calculated using the following equations:

$$f(t) = f(t-1) + \Delta f(t), \tag{A2}$$

where

$$\Delta f(t) = m[p(t) - f(t-1)], \tag{A3}$$

where $f(t)$ is the frequency of outgoing pulses at time t, $p(t)$ is the output (1 or 0) of the neuron at time t, and m is a small positive constant representing a leakage rate for the frequency calculation.

Following Klopf (1988), changes in synapse efficacy occur according to

$$\Delta w_i(t) = \Delta y(t) \sum_{j=1}^{\tau} c_j \, |w_i(t-j)| \, \Delta x_i(t-j), \tag{A4}$$

where

$$\Delta w_i(t) = w_i(t+1) - w_i(t),$$

and $\Delta y(t)$ and $\Delta x_i(t)$ are calculated analogously to $\Delta f(t)$, τ is the longest interstimulus interval (ISI) over which delay conditioning is effective, and c_j is an empirically established set of learning rates that govern the efficacy of conditioning at an ISI of j.

Changes in $w_i(t)$ are governed by the learning rule in Equation A4, which alters $v(t)$ via Equation A1.