A Simplified Drive-Reinforcement Model for Unsupervised Learning in Artificial Neural Networks

David B. Suits
Department of Philosophy
Rochester Institute of Technology
Rochester, NY 14623

david.suits@rit.edu

(Originally part of a Master's Thesis, May, 1992;
revised for placement on the web, May 2004.

1. Background

A. H. Klopf (1982) proposed to treat neurons (or a class of neurons) as "hedonists". If depolarization (i.e., excitation) is considered "pleasure", and hyperpolarization (i.e., inhibition) "pain", then we may imagine a neuron which seeks to maximize depolarization and minimize hyperpolarization (or, rather, maximize their difference). Klopf calls the strategy which such a neuron might use *heterostatic adaptation*. It is quite straightforward: whenever the neuron fires (because its excitations less its inhibitions exceed some threshold), it will "notice", during some subsequent interval, τ (probably a few seconds), whether the difference between its excitations and inhibitions changes. Positive changes result in the neuron's greater tendency to fire on subsequent occasions, and negative changes have the opposite effect. Such tendencies will be implemented by means of changes to the neuron's postsynaptic efficacy — increasing the efficacy of excitatory connections in the case of positive changes, and increasing the efficacy of inhibitory connections in the case of negative changes.

> For a neuron, temporal and spatial configurations of active synapses represent conditioned stimuli (CS), firing represents a conditioned response (CR), and the excitation or inhibition that arrives during a limited period of time after firing constitutes the unconditioned stimulus (US) [Klopf, 1982, p. 5]

which then acts as CS for signals which arrive still later.

On Klopf's drive-reinforcement model, earlier changes in synapse activation are to be correlated with later changes in neuronal activity. This is accomplished by discretizing a certain interval of time and assigning "eligibility" values to each discrete time — in effect, an array of values for the past τ time units. A

second requirement is in effect a second array, namely, the record of changes of presynaptic activations, so that when neuronal activation changes, the past presynaptic activations can be correlated with the past eligibility values so as to create changes in synaptic weight.

But these requirements impose a computational burden (especially in large networks), and so a simplifying alternative to one or both of them would be desirable. In addition, lists of eligibility values and presynaptic changes are necessarily finite, so that unless the lists are changed, we cannot take a more fine-grained look at weight changes, nor investigate conditioning effects for interstimulus intervals greater than the last discrete time unit, nor investigate conditioning effects for different eligibility values.

Klopf (1986) makes several not unreasonable simplifying stipulations: that the optimum interstimulus interval (ISI) is 500 msecs; that ISIs shorter than the optimum are not interesting; and that ISIs greater than about 2500 msecs are not interesting. But these stipulations are very restricting. In particular, in animal learning experiments the optimum ISI varies depending on the experimental preparation and the species of the animal to be conditioned (Rachlin, 1976; Bitterman, 1965; Ost & Lauer, 1965; Razran, 1965; Gormezano, 1972; Alkon, 1983; Alkon *et al.*, 1989), and conditioning effects occur well beyond 2.5 seconds (Ost & Lauer, 1965; Kehoe, 1990). In one experiment the ISI was 60 seconds (Garcia, McGowan & Green, 1972).

Since Klopf's eligibility values fall roughly on an exponential curve, my simplified drive reinforcement (SDR) model uses a single value which acts as an exponentially decaying impression of past presynaptic changes, thereby both simplifying and generalizing Klopf's model. The exponential decay may be followed out to an arbitrarily large ISI; there is no requirement to keep lists of eligibility values and past presynaptic changes; and consequently the computational burden is considerably lessened.

2. The SDR model

$$\Delta w_i(t) = \beta e_i(t)\, \Delta y(t) \tag{1}$$
$$e_i(t) = \alpha e_i(t-1) + |w_i(t-1)|\, \min[0, \Delta x_i(t-1)] \tag{2}$$
$$e_i(0) = 0 \tag{3}$$

where $\Delta w_i(t) = w_i(t+1) - w_i(t)$; $\Delta y(t) = y(t) - y(t-1)$; $y(t)$ is the node output, defined as the sum of all inputs, x_i, times their respective weights, w_i, and is bounded to $0 \le y(t) \le YMAX$ (probably any convenient function will do; I have used both hard-limiting and an exponential function); $\Delta x_i(t) = x_i(t) - x_i(t-1)$; and all weights, w_i, have initial non-zero values and minimum absolute values, $0 < WMIN \le |w_i|$. β is a positive constant which influences the rate of weight changes. The e_i, always non-negative, are the "eligibility" values and act as exponentially decayed impressions of past presynaptic changes with decay rate α, $0 < \alpha < 1$.

The equations above can be described in terms of a simple algorithm for updating the synapses of a given node:

$y = \text{bound}(\sum_i w_i x_i)$
$\Delta y = y - \text{previous\_y}$

for each synapse, i {

$e_i = e_i \, \alpha$

if $\Delta x_i > 0$ then $e_i = e_i + \Delta x_i \, |w_i|$

$w_i = w_i + \beta e_i \Delta y$

if $|w_i| <$ WMIN then $|w_i|$ = WMIN

$\Delta x_i = x_i -$ previous_x_i

previous_$x_i = x_i$

}

(Although the SDR model was developed principally from Klopf's model, it may also be derived from the Sutton-Barto model (Sutton and Barto, 1981, 1990) in

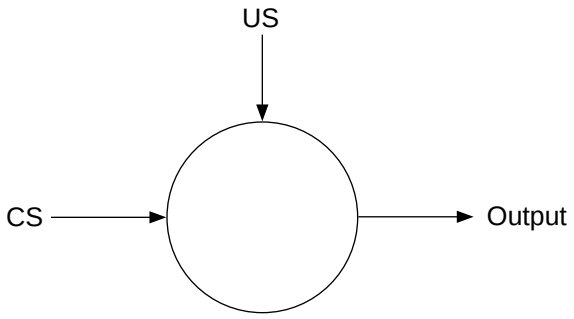

Figure 1
An SDR node has one or more CSs (conditioned stimuli), each with a modifiable efficacy (weight), one US (unconditioned stimulus), whose synapse is unmodifiable, and an output.

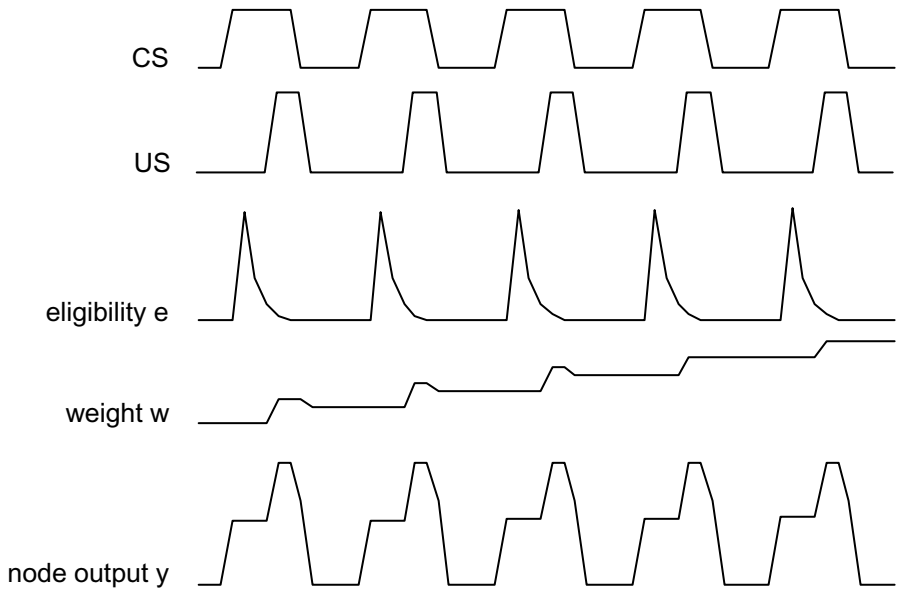

Figure 2
A time trace of the SDR model during several conditioning trials. Except for the eligibility values, the model acts approximately like Klopf's.

a straightforward way.)

Figure 2 shows a time trace of the main variables of the SDR model during some typical conditioning trials. Notice that the eligibility, e, is triggered by the rising edge of the conditioned stimulus (CS) (i.e., the node input x). Neither the falling edge of CS nor a steady level of CS (whether low or high) has any effect on e, although the node output y, and therefore also the change in output, are affected. The SDR model, following Klopf, requires that weights have minimum absolute values. Why should we begin with non-zero weights? For two reasons. First, the equations of the model will not produce weight changes when weights are zero. For this reason also, weight changes are not allowed to cross zero: excitatory connections remain excitatory, and inhibitory connections remain inhibitory. Second, a CR cannot be reinforced unless the CR has some possibility of occurring (prior to US onset).

3. Classical conditioning experiments with the SDR model

3.1 Trace conditioning, delay conditioning effects, and extinction

Figure 3 shows the progress of synaptic weight changes in an SDR unit over 150 trials using trace conditioning, delay conditioning, and "simultaneous conditioning". For CS_1, onset is at 2, offset at 6. Four separate versions of delay conditioning are presented, with four CSs, each four time units in duration, but with shifted onsets: CS_2 onset = 3, offset = 7; CS_3 onset = 4, offset = 8; CS_4 onset = 5, offset = 9; and CS_5 onset = 6, offset = 10. In all cases, US onset is at 7 and US offset is at 9. US amplitude is 0.7 and amplitude of all CSs is 0.2. α, the decay rate constant,

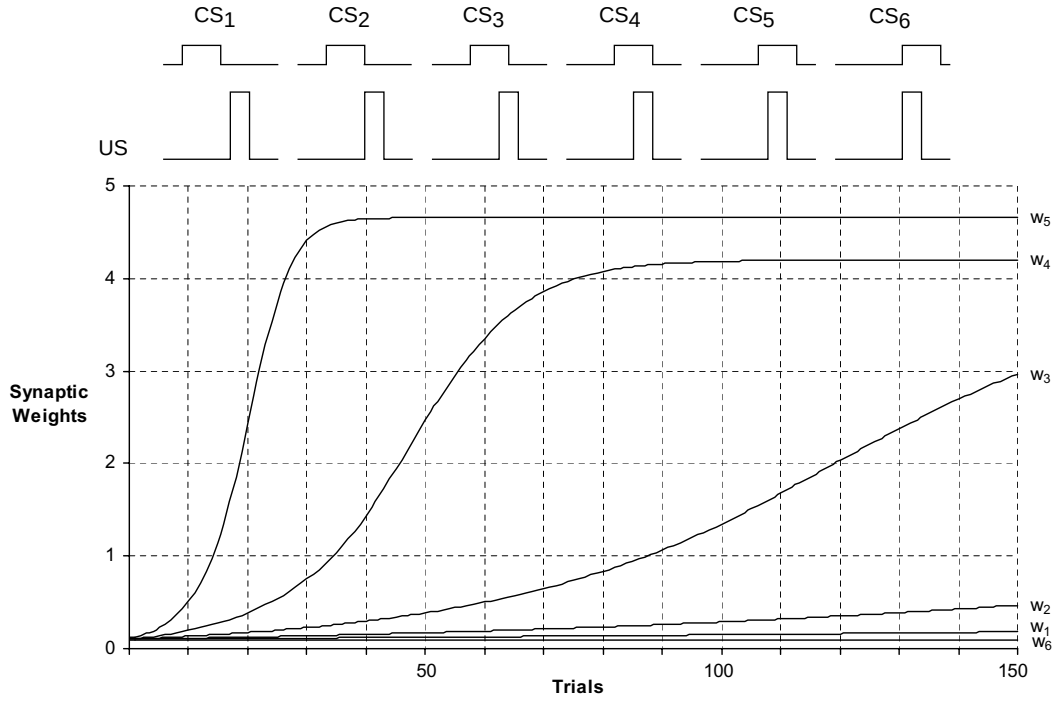

Figure 3
The SDR model under six conditioning experiments. Parameter values are given in the text.

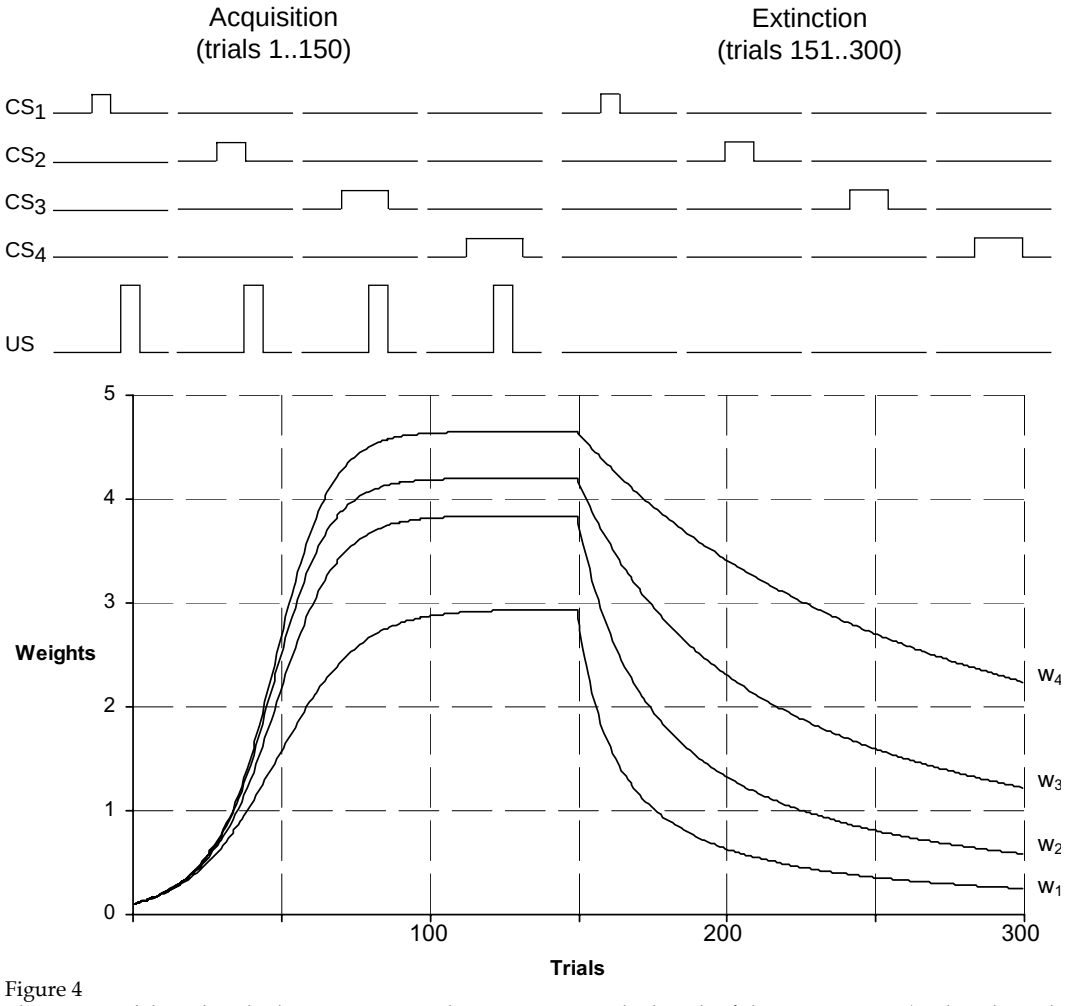

Figure 4
The SDR model predicts higher synaptic weight asymptotes as the length of the CS increases (trials 1 through 150). The model also accounts for extinction of the CR in the absence of the US (trials 151 through 300).

is 0.4, and β, the learning rate constant, is 1.5.

As expected, a synaptic weight (w_1, w_2, w_3, w_4, and w_5 in figure 3) is shown to change more rapidly and approach a higher asymptote when its onset is closer to US onset. Each curve is S-shaped, corresponding to acquisition curves obtained in animal learning experiments. In the simulations, node output is clipped at a maximum (1.0). Different node output functions may be employed (for example, $y = 1 - \exp(-y)$) without disturbing the basic relationships of the synaptic weight curves.

Not only is the CS onset–US onset interval important, but the length of the CS also determines relative synaptic weight change differences. Figure 4 shows the same constants as before, except for the CSs: this time all CSs have onset at 4; CS_1 offset = 6, CS_2 offset = 7, CS_3 offset = 8, and CS_4 offset = 9. Figure 4 also shows that the SDR model predicts extinction phenomena (trials 151–300).

Figure 5 graphs the relationship between the synaptic weight asymptote and the decay rate constant, α.

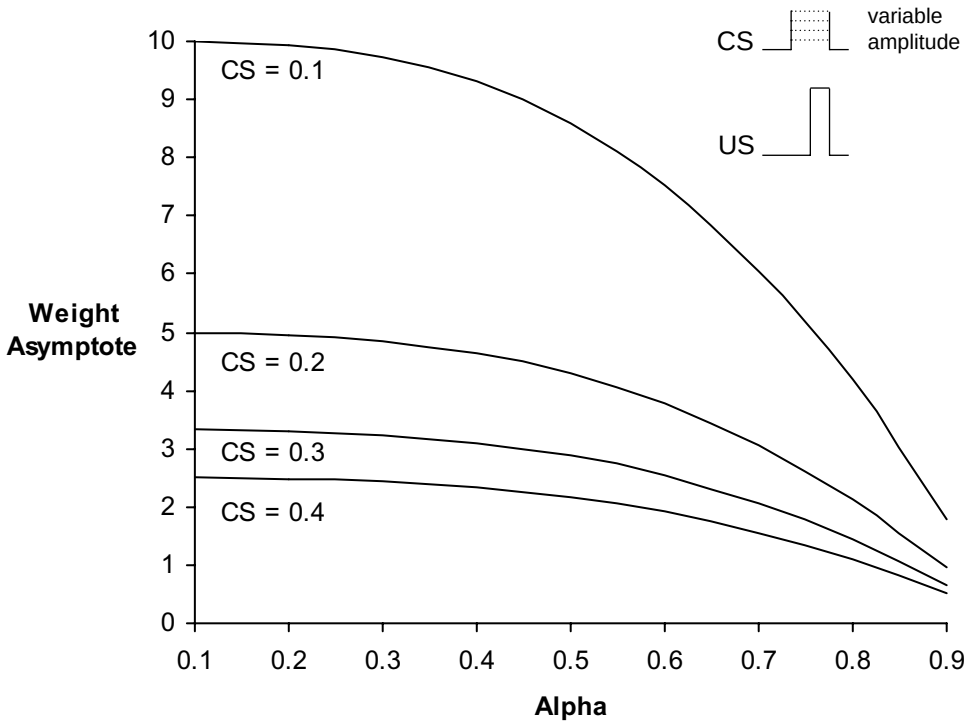

Figure 5
Relationship between weight asymptote and the decay rate α for various CS amplitudes, given a fixed US.

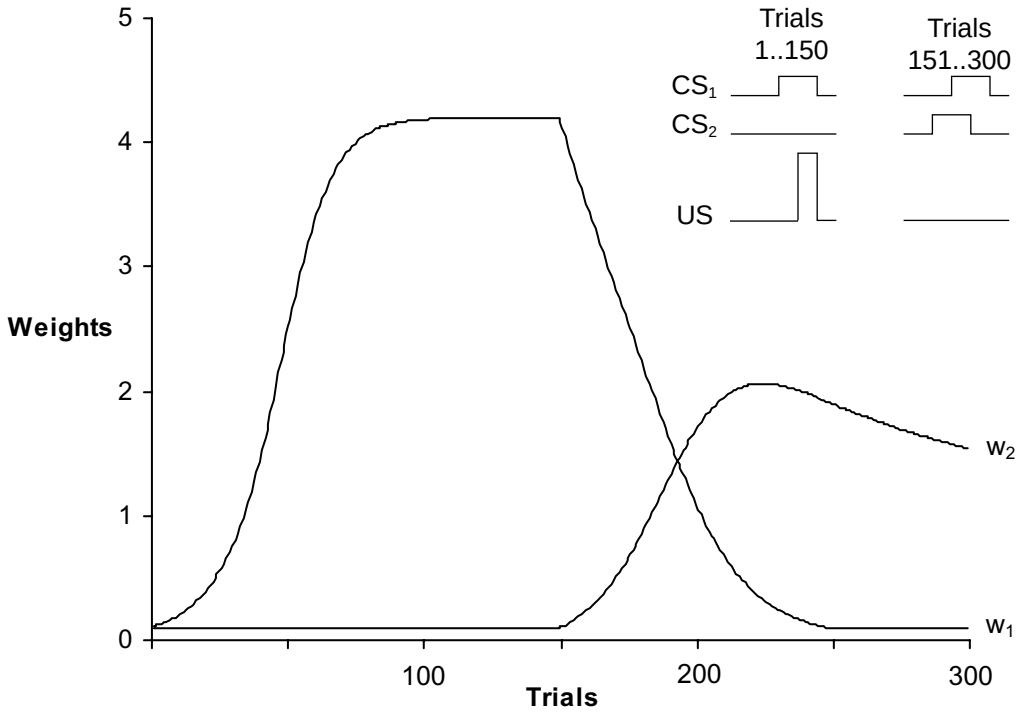

Figure 6
The SDR model's predictions of second order conditioning. Note that during trials 151 through 300 the US is absent, and so the response to CS_1 is undergoing extinction.

3.2. Second order conditioning

Figure 6 graphs the SDR model's predictions about second order conditioning. During trials 1–150, CS_1 is paired with the US, and the synaptic weight w_1 rises. During trials 151–300, the US is removed and at the same time CS_2 is introduced, with onset slightly preceding CS_1 onset. CS_1 therefore acts as US for CS_2. Second order conditioning is rather weak and sometimes ephemeral, because CS_1 is undergoing extinction; when w_1 finally falls to its lower limit (0.1), CS_2 begins to extinguish as well.

3.3 Blocking

Figure 7 graphs a simulation of the phenomenon of blocking. During the first 150 trials, CS_1 is paired with the US, and the synaptic weight (w_1) increases. During subsequent trials, a second stimulus, CS_2 is compounded with CS_1, and both are reinforced by the US. As expected, there is little, if any, gain in synaptic efficacy in CS_2, and only a small loss of efficacy in CS_1.

3.4 Compound conditioning and overshadowing

Figure 8 shows the SDR model in a compound conditioning experiment. Both CS_1 and CS_2 are presented together. Notice that both w_1 and w_2 remain equal throughout when the amplitudes of CS_1 and CS_2 are equal, but the weights do not rise as high

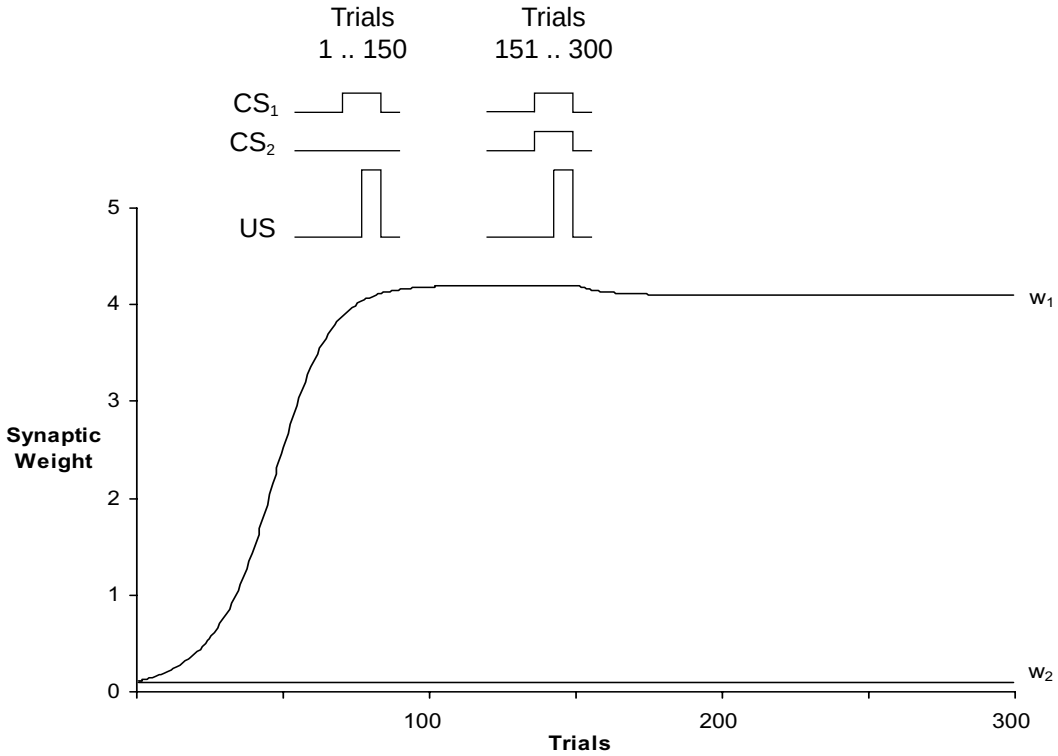

Figure 7
The SDR model's prediction of blocking. The second CS, introduced after the first has been conditioned, gains almost no strength.

as when a single CS is used, everything else being equal (for example, CS_4 in figure 3). When the amplitudes of the two CSs are different, the phenomenon is called overshadowing: a stronger — more salient — stimulus gains proportionally more synaptic efficacy than (i.e., tends to overshadow) a weaker stimulus (figure 9).

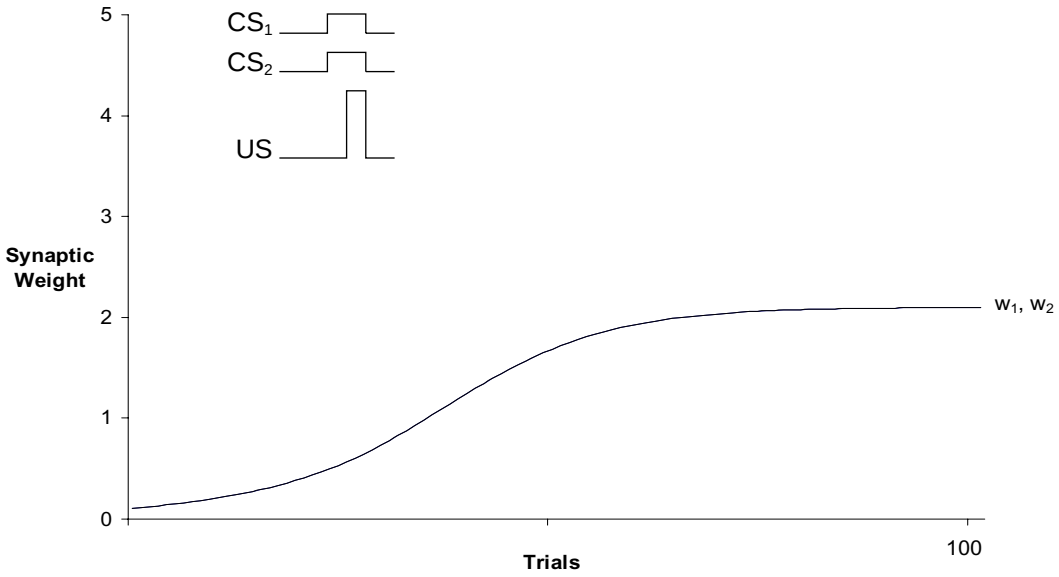

Figure 8
Consistently with the Rescorla-Wagner model and empirical results, the SDR model predicts that two equally salient CSs presented simultaneously will generate equal CRs.

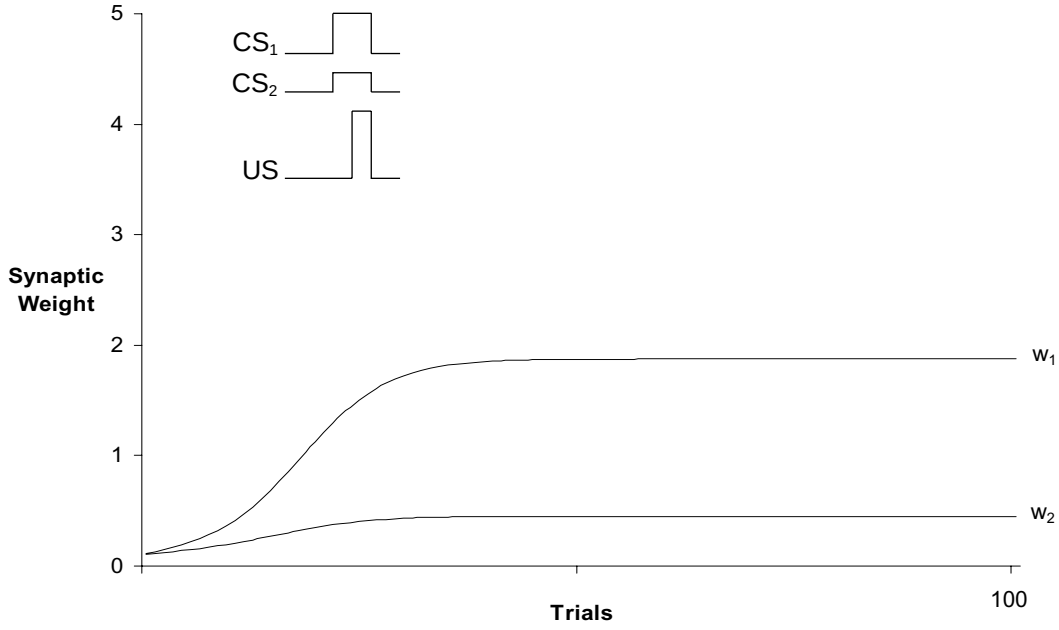

Figure 9
The SDR model's prediction in a compound conditioning experiment when one CS is stronger than a second. The first is said to *overshadow* the second.

3.5 The Wagner-Saavedra experiment

Consider an experiment by Wagner and Saavedra (Rescorla and Wagner, 1972). Three groups of rabbits underwent slightly different conditioning regimens (on the eyelid nictitating membrane, wherein a CS such as a tone or a light is followed by a puff of air to the cornea — the US — which causes the membrane to close). On every other trial, each group received conditioning to the compound $CS_1 + CS_2$. (Assume equal salience of the two stimuli.) The three groups differed in what took place during the alternate trials. For group 1, CS_2 was eliminated (i.e., conditioning continued with CS_1 alone). For group 2, CS_1, CS_2 and US were absent (i.e., there was a rest period). And for group 3, CS_1 was presented without the US (i.e., CS_1 underwent an extinction trial). The three groups were treated identically with respect to CS_2, yet CS_2 emerged with quite different properties. Figure 10 shows the SDR model's predictions of the Wagner-Saavedra experiment. The results accord well with the results from animal learning experiments.

Figure 11 shows the results of a similar experiment. In this case, CS_1 is conditioned during the first 150 trials. On the next 150 trials CS_2 is introduced, but in addition, reinforcement trials with both CSs and the US are alternated with CS_1 extinction trials. As the graph makes clear, CS_2 now tends more and more to be a better predictor of the US, and so CS_1's salience decreases.

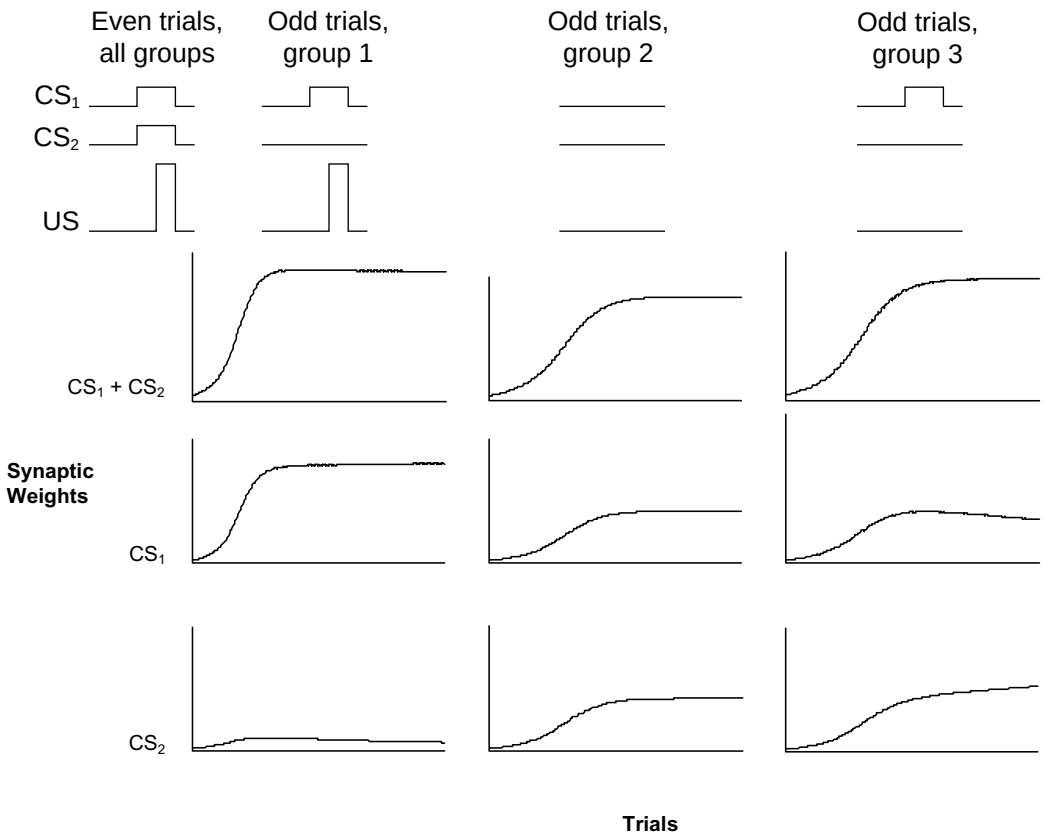

Figure 10
The SDR model's predictions for the Wagner-Saavedra experiment.

3.6 Temporal primacy effects

Empirical data from animal learning experiments indicate that the closer in time a CS is to the US onset, the more powerful the CS's effects will be — i.e., the higher its asymptote, and the quicker it will approach that asymptote. Nevertheless, an earlier CS will eventually win out over a later CS, as figure 12 shows. The SDR model tends to reward the earliest predictor of the US. This temporal primacy effect can even undo the effects of blocking (figure 13) and overshadowing (figure 14).

4. Limitations of the model

In animal learning experiments, as conditioning progresses, the CR begins to appear earlier in the interstimulus interval (ISI). And if a CR is acquired with a random mix of two different ISIs, the CR will eventually appear with two peaks. But there is no provision in temporal models of conditioning (of which the SDR model is one) to account for this. Instead, the SDR model produces a CR precisely at CS onset; the CR remains constant until CS offset.

Nor does the SDR model take into account the frequency of conditioning trials, i.e., the intertrial interval (ITI). Long ITIs seem to promote more rapid conditioning (in terms of number of trials) than shorter ITIs. (Just why this phenomenon appears in animal learning experiments — and whether it is related to some kind of memory consolidation — is an interesting puzzle. The phenomenon also appears

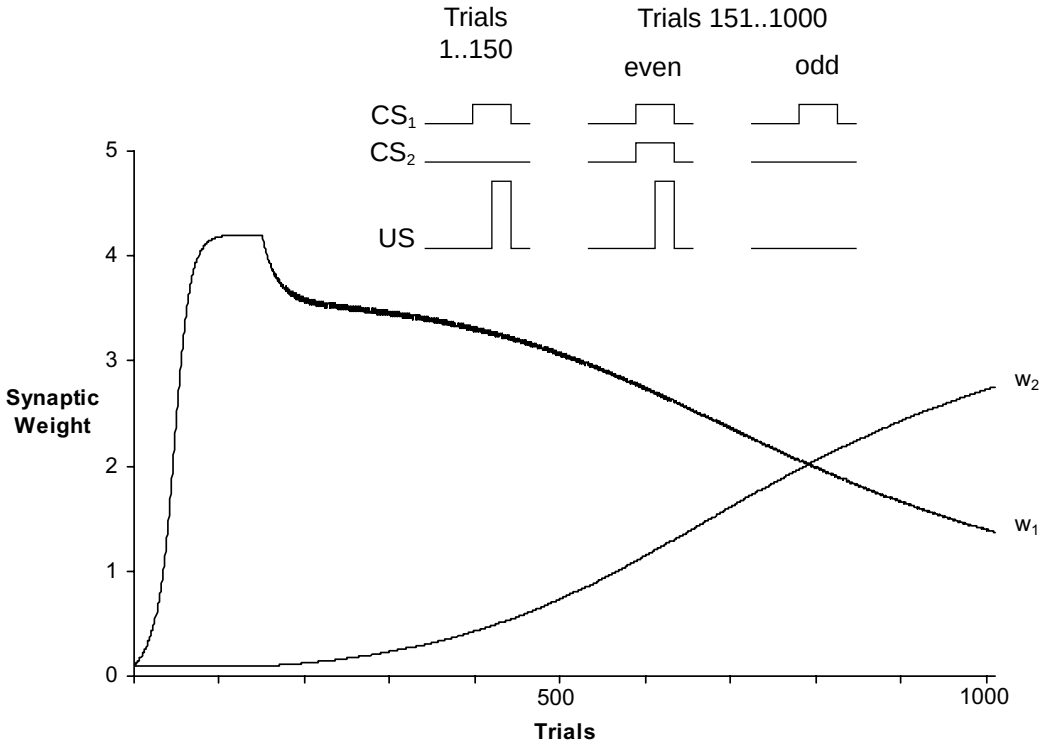

Figure 11
After conditioning on CS_1, CS_2 is introduced on trials alternating with CS_1 extinction trials. According to the SDR model, CS_2 eventually becomes a better predictor of the US.

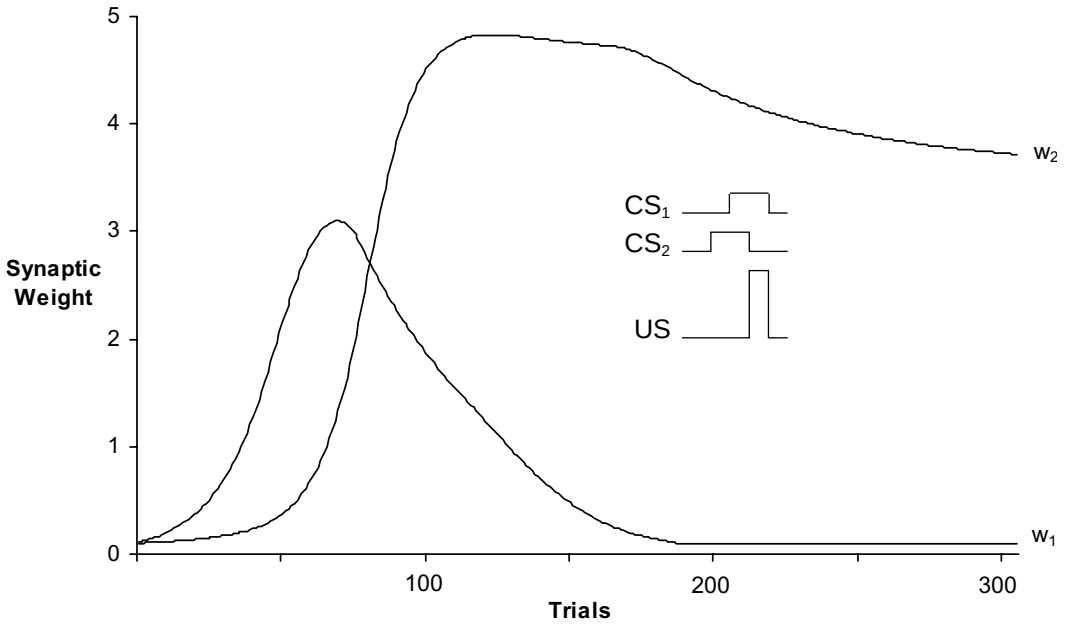

Figure 12
The SDR model's predictions about temporal primacy effects. CS_1, being temporally closer to the US, will at first gain more weight than an earlier CS. But CS_2, as an earlier predictor of the US, eventually dominates.

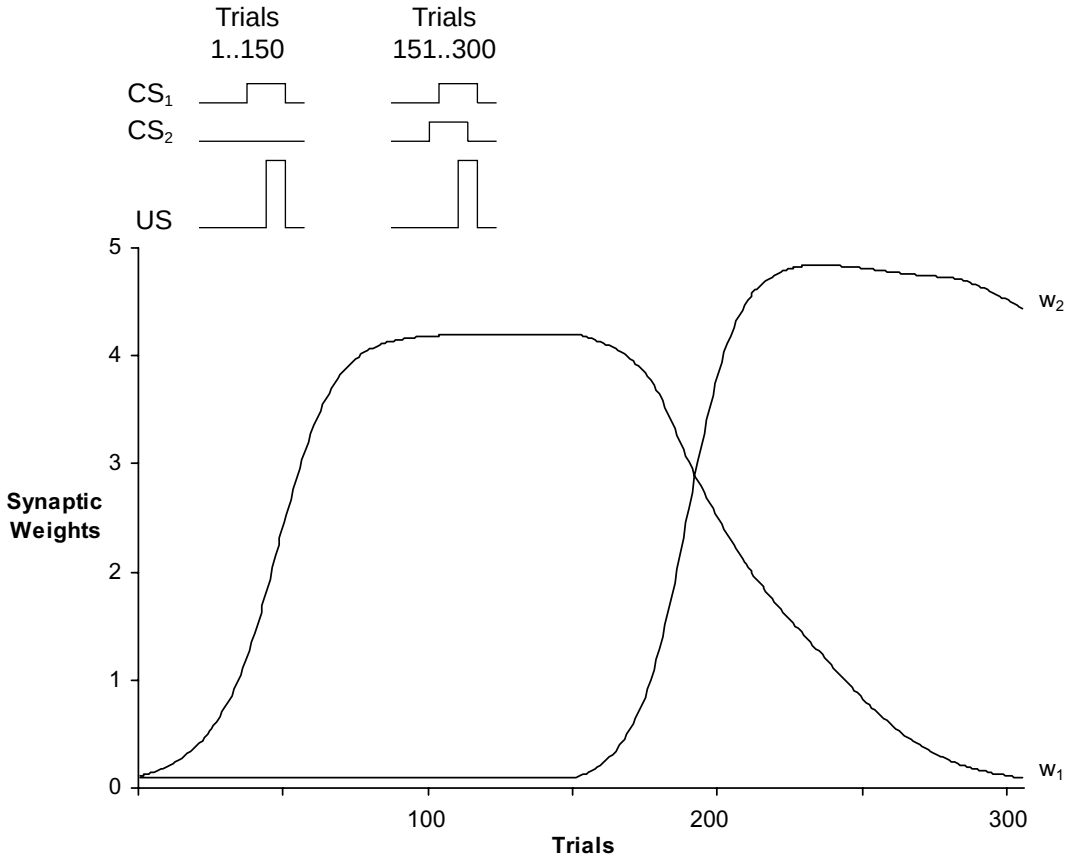

Figure 13
The SDR model predicts that temporal primacy can undo the effects of blocking.

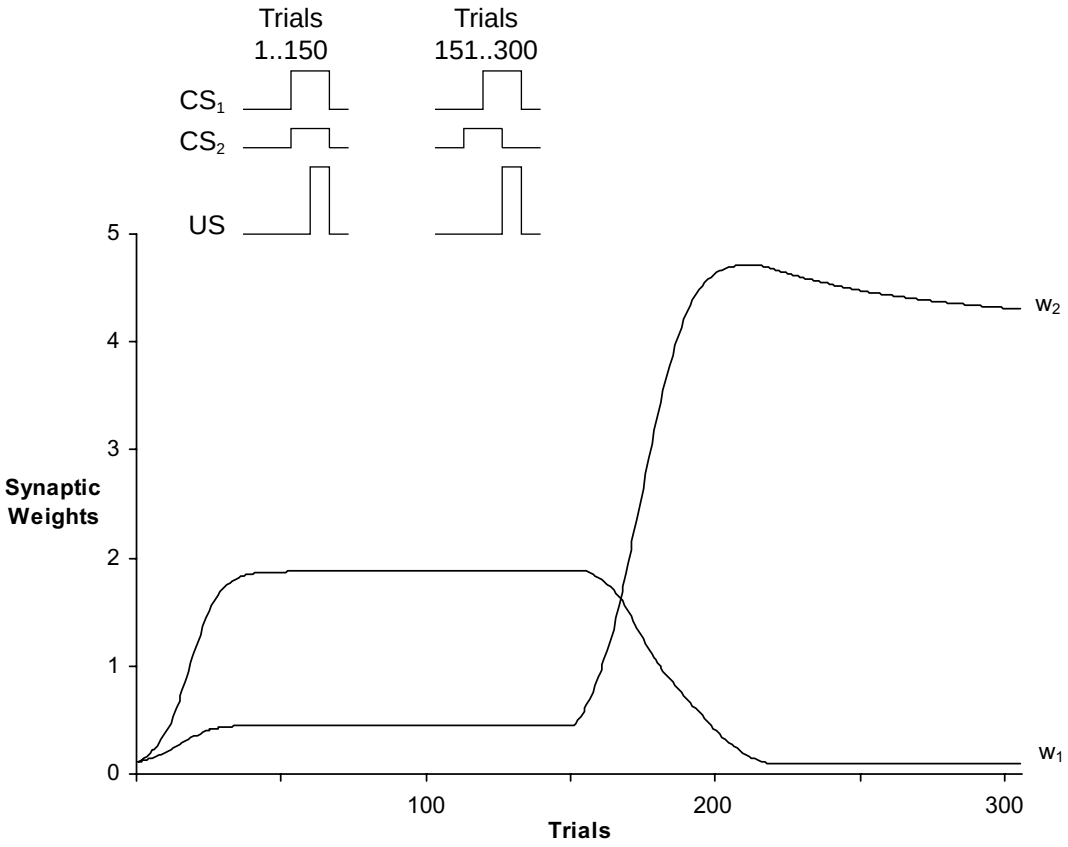

Figure 14
The SDR model predicts that temporal primacy can undo the effects of overshadowing.

in instrumental conditioning.)

Sutton and Barto (1990) point out a problem with Klopf-like models (and hence also the present SDR model) employing an eligibility trace which is initiated by CS onset, but which then proceeds independently of the CS: Changes in synaptic weight depend on changes in neuronal output (which may be caused by CS offset or US onset and offset). But extinction of CR occurs in the absence of the US, so in that case it is the CS offset which provides the negative change in output which causes the loss of synaptic efficacy, depending on the eligibility curve at that moment. But for very long CSs, the eligibility curve will have approached zero (or some small minimum), and so the decrement in synaptic weight will approach zero, which is to say that the model predicts no (or at least extremely slow) extinction for very long CSs. Sutton and Barto say that "the empirical data currently available do not directly contradict this prediction, but they are not supportive of it" (p. 514).

Nor does the present SDR model account for reacquisition effects or spontaneous recovery (of CR following extinction). The SDR model might be made to accommodate spontaneous recovery by positing (at least for some units, if not for all) a tonic response level, such that in the absence of any conditioning, responses will tend to rise to that level over time, even after being forced below that level by extinction trials. Alternatively, inhibitory synapses might play a role.

There are some kinds of changes in responses which are traditionally discussed in animal learning texts, but which are not always said to participate in learning — or to be a part of the phenomena of associative learning. Muscle fatigue, for example, might be a cause of some behavioral changes during conditioning, but it is not clear that we ought to classify it as a kind of learning, although at the level of the neuron there may be some kind of fatigue as well. (There is, for example, a limiting *rate* at which a reuron can fire; the SDR model acknowledges this by assigning the value of 1.0 to an SDR unit's limit.)

Sensitization is said to occur when an input stimulus of small amplitude produces a greater than normal response — or a response where before there was none. The SDR model does not account for this.

Habituation is said to occur when an organism ceases to respond to a stimulus repeated monotonously — the animal "gets used to it". But it is not always said to participate in associative learning, even though it might be classified as a response undergoing extinction; in fact, it shares some of the features of conditioning, including spontaneous recovery (Bower & Hilgard, 1981). Aparicio & Strong (1992) suggest that habituation is integral to any complete model of Pavlovian conditioning, but, unfortunately, it is too often neglected.

SDR units share some features with perceptrons and other simple learning devices, namely, they are limited in their ability to distinguish input stimuli (and, hence, limited in their ability to selectively respond). One-layer perceptrons, for example, are incapable of solving the exclusive-OR problem. A network with more than one layer presents opportunities for more complex discrimination, provided the credit assignment problem — how to change synaptic weights in an earlier layer so as to produce the required output of a subsequent layer — can be solved (for example, by back propagation).

SDR units, as I have been using them, are one-layer devices, and therefore can be expected to fail exclusive-OR type problems. Consequently, multi-layered networks of SDR units ought to be thoroughly investigated.

I have presented experiments with the SDR model using only positive (excitatory) synaptic weights. Yet the SDR model allows for negative (inhibitory) weights as well. Equation (2) bears repeating here:

$$e_i(t) = \alpha e_i(t-1) + |w_i(t-1)| \min[0, \Delta x_i(t-1)] \qquad (2)$$

In addition, the model specifies that $0 < \text{WMIN} \leq |w_i|$, which is to say that weights do not cross zero; excitatory weights remain excitatory, and inhibitory weights remain inhibitory.

What changes in an SDR unit's behavior will occur if we add an inhibitory synapse for each excitatory synapse? (See figure 15.) Weight increases occur on the rising edge of the output (following a positive change in input). Weight decreases occur only on the falling edge of the output (following a positive change in input). But the falling edge is temporally farther than the rising edge from the positive input change. Consequently, inhibitory synaptic weights will gain efficacy slower than excitatory weights, as shown in figure 16.

Even greater changes in both excitatory and inhibitory synapses are possible if negative USs are allowed. But since the amplitude of a stimulus is a representation of its firing rate, it is not clear how there could be negative rates. So we might suppose instead that there is some level other than zero for inactive USs. Perhaps

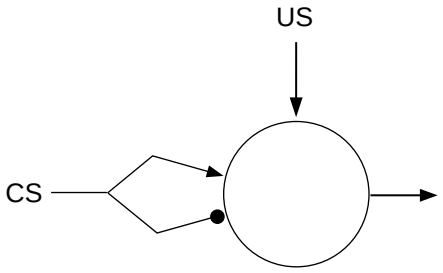

Figure 15
The SDR unit with both excitatory (arrow) and inhibitory (circle) synapses.

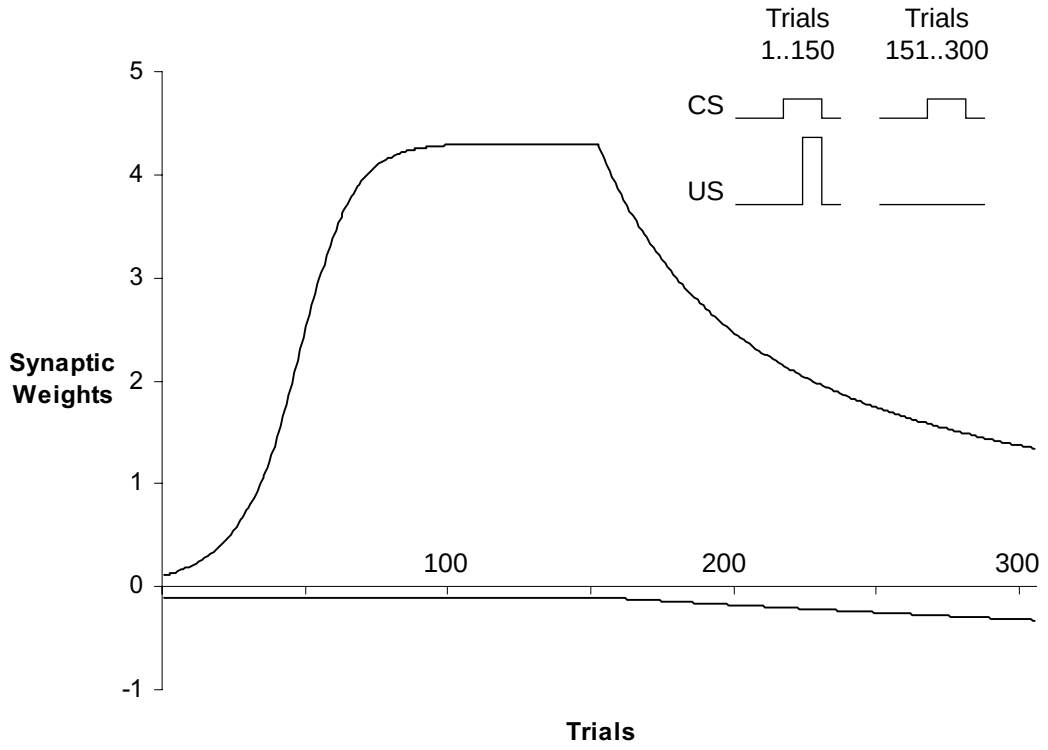

Figure 16
Excitatory and inhibitory synapse weights during acquisition and extinction.

the inactive US value is 0.5, excitatory (reward) USs range from 0.5 to 1, and inhibitory (aversive) USs range from 0 to 0.5. Or, equivalently, let there be a *tonic* level of an SDR unit's activity, such that 0 < tonic < 1, and the unit's output is tonic when there is no input. This would allow for both positive and negative USs.

All the classical conditioning experiments discussed above ought to be repeated using the added inhibitory synapses to make sure that no behavior inconsistent with empirical results occurs.

A threshold is sometimes added to artificial neural net models such that node activation below the threshold results in output of zero. Klopf (1986) explicitly provides for a threshold in his drive-reinforcement model, but then effectively discards it by giving its value as zero in his experiments.

So far, nothing at all has been said of threshold values for the SDR model. Perhaps we might say that we have assumed its value to be zero. But suppose it is greater than zero. Might there be a mechanism for changing a threshold value?

What effect will that have on the classical conditioning experiments? Suppose an input stimulus is below the threshold and therefore produces no output. If subsequently the threshold is lowered, then what before was too weak to evoke a response will now produce a response. This may be part of what occurs in sensitization.

Works Cited

Alkon, D. L. 1983. "Learning in a Marine Snail", *Scientific American*, 249: 70–84.

Alkon, D. L., Quek, F., and Fogl, T. P. 1989. "Computer Modeling of Associative Learning", in D. Touretzky (Ed.), *Advances in Neural Information Processing Systems*, Vol. 1 (San Mateo, CA: Morgan Kaufmann): 419–435.

Aparicio, M., and Strong, P. N. 1992. "Propagation Controls for True Pavlovian Conditioning", in D. S. Levine and S. J. Leven (Eds.), *Motivation, Emotion, and Goal Direction in Neural Networks* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Bitterman, M. E. 1965. "The CS-US Interval in Classical and Avoidance Conditioning", in W. F. Prokasy (Ed.), *Classical Conditioning: A Symposium* (New York: Appleton-Century-Crofts): 1–19.

Bower, G. H., and Hilgard, E. R. 1981. *Theories of Learning*, 5th edn. (Englewood Cliffs, NJ: Prentice-Hall).

Garcia, J., McGowan, B. K., and Green, K. F. 1972. "Biological Constraints on Conditioning", in A. H. Black and W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (New York: Appleton-Century-Crofts): 3–27.

Gormezano, I. 1972. "Investigations of Defense and Reward Conditioning in the Rabbit", in A. H. Black and W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (New York: Appleton-Century-Crofts): 151–181.

Kehoe, E. J. 1990. "Classical Conditioning: Fundamental Issues for Adaptive Network Models", in M. Gabriel and J. W. Moore (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (Cambridge, MA: MIT Press): 389–420.

Kehoe, E. J. 1992. "Versatility in Conditioning: A Layered Network Model", in D. S. Levine and S. J. Leven (Eds.), *Motivation, Emotion, and Goal Direction in Neural Networks* (Hillsdale, NJ: Lawrence Erlbaum Associates): 63–90.

Klopf, A. H. 1982. *The Hedonistic Neuron* (New York: Hemisphere).

Klopf, A. H. 1986. "A Drive-Reinforcement Model of Single Neuron Function: An Alternative to the Hebbian Neuronal Model", in J. S. Denker (Ed.), *Neural Networks for Computing, AIP Conference Proceedings, 151* (New York: American Institute of Physics): 265–270.

Ost, J. W. P., and Lauer, D. W. 1965. "Some Investigations of Classical Salivary Conditioning in the Dog", in W. F. Prokasy (Ed.), *Classical Conditioning: A Symposium* (New York: Appleton-Century-Crofts): 192–207.

Rachlin, H. 1976. *Behavior and Learning* (San Francisco: W. H. Freeman).

Razran, G. 1965. "Empirical Codifications and Specific Theoretical Implications of Compound-Stimulus Conditioning: Perception", in W. F. Prokasy (Ed.), *Classical Conditioning: A Symposium* (New York: Appleton-Century-Crofts).

Rescorla, R. A., and Wagner, A. R. 1972. "A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement", in A. H. Black and W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (New York: Appleton-Century-Crofts): 64–99.

Sutton, R. S., and Barto, A. G. 1981. "Toward a Modern Theory of Adaptive Networks: Expectation and Prediction", *Psychological Review*, 88 (2): 135–170.

Sutton, R. S., and Barto, A. G. 1990. "Time-Derivative Models of Pavlovian Reinforcement", in M. Gabriel and J. W. Moore (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (Cambridge, MA: MIT Press): 497–537.