To appear in D.A. Rosenbaum & C.E. Collyer (Eds.), *Timing of behavior: Neural, computational, and psychological perspectives.* Cambridge, MA: MIT Press (HTTP://www-mitpress.mit.edu)

Predictive Timing Under Temporal Uncertainty: The TD Model of the Conditioned Response

John W. Moore, June-Seek Choi, and Darlene H. Brunzell
Neuroscience and Behavior Program
University of Massachusetts - Amherst

Abstract

Classical conditioning procedures instill knowledge about the temporal relationships between events. Conditioned stimuli (CSs) are regarded as predictive signals and triggers for action. The unconditioned stimulus (US) is the event to be timed. The conditioned response (CR) is viewed as a prediction of the imminence of the US. Knowledge of the elapsed time between CSs and US delivery is expressed in the topological features of the CR. The peak amplitude of the CR typically coincides with the timing of the US. A simple connectionist network based on Sutton and Barto's (1990) Time Derivative (TD) Model of Pavlovian Reinforcement provides a mechanism that can account for and simulate virtually all known aspects of conditioned response timing in a variety of protocols. This chapter describes extensions of the model to predictive timing under temporal uncertainty. The model is expressed in terms of equations that operate in real time according to Hebbian competitive learning rules. The unfolding of time from the onsets and offsets of events such as conditioned stimuli is represented by the propagation of activity along a sequence of time-tagged elements. The model can be aligned with anatomical circuits of the cerebellum and brainstem that are essential for learning and performance of eyeblink conditioned responses. The eyeblink conditioned response is a simple skeletal response in that lid movement can be expressed as a scalar quantity. For the purposes of this chapter, the form of this response is portrayed as an index of an actor's expectation of the timing of the target event. Hence, this chapter uses eyeblink conditioning as a model system for understanding the acquisition and expression of knowledge for timing and action at behavioral, computational, and physiological levels.

Introduction

Imagine a task in which an actor is required to predict not only the timing of some target event but also the degree of certainty or confidence that the prediction is correct. In its

simplest form, the actor is presented with a signal or cue that initiates the timing. This cue is followed some fixed time later by the to-be-predicted (target) event. The actor's prediction of the target at each point in time is expressed as a scalar quantity. Performance is assessed by the degree to which the peak or maximum of the prediction coincides with the onset of the target. The simplest variation of this task is one in which the actor predicts the offset of the timing cue. In this form, the task reduces to duration prediction or judgment. The onset of the timing cue can be regarded as a conditioned stimulus (CS) and its offset can be regarded as the unconditioned stimulus (US). The TD model of the conditioned response (CR), the main focus of this chapter, can be applied to duration prediction by regarding the offset of the timing cue as the target event.

If the actor responds to the timing cue at full strength with the shortest possible reaction time, then performance would be assessed as being rather poor because the prediction lacks precision. Furthermore, the longer the timing interval, the poorer the performance would be. Performance would be assessed more favorably were the actor to withhold its prediction until precisely the time when the target occurs and then respond with a step-impulse of maximum amplitude. However, because of inertia and delays in effector systems, the actor may need to begin responding (predicting) before the target event occurs, so that the prediction response can rise smoothly to a maximum that coincides with onset of the target.

Although too much anticipation can spoil prediction accuracy, some anticipation can be a desirable feature of a timing device or controller. For example, foreknowledge allows an actor to marshall whatever resources are needed for an eminent target event or perturbation. Because the prediction of the moment-by-moment prediction of the target is a scalar quantity, experimental tests of theories of prediction in the domain of action must rely on behaviors such as the eyeblink that can be expressed as a scalar (single degree of freedom) quantity. Because of its rich literature and ties with core issues of learning theory, eyeblink conditioning has become the focus of work on predictive timing and prediction by several groups of investigators (Gabriel & Moore, 1990).

Sutton and Barto (1990) developed a computational model based on Time Derivative (TD) learning that is capable of generating basic features of predictive timing in the context of classical conditioning.[1] This chapter shows how TD learning can be extended to predictive timing under temporal uncertainty. Here, temporal uncertainty refers to procedures in which the timing of the target event, the US, with respect to the CS varies randomly from trial to trial.

As in the case of other models of conditioned response timing (e.g., Desmond & Moore,

[1]TD methods have been used with considerable success in training connectionist networks (e.g., Barto, 1995; Barto, Sutton, & Anderson, 1983; Sutton, 1992).

1988), implementation of the TD models assumes that time is represented as an ordered sequence of time-tagged components or elements. The core assumption is that a CS event initiates a cascade of activation such that one component excites that next, with some delay. The target event might occur at any point in the cascade. When it does occur, a connection is established between elements of the cascade and the target. A representation of the target is evoked the next time these elements are activated. Thus, a nominal CS is simply an event that triggers a cascade of activation among hypothetical components, which are the actual CSs at the level of the nervous system. A neural network representation of time-tagged CS components and their connections to representations of the target and action generators are described later on.

The TD Model of the Conditioned Response

Before stating the TD model in mathematical form, let us introduce some notation and key assumptions.

1. The symbol X refers to the activation level of a cue such as a CS. In the TD model each time step after the initiation of the CS cascade is regarded as a separate stimulus. (The TD model treats time, t, as a discrete variable.) We refer to such stimuli as serial component CSs. In this role (as a time-tagged serial component), X is given a subscript that denotes its ordinal position in the cascade: X_1 denotes the activation of the first serial component CS, X_2 denotes the activation of the second serial component CS, and so forth. On the first time step X_1 has a positive real value (typically assumed to equal 1.0 in simulations), and X_2 equals 0. On the second time step X_2 becomes positive (typically assumed to equal 1), and X_1 becomes 0 because it is no longer active. On the next time step X_3 becomes active, and X_2 reverts to 0.

2. CS triggered cascades, which result in the activation of successive serial components, do not last indefinitely. There is a limit on how many serial components can be activated by a nominal CS event. However, this limit must span at least the time (number of time steps) between the onset of the cascade and the target for any learning to occur.

3. The salience of a CS refers its *associability*, not its ability to control behavior, although they are often correlated. One CS is more salient then another if it lends itself more readily to conditioning. Salience is partially determined by a CS's intensity, but salience can also depend on attentional processes that are themselves subject to cognitive factors such as attention. The Greek letter α, introduced later on, is a parameter that specifies a CS's salience. The variant of the TD model presented here assumes that all serial component CSs of a cascade are equally salient.

4. The symbol $\overline{X}_i$ represents the decaying trace of the ith serial component CS. This

trace is referred to as the CS's *eligibility* for changing the weight of its connection to the target event. $\overline{X}_i$ is not a stimulating trace because it does not evoke a response. Its role is to allow a serial component CS to participate in computations of connection weights (Vs) even though it is no longer active. (See Sutton & Barto, 1981, 1990; Desmond, 1990).

5. The symbol V_i (associative value) refers to the strength or weight of the connection between the ith serial component CS and the target, but it is equally correct to think of it as the connection between the CS and the conditioned (prediction) response.

6. Because V_i is a function of time, the notation $V_i(t)$ denotes the value of V_i at time t, where t denotes the time step after onset of the CS cascade.

7. The symbol $Y(t)$ refers to the strength or magnitude of the response at time t. In conditioning, Y reflects the effects of contiguous pairing of a CS and US. Therefore, $Y(t)$ is equal to $X(t) \times V(t)$. In the TD model, $V(t)$ is estimated by $V(t-1)$ because $V(t)$ is not immediately available, and $X(t) \times V(t-1)$ is therefore an estimate of $Y(t)$. This estimate is attenuated (discounted) by the factor γ, a parameter of the model.

8. The symbol $\dot{Y}$ (Y-dot), the first time derivative of Y, represents the change in the value of Y from one time step or computational epoch to the next: $\dot{Y} = Y(t) - Y(t-1)$.

9. The Greek letter Δ refers to a change in a variable such as V from one time-step to the next. Hence, $\Delta V_i(t)$ refers to the change in the connection between the ith serial component CS and the target event computed on time step t. Computation of $\Delta V_i(t)$ requires $\dot{Y}(t)$. Because the TD model treats time as discrete, $\dot{Y}$ is computed using connection weights from the preceding time step $t-1$. This is made explicit in the formal statement of the model given below.

10. The Greek letters α, β, γ, are δ are parameters in equations describing the TD model.

Formal Statement of the TD Model

The TD model is a member of a class of computational models that Sutton and Barto (1990) refer to as $\dot{Y}$ or time-derivative theories of reinforcement learning. Such theories take the form of Equation 1, which specifies the moment-by-moment changes in associative value, V_i, the weight of the connection between CS_i, the ith of a potential set of CSs, and the US.

$$\Delta V_i = \beta \dot{Y} \times \alpha_i \overline{X}_i \tag{1}$$

As with Hebbian learning rules generally, changes in associative value, ΔV_i for CS_i, are computed as the product of two factors. The coefficents α_i and β are rate parameters ($0 < \alpha_i, \beta \leq 1$). α_i is the salience of the ith serial component CS. The factor $\overline{X}_i$ represents the eligibility of the connection between CS_i and the US for modification. Eligibility is a weighted average of previous and current levels of activation of CS_i. The other factor, $\dot{Y}$, represents *reinforcement*. Reinforcement in time-derivative models is a function of the difference (time derivative) between the response or output at time t, $Y(t)$, and the response or output at some previous time, $Y(t - \Delta t)$ (Equation 2).

$$\dot{Y} = Y(t) - Y(t - \Delta t) \tag{2}$$

With time treated as a discrete variable, $\dot{Y} = Y(t) - Y(t - 1)$, as noted above.

Any system or device that would implement a time-derivative learning rule must be capable of monitoring the actor's output on both current and immediately preceding time steps. Later on, we discuss how this might be accomplished within the cerebellum, the putative site of learning in classical eyeblink conditioning.

Sutton and Barto (1990) review evidence that the TD model is superior to other computational theories of classical conditioning because it encompasses problematic phenomena such as the form of CS–US interval functions and higher-order conditioning. CS–US interval functions are empirically derived relationships between the efficacy of conditioning and the CS–US interval. Higher-order conditioning refers to conditioning derived from the pairing a novel stimulus with a previously established CS. In addition to overcoming these shortcomings of other models, the TD model can describe the appropriate timing and topography of eyeblink CRs, provided the CS–US interval is segmented into a sequence of time-tagged elements. Each of these elements develops its own associative value with training. Sutton and Barto (1990) have referred to this representation of the CS as a complete serial compound (CSC).[2]

The following equation expresses the TD learning rule for classical conditioning.

[2]This representation resembles the approach to conditioned response timing and topography employed by Desmond and Moore's VET model (Desmond, 1990; Desmond & Moore, 1988, 1991a, 1991b; Moore, 1991, 1992; Moore & Desmond, 1992; Moore, Desmond, & Berthier, 1989). VET is an acronym for the process of mapping associative *value* onto action based on *expectancies* about *timing*. The main advantage of the TD model over the VET model is that it generates higher-order associative connections, such as those underlying secondary reinforcement, a feature of $\dot{Y}$ models that is lacking in the VET model. The timing structure employed in the VET model is here extended to the TD model.

$$\Delta V_i(t) \;=\; \beta[\lambda(t) + \gamma Y(t) - Y(t-1)] \times \alpha \overline{X}_i(t) \tag{3}$$

where

$$Y(t) \;=\; \sum_j V_j(t) X_j(t) \tag{4}$$

.

In Equation 3, the Y-dot factor of Equation 2 becomes $\lambda(t) + \gamma Y(t) - Y(t-1)$. $\lambda(t)$ represents the strength of the US at time t.

α and β are rate parameters, as in Equation 1. Notice that we have dropped the subscript from α, so that $\alpha_i = \alpha$ for all serial component CSs.

In Equation 4, the subscript j includes all serial CS components, and $X_j(t)$ indicates the on-off status of the jth component at time t. $Y(t)$ corresponds to CR amplitude at time t.

$\overline{X}_i(t)$ is the eligibility of the ith CS component for modification at time t, given by the following expression.

$$\overline{X}_i(t+1) \;=\; \overline{X}_i(t) + \delta[X_i(t) - \overline{X}_i(t)] \tag{5}$$

where $0 < \delta \leq 1$.

A key feature of the TD model is the parameter γ ($0 < \gamma \leq 1$). This parameter determines the rate of increase of CR amplitude, $Y(t)$, as the US becomes increasingly imminent over the CS–US interval. With the CSC representation of CSs, the TD model generates realistic portraits of CRs as they unfold in time. Realistic CRs resemble goal gradients in that CR amplitude increases progressively to the onset of the US. The behavior of the model with variations in γ is illustrated in Figure 1.

γ is referred to as the *discount* parameter in applications of the TD learning rule. γ is applied to $Y(t)$ in Equation 3 because $Y(t)$ is actually an estimate or prediction. On time step t, $Y(t)$ is not known with certainty until after the fact. However, $Y(t)$ can be estimated by the sum of products of the form $X_j(t) \times V_j(t-1)$. That is, the connection weights from the preceding time step are used to estimate the value of $V(t)$ for the current time step. γ can be regarded as the penalty for using $V_j(t-1)$ instead of the true connection weight, $V_j(t)$. To reiterate the point, $V_j(t)$ does not become known until time step $t+1$.

Basically, Equation 3 says that the connection between the ith serial component CS and the US is modified to the extent that there exists a discrepancy (algebraic difference) between the value of the US, $\lambda(t)$, plus the predicted output $Y(t)$ discounted or attenuated by γ, and the output on the preceding time step, $Y(t-1)$. Equation 4 states that output

is the algebraic sum of weighted input. Equation 5 states that $\overline{X}$ declines geometrically at a rate determined by δ. The larger the value of δ, the faster the decline. A low value of δ implies that connections between serial component CSs and the US remain eligible for modification for several time steps.

Simulation Form of the TD Model

For technical reasons, Equations 3–5 are not readily applicable for simulations. For simulations, it is necessary to decompose $Y(t)$ into its constituent parts, $X(t)$ and $V(t)$. Although it is straightforward to substitute $X(t-1) \times V(t-1)$ for $Y(t-1)$, for the preceding time step, for the current time step (t) the simulation must use the value of V from the preceding time step. As discussed above, $V(t-1)$ is multiplied by $X(t)$ in order to obtain $Y(t)$. This makes sense because, as a conditioned response, $Y(t)$ reflects prior experience. It is also essential (to ensure convergence) that the sums of products of X and V not be negative. The notation $\lfloor \sum XV \rfloor$ specifies this constraint: If this quantity is less than 0, it is set equal to 0. (See Sutton & Barto, 1990, p 533). These two constraints have been incorporated into Equation 6, the simulation form of Equation 3.

$$\Delta V_i(t) = \beta\{\lambda(t) + \gamma\lfloor\sum_j X_j(t)V_j(t-1)\rfloor - \lfloor\sum_j X_j(t-1)V_j(t-1)\rfloor\} \times \alpha\overline{X}_i(t) \qquad (6)$$

Equation 6 emphasizes the fact that changes in the strength of connections between serial components CS and the US are functions of their level of activation.

Simulations of CR Timing by the TD Model

Figure 1 shows a family of asymptotic CR waveforms with different values of γ and δ. The figure shows that CR topography depends primarily on γ: The smaller the value of γ, the lower the peak value of CR amplitude, $Y(t)$. Lower values of γ also increase the positive acceleration of CR amplitude, $\Delta\dot{Y}(t)$, without compromising the accuracy of $Y(t)$'s prediction of the timing of the US.

In simulations, $X_j(t)$ takes on values of 1 (when activated) or 0 (when not activated), with activation lasting for one time-step (10 milliseconds in the simulations). Under these circumstances response topography depends only on the model's other parameters and on constraints on the effector system.[3] In the case of classically conditioned eyelid movements,

[3]In general, CR topography depends on the physical characteristics of CSs and their serial components. These characteristics, such as acoustic frequency and intensity, can be captured by the variables $X_j(t)$ in Equations 4-6, as suggested by Kehoe, Schreurs, Macrae, & Gormezano (1995), but these complexities are suppressed here.

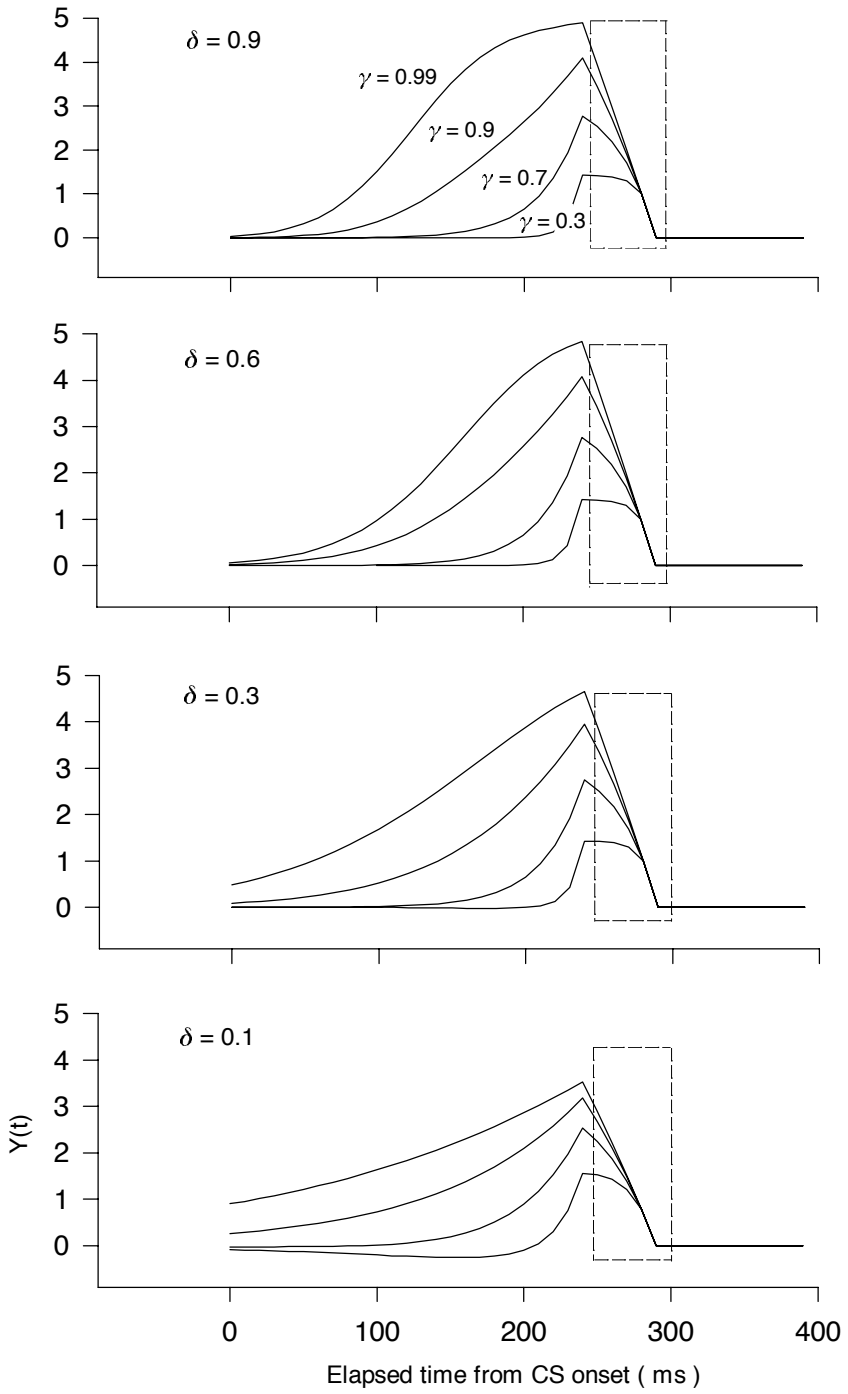

Figure 1 Simulated CRs, Y(t), after 200 trials as a function of γ and δ. Time-steps in this and other simulations are 10 ms, α = 0.05, β = 1.0, and λ = 1.0. The rectangle in each panel indicates the duration of the US, which is 50 ms. Note that CR timing and amplitude are determined primarily of the discount factor γ.

the eyelids are normally open. In this position, CR amplitude has a value of 0. A fully developed CR is one in which the eyelid's position moves from open to completely closed. Yet, no matter how strong the prediction that the US will occur, the eyelids can only close so far. This constraint implies that the progressive closure of the lids in the course of CR production can saturate before the US's anticipated time of occurrence. In addition, these constraints on eyelid position render it impossible for negative predictions of the US to be expressed directly in eyelid movement, predictions that the US will not occur at some time when it would otherwise be anticipated. No matter how strong the prediction that the US will not occur, the eyelids can only open so far and no farther.

Simulating Predictive Timing Under Uncertainty

Consider now a variation of the predictive timing task discussed in the Introduction in which the timing of the target event varies from one occasion or trial to the next. Here, a trial is the initiation of a CS cascade, and uncertainty arises from random variation in the timing of the target event, the US. That is, target timing from one trial to the next is a random variable.

Predictions of target timing, reflected in CR topography, will clearly be affected by the probability distribution of target times. Simulations were confined to the case where the target occurs at one of three times after onset of the CS. In order to connect with data presented in the next section, these times are 300, 500, and 700 milliseconds, each CS–US interval occurring equally often and randomly over a sequence of trials.

Prediction Strategies and CR Topography

There are a number of prediction strategies that might be adopted by an actor. Predictions generated by TD learning are based on the actor's experience with the timing cue and target. However, the actor might expect that target timing can change abruptly and without warning. The actor might then generate predictions based on the possibility that the target will occur at some unprecedented time or that the target will occur more than once. Although exogenous to the model, processes that produce such expectancies could alter the parameters of the model or the structure of timing. With the appropriate parametric alterations or timing structure, the TD model can simulate any of a number of prediction strategies.

The TD model does not state *how* a given prediction strategy is selected.[4] Nor does the model specify *why* a strategy is selected, although its consequences (values and costs) would

[4]We use the expression 'strategy selection' to facilitate communication. We do not assume, nor do we rule out, the contribution of cognitive processes. Strategies may reflect high level top-down control, low level bottom-up control, or a combination of both.

be relevant considerations. These topics are beyond the scope of this chapter. The TD model provides the means or mechanisms for *implementing* a strategy through a combination of parameter setting and structuring of the timing mechanism.

Fail-Safe Predicting

One prediction strategy would be to respond with a maximal prediction response before the first possible target time and to sustain this prediction until the target event occurs or the trial terminates. This is a *fail-safe* strategy in that the actor's response is appropriate for all possible times that the target *might* occur. In the example, the prediction response would begin before the earliest possible target time, 300 milliseconds in the example. It would then plateau to a maximum at that point in time, as illustrated in Figure 2. The TD model simulates this type of response by assuming that the target event, the US, occupies every time-step between 300 ms and some later time-step beyond the maximum possible target time of 700 milliseconds.

A fail-safe strategy makes sense if there is a lot of inertial or resistive force to be overcome by the effector system in moving from its starting position to its maximum displacement and back again. Fail-safe prediction also makes sense if the costs of error are low or if the actor believes that the past is not a reliable guide to the present, i.e., that target timing or number of occurrences per trial might change without warning. Fail-safe prediction involves considerable error, as the target is predicted with minimal precision.

Hedging

A second strategy would have the actor respond maximally to each of the three possible target times it has experienced on previous occasions, as illustrated in Figure 3. The TD model simulates this type of response through any of a variety of parametric adjustments that produce large-amplitude waveforms, such as inflating the value of λ or selecting a large value of the discount factor, γ (see Figure 1).

This is a *hedging* strategy in that the actor's predictions are appropriate for the possibility that the target might appear at any one, or indeed all, of the three target times. Instead of making one prediction, the actor makes three, one for each target time. Unlike a fail-safe strategy, a hedging strategy ensures that the target event is predicted with accuracy (a 'hit' in the vernacular of signal detection theory), but at the cost of 'false alarms.' Hedging makes sense if the costs of moving from baseline and back again are negligible and the penalties for false alarms are low.

Proportionate Hedging

A third strategy would be to partition costs of predictions equally among the three

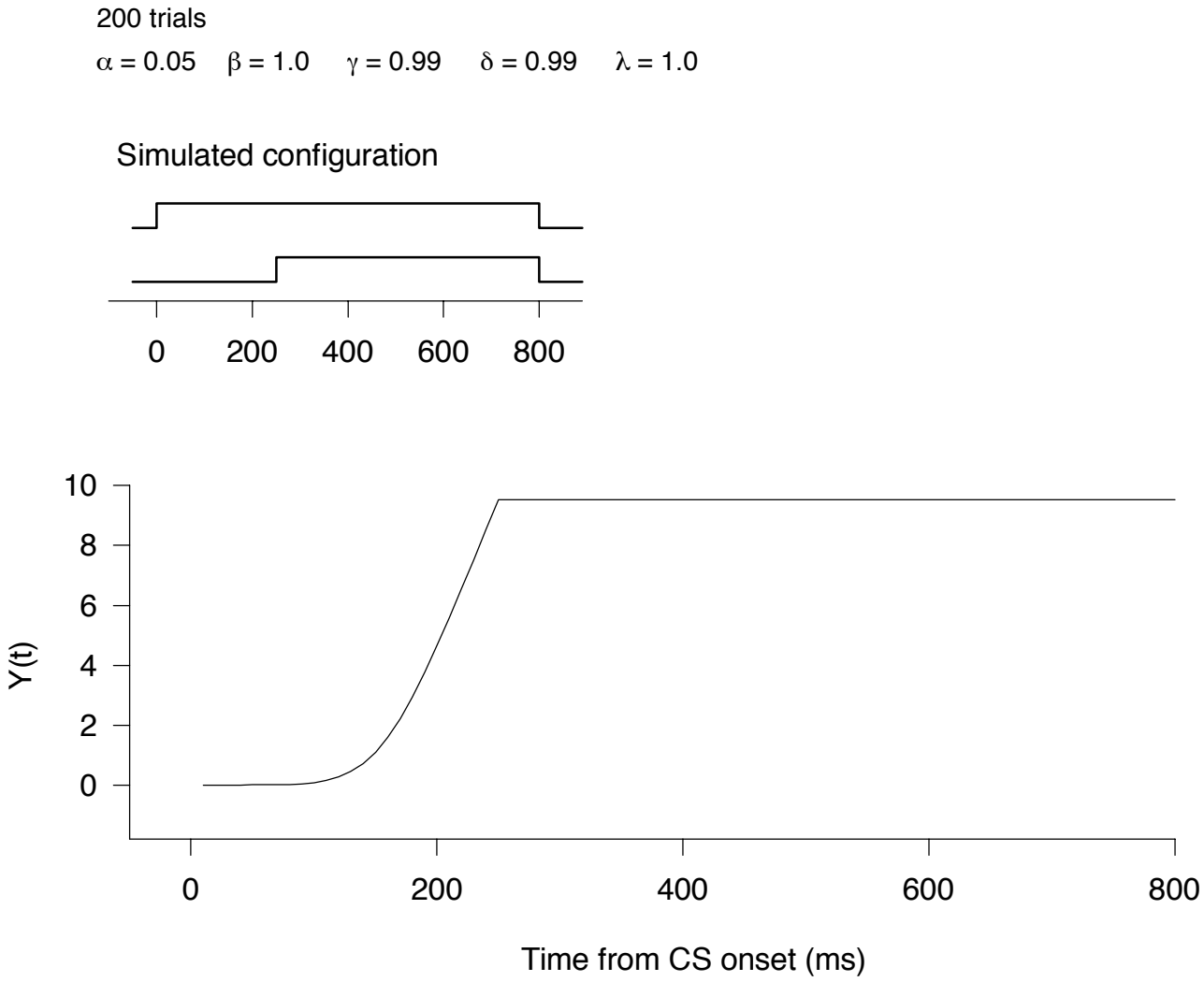

Figure 2 Simulated fail-safe prediction response under temporal uncertainty.

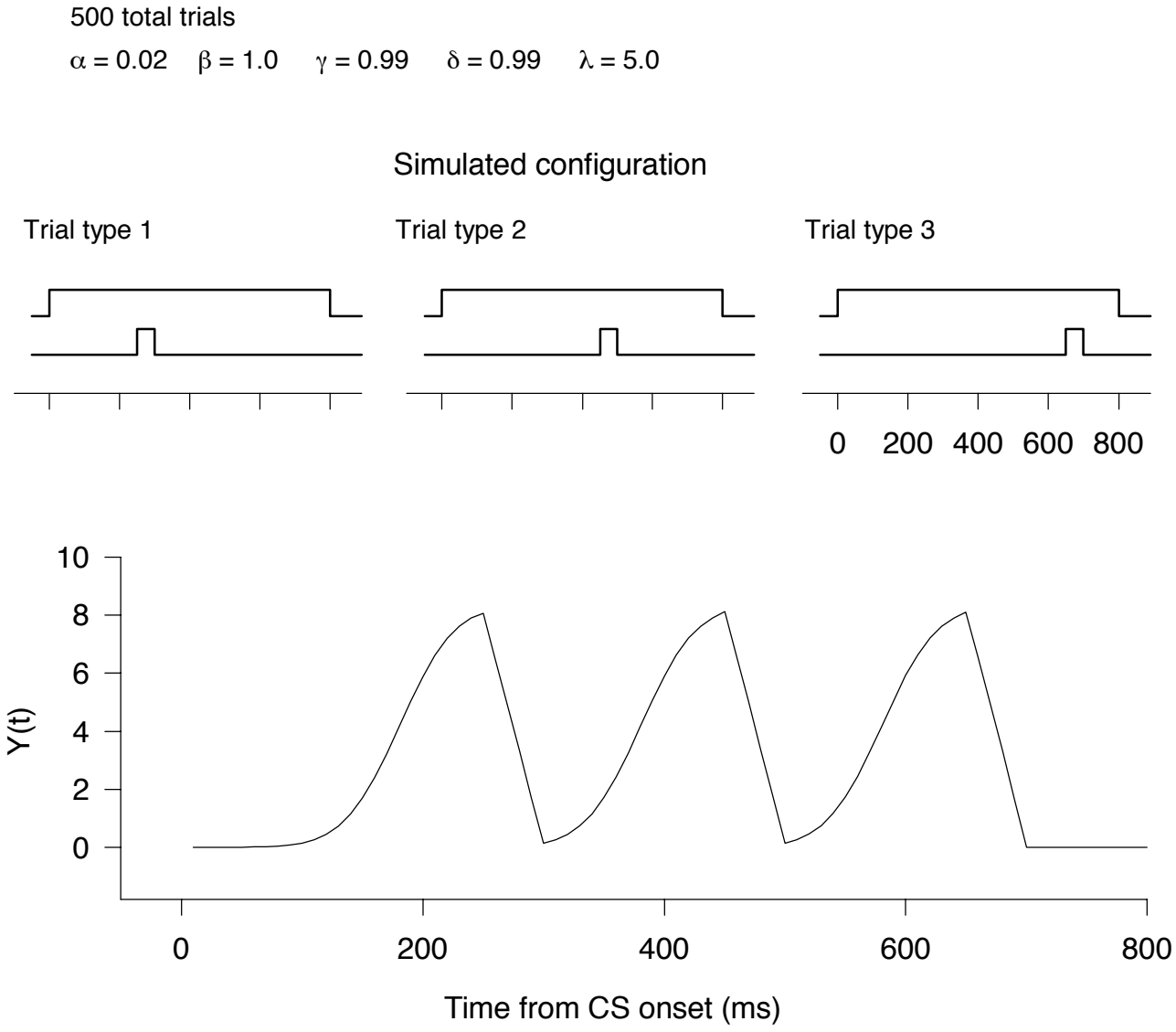

Figure 3 Simulated hedging prediction under temporal uncertainty.

target times, as with the hedging strategy, but with less than maximal amplitude. This is illustrated in Figure 4. The TD model simulates this prediction strategy without parametric adjustments.

This strategy is *proportionate hedging* because the actor's predictions are proportionate to the percentage of times the target occurs at each of the three CS–US intervals. This strategy makes sense when the costs of full-fledged false alarms outweigh the benefits of hits.

Conditional Expectation

A fourth strategy would be one based on conditional expectations of target timing, as illustrated in Figure 5. The TD model simulates this strategy only with additional assumptions about the structure of timing. These additional assumptions are spelled out later on.

Using a *conditional expectation* strategy, the actor's predictions reflect the fact that the probability of the target occurring at later times increases if it has not occurred at earlier times. This strategy makes sense if both the costs of false alarms and the benefits of hits are high. The conditional expectation strategy is a reasoned compromise between the hedging and proportionate hedging strategies.

All of the aforementioned prediction strategies are feedforward actions triggered by the CS. As shown by Figures 2-5, they can all be generated by the TD model of the CR, using different parameters or assumptions about the structure of timing. The next section presents data bearing on predictive timing strategies expressed by CR topography.

CR Topography Under Predictive Uncertainty

We trained eight rabbits to make conditioned eyelid movements under temporal uncertainty. The CS was a compound stimulus consisting of a tone and light. It lasted 800 milliseconds. The US was a 50-millisecond train of pulses from a dc source applied to the periocular region of the right eye by steel sutures. The US was applied at either 300, 500, or 700 milliseconds following CS onset. Because the time of the US was selected randomly, its timing from trial to trial was uncertain. Rabbits were not cued as to whether the US would occur at 300, 500, or 700 milliseconds. The interval between CS presentations was 30-40 seconds, and there were 60 such trials per day for 10 days. Movements of the right superior eyelid were recorded with a low torque potentiometer.

We examined the waveforms depicting eyelid movement for each rabbit for the subset of trials on which the US occurred at 700 milliseconds. These trials provided waveforms that covered the entire interval of uncertainty without contamination by the reflexive response to the US occurring earlier in the interval.

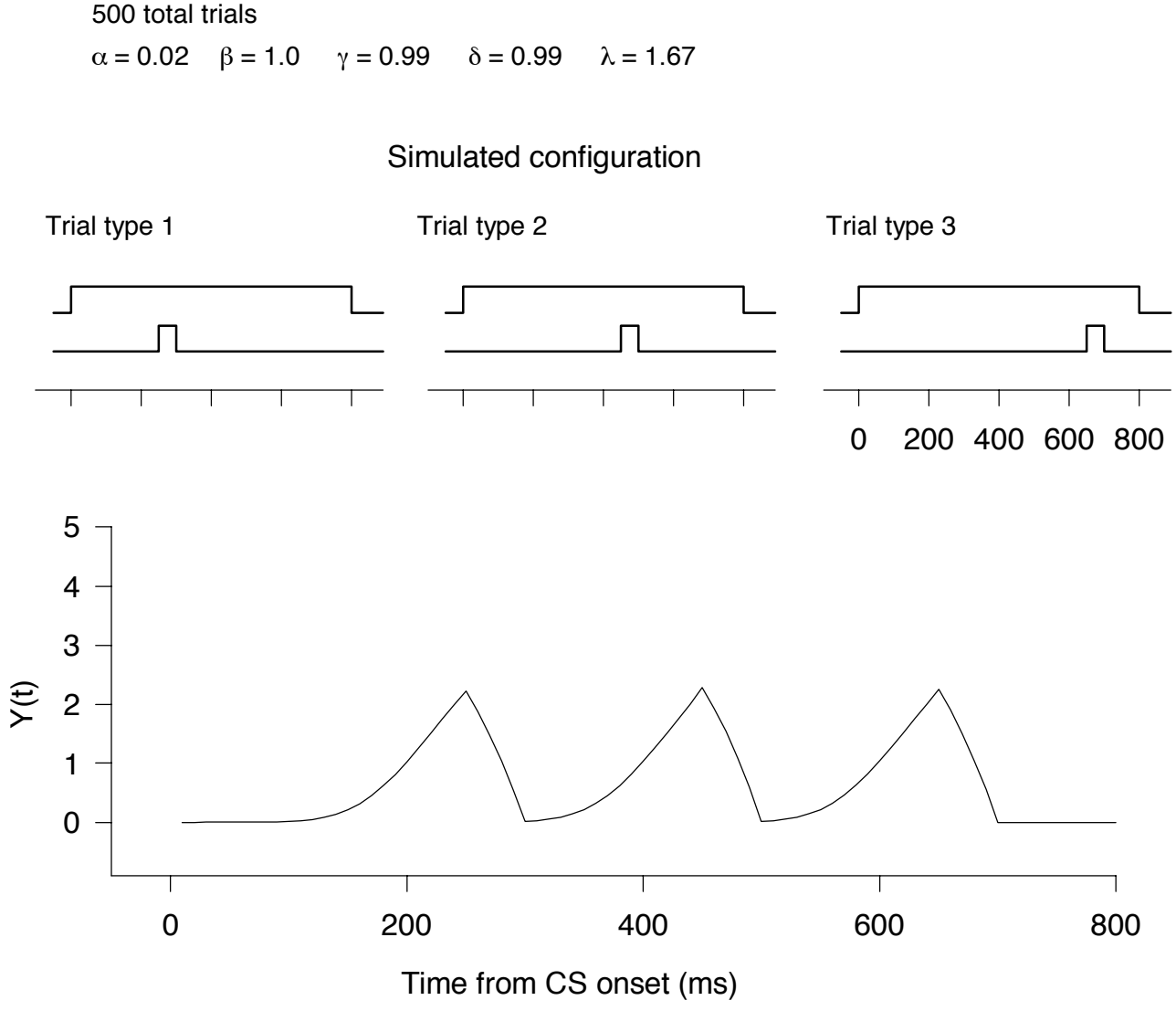

Figure 4 Simulated proportionate hedging prediction under temporal uncertainty.

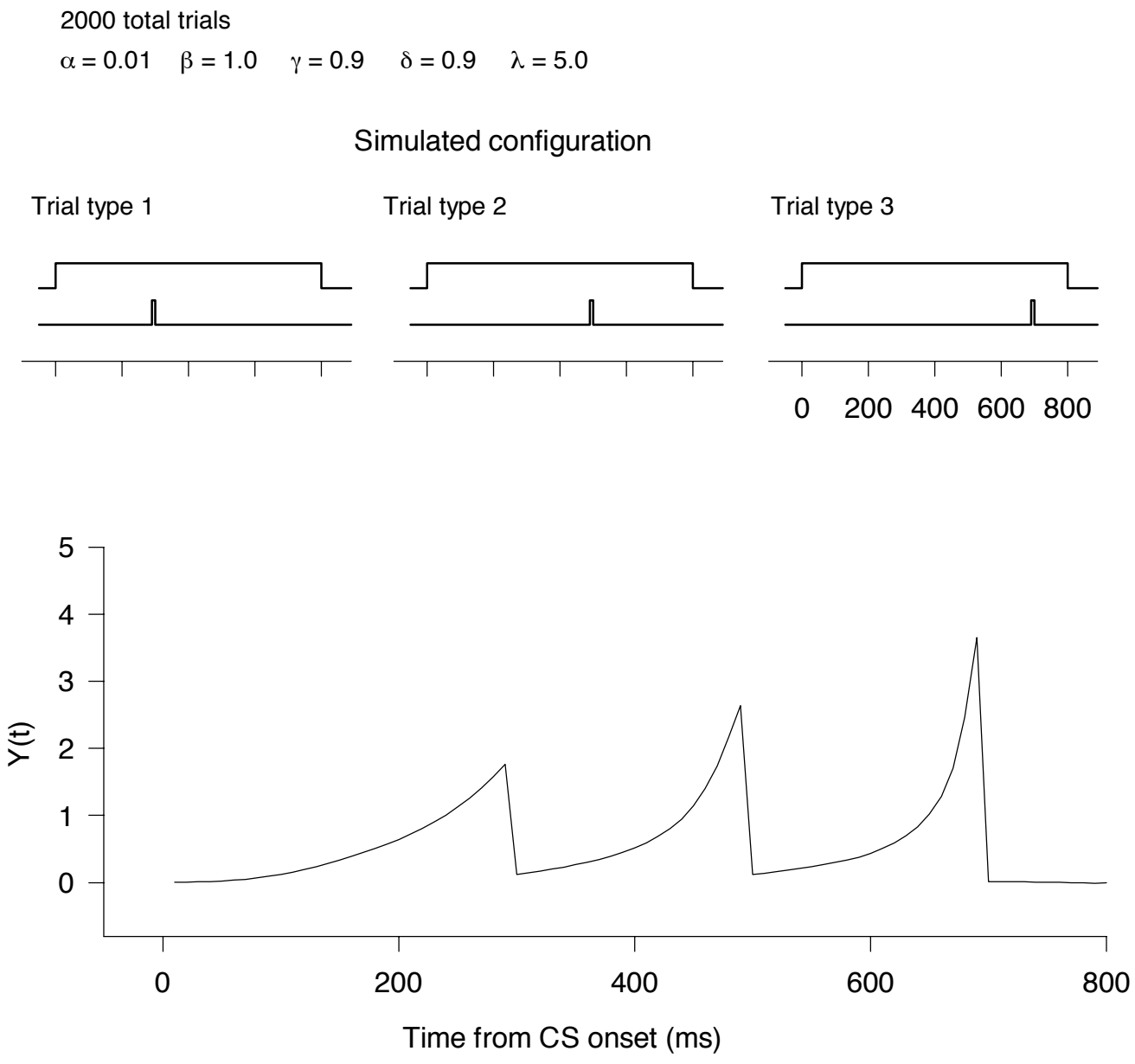

FIGURE 5 Simulated conditional expectation prediction under temporal uncertainty.

The data revealed that half the animals ($n = 4$) developed a fail-safe strategy and the remainder ($n = 4$) developed a conditional expectation strategy. Response patterns that reflect hedging strategies were not observed. (We have observed hedging when the number of possible times of US occurrences number two instead of three and the separation is 400 milliseconds instead of the 200 millisecond separation employed in this study.)

Figure 6 shows a single trial record from a typical fail-safe responder (Subject 11, Day 4, Trail 14). The CR begins just after 200 milliseconds has elapsed from CS onset and peaks quickly at 300 milliseconds. This peak is largely sustained through the remaining duration of the trial.[5]

Figure 7 shows a single trial record from an animal employing a conditional expectation strategy (Subject 3, Day 4, Trial 45). Although the timing of the succession of peaks of increasing amplitude does not match the actual times that the US might occur, one nevertheless gets the impression that the increasing likelihood of the eye shock as the CS unfolds controls this animal's CR topography.

By the end of training, all eight rabbits respond with a fail-safe strategy, likely indicating saturation of the effector system at all serial component CSs.

The conditional expectation strategy is interesting because it contradicts expectations from experiments in which the CS–US interval remains constant throughout training. Such studies show that relatively brief intervals are more favorable for conditioning those of greater duration (Desmond, 1990; Desmond & Moore, 1988; Sutton & Barto, 1990). For the rabbit eyeblink preparation, this 'optimal interval' is on the order of 250 milliseconds, and declines progressively with longer intervals. From this perspective, response peaks should first appear at 300, then 500, and finally 700 milliseconds. The reversal of this sequence implied by the conditional expectation strategy was therefore a surprizing observation.

The strategies for predictive timing and CR topography discussed above and illustrated in Figures 2-7 have a clear cognitive flavor, in that they can be implemented as rule-based actions. For example, the fail-safe strategy of conditioned eyelid movement follows from the rule: *Close eye quickly and keep it closed until probability of US is minimal.* Indeed, the response topography shown in Figure 6 is what researchers in this field traditionally call a voluntary response (Coleman & Webster, 1988). However, it is equally clear that this strategy can also be implemented by low-level processes involving a combination of interval-optimality and generalization to other intervals sufficient to produce saturation of the effector system. Similarly, the conditional expectation strategy follows from the rule:

[5]There is no special significance in the fact that the maximum eyelid movement in Figure 6 is just over 2 millimeters, as rabbits vary in the maximum blink they produce.

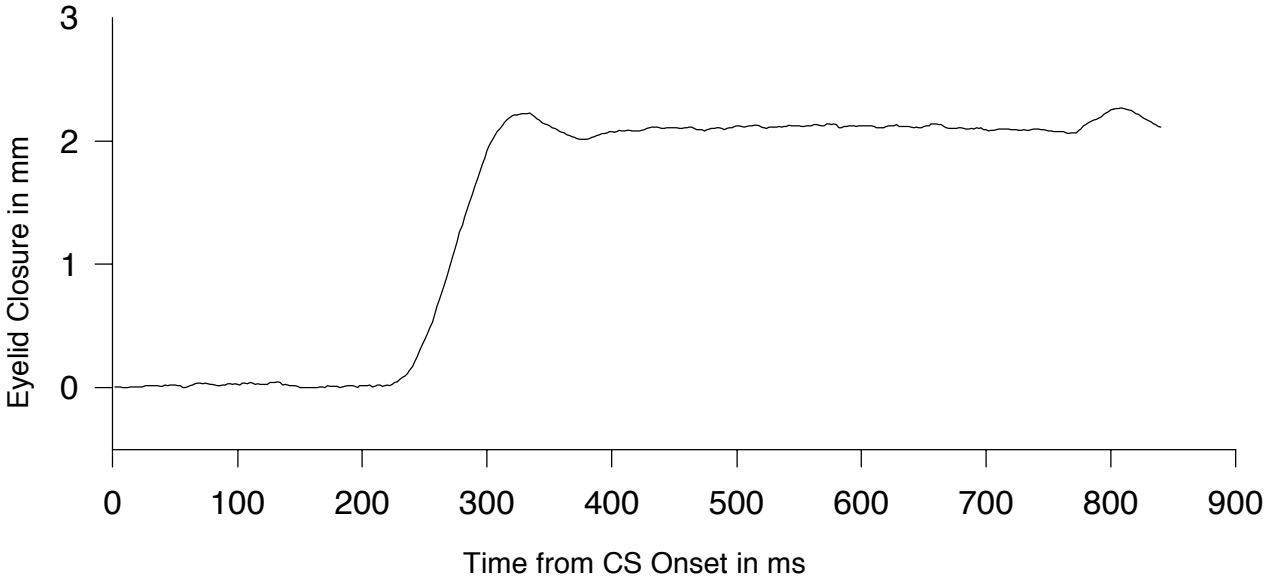

Figure 6 Eyeblink conditioned response consistent with a fail-safe prediction strategy.

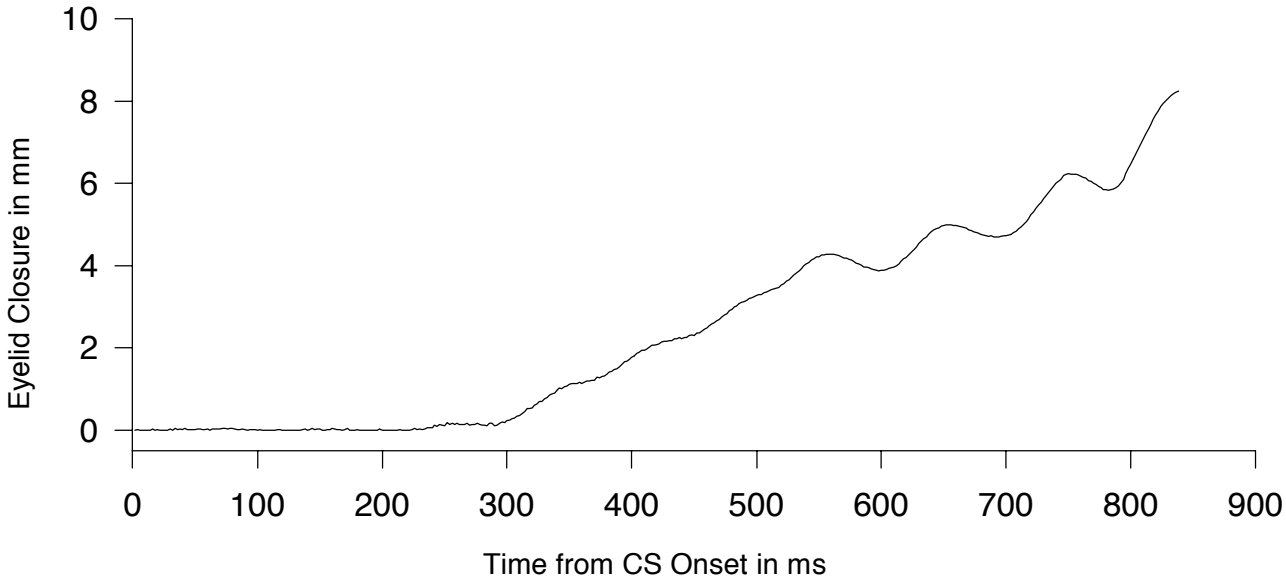

Figure 7 Eyeblink conditioned response consistent with a conditional expectation strategy.

Close eye progressively as the conditional probability of the US increases to maximum. Is there a low-level mechanism consistent with the TD model for implementing this rule? The next section develops ideas that underlie such a mechanism.

Structure of Timing

This section reviews assumptions about the structure of timing that underlie application of the TD model to predictive uncertainty. This structure allows the model to simulate the conditional expectation strategy.

1. Elapsed time (duration) from the onset of a timing cue or CS can be segmented into a sequence of time-tagged elements or units. Above, we referred to these units as serial components, in discussing the TD model. Let CS_{on_i} be the ith element of the onset-triggered timing cascade. The subscript *on* distinguishes this element from the ith element of an offset-triggered cascade, discussed below. These time-tagged units, when activated, provide input to effector systems that generate timing predictions. The TD model assumes that connections between these elements and the effector systems are modifiable, according to the basic equations of TD learning.

2. The TD model does not specify the real-time that elapses between activation of successive elements of a timing cascade. The 'temporal grain' or segmentation rate is simply a parameter of the simulation. The segmentation rate can depend on processes outside the learning device that depend on the state of the system. Thus, the TD model does not specify the number of time-tagged elements that constitute a given elapsed time. Nor does it specify the speed of propagated activation along the cascade. These factors have testable consequences, but the model does not explicitly take them into account. In order to implement the model, the learning device need only differentiate between its output at two successive time steps (Equation 2).

3. Stimulus offset can trigger a timing cascade. The offset cascade involves *different elements* from those involved in an onset cascade. Introducing notation, CS_{off_i} designates the ith element of an offset cascade, which distinguishes it from the ith element of the corresponding onset cascade, CS_{on_i}. The two cascades are otherwise independent. The output of offset elements summate with onset elements to determine the magnitude of the actor's prediction responses. It is important to appreciate that onset cascades can persist after offset of the triggering stimulus. Desmond and Moore (1991a) demonstrated the validity of this assumption, in a study of classically conditioned eyeblinks in rabbits following a trace conditioning procedure.

4. Salient events such as CS onsets and offsets initiate (trigger) cascades of activation, thereby marking the onset of the to-be-timed interval. The likelihood and degree of

activation depend on a host of variables, including physical and psychophysical factors and the state of the actor. The question of what constitutes a salient event can be finessed for the time being. A salient event is anything that triggers timing, but it might also be aligned with notions of detectable signals, affordances, and invariances among temporal and spatial components. The degree to which a cue or complex of cues is adequate for timing is basically an empirical question that can be assessed in the actor's performance.

5. Timing elements are activated sequentially, such that an element active at time t will activate its next-in-line neighbor at some later time $t+1$. This is the *delay line* assumption, but the notion can be extended to arrays (Desmond, 1990). If these elements are neurons (or strings of neurons), then the delay between activation of element CS_i and element CS_{i+1} represents neuronal processes of recruitment, propagation, and transmission. These processes run forward in time, which simply means that activation of element CS_i can directly activate element CS_{i+1} (with a delay) but not element CS_{i-1}.

6. Different timing cues or CSs are capable of triggering different onset and offset cascades. These timing cascades summate to determine the amplitude of the actor's prediction response. There is good evidence that this summation occurs in rabbit eyeblink conditioning (e.g., Kehoe, Graham-Clarke, & Schreurs, 1989).[6]

7. Any element in an onset-triggered timing cascade can potentially initiate another timing cascade through a branching process we shall refer to as *marking*. Branching (or marking) can occur at any point along a timing cascade. Thus, activation of the ith element of a cascade might trigger the activation of another, independent cascade, as well as activation of the next-in-line element of the original cascade. By incorporating *marking*, as in marking time, the TD model can generate the conditional expectation strategy simulated in Figure 5. The next section develops this idea further.

Marking

Marking is a mechanism that can initiate a separate cascade of serial component CSs that parallels the cascade triggered by the nominal CS or its offset. The basic idea is that serial components that have been active at the same time the US occurs can acquire the ability to evoke an additional set of serial component CSs that contribute to the form of the prediction response.

[6]Although this assumption does not imply an unmanageable 'combinatorial explosion' in complexity, it again begs the question of what processes contribute to stimulus salience and how stimulus components cohere to become timing cues.

Suppressing the distinction between onset and offset cascades for the time being, it should be clear at this juncture that a timing cue or CS triggers a cascade by activating element CS_1, the first element of the cascade. Its level of activation, X_1, increases within a computational cycle or time-step from some below-threshold value (typically 0 in simulations) to an above-threshold value (typically 1.0 in simulations). Activation of CS_1 activates the next element in the sequence CS_2, with some delay equal to the assumed temporal grain or segmentation rate. In marking, we assume that activation of any element of the cascade, CS_i, can *in principle* cause the activation of another cascade that proceeds from that point in time in parallel with activation of the remaining elements of the original cascade of serial component CSs.

Under what circumstances does marking occur? We suggest that marking is an acquired property of the timing structure that follows a simple *delta* learning rule. Let μ_i denote the weight of a connection from CS_i to M_{i_1}, the first element of the marking cascade activated by CS_i. Increments in μ_i are given by the following equation:

$$\Delta\mu_i = \eta(\lambda - \mu_i) \times X_i \tag{7}$$

where $0 < \eta \leq 1.0$ is a rate parameter. μ_i is the likelihood (up to the maximum value, $\lambda \leq 1.0$) that activation of element CS_i will trigger a marking cascade. X_i denotes the level of activation of the marker, CS_i. λ, donated by the target or US, is the asymptotic likelihood of marking. In sum, marking develops whenever the target event occurs contiguously with activation of an element. If the target is a US, then the likelihood of marking increases asymptotically to λ as a function of repeated occurrences of the US that are contiguous with activation of the marker.

It is necessary to make explicit what happens when element CS_i is activated but the target or US does not occur contiguously. In keeping with assumptions of related delta learning rules, we shall assume that this causes a down-regulation of μ_i according to the following equation.

$$\Delta\mu_i = -\theta\mu_i \times X_i \tag{8}$$

where $0 < \theta \ll \eta$.

Let us now consider what happens when more than one element of a timing cascade branches through marking. We assume that each branch consists of a sequence of unique elements. That is, the element M_{i_k} is not the same as the element M_{k_i}. The former is the kth element of a marking cascade triggered by CS_i; the latter is the ith element of a

separate marking cascade triggered by CS_k. The output of these elements, M_{i_k} and M_{k_i}, summates with elements of the original CS-onset and CS-offset cascades to determine the actor's response. Extending our notation further, we allow offset elements as well as onset elements to participate in marking. CS_{on_i} denotes the ith element of an onset-triggered cascade. CS_{off_j} denotes the jth element of an offset-triggered cascade. $M_{on_{i_k}}$ denotes the kth element of a marking cascade triggered by the ith element of an onset cascade and $M_{off_{k_l}}$ denote the lth element of a marking cascade triggered by kth element of an offset cascade.

Let us now consider the implications of the marking mechanism for predictive timing under temporal uncertainty. In our example, we imagine a timing cue or CS that signals that the target or US might occur at any one of three intervals, 300, 500, or 700 milliseconds after onset of the cue. Whenever the target occurs at 300 milliseconds, the contiguously activated element, here designated CS_{300}, acquires some capacity to trigger a marking cascade. With a sufficient number of such coincidental events, the likelihood that this will happen saturates (asymptotically) to the value λ, the magnitude or intensity of the target, i.e., $\mu_{300} \doteq \lambda$. The first element of the marking cascade is designated M_{300_1} using notation developed in the preceding paragraph. When the target or US occurs at 500 milliseconds, another marking cascade develops. This one is triggered by CS_{500} and the first element of its marking cascade is M_{500_1}.

To mitigate combinatorial explosion, we assume that marking cascades can only branch from elements of CS-triggered cascades. That is, CS_{500} can initiate marking when the target occurs contiguously with its activation, but element $M_{300_{200}}$, which can also be contiguously activated when the target occurs at 500 milliseconds, does not spawn marking. Furthermore, elements of this cascade, whose first element is M_{500_1} in our notation, are *not* also elements of the first marking cascade, the one triggered by CS_{300}. That is, M_{500_1} is different from $M_{300_{200}}$ because they exist on different timing cascades. Extending this principle, when the target or US occurs at 700 milliseconds, a third marking branch develops. This timing cascade is triggered by CS_{700}, and its first element is denoted M_{700_1}. It is different from elements $M_{300_{400}}$ and $M_{500_{200}}$ because it exists on a separate timing cascade.

The basic idea in marking should now be clear. Each element of a CS-triggered timing cascade, be it onset or offset, has the capacity to mark the beginning of another cascade operating in parallel. The mechanism of marking is contiguous activation arising from the timing cascade and the target event. Whenever target timing is uncertain (stochastic), the contribution of marking elements to the prediction response is proportional to the probabilities that their first elements have acquired the capacity to trigger cascades. Because more marking cascades would be recruited as time elapses, the number of elements contributing to the actor's prediction increases across the CS–US interval. And because the output of timing elements are assumed to summate to determine output of the system, the scheme

predicts larger prediction responses as elapsed time unfolds. If the probability distribution of target times is uniform across the CS–US interval, our assumptions of the structure of timing leads to the prediction that the response increases linearly across the interval. If, as in our original example, the target occurs equally often at the three interval of 300, 500, and 700 milliseconds, then the prediction response will have three peaks, one for each target time, but they would be successively larger. That is, the response would correspond to a conditional expectation strategy, as illustrated in Figures 5 and 7.

Connectivity of Elements

We have discussed two types of connections of the elements of a timing cascade. These are afferent projections from one element to the next in line, $CS_i \rightarrow CS_{i+1}$, and projections that can activate additional cascades acting in parallel through marking, $CS_i \rightarrow M_{i_1}$, $CS_{i+k} \rightarrow M_{i+k_1}$, and so forth. Because of branching of a timing cascade through marking, it is also evident that elements receive afferent connection from the target or US. It is important to realize that input from the target serves only to promote marking, as given by Equations 7-8, and the strength of the connection between the timing element and the response, as given by Equations 3-5. It does not affect the element's activation of the next-in-line element. In order to realize this assumption in the brain, it is necessary to view a timing element as a multicellular ensemble.

Before considering the possible neuronal organization of a timing element, there are other types of connections that must be considered. Foremost of these are inputs that deliver $\dot{Y}$ information to elements so that changes in associative values (V_is) can occur. Together with the target or US, $\dot{Y}$ determines the strength of the connection between the element and the response (Equations 1-3). It does not affect the element's activation of the next-in-line element. Another possible input to a timing element would come from one of the brain's oscillators. This input would not affect the strengths of connections, but it would influence the temporal coherence of the timing cascade of which the element is a member. Oscillation pulses can entrain timing cascades, determining the inter-element activation rate and the degree to which delays in transmission from one element to the next are correlated. Such entrainment would affect timing accuracy and consistency.

Timing Accuracy

Although the TD model is deterministic, the structure of timing need not be, and therefore timing predictions are subject to error that arises from stochastic processes. The delay in the spread of activation from one timing element to the next is a random variable with some mean and variance. It might be reasonable to assume that the same distribution applies to the delay between any two successive elements and that this distribution remains stationary. Under these circumstances, the mean elapsed time from the onset of a timing cue or CS to

the peak prediction would be equal to the mean inter-element delay in activation multiplied by the number of elements in the cascade. If the distributions governing activation delay for each of the elements are independent, then the variance of the peak predictions would be given by the sum of the variances of the elements of the cascade. This implies that the coefficient of variation (standard deviation/mean) would decrease as the timing interval increases. Such an observation would be at odds with the Weberian property (coefficient of variation is constant) that holds for many timing tasks (Wearden, 1994). However, this is likely not the case, as the spreading activation of elements would be correlated, particularly under entrainment by oscillators. Covariation among the inter-element activation delays would contribute to peak variance, perhaps enough to maintain a constant coefficient of variation. This covariation can be estimated as the proportion of variance in peak timing that cannot be accounted for under the independence assumption.

Neuronal Realization of Timing Elements

Each element of a timing cascade should be regarded as an ensemble of neuronal elements. The various connections discussed in the preceding sections would influence different parts of the ensemble. Because the cerebellum has been characterized as a prediction device (Maill, Weir, Wolpert, & Stein, 1993) and is the primary neural substrate of classical eyeblink conditioning (Rosenfield & Moore, 1995), it is here that we shall speculate about the anatomical basis of timing and implementation of the TD model.

Implementation of TD Learning in the Cerebellum

TD learning can be implemented in the cerebellum by aligning known anatomical ingredients with elements of the learning rule. In TD learning, we assume that each computational time step after the onset or offset of a CS is represented by an anatomically distinct input to the cerebellum. We have suggested that the onset of a CS initiates a spreading pattern of activation among neurons tied to whatever sense modality is involved. This spreading of activation, possibly under entrainment from an oscillator, would sequentially engage pontine nuclear cells, which are the primary source of cerebellar mossy fibers and their associated granule cells. Under this assumption, timing elements would consist of an ensemble that includes pontine nuclear cells, mossy fibers, granule cells, parallel fibers and influences from intrinsic cerebellar neurons such as Golgi cells. Entrainment by oscillators would likely occur at the level of the pontine nuclei, as these are the nexus of neural influences from the lemniscal systems, midbrain, and forebrain (Wells, Hardiman, & Yeo, 1989).

This implementation scheme relies on evidence from rabbit eyeblink conditioning that CR topography is formed in cerebellar cortex through converging contiguous action of parallel fiber and climbing fiber input to Purkinje cells. This action produces synaptic changes known as long term depression (LTD). Mechanisms of LTD in the cerebellum have been spelled out

in recent articles (Eilers, Augustine, & Konnerth, 1995; Ghosh & Greenberg, 1995; Kano, Rexhausen, Dreesen, Konnerth, 1992; Konnerth, Dreesen, & Augustine, 1992).

Figure 8, adapted from Rosenfield and Moore (1995), summarizes the neural circuits that are likely involved in rabbit eyeblink conditioning. The figure shows that CS information ascends to granule cells in the cerebellar cortex (Larsell's lobule H-VI) via mossy fibers originating in the pontine nuclei (PN). Information about the US ascends to cerebellar cortex by two routes, mossy fiber (MF) projections from the sensory trigeminal complex, spinal oralis (SpO) in the figure, and climbing fiber (CF) projections from the inferior olive (IO). A CR is generated within deep cerebellar nucleus interpositus (IP), where the CR is formed by modulation from Purkinje cells (PCs). A full-blown CR is expressed as an increased rate of firing among IP neurons (e.g., Berthier & Moore, 1990; Berthier, Barto, & Moore, 1991). This activity is projected to the contralateral red nucleus (RN). From RN, activity is projected to motoneurons (MN) that innervate the peripheral musculature controlling the position and movements of the eyelids and eyeball (Desmond & Moore, 1991a). The RN also projects to SpO, giving rise to CR-related activity among these neurons (Richards, Ricciardi, & Moore, 1991).

Figure 8 depicts an inhibitory projection from IP to IO. The consequence of this arrangment is that olivary signals to PCs are suppressed when the CR-representation within IP is robust. This anatomical feature suggests that climbing fibers are only excited when the US occurs *and* the CR is weak or absent, implementing a simple *delta* learning rule. The TD learning rule is not a simple delta rule because of the $\gamma Y(t)$ term in Equation 3.

The TD learning rule is implemented by a combination of two reinforcement components. The first is donated by the US, λ in the model's learning rule. The implementation scheme assumes that λ can be aligned with climbing-fiber activation of PCs, which functions to produce LTD among coactive parallel fiber (PF) synapses, as depicted in the figure. The second reinforcement operator is donated by the $\dot{Y}(t)$ terms in the learning rule, $\gamma Y(t) - Y(t-1)$.

Figure 9 shows circuit elements, not shown in Figure 8, for implementing the $\dot{Y}(t)$ component of the learning rule. These components include the projections to cerebellar cortex from the RN and SpO indicated in Figure 8. We hypothesize that the RN projection carries information about $Y(t)$ to cerebellar cortex as efference copy. Parallel fibers project this information to PCs that have collaterals to a set of Golgi cells (Go). Because these projections are inhibitory (Ito, 1984), these PCs invert the efference signal from the RN. In addition, the interpositioning of the PCs between the RN and Golgi cells attenuates the signal and implements the TD model's discount factor, γ.

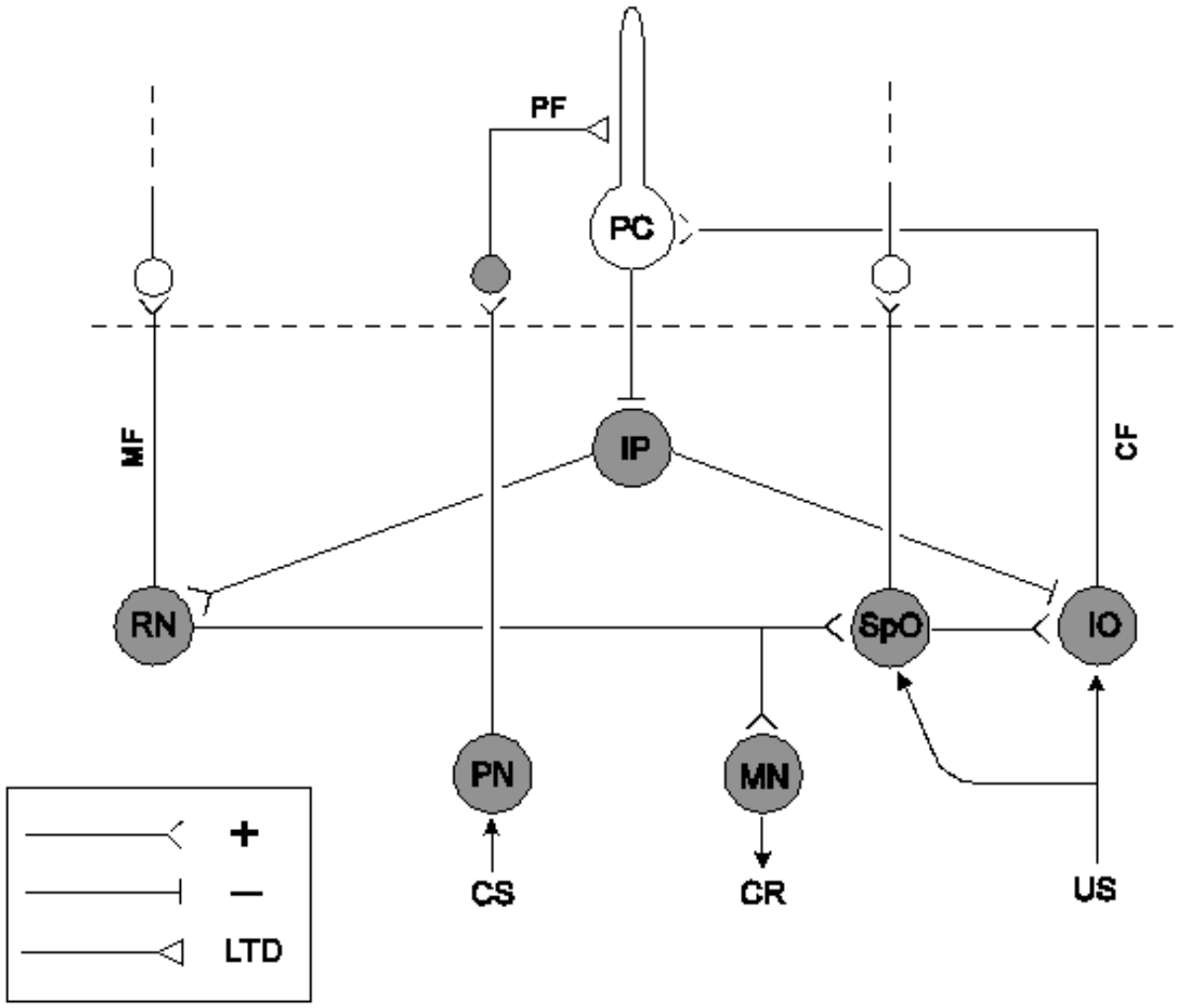

Figure 8 Cerebellar and brain stem circuits underlying eyeblink conditioning, after Rosenfield and Moore(1995). MF = mossy fibers; PF = parallel fibers; PC = Purkinje cell; RN = red nucleus; IP = interpositus nucleus; SpO = spinal trigeminal nucleus pars oralis; CF = climbing fiber; IO = inferior olivary nucleus; PN = pontine nucleus; MN = motoneurons; LTD = long term depression; CS = conditioned stimulus; CR = conditioned response; US = unconditioned stimulus.

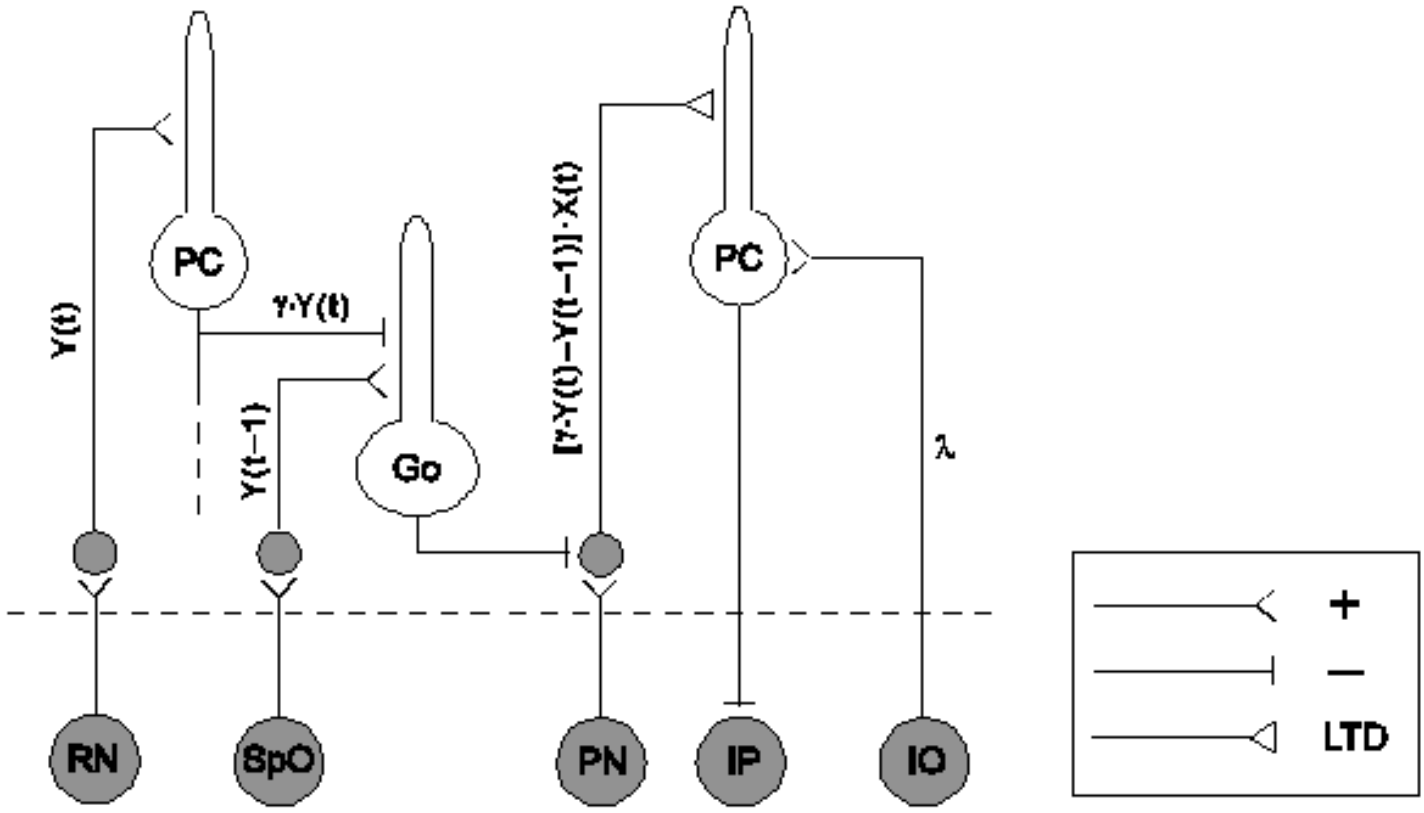

Figure 9 Neural circuits implementing $\gamma \mathbf{Y(t)}$ and other variables of the TD learning rule.

Because Golgi cells are inhibitory on granule cells, the consequence of their inhibition by PCs receiving efference from the RN would be to disinhibit activity of granule cells. In other words, since granule cells relay CS-information from the PN to PCs involved in LTD and CR generation, disinhibition of granule cells by Golgi cells enhances the information flow from active CS components. Mathematically, the implementation scheme assumes that the variables X_j in Equation 4 engage granule cells. The PFs arising from these granule cells engage LTD-PCs in proportion to $\dot{Y}(t) \times X_j$.

In this scheme, PCs driven by projections from the RN would increase their firing rate so as to mimic the representation of the CR as it passes through the RN enroute to MN and SpO. Berthier and Moore (1986) recorded from several H-VI PCs with CR-mimicking increases in firing. Since increases in firing during a CS is incompatible with CR formation through LTD, it is likely that these PCs were inhibiting motor programs incompatible with CR generation, e.g., eyelid and eyeball musculature that would lead to eyelid opening instead of eyelid closure. Here, we are suggesting an additional function of these PCs, that of projecting inverted and discounted CR efference from the RN to Golgi cells.

The implementation scheme assumes that the Golgi cells that receive the inverted efference from the RN also receive a direct, non-inverted, excitatory projection from SpO. This projection carries information about the CR at time $t - \Delta t$. Therefore, the Golgi cell in Figure 9 fires at a rate determined by the differential between two inputs: $\gamma Y(t)$ donated by the RN and $Y(t - \Delta t)$ donated by SpO. Hence, Golgi cells act as $\dot{Y}(t)$ detectors. In terms of Equation 3, $Y(t)$ is transmitted to cerebellar granule cells by the RN, and $Y(t - 1)$ is transmitted to granule cells from SpO. The RN input engages PCs that inhibit Golgi cells responsible for gating inputs from CSs to PCs. Efference from SpO engages the same Golgi cells directly. Because Golgi cells are inhibitory on granule cells, the bigger the RN input relative to SpO input, the bigger the signal from serial component CSs active at that time, be they from onset, offset, or marking cascades. Enhanced throughput from active CS elements in the granular layer would lead to local recruitment of other active PFs that synapse on PCs involved in LTD and CR generation.

In this way, the Golgi cells which implement $\dot{Y}(t)$ reinforce and maintain the down-regulated state of active PF/PC synapses subject to LTD. Parallel fiber/PC synapses that are activated by a CS element are down-regulated by the contiguous US-triggered activation of climbing fiber input from the inferior olive. As CS-elements earlier in the sequence of elements become capable of evoking an output that anticipates the US, inhibition is relayed to the olive and the US loses its capacity to trigger a climbing fiber volley, as shown is Figure 8. However, the down-regulation of these synapses is maintained, and still earlier CS-elements are recruited, by PFs carrying $\dot{Y} \times X_j$ to LTD-PCs, as indicated Figure 9.

In a single-unit recording study, Desmond and Moore (1991a) observed an average lead-

time of 36 milliseconds from the onset RN cells with highly CR-related firing patterns and the peripherally observed CR. The average lead-time of SpO cells with CR-related activity was 20 milliseconds. Therefore, the time difference in CR-related efference arising from the two structures is probably on the order of 15-20 milliseconds. This difference spans one 10-millisecond time-step used in our simulations with the TD model. This temporal difference is consistent with a conduction velocity of 2 meters/second for the 10 millimeter trajectory of unmyelinated axons from the RN to rostral portions of SpO. The 10-millisecond grain also ensures high-fidelity resolution of fast transients. The fastest transients in eyeblink conditioning occur during unconditioned responses (URs). At its fastest, the eyelids require 80 milliseconds to move from completely open to completely closed, with a peak velocity of approximately 4-5 millimeters/20 milliseconds.

Efference from SpO neurons recorded among H-VI PCs would tend to lag behind the peripherally observed CR, if it arises from more caudal portions of the structure. Berthier and Moore (1986) observed a continuum of lead and lag times among PCs that increased their firing to the CS. Purkinje cells that receive projections from SpO (not shown in the figure) would be expected to increase their firing, but with a lag relative to those receiving projections from the RN. It makes sense that the proportion of CR-leading PCs observed by Berthier and Moore (1986) matched the number of CR-lagging PCs, since these two populations would merely be reflecting CR efference from two temporal vantage points.

Figure 10 is an expanded version of Figure 9 showing three sets of granule cells associated with three serial component CSs. These components include those arising from CS onset and offset, as well as those that might arise from marking processes. The degree to which information from any of these serial CS components reaches the PCs to which they project is determined by Golgi cells firing in proportion to $\dot{Y}(t)$, as just described. Figure 1 shows that, depending on γ, TD-simulated CRs are positively accelerating in time up to the occurrence of the US, so $\dot{Y}(t)$ increases progressively over the CS–US interval. Therefore, those PF/PC synapses activated near the time of the climbing fiber signal from the US would have the greatest impact in establishing and maintaining LTD of PF/PC synapses that ensure the appropriate form and timing of CRs. The spatial arrangement of PF/PC synapses has no significance for CR timing.

Implications of the Implementation

The implementation scheme has several testable implications. One that has already been mentioned is that the firing pattern of most H-VI PCs with CR-related firing resembles the CR. We maintain that this pattern of firing reflects CR efference. Since this efference cannot arise from proprioceptors, which are absent in muscles controlling the eyeblink, and since the axons of motoneuons innervating these muscles do not possess recurrent collaterals, this efference must arise from premotor centers. The RN and SpO are the prime candidates.

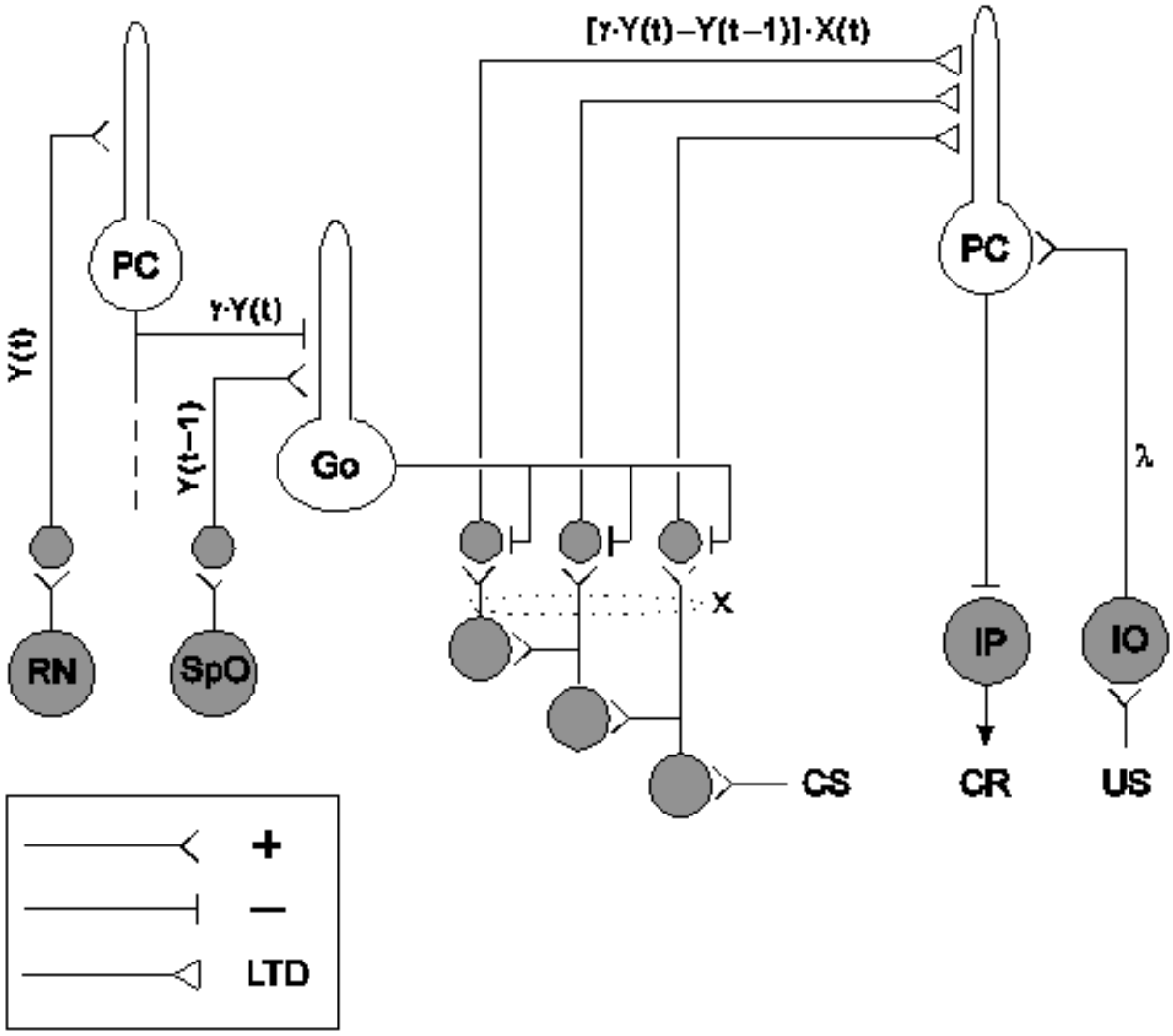

Figure 10 The complete TD implementation scheme showing three sequentially activated CS components, representing both onset and offset cascades in the manner of Desmond and Moore's(1988) VET model.

One can only speculate as to the functions of this efference. We suggest that one function is to activate Golgi cells that modulate information flow through the granule cells. Another function is to excite PCs that ultimately control muscles that, if engaged, would preclude or interfere with CR performance, such as those producing eye opening or saccadic movements.

The implementation scheme also requires that Golgi cells that modulate information flow from serial component CSs fire in relation to *changes* in eyelid position, i.e., they fire in relation to $\dot{Y}$. This property of Golgi cell firing patterns has been reported by van Kan, Gibson, and Houk (1993), in a study of monkey limb movements, and Edgley and Lidierth (1987), in a study of cat locomotion.

A third implication of the implementation scheme concerns the effects of inactivation of the RN through the use of pharmacological agents or cryostatic probes. Although inactivation of the RN would cause a temporary interruption of information flow that results in a conditioned response, it would not prevent learning of the primary association between components of the CS and the US. This association would proceed with little disruption because the pontine nuclei and the IO would still be able to convey CS and US information to cerebellar cortex. Evidence for this proposition comes from a study of rabbit eyeblink conditioning by Clark and Lavond (1993). They demonstrated that inactivation of the RN by cooling did not prevent learning, as CR magnitude recovered immediately upon reactivation of the RN.

However, inactivation of the RN would interrupt efference about the position of the eyelid at times t and $t - \Delta t$ from the RN and SpO. Thus, $\dot{Y}$ would not be available to cerebellar cortex. According to the TD model, $\dot{Y}$ allows for increments of predictive associations in the absence of the US, as would occur in second-order conditioning (Kehoe, Feyer, & Moses, 1981). This being the case, inactivation of the RN would interfere with second-order conditioning. Animals trained simultaneously in first- and second-order conditioning with the RN inactivated, would be expected to show first-order learning, as in the Clark and Lavond (1993) study, but little or no second-order learning.

Interpretations of $\dot{Y}$: Efference or Afference

Equation 3 emphasizes interpretations of $\dot{Y}$ as efference, but it is equally correct to interpret changes in associative values in terms of afference, as discussed above in connection with Equation 6. A recent study by Ramnani, Hardiman, and Yeo (1995) suggests that the efference interpretation of $\dot{Y}$ is correct. This experiment shows that temporary inactivation of IP by muscimol application prevents extinction of the CR. That is, CS-alone trials that would normally lead to a gradual elimination of the CR, instead had no effect whatsoever. When tested later, after the muscimol blockade had been removed, the previously established CR was at full strength. It did extinguish with continuing presentation of CS-alone trials.

This finding is consistent with the efference interpretation of the model because inactivation of IP eliminates the CR and therefore prevents efference from the RN and SpO from affecting learning. Under an afference interpretation, inactivation of IP would not prevent CS information from ascending to cerebellar cortex, where extinction would proceed normally, as this information arises from the PN.

Prediction Strategies and the Cerebellar Implementation

The anatomical and physiological relationships reviewed in the preceding sections do not exhaust all of the potentially important circuits and factors that are likely involved in predictive timing by the cerebellar system (see Ito, 1984; Rosenfield & Moore, 1995). Some of these influences could be important for implementing specific prediction strategies. Before considering some of these options, it should be clear that the cerebellum need not participate in any of them. The fail-safe strategy, for example, resembles a voluntary response (Coleman & Webster, 1988). Such a response could come about through the direct action of cerebral motor cortex on motoneurons, bypassing the cerebellum altogether. The same is true of the other prediction strategies reviewed previously. But there is no evidence that cerebral motor cortex is involved in eyeblink conditioning. Specifically, CR timing and topography are normal following aspiration of this part of the brain. However, learning-dependent timing of conditioned eyeblink responses is disrupted by lesions of the cerebellar cortex (Perrett, Ruiz, & Mauk, 1993).

Within the context of the cerebellum, a fail-safe strategy can be implemented in any of number of ways. Perhaps the most straightforward from the standpoint of the TD model would be to assume that timing uncertainty promotes a marking cascade that is directed to the inferior olive, a putative site of representation of the US (Figure 8). In fail-safe mode, this marking provokes olivary neurons to oscillate in such a way as to saturate the system and prevent return to baseline until after the trial terminates (e.g., Yarom, 1989).

A hedging strategy would not involve this sort of marking-derived prolonged activation of the inferior olive, and there would be no prolonged saturation of Purkinje cells. The degree of hedging would depend on the value of the discount parameter of the TD model, γ, and this could be controlled through of aminergic modulation of Purkinje cells, as discussed by Ito (1984, p 60). For example, a high state of noradrenergic 'arousal' would inhibit Purkinje cells and this would decrease the value of $\dot{Y}$ computed by Golgi cells. Assuming no other effects, this attenuation of γ would lower response amplitude (see Figure 1), as in proportionate hedging. Removing the aminergic modulation would have the opposite effect on γ and response amplitude.

A conditional expectation strategy requires the recruitment of additional timing elements through marking. This marking involves interactions with the US that occur in precerebel-

lar neurons, and brain stem reticular formation would seem to be a likely candidate, on anatomical and physiological grounds (Richards, Ricciardi, & Moore, 1991).

Entrainment of Timing and the Cerebellar System

Previously, we speculated that timing cascades might have the capacity to become entrained by one of the brains oscillators. The inferior olive is an oscillator, with a normal rhythm of 10 Hz. Because the primary projection of the inferior olive is to Purkinje cells, the entrainment of timing elements could well occur at this level, but a more plausible scenario would be entrainment by oscillations within catecholamine systems that engage brainstem reticular formation neurons and ultimately the pontine nuclei.

Summary and Conclusions

This chapter considered how the TD theory of reinforcement learning, which lies at the heart of promising applications in adaptive control in both real and artificial systems, might be adapted to training protocols in which behavior is controlled by cascades of activation of time-tagged serial elements. The TD model with the CSC assumption generates appropriate CR waveforms in simple protocols, but it can also be extended to predictive timing under temporal uncertainty.

The chapter also suggests an implementation scheme for TD learning within the cerebellum. The implementation draws on neurobiological evidence regarding how LTD is established, reinforced, and maintained among Purkinje cells that form the CR. The implementation incorporates recent anatomical findings, reviewed by Rosenfield and Moore (1995), that allow these Purkinje cells to receive both components of the TD model's reinforcement operator, one component donated by the US and another component donated by $\dot{Y}(t) = Y(t) - Y(t - \Delta t)$. The implementation scheme lays the foundation for network simulations at the cellular level.

The entire exercise reinforces the synergy that has enlightened and invigorated behavioral and neurobiological studies of reinforcement learning. By providing a comprehensive rendering of classical conditioning as it occurs in real time, the TD model provides a framework for novel insights about sequencing and timing behavior.

References

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, D. C. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215-232). Cambridge, MA: MIT Press.

Barto, A. G., Sutton, R. W., & Anderson, C. W. (1983). Neuronlike elements that can solve difficult control problems. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13,* 834-846.

Berthier, N. E., Barto, A. G., & Moore, J. W. (1991). Linear systems analysis of the relationship between firing of deep cerebellar neurons and the classically conditioned nictitating membrane response in rabbits. *Biological Cybernetics, 65,* 99-105.

Berthier, N. E., & Moore, J. W. (1990). Activity of deep cerebellar nuclear cells during classical conditioning of nictitating membrane extension in rabbits. *Experimental Brain Research, 83,* 44-54.

Berthier, N. E., & Moore, J. W. (1986). Cerebellar Purkinje cell activity related to the classically conditioned nictitating membrane response. *Experimental Brain Research, 63,* 341-350.

Clark, R. E., & Lavond, D. G. (1993). Reversible lesions of the red nucleus during acquisition and retention of a classically conditioned behavior in rabbits. *Behavioral Neuroscience, 107,* 264-270.

Coleman, S. R., & Webster, S. (1988). The problem of volition and the conditioned reflex. Part II. Voluntary-responding subjects, 1950-1980. *Behaviorism, 16,* 17-49.

Desmond, J. E. (1990). Temporally adaptive responses in neural models: The stimulus trace. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 421-461). Cambridge, MA: MIT Press.

Desmond, J. E., & Moore, J. W. (1988). Adaptive timing in neural networks: The conditioned response. *Biological Cybernetics, 58,* 405-415.

Desmond, J. E., & Moore, J. W. (1991a). Activity of red nucleus neurons during the classically conditioned rabbit nictitating membrane response. *Neuroscience Research, 10,* 260-279.

Desmond, J. E., & Moore, J. W. (1991b). Altering the synchrony of stimulus trace processes: tests of a neural-network model. *Biological Cybernetics, 65,* 161-169.

Edgley, S. A., & Lidierth, M. (1987). Discharges of cerebellar Golgi cells during locomotion in cats. *Journal of Physiology, London, 392,* 315-332.

Eilers, J., Augustine, G. J., & Konnerth, A. (1995). Subthreshold synaptic Ca^{2+} signaling in fine dendrites and spines of cerebellar Purkinje neurons. *Nature, 373,* 155-158.

Gabriel M., & Moore, J. (1990). *Learning and computational neuroscience: Foundations of adaptive networks.* Cambridge, MA: MIT Press.

Ghosh, A., & Greenberg, M. E. (1995). Calcium signaling in neurons: Molecular mechanisms and cellular consequences. *Science, 268,* 239-247.

Ito, M. (1984). *The cerebellum and neural control.* New York: Raven Press.

Kano, M., Rexhausen, U., Dreesen, J., & Konnerth, A. (1992). Synaptic excitation produces a long-lasting rebound potentiation of inhibitory synaptic signals in cerebellar Purkinje cells. *Nature, 356,* 601-604.

Kehoe, E. J., Feyer, A. M., & Moses, J. L. (1981). Second-order conditioning of the rabbit's nictitating membrane response as a function of the CS2–CS1 and CS1–US intervals. *Animal Learning & Behavior, 9,* 304-315.

Kehoe, E. J., Graham-Clarke, P., Schreurs, B. G., (1989). Temporal patterns of the rabbit's nictitating membrane response to compound and component stimuli under mixed CS–US intervals. *Behavioral Neuroscience, 103,* 283-295.

Kehoe, E. J., Schreurs, B. G., Macrae, M., & Gormezano, I. (1995). Effects of modulating tone frequency, intensity, and duration on the classically conditioned rabbit nictitating membrane response. *Psychobiology, 23,* 103-115.

Konnerth, A., Dressen, J., Augustine, G. T. (1992). Brief dendritic signals initiate long-lasting synaptic depression in cerebellar Purkinje cells. *Proceedings of the National Academy of Science USA, 89,* 7051-7055.

Maill, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the cerebellum a Smith Predictor? *Journal of Motor Behavior, 25,* 203-216.

Moore, J. W. (1991). Implementing connectionist algorithms for classical conditioning in the brain. In M. L. Commons, S.Grossberg, & J. E. R. Staddon (Eds.), *Neural Network Models of Conditioning and Action* (pp. 181-191). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moore, J. W. (1992). A mechanism for timing conditioned responses. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, Action and Cognition* (pp. 229-238). Dordrecht,

The Netherlands: Kluwer Academic Publishers.

Moore, J. W., & Desmond, J. E. (1992). A cerebellar neural network implementation of a temporally adaptive conditioned response. In I. Gormezano & E. A. Wasserman (Eds.), *Learning and Memory: The Behavioral and Biological Substrates* (pp. 347-368). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moore, J. W., Desmond, J. E., & Berthier, N. E. (1989). Adaptively timed conditioned responses and the cerebellum: A neural network approach. *Biological Cybernetics, 62,* 17-28.

Perrett, S. P., Ruiz, B. P., & Mauk, M. D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience, 13,* 1708-1718.

Ramnani, N., Hardiman, M. J., & Yeo, C. H. (1995). Temporary inactivation of the cerebellum prevents the extinction of conditioned nictitating membrane responses. *Society for Neuroscience Abstracts, 21,* 1222.

Richards, W. G., Ricciardi, T. N., & Moore, J. W. (1991). Activity of spinal trigeminal pars oralis and adjacent reticular formation units during differential conditioning of the rabbit nictitating membrane response. *Behavioural Brain Research, 44,* 195-204.

Rosenfield, M. E., & Moore, J. W. (1995). Connections to cerebellar cortex (Larsell's HVI) in the rabbit: A WGA-HRP study with implications for classical eyeblink conditioning. *Behavioral Neuroscience, 109,* 1106-1118.

Sutton, R. S. (1992). Guest Editor: Special issue on reinforcement learning. *Machine learning, 8,* 1-171.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497-537). Cambridge, MA: MIT Press.

van Kan, P. L. E., Gibson, A. R., & Houk, J. C. (1993). Movement-related inputs to intermediate cerebellum of the monkey. *Journal of Neurophysiology, 69,* 74-94.

Wearden J. (1994). Prescriptions for models of biopsychological time. In M. Oaksford & G. D. A. Brown (Eds.), *Neurodynamics and psychology* (pp. 215-236). San Diego, CA: Academic Press.

Wells, G. R., Hardiman, M. J., & Yeo, C. H. (1989). Visual projections to the pontine nuclei of the rabbit: Orthograde and retrograde tracing studies with WGA-HRP. *Journal of Comparative Neurology, 279* 629-652.

Yarom, Y. (1989). Oscillatory behavior of olivary neurons. In P. Strata (Ed.), *The olivocerebellar system in motor control* (pp. 209-220). Berlin: Springer-Verlag.