Journal of Computational Neuroscience 6, 191–214 (1999)

A Predictive Reinforcement Model of Dopamine Neurons for Learning Approach Behavior

JOSÉ L. CONTRERAS-VIDAL
Motor Control Laboratory, Arizona State University, Tempe, AZ 85287-0404, USA
pepe@asu.edu

WOLFRAM SCHULTZ
Institute of Physiology, University of Fribourg, Rue du Musée 5, CH-1700 Fribourg, Switzerland
wolfram.schultz@unifr.ch

Received December 4, 1997; Revised September 4, 1998; Accepted September 22, 1998

Action Editor: Charles Wilson

Abstract. A neural network model of how dopamine and prefrontal cortex activity guides short- and long-term information processing within the cortico-striatal circuits during reward-related learning of approach behavior is proposed. The model predicts two types of reward-related neuronal responses generated during learning: (1) cell activity signaling errors in the prediction of the expected time of reward delivery and (2) neural activations coding for errors in the prediction of the amount and type of reward or stimulus expectancies. The former type of signal is consistent with the responses of dopaminergic neurons, while the latter signal is consistent with reward expectancy responses reported in the prefrontal cortex. It is shown that a neural network architecture that satisfies the design principles of the adaptive resonance theory of Carpenter and Grossberg (1987) can account for the dopamine responses to novelty, generalization, and discrimination of appetitive and aversive stimuli. These hypotheses are scrutinized via simulations of the model in relation to the delivery of free food outside a task, the timed contingent delivery of appetitive and aversive stimuli, and an asymmetric, instructed delay response task.

Keywords: neural network, prefrontal, reinforcement learning, striatum, timing

1. Introduction

Anatomical, neurophysiological, and pharmacological data support the notion of parallel, functionally segregated, basal ganglia-thalamo-cortical circuits (Eblen and Graybiel, 1995; Strick et al., 1995; Alexander et al., 1986). These parallel systems are based on a broad division at the level of the striatum in terms of its source of inputs from sensory-motor, association, and limbic cortical areas and in terms of neurochemical subsystems that may be modulated by the behavioral context (Graybiel, 1990; Contreras-Vidal and Stelmach, 1995). These circuits are thought to subserve motor control, cognitive, and motivational aspects of goal-directed behavior, respectively. However, little is known about how these systems interact during goal-directed behavior.

Recent experimental data suggest that the dopaminergic (DA) neurons of the substantia nigra (A9) and adjoining groups A8 and A10 may be involved in establishing and sustaining conditioned approach behavior through stimulus-reward associations. Specifically, DA cells are activated by novel, unexpected stimuli eliciting orienting reactions, by primary reward during self-initiated movements in the absence of reward-predicting phasic stimuli (Romo and Schultz, 1990), and by conditioned, reward-predicting stimuli (Ljungberg et al., 1992; Schultz et al., 1993;

Mirenowicz and Schultz, 1994). Moreover, Mirenowicz and Schultz (1996) have shown that the DA responses occur preferentially due to the temporally unpredicted occurrence of appetitive or potentially appetitive rather than aversive stimuli. However, these DA responses tend to decrease progressively with continuous presentation of behaviorally irrelevant stimuli, with repetitive presentation of primary reward delivered at regular intervals without performance of any task, and with overtraining (Ljungberg et al., 1992). Schultz and colleagues suggested that this decrease in DA responsiveness may reflect a measure of automaticity in the performance of the task (Ljungberg et al., 1992; Schultz et al., 1993), perhaps as a function of reducing the predictibility of the behavioral outcome.

Recently, Montague et al. (1996) and Schultz et al. (1997) have suggested that dopamine neurons in the ventral tegmental area (VTA) and the substantia nigra compute ongoing reward prediction errors. Their proposal is based on the temporal difference (TD) model of Sutton and Barto (Sutton and Barto, 1981; Sutton, 1988). The TD model suggests that Pavlovian reinforcement is the time derivative of a composite association combining primary reinforcement (that is, juice or food reward) and previously acquired associations. Montague et al. (1996) have shown that the TD model can account for the responses of DA neurons under various task conditions in the monkey. However, the TD model does not address the effects of attention, salience or novelty, and generalization or discrimination in DA responses (see discussion in Schultz et al., 1997).

The present article introduces a model of how DA activity modulates short- and long-term neuronal activity within the cognitive-motivational cortico-striatal loops during learning of approach behavior. We postulate two types of physiological signals for reward prediction, which are critical for learning: a signal representing an error in the timing of reward prediction and a signal coding for errors in the type and amount of reward prediction (or stimulus expectancies). Thus, reward prediction is decomposed into its temporal and featural components. The former may be related to the TD model, and the latter may be related to the adaptive resonance theory (ART) of Carpenter and Grossberg (1987). In particular, it is shown how a neural network architecture that satisfies the design principles of ART systems may account for the DA responses to novelty, generalization, and discrimination of appetitive and aversive stimuli (Mirenowicz and Schultz, 1996). In the proposed model reinforcement-based error signals constrain the learning of recognition codes in an ART system and help evaluate the biological significance of arbitrary stimuli by shifting attention to only those environmental events that are or can be rewarding based on previous experience (Carpenter and Grossberg, 1990). Moreover, it is proposed that the striatal matrix may be involved in evaluating and learning the motivational significance of arbitrary sensory stimuli in relation to the delivery of primary reward, while the striatal patch may be involved in computing the expected time of occurrence of primary reward. In the next sections, it is shown (1) how an adaptive striatal spectral timing circuit (cf. Grossberg and Merrill, 1992) can learn to predict the expected time of ocurrence of appetitive stimuli, and (2) how the evaluation and discrimination of appetitive and aversive stimuli in cortico-striatal loops may be gated by DA activity during learning of approach behavior. These desirable model competencies have been recently reviewed in Schultz et al. (1997). Some of the data have been presented briefly as an abstract (Contreras-Vidal and Schultz, 1996).

2. Methods

The neural network model shares fundamental design principles postulated in ART systems (Carpenter and Grossberg, 1987) and in the spectral timing model of Grossberg and Merrill (1992). The proposed model, depicted in Fig. 1, consists of a predictive timing network that computes the expected time of delivery of primary reward and a sensory-motivational associational network that evaluates and learns the biological significance of arbitrary sensory stimuli based on the current context and previous experience. It is postulated that the short-term (STM) and long-term memory (LTM) activities in the cortico-striatal pathways are gated by DAergic pathways during learning of approach behavior.

2.1. Predictive Timing Circuit

A network model of simplified striatal principal neurons with lateral inhibition was modeled as

$$\epsilon \frac{d}{dt} S_j = -S_j + 2(0.6 - S_j) \times \left(\sum_m f(P_m(t)) + f_n(S_j) \cdot T_j \right) - 15(0.7 + S_j) \sum_{k \neq j} ([S_k]^+ \cdot T_k), \quad (1)$$

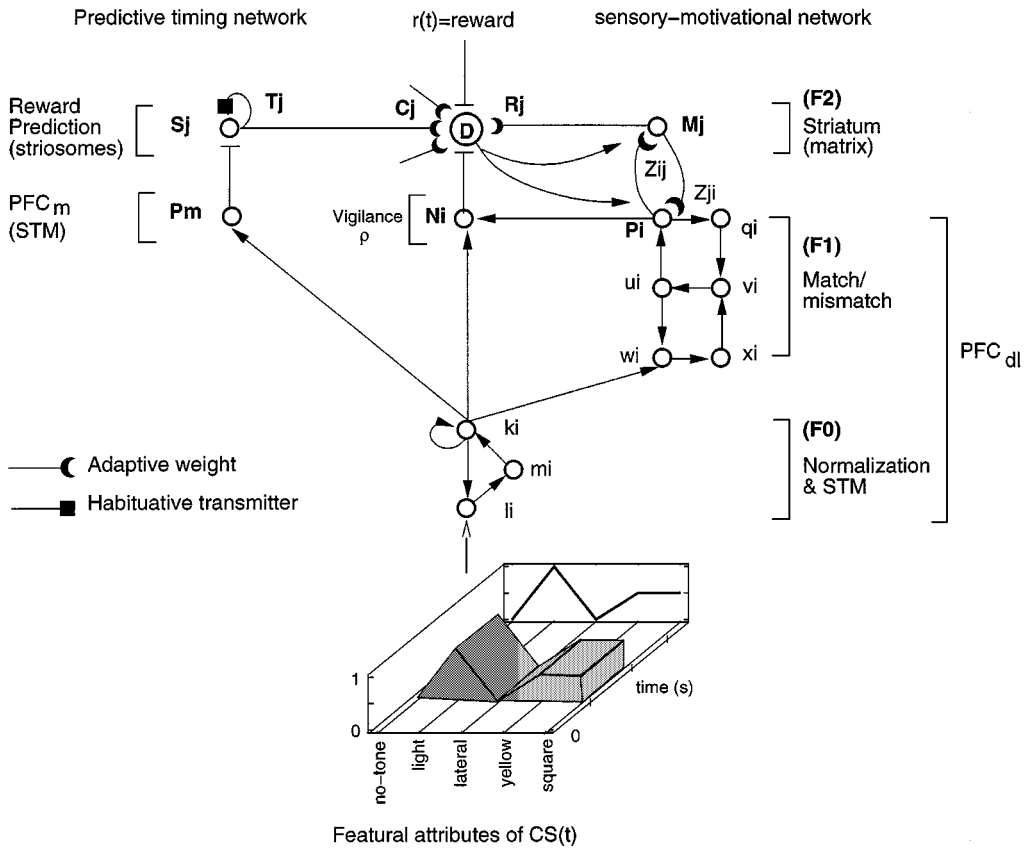

Figure 1. Adaptive resonance theory architecture used for simulating the sensory-motivational and the timing circuits. Learning within the cortico-basal ganglia pathways in the proposed model is modulated by dopamine activity, which constrains the access of arbitrary sensory stimuli to long-term memory to only those stimuli that are motivationally significant to the animal. Biologically significant stimuli are signaled by changes in dopamine activity during learning of goal-directed behavior. A typical input pattern is shown at the bottom illustrating the featural attributes of a visual stimulus (yellow square placed in a lateral position with respect to the midline of a monkey not shown). Keys: C_j, adaptive timing weight; D, dopamine cells; M_i, matrisomal neurons; N_i, mismatch or novelty signal; P_i, output of PFC$_{dl}$ neurons; P_m; output of PFC$_m$ neurons; PFC_{dl}, dorsolateral prefrontal cortex; PFC_m, medial PFC; R_j, conditioned reinforcer weight; S_j, striosomal neurons; STM, short-term memory gated by dopamine activity; T_j, habituative transmitter; Z_{ij} adaptive weights in the cortico-striatal (F1 $\rightarrow$ F2) pathway; Z_{ji} adaptive weights in the striato-cortical (F2 $\rightarrow$ F1) pathway; ρ, vigilance parameter; l_i, input neuron; k_i, m_i, q_i, w_i, x_i, u_i, and v_i represent computations performed by interneurons or processing layers within PFC.

where S_j represents the average firing rate of striatal spiny neurons. The first term in the right-hand side of Eq. (1) represents the passive decay of neural activity, and the second and third terms represent shunting excitation and shunting inhibition respectively. The activity of neuron S_j is bounded in the interval $[-0.7, 0.6]$. $P_m(t)$ represents the onset timing of the input that is turned on by the presentation of a conditioned stimulus (CS(t)); T_j are variables that represent the availability of neurotransmitter within the striatal patches; $\epsilon = 0.0833$; $f(x) = 0.5$ for $x \geq 0$ and zero otherwise; $f_n(x) = 10(x - 0.2)$ if $x \geq 0.2$ and zero otherwise, represents a thresholded nonlinear function of striatal activity; and $[x]^+ = x$ if $x > 0$ and zero otherwise.

The levels of neurotransmitters in the striatal output pathways (T_j) are modeled as (Contreras-Vidal and Stelmach, 1995)

$$\frac{d}{dt}T_j = \alpha(1 - T_j) - \beta_j f_g(S_j)T_j, \qquad (2)$$

where $\alpha(1 - T_j)$ represents accumulation to a constant target level 1, $\beta_j f_g(S_j)T_j$ depletion or habituation by signal pathways, $f_g(x) = x$ if $x > 0.2$ and zero otherwise, $\alpha = 0.1$, and $\beta_j = 500/(1 + j)$ represents a spectrum of depletion rates as in Grossberg and Merrill (1992). Each striatal activation signal $f_g(S_j)$ interacts with this spectrum of habituative chemical

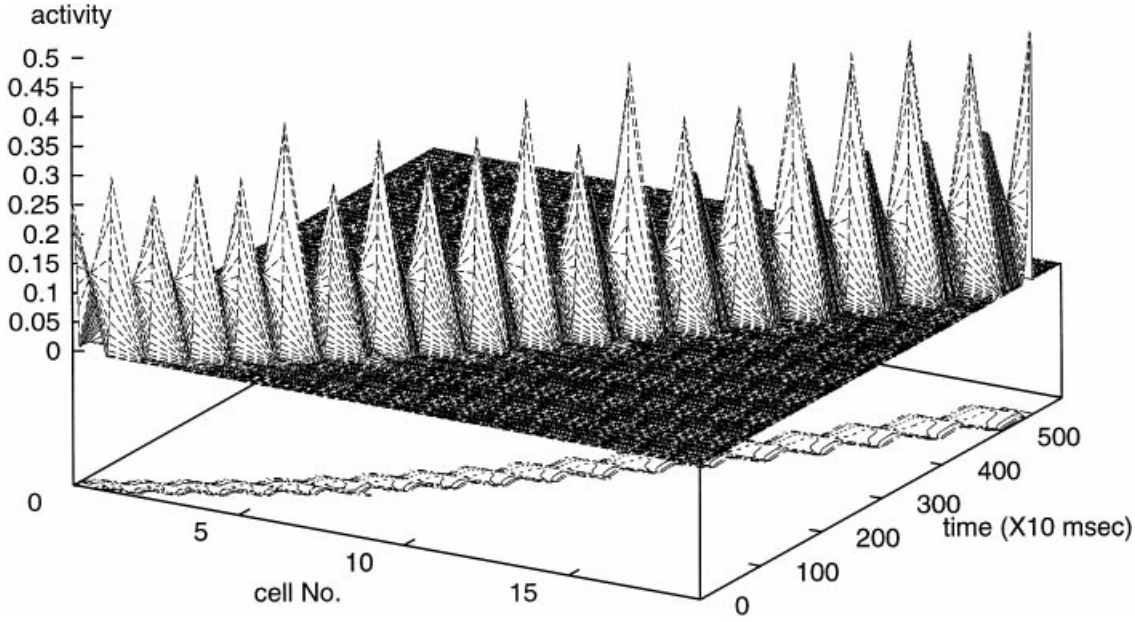

Figure 2. Temporal representation of the timing of stimulus events. It is hypothesized that different subsets of striatal cells (striosomes) are activated successively to represent sensory stimuli through time. These activations are adaptively and selectively reinforced to develop a prediction signal that encodes the expected time of reward delivery. The internal timing displays the scalar property (for example, the width of the activation grows with the interval being timed). The simulated striosomal responses resemble the striatal responses at different latencies recorded by Apicella et al. (1991, 1992).

transmitters via Eq. (2). This spectrum causes some striosomal populations (S_j) to deplete neurotransmitter (T_j) levels faster than other striosomes with smaller depletion rates. A spectrum of depletion rates (such as β_j) is critical in generating the striosomal clock; however, any mechanism that could generate a family of depletion rates would also work, including other instantiations of β_j with different constants. It is assumed that a new input pattern resets to zero all the neural activities. Thus, the striosomal clock is reset any time a new input arrives to layer F0.

Figure 2 depicts a simulation of a striosomal temporal representation in response to an input stimulus. It shows the activity of 20 striosomal cells over a period of five simulated seconds. Note an ordered activation of cells with time, such that different subsets of striosomal cells are activated successively to represent the sensory stimulus through time. These simulated neural activations may correspond to the dorsal and ventral striatal responses at different latencies found by Apicella et al. (1991, 1992) during Go/NoGo and instructed delay response tasks. It is hypothesized that these striatal activations are adaptively and selectively reinforced depending on DA activity to develop a prediction signal that encodes the expected time of reward delivery.

The inhibitory striosomal temporal representation is sent to the model DA cells through an adaptive weight (C_j). Striosomes that are simultaneously active with DA cells have their weights C_j increased to select the striosomal population that best signals the expected time of reward delivery (such as predictive timing, $S_j \cdot C_j$). The adaptive timing weights (C_j) are computed as

$$\frac{d}{dt}C_j = [S_j]^+(-C_j + 20(1 - C_j)[D]^+), \quad (3)$$

where S_j represents presynaptic activity from one or more striosomal cells and D is the nigrostriatal DA activity defined below. Equation (3) can undergo long-term potentiation (LTP) if both dopamine (D) and striatal activity (S_j) are above zero, long-term depression (LTD) when dopamine activity is zero and striatal activity is above zero, or no change if both dopamine and striatal activities are zero. We implemented the adaptive timing computations assuming synaptic plasticity between striatal and DA cells; however, this does not

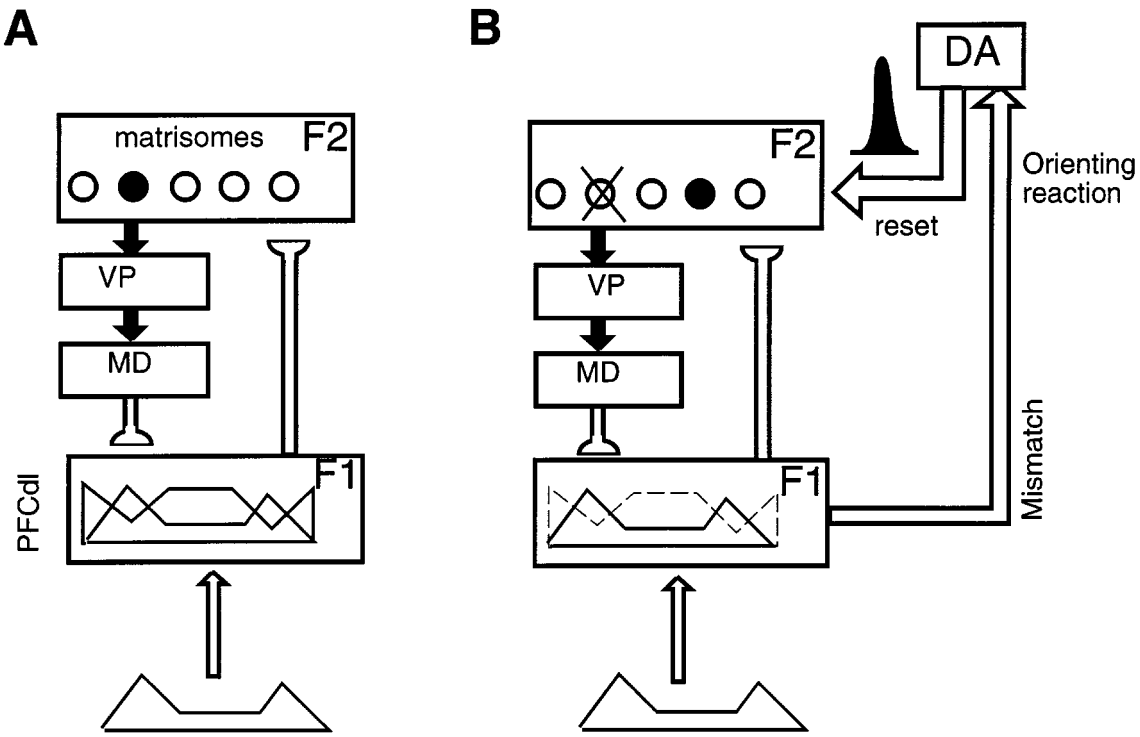

Figure 3. Match and mismatch (novelty) computations by prefrontal cortex cells. A: Novel unexpected stimuli is registered at prefrontal cortex (F1) layer of neurons (Carpenter and Grossberg, 1987). PFC_{dl} cells in turn send weighted projections to the extrastriosomal matrix where a particular sensory-motivational representation is selected via winner-take-all computations at F2. The selected representation in turn sends, via the pallidum and thalamic neurons, an expectation based on the current sensory stimulus. As the evaluation of the current stimulus differs from its expected biological significance, a mismatch is generated at PFC_{dl} cells. B: This mismatch or novelty signal activates dopamine neurons, which in turn reset or inhibit the current matrisomal representation to allow an orienting reaction to the new stimulus and to allow new learning. The model suggests that the reset wave is mediated by DA activity in response to prefrontal input.

rule out the possibility of synaptic plasticity within the striatum itself. In this regard, Wickens et al. (1996) have reported both LTD and LTP in corticostriatal synapses of the rat in vitro.

2.2. Reinforcement ART

The adaptive sensory-motivational circuit is largely based on the mathematical description of the ART-2 architecture proposed by Carpenter and Grossberg (1987), except for the weight-change learning rules, which in the present instantiation of the analog ART (that is, reinforcement ART-2) model are hypothesized to be modulated by dopamine activity. The ART system, and the model presented herein, consist of an attentional subsystem and an orienting subsystem. The attentional subsystem is comprised by the prefrontal (F1) and striatal (F2) networks, as well as the cortico-striatal and striato-cortical adaptive filters depicted in Fig. 1. The orienting subsystem becomes activated when a cortico-striatal pattern generated by inputs at F1 does not match the learned striato-cortical expectation produced by the currently active category at F2. In this case, the orienting subsystem causes a reset of the current category at F2, and alternative categories are tested until a match is found or a new category is formed (Fig. 3). The particular instantiation of ART-2 used in the present simulations is based on the Fig. 10 of Carpenter and Grossberg (1987), which uses a preprocessing stage (that is, F0 in Fig. 1) to normalize (and contrast-enhance) the inputs to the F1 network and help stabilize the orientating subsystem during category search.

2.2.1. Contrast-Enhancement of Prefrontal Inputs. The input to the network are first normalized (these

equations follow those of Carpenter and Grossberg, 1987):

$$k_i = \frac{(f(m_i) + 5.0\, f(k_i/\|k\|))}{\|f(m)\|} \tag{4}$$

$$l_i = I_i + 5.0 k_i \tag{5}$$

$$m_i = \frac{l_i}{\|l\|}, \tag{6}$$

where $\|x\|$ denotes the L_2 norm of a vector x, I_i is the input vector, k_i is the output of the preprocessing layer (F0 level in Fig. 1), l_i and m_i represent interneuron activities, and $f(x) = 0$ if $0 \leq x < 0.1$ or $f(x) = x$ if $x \geq 0.1$. Normalization of input patterns is important for increasing the signal-to-noise ratio of analog input patterns; however, as Carpenter and Grossberg (1987) have shown, it is not critical for ART systems to operate.

2.2.2. Attentional Subsystem. The output activity at network layer F1 in Fig. 1 is computed (at equilibrium) as in Carpenter and Grossberg (1987)—

$$P_i = u_i + \sum_{j=1}^{M} g(M_j) Z_{ji}, \tag{7}$$

where P_i represents the output of the F1 neurons, Z_{ji} is an adaptive weight from pathways originating at M_j and contacting neuron P_i, $M = 45$ is the number of striatal matrix neurons at F2, and $g(M_j) = 0.9$ if neuron M_j is the winner (winner-take-all) cell and it has not been reset in the current trial (only one neuron is allowed to win the striatal competition in the present simulations, although a distributed representation is also possible; see Raijmakers et al., 1996). The variables q_i, u_i, v_i, and x_i depicted in Fig. 1 and defined below may be related to computations within distinct cortical layers subserved by interneuronal and pyramidal cell interactions at PFC (for each of these variables, the number of neurons is equal to the size of the input vector—that is, $i = 1, \ldots, 5$)—

$$q_i = \frac{P_i}{\|P\|} \tag{8}$$

$$u_i = \frac{v_i}{\|v\|} \tag{9}$$

$$v_i = f(x_i) + 5\, f(q_i), \tag{10}$$

where $f(x) = 0$ if $0 \leq x < 0.1$ or $f(x) = x$ if $x \geq 0.1$, and

$$x_i = \frac{w_i}{\|w\|}. \tag{11}$$

The input to the attentional subsystem at PFC is given by

$$w_i = k_i + 5u_i. \tag{12}$$

The activity in the striato-cortical pathway is initiated by a winner matrisomal cell (M_w) at level F2 as follows:

$$M_w(\text{winner}) = \max\left\{ M_j = \sum_{i=1}^{N} P_i Z_{ij} : j = 1, \ldots, M \right\}, \tag{13}$$

where $N = 5$ is the dimensionality of the PFC input vector and $M = 45$ is the number of matrisomal cells.

The above equations follow the mathematical description of ART-2 of Carpenter and Grossberg (1987). However, the asymmetrical, non-Hebbian learning rules for the adaptive cortico-striatal (Z_{ij}) and striato-cortical (Z_{ji}) pathways are modified in the present model so that dopamine activity gates the learning process. For category (F1 $\rightarrow$ F2) learning,

$$\frac{d}{dt} Z_{ij} = 0.1 \cdot M_j (P_i - Z_{ij}) D, \tag{14}$$

such that the weight vector $(Z_{1J}, \ldots, Z_{MJ})$ converges to the PFC signal vector $(P_1, \ldots, P_M)$ when the Jth category (M_J) node is active and dopamine activity (D) are present; but that weight vector remains unchanged when a different matrisomal category is active or when DA neurons are silent. For pattern learning (F2 $\rightarrow$ F1),

$$\frac{d}{dt} Z_{ji} = 0.1 \cdot M_j (P_i - Z_{ji}) D, \tag{15}$$

such that the weight vector $(Z_{J1}, \ldots, Z_{JM})$ converges to the activity pattern vector $(P_1, \ldots, P_M)$ when the category node (M_J) is active and dopamine (D) neurons are active simultaneously.

2.2.3. Orienting Subsystem. The winner matrisomal activity (M_w) can, however, be reset by a mismatch between current stimulus-activated representations and

learned motivational expectancies (Fig. 3). The degree of match and mismatch between the signal vector P_i at F1 and an active long-term memory (LTM) pattern (Z_{ji}) elicited by neuron M_j at F2 is determined as in Carpenter and Grossberg (1987) by

$$N_i = \frac{k_i + 0.1P_i}{1.1}, \tag{16}$$

where N_i are the elements of the vector of match and mismatch activities, P_i is the output of the F1 neurons, and k_i is the preprocessed input. A reset is generated whenever an input pattern is active and

$$\|N\| < \rho, \tag{17}$$

where ρ is the vigilance parameter that determines the tradeoff between discrimination and generalization (the larger the ρ value, the lesser the generalization to similar input patterns). This reset signal activates DA cells, and these in turn cause a strong, lasting inhibition of the current matrisomal representation (M_i).

2.2.4. Conditioned Reinforcer Learning. Once a sensory-motivational representation (M_j) is repetitively selected and becomes stable during learning, it can become a predictor of appetitive stimuli through a conditioned reinforcer adaptive (R_j) pathway (see Fig. 1)—

$$\frac{d}{dt}R_j = M_j(-0.1R_j + 5(1 - R_j)D). \tag{18}$$

2.2.5. Dopamine Dynamics. Dopamine cell (D) dynamics are modeled as

$$\frac{d}{dt}D = -6D + (0.5 - D) \times \left(1 + N + r + 10\, f_s \sum_j (M_j \cdot R_j)\right) - 10(0.5 + D)\sum_j ([S_j]^+ \cdot C_j), \tag{19}$$

where N is the mismatch or novelty signal (Eq. (17)), $r = 3.0$ indicates the onset timing of primary reward, M_j is the activity of matrisomal representations, S_j is the output of the striosomal circuit, R_j is a weight that signals the amount of conditioned reinforcer associated with matrisomal population M_j, $[x]^+ = x$ if $x > 0$ or zero otherwise, $f_s(t) = u(t) - u(t - 0.1)$ is a unitary phasic pulse function of time, and C_j is a weight that selects the striosomal populations that best signal the expected time of reward delivery.

2.2.6. Coding of Input Patterns and Temporal Sequence of Inputs. The behavioral tasks simulated were free liquid delivery outside a task (Schultz et al., 1997), the timed, contingent delivery of appetitive and aversive stimuli (Mirenowicz and Schultz, 1996), and an asymmetric, instruction-dependent Go/NoGo task (Schultz and Romo, 1990). The sensory stimuli were represented as 1D vectors of length $N = 5$ that combined binary and analog elements. Each element of the input pattern coded whether the stimulus was (1) auditory or (2) visual, (3) its spatial location (e.g., medial, lateral, center, or bottom, top depending on the task), (4) its shape (square, round, none), or (5) its color (yellow, green, red, none). Other representation schemes could also be used as the categorization process is self-organized (that is, number of categories does not need to be specified a priori, although in the present simulations an upper value of 45 representations was chosen for computational reasons), and the level of pattern discrimination can be varied with the vigilance parameter (Eq. (17)). The temporal sequence of input patterns were simulated according to each experimental paradigm.

Timed, Contingent Delivery of Appetitive and Aversive Stimuli. The appetitive round green light 10 mm above lever was coded as $I^+ = [0.0, 1.0, 1.0, 0.0, 1.0]$; the aversive square yellow light 40 mm above lever was coded as $I^- = [0.0, 1.0, 0.0, 0.5, 0.5]$; the appetitive sound located 10 mm above lever was coded as $I^+ = [1.0, 0.0, 1.0, 0.0, 0.0]$; and the aversive square red light located 40 mm above lever was coded as $I^- = [0.0, 1.0, 0.0, 0.5, 0.5]$. These patterns are illustrated in Fig. 4A. Inputs to the simulation were given in randomly alternating appetitive and adversive trials according to the experimental paradigm described in Fig. 3 of Mirenowicz and Schultz (1996). Simulated appetitive trials were followed by a reward input ($r = 3.0$ for 10 time steps or 100 ms) to DA neurons given 1.3 simulated seconds after the onset of the tone or light. Simulated aversive trials were not rewarded (that is, reward = 0.0). The intertrial interval was randomized (range = 5.4 to 9.4 s). Transient visual or auditory (100 ms duration) stimuli were stored in short-term

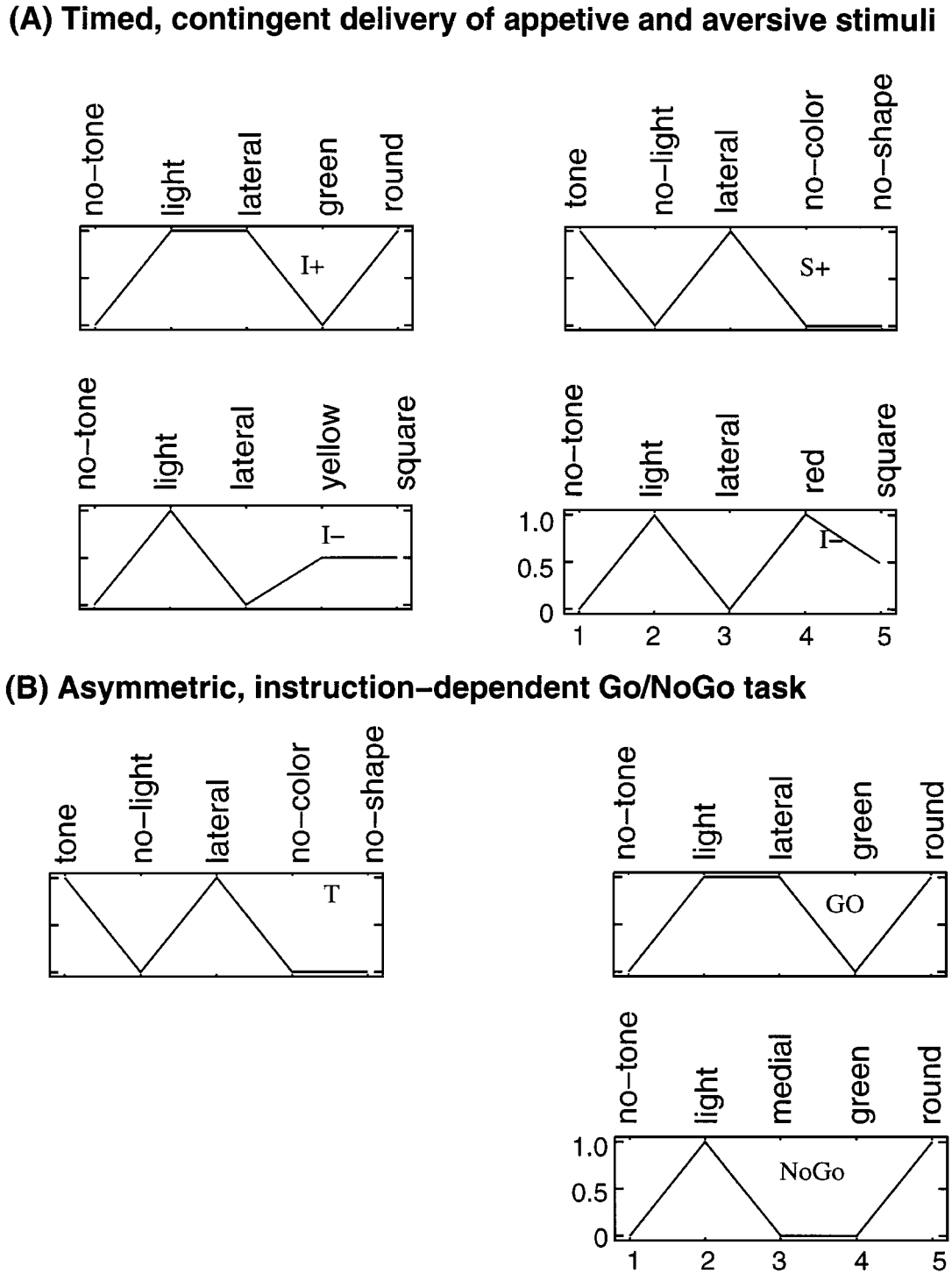

Figure 4. Coding for the analog input patterns used in the simulations.

memory (STM) at level F0 neurons during the duration of one trial or until a new input vector was presented to the system.

Asymmetric, Instruction-Dependent Go/NoGo Task. Trigger sound (door opening) was represented by the vector $I = [1.0, 0.0, 1.0, 0.0, 0.0]$, whereas the Go-instructed (lateral, green, round) light was coded by $I^+ = [0.0, 1.0, 1.0, 0.0, 1.0]$, and the NoGo-instructed (medial, green, round) light was given by $I^- = [0.0, 1.0, 0.0, 0.0, 1.0]$. The instructed stimuli differed only in the horizontal spatial location. Inputs to the simulation were given in randomly alternating Go/NoGo trials, first with the door-opening sound alone and then with the instructed lights followed by the door (trigger) sound as described in Schultz and Romo (1990). The inter-stimulus-interval (ISI) was randomized (range = 2 to 3 s). The intertrial interval was randomized (range = 4 to 8 s) after the reward acquisition in appetitive Go trials or after 4 s in nonrewarding NoGo trials. These patterns are illustrated in Fig. 4B.

In the model, all the processes—except the striatal competition, neurotransmitter dynamics, DA activity, and adaptive weights—are assumed to react so quickly that they can be represented at equilibrium as algebraic equations. However, at each time step, both algebraic and dynamical equations were recalculated. All simulations were performed using C on a Silicon Graphics Indy workstation. The initial levels of neurotransmitter in the striosomal network (that is, the initial condition) were computed from a uniform distribution between 0.4 and 1.0. The differential equations were solved using the classical fourth-order Runge Kutta method with a time step of 0.01.

3. Results

Three main sets of results are presented. First, we show simulations that illustrate the role of DA cells in signaling errors in the *timing* of reward prediction by comparing the onset timing of a striosomal reward prediction signal and the actual time of reward delivery. Second, we study why DA cells (1) appear to be preferentially activated by appetitive rather than aversive stimuli and (2) show generalization to physically similar but motivationally different stimuli. Here, we propose that reward responses observed in PFC_{dl} (Watanabe, 1996) may signal errors in the quality (such as type and amount), as opposed to the timing of delivery, of reward prediction. Finally, the role of the environmental context in shaping the response of DA cells is shown in simulations involving an asymmetric, instruction-dependent Go/NoGo task (Schultz and Romo, 1990).

3.1. *Dopamine Signals Errors in the Timing of Reward Prediction*

Figure 5 shows experimental (left) and simulation (right) data depicting the DA responses to (A, D) the unexpected delivery of primary rewards (R) outside a task, (B, E) conditioned stimulus (CS) predicting a reward after established task performance, when the reward occurs; (C, F) conditioned stimulus when the reward fails to occur (no R) due to the lack of reaction of the animal; and (G) when the reward occurs later than expected, or (H) earlier than expected. The first three panels correspond to data shown in Schultz et al. (1997), while the last two panels represent the model's predictions when the timing of reward delivery is unexpectantly changed. The experimental data and the simulations show that the DA response transfers from the primary reward to the conditioned, reward-predicting stimulus (CS) during learning. These data show that the CS becomes a predictor of reward, and indeed it becomes a substitute for the primary reward in eliciting a DA response. The lack of DA response to the reward reflects the occurrence of the reward at the time predicted by striosomal cells.

Furthermore, the DA response is independent of the behavioral reaction to the CS as it also occurs even when the animal erroneously fails to act (as depicted in Fig. 5C) or when the reward appears later or earlier than expected (depicted in Figs. 5G and H). The depression in DA responses in Figs. 5F and G in the simulations (and presumably in the experimental data shown in Fig. 5C) are due to the inhibitory activations from the simulated striosomal timing circuit, which predicts the *timed* occurrence of a reward, but the reward either does not appear at all (as in Fig. 5C) or occurs at a later time. In the latter case, the delivery of the reward provokes a DA activation due to both the reward's appetitive value and its unexpected occurrence at that point in time. Finally, Fig. 5H shows an activation due to the CS and an activation due to the earlier occurrence of the reward. In this case, no DA depression is seen at the expected time of reward delivery because the earlier-than-expected reward resets the predictive timing circuit and the current sensory-motivational

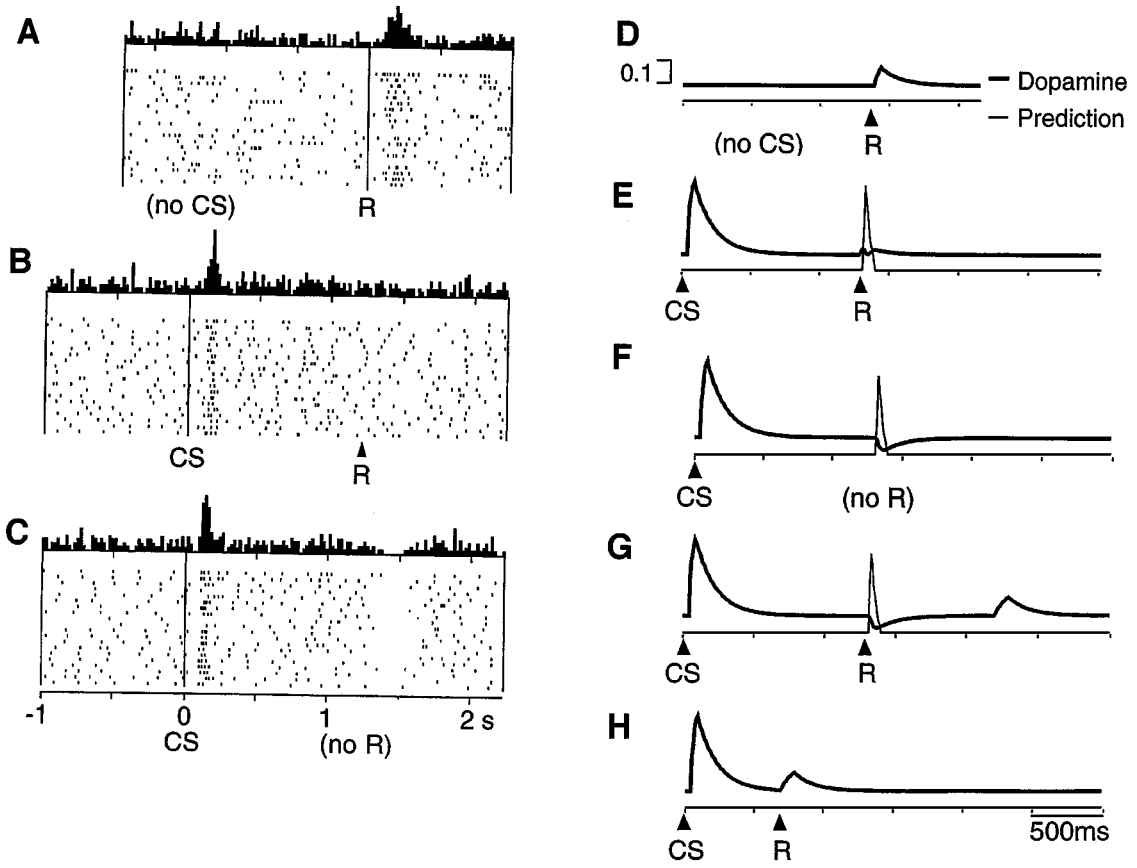

Figure 5. Dopamine cells signal errors in the timing of reward prediction. Experimental (left panels) and simulated (right panels) DA data are shown. Simulations also show the computed prediction of the expected time of reward delivery. Dopamine activity is shown in the case of the following. A, D: Free delivery of reward in the absence of behavioral task. B, E: Reward prediction (after learning) accompanied by actual delivery of the reward. C, F: Unexpected nonoccurrence of reward when reward is predicted, in which case the dopamine cell is activated by the conditioned stimulus (CS), which predicts a reward. The reward does not occur due to lack of reaction of the animal, and a depression in DA cell activity appears at the expected time of reward delivery (such as prediction). G: Later-than-expected reward delivery showing an activation by the reward-predicting CS followed by a depression at the expected time of reward delivery and an activation due to the actual later-than-expected delivery of the reward. H: Earlier-than-expected reward delivery in which case the reward-predicting stimulus causes a dopamine activation but the earlier-than-expected reward results in subsequent dopamine activation that is not followed by a depression at the expected time of reward delivery. The last two cases are predictions of the model system shown in Fig. 1 and have been recently tested and supported experimentally by data from Hollerman and Schultz (1996). Experimental data reprinted with permission from Schultz et al. (1997), A neural substrate of prediction and reward, *Science* 275:1593–1599, Copyright 1997 American Association for the Advancement of Science.

representation. Recently, Hollerman and Schultz (1996) have corroborated experimentally the model's predictions shown in Figs. 5G and H, which support the view that DA cells signal an error in the *timing* of reward prediction.

3.2. Dopamine Cells Respond Preferentially to Appetitive Stimuli

Mirenowicz and Schultz (1996) have compared DA cell responses to appetitive and aversive stimuli of similar and different modalities to study whether DA responses are concerned with specific motivational attributes or reflect more general stimulus salience properties. Figure 6 depicts experimental data (left panel) from Mirenowicz and Schultz (1996) and the simulated data (right panel). The latter assumed that the aversive stimuli (such as an air puff delivered to the monkey's hand or saline solution delivered to the mouth) did not require learning of a specific timing for expecting the punishment as the monkey was required to withdraw its hand from the resting key *within* 1.5 s after the visual

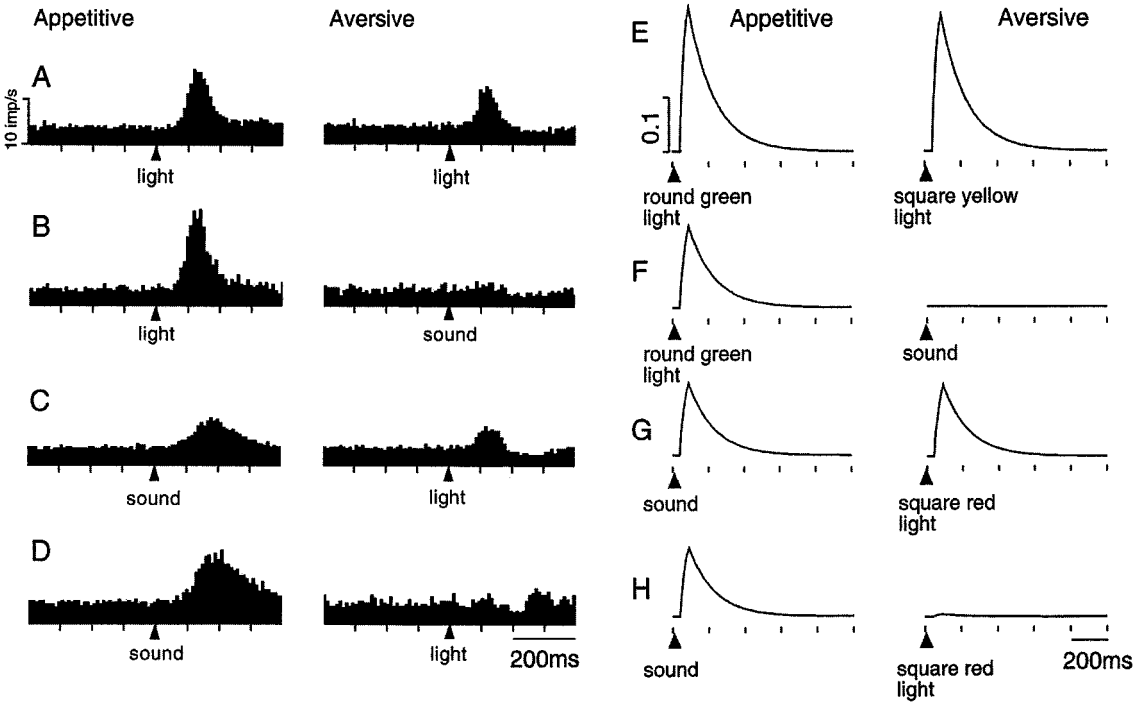

Figure 6. Chronology of experimentation from A to D and the corresponding simulations from E to H. Population histograms depicted in the left panels show neuronal data from the four combinations of randomly alternating appetitive and aversive trials. Simulated average membrane potentials depicted in the right panels show simulated data from the same four combinations of randomly alternating appetitive and aversive trials. Number of neurons tested experimentally varies from A to D. Data are separated according to trial type. A: Activations in the first monkey elicited by two similar lights serving as appetitive (round green light) and aversive (square yellow light) conditioned stimuli in randomly alternating trials. B: Subsequent separation of sensory modalities revealed an exclusive activation for appetitive (round green light), thus implying the DA response to aversive stimulus shown in panel A is due to generalization to appetitive visual stimulus of different shape and color. C: Subsequent reversal of modalities shows the efficacy of appetitive sound stimuli for activating dopamine cells and an activation by the aversive (square red) light due to the previous appetitive training with lights. D: Lack of activation by an aversive square red light in a second monkey, not experienced with visual appetitive stimuli but activated by the appetitive sound, suggests that the visual aversive responses in A and C were due to stimulus generalization. The right panel shows the simulated sequence of trials with the same chronology of experimentation as in the left panel. For each display, histograms from each neuron tested were normalized for trial number and added together, and the resulting sum was divided by the number of neurons. Each simulated trace is the average of 50 trials. The vigilance threshold was set equal to $\rho = 0.96$. Experimental data reprinted with permission from Mirenowicz and Schultz (1996), Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli, *Nature* 379:449–451, Fig. 3, Copyright 1996 Macmillan Magazines Limited.

stimulus to avoid the mild air puff to the hand or the delivery of the saline solution to its mouth. Therefore, it is highly likely that predictive timing was not required. As neither reward nor predictive timing inputs were delivered to the model DA cells, no activation was generated by DA cells except for the responses due to generalization in the cases where the aversive stimuli were similar to previously seen appetitive stimuli (e.g., Figs. 6A, C, E, and G).

Both the experimental and the simulated data show a preferential activation of DA cells by appetitive stimuli, albeit some DA responses are also obtained in response to aversive stimuli sharing the same modality as the appetitive stimuli (Figs. 6A, C, E, and G). Thus, these results show that DA cells are broadly tuned to sensory stimuli of the same modality and that they are capable of discriminating stimuli from different modalities (e.g., Figs. 6B, D, F, and H). It is likely that the generalization responses seen within stimuli of the same modality but different motivational significance (e.g., Fig. 6A) or stimuli across modalities (e.g., Fig. 6C) resulted from the context of previous learning that may suggest to the animal that *similar* stimuli have been followed by reward in the recent past. In these simulations, similar stimuli (such as the same modality) do not cause a reset for the given value of vigilance threshold chosen ($\rho = 0.96$; see Eq. (17)) resulting in generalization. When the vigilance parameter is small,

no reset can ever occur, whereas a high vigilance would cause the network to discriminate subtle stimulus features. Thus, the vigilance threshold regulates the tradeoff between generalization and discrimination. Furthermore, the neural threshold for novelty or mismatch detection may depend on the attentional level of the animal (see Carpenter and Grossberg, 1987). A value of ρ close to 1 was used in the simulations to favor discrimination over generalization.

The generalization and discrimination responses are explored in the simulations shown in Fig. 7, which depicts the simulated sensory-motivational representations or categories learned during the appetitive-aversive trials shown in Fig. 6. The simulations shown in Fig. 7 are divided in two parts as in the experimental paradigm designed by Mirenowicz and Schultz (1996) depicted in Fig. 6. Figure 7A depicts the selected matrisomal representations and their associated conditioned reinforcer weights (R_j; see Fig. 1) corresponding to the simulations shown in panels E to G of Fig. 6. These simulations correspond to the first "monkey" data of Mirenowicz and Schultz (1996). Simulations for the second "monkey" are shown in Fig. 7B for the 100 simulated trials starting from H (see Fig. 6H). During the first 100 simulated trials from E to F, categories (matrisomes) 25, 26, and 31 are repetitively selected to code for the appetitive value of the sensory stimuli (see Fig. 6E). Selection of these categories during appetitive trials lead to a strenghtening of the conditioned reinforcer weights associated with these categories (e.g., R_{25}, R_{26}, and R_{31}). Early during this learning phase, matrisomal cell 25 is selected, but after about 25 learning trials the number of categories is reduced, and only categories 26 and 31 are activated by the appetitive visual stimulus. As a result, the conditioned reinforcer weight associated with category 25 (e.g., R_{25}) decreases to zero. Learning associated with categories R_{26} and R_{31} is responsible for the generalization responses to the aversive square yellow light (Fig. 6E), as this aversive stimulus is not discriminated from the appetitive round green light for the given value of vigilance ($\rho = 0.96$) used.

During the next 100 simulated trials from F to G, the lack of response of the DA cell to the new aversive stimulus does not allow the sound stimulus to gain access to an appetitive sensory-motivational representation and no new representations are formed. However, the previous categories, represented by matrisomes 26 and 31, formed in response to the appetitive light continue to be selected during these trials. Moreover, during the final 100 simulated trials starting from G, category 31 is gradually weakened, and its associated conditioned reinforcer weight decays to zero by the end of this phase. Furthermore, category 26 remains activated, and a new category 37 is formed to code for the appetitive sound for the last 100 trials. The corresponding conditioner reinforcer weight (R_{37}) associated with the "appetitive" sound coded by category 37 gradually increases during learning.

The second part of the simulation corresponds to new learning in the model (the second "monkey" of Mirenowicz and Schultz, 1996). For this simulation, the model was started again from initial conditions (no previous learning). The plots depicted in Fig. 7B show that during initial learning three matrisomal representations are selected—namely, 2, 25, and 37. However, as learning progresses and the system learns to predict the occurrence of appetitive stimuli, only category 37 is established to code for the appetitive sound as in the first part of this experiment (from G on). The associated conditioned reinforcer weight gradually increases in strength during this phase.

3.3. Instructed Go/NoGo Task

Schultz and Romo (1990) have compared the effect of preparatory instruction cues presented several seconds before overt behavioral reactions to stimuli eliciting immediate behavioral reactions. In an asymmetric instruction-dependent Go/NoGo task, a green instruction light located above a lateral or medial food box was illuminated 2 to 3 s before door opening to prepare the monkey to perform a reaching movement on door opening (that is, trigger signal or Go task) or to refrain from moving (No/Go task), respectively. Only Go trials were rewarded if the reaching movement was performed within 800 to 1000 ms after box opening. The trigger signal was the sliding noise from door opening and did not indicate which box had opened nor the expected behavioral reaction. Figures 8A and B depict experimental data from Fig. 14 of Schultz and Romo (1990). The experimental data show the influence of preparatory instruction lights on responses to door opening in 1 DA cell. Both perievent time histograms and dot displays of neuronal impulses recorded in the same trials are shown, as well as the horizontal electrooculograms. Figures 8C and D show the simulated response for 1 model DA cell (eye movements were not simulated).

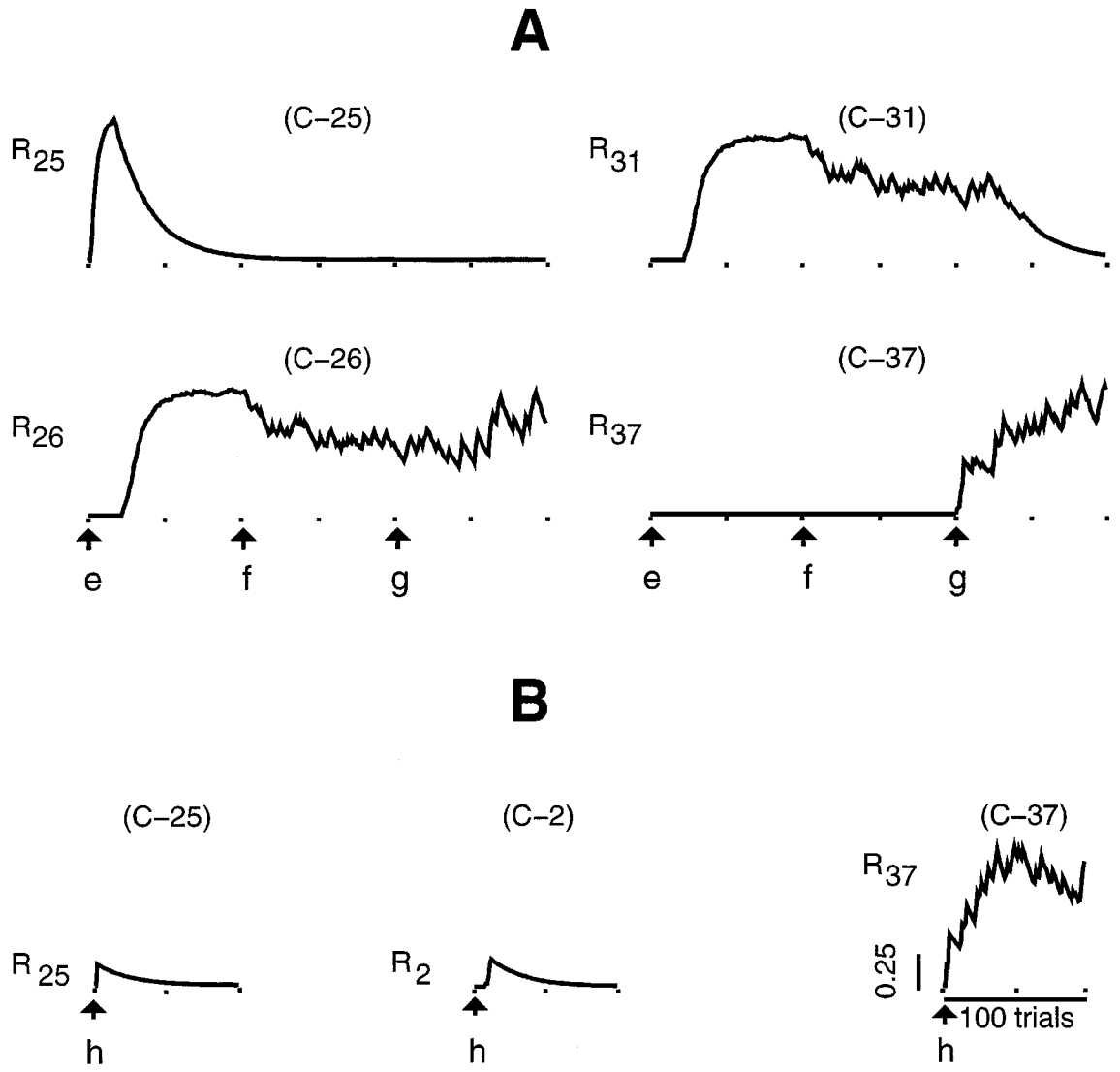

Figure 7. Selected sensory-motivational matrisomal representations ("categories," C) and conditioned reinforcer weight (R_j in Fig. 1) learning during the simulated experimental chronology shown in Fig. 6. These matrisomal categories represent simulated electrophysiological responses to unique biologically significant stimuli, and their associated conditioned reinforcer weights represent hypothesized changes in synaptic plasticity in the pathway between the striatal matrisomes and dopamine cells. A: These panels depict the selected matrisomal categories and the conditioned reinforcer weight traces corresponding to the first simulated "monkey" of Mirenowicz and Schultz (1996) depicted from panels E through G in Fig. 6. During appetitive trials with the round green light from E to G, categories 26 and 31 are formed and consistently selected. This leads to a strengthening of the conditioned reinforcers (R_{26} and R_{31}) associated to these categories. However, during the appetitive trials with the sound, a new category 37 is formed, and a previous category (31) is not longer associated with the light resulting in weakening of its associated conditioned reinforcer weight. The remaining category (26) associated previously with the appetitive light is responsible for the generalization shown to the aversive light stimulus in Fig. 6G. B: These panels depict the categories and conditioned weight learning corresponding to their second "monkey" shown in panel H in Fig. 6. During reversal learning with the second "monkey," category 37 is selected, and the conditioned reinforcer (R_{37}) associated with this category is strengthened during the appetitive sound trials. However, no categories for the new aversive light stimulus are formed. Only categories with conditioned reinforcer weight learning are shown.

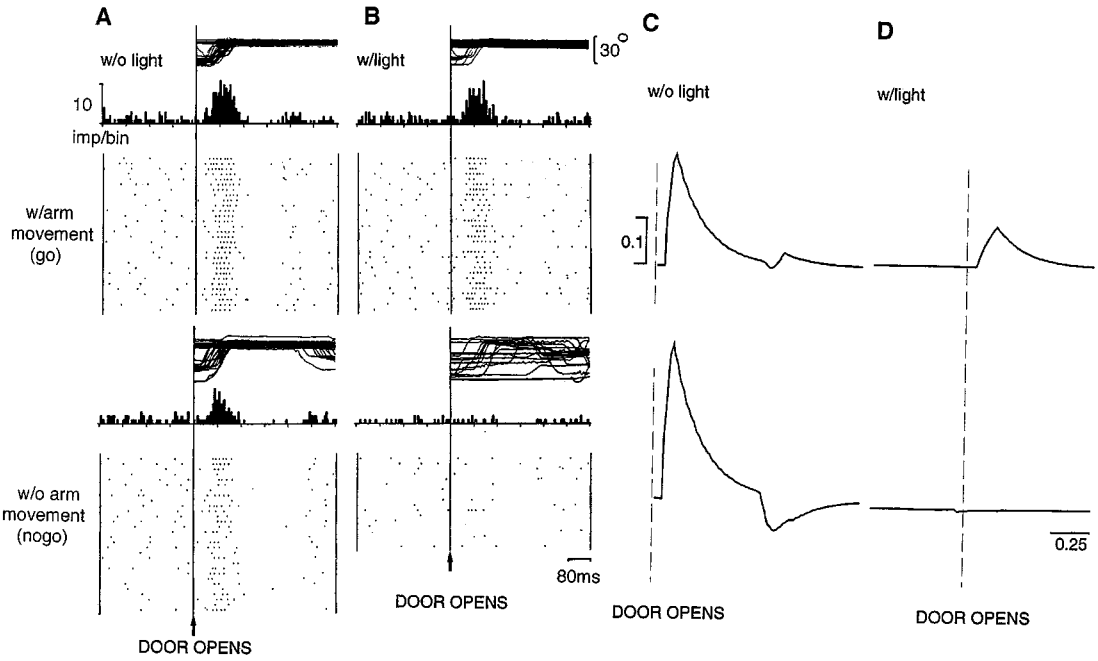

Figure 8. Influence of preparatory instruction lights on responses to noise of door opening in 1 dopamine neuron (LEFT) and 1 simulated DA neuron (RIGHT). A: Responses to door opening in Go and NoGo trials without instruction lights. Sound of door opening did not indicate which food box had opened. B: When the green instruction lights located above each box were used while testing the same neuron in a separate session, responses to door opening disappeared in NoGo trials (bottom) but esentially remained in the Go task (top). In parallel, the eyes showed a higher tendency for fixating on the food box at the same time the door opened in Go trials. The experimental and modeling data suggest that the contextual information provided by the instruction light is responsible for the absence of DA responses to door opening as this stimulus is no longer capable of triggering behavioral reaction. The vigilance threshold was set equal to $\rho = 0.99$. Experimental data reproduced with permission from Schultz and Romo (1990), Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioral reactions, *J. Neurophysiol.* 63(3):607–624, Copyright 1990 The American Physiological Society.

Figure 8A shows the DA responses to stimuli eliciting immediate behavioral reaction in the absence of preparatory stimuli. In this case, the DA neuron responded to both the Go and the NoGo tasks with an activation followed by a depression that was larger for the NoGo condition. The prominent DA depression in the NoGo condition was presumably due to the error in reward prediction, as the reward did not occur. The simulated DA response shown in Fig. 8C qualitatively reproduces the experimental data in Fig. 8A. Figure 8B shows the DA responses when the preparatory instruction signals precede by 2 to 3 s the trigger noise. When the preparatory light (lateral light) instructed a Go trial, the DA cell was activated by the trigger signal; however, when the instruction light corresponded to the NoGo trials (medial light), the DA cell was not activated by the trigger signal. Moreover, the electrooculogram in the latter condition showed random eye movements, as the monkey's attention was not attracted to the food box. The simulations shown in Fig. 8D are qualitatively equivalent to the experimental data and illustrate that contextual information is used during the delay period to guide appropriate approach behavior.

4. Discussion

4.1. Biological Correlates

The present model may be related to anatomical, neurophysiological, and neurochemical data underlying the cortico-striatal circuits and their role in behavioral learning (Alexander et al., 1986; Gerfen, 1992; Eblen and Graybiel, 1995). It is proposed that a striosomal-based (Gerfen, 1992), predictive timing circuit generates temporal activations that are used to signal the expected time of occurrence of appetitive or aversive stimuli. These activations may be developed during

learning in response to the delivery of motivationally significant stimuli that predicts a specific, timed, behavioral response. The predictive timing circuit may involve the posterior orbitofrontal/anterior insular cortex and the mediofrontal prelimbic/anterior cingulate cortex, which jointly innervate the striosomal-rich anterior and ventromedial striatum (Eblen and Graybiel, 1995). The striosomal neurons are known to project to DA neurons in the SNc and the islands in the pars reticulata (SNr) (Jimenez-Castellanos and Graybiel, 1987, 1989; Graybiel, 1990; Gerfen, 1992) and to the ventral pallidum, VP (Haber et al., 1990).

It is proposed that inhibitory axon collaterals, interneuronal subsystems within the striatum, or opposite effects of dopamine in the activity of different populations of striatal neurons may provide winner-take-all computations by which the neuronal population with the highest activation is selected while the remaining cells are suppressed (Wickens et al., 1991; but see Section 4.3, Limitations of the Model, below). These competitive interactions in the presence of neurotransmitter dynamics (such as accumulation and depletion) of striatal peptides may underly a mechanism by which a given cell population can be active only temporarily, therefore allowing other cells to become active at successive times. In general, it is possible that more than one striosomal neuron be active at any given time, resulting in a population coding of time. Several other mechanisms for computing temporal representations have been proposed in the literature (e.g., Fiala et al., 1996; Montague et al., 1996; Buonomano and Mauk, 1994; Grossberg and Merrill, 1992); however, those mechanisms have been mostly based on cerebellar or hippocampal data. Furthermore, it has been suggested that cerebellar or hippocampal timing circuits may operate over a relatively short time window and that the basal ganglia may operate in tasks that require longer time windows (Ivry, 1996). Recently, Meck (1996) has proposed that the internal clock used to time durations in the seconds to minutes range appears to be linked to DA function in the basal ganglia. In this regard, Meck suggested that DA D2 receptors may play a role in determining the rate of temporal integration for time perception as DA agonist, or antagonists can speed up or down the clock speed, respectively.

In the model, PFC_{dl} cells can gain access to DA cells through two projections: (1) via indirect projections to cells in the striatal matrix at layer F2, which in turn project to DA cells, and (2) through direct projections to the midbrain DA cells from layer F1 (Kornhuber et al., 1984). Sensory-motivational representations in the striatal matrix in turn can form adaptive links (through conditioned reinforcer weights, R_j) to DA neurons during learning of approach behavior. Thus, these links are responsible for the DA activations to conditioned stimuli after learning. Conversely, the direct projection from PFC to DA cells are prewired and are used by PFC neurons to signal mismatches between motivational expectancies and behavioral outcome (such as novelty). These projections are critical for organisms to generate orienting reactions to new, salient stimuli, which itself may lead to new learning in the cortico-striatal (indirect) pathway.

It is also known that neurons from the striatal matrix send projections to the GABAergic interneurons and principal neurons of the SNr. It may be that the phasic excitatory input from the matrisomal cells to DA neurons in the present model are mediated by SNr axon collaterals that terminate on DA neurons within the zona compacta (for a review, see Pucak and Grace, 1994). In this regard, neurons in SNr are known to transiently increase or decrease their activity shortly after a preparatory or instructive auditory or visual stimulus indicating a behavioral situation (Schultz, 1986).

The simulations shown in Figs. 6 and 7 suggest that there should be some populations of neurons within the striatal matrix that code for the biological significance of arbitrary sensory stimuli. In this regard, Nishino et al. (1981) have found responses of caudate nucleus neurons that were food-specific (for orange, bean, or cookie). Furthermore, some neurons responded not only to the sight of food but also during bar pressing to obtain food, which may represent motivational activity related to the bar pressing to obtain food. The latter activity was fairly constant throughout the bar-pressing period. Neurons in the monkey ventral striatum show both long-term tonic changes in activity in anticipation of reward (Schultz et al., 1992) as well as phasic activity in response to the delivery of primary reward (Apicella et al., 1991; Bowman et al., 1996). These learned activations seen in ventral striatum may be the neurophysiological correlate of our hypothesized long-term sensory-motivational representations. Watanabe (1996) reported that some PFC_{dl} cells show reward-dependent, sustained activity during the delay period between the presentation of a cue and the response in a delayed response task. Furthermore, these cells responded differentially to the type of food (cabbage, potato, raisin) as in the Nishino et al. (1981) data. The experiment of Watanabe (1996) supports the

view that PFC neurons compute a match and mismatch operation between motivational expectancies and the outcome of behavior. Specifically, Watanabe found that reward-related delay neurons showed responses after delivery of unexpected rewards (for example, after reward change from grape juice to potato). The model also predicts that PFC neurons should also respond to novel cues (such as after a cue change from a red light to a sound). Overall, the differential responses of striatal and PFC cells to the sight of or predicted type of food and the responses after delivery of unexpected rewards suggest that these neurons are involved in evaluating and learning the motivational significance of arbitrary stimuli.

4.2. Predictions of the Model

Several testable predictions for future experiments are put forward by our model. Perhaps the most critical prediction is the existence of two reward-related error prediction signals: a signal that conveys information about errors in the timing of reward prediction and a signal that codes for errors in the amount and type of reward prediction. The former is consistent with the responses of DA neurons (Mirenowicz and Schultz, 1996), while the latter is consistent with reward expectancy responses observed in PFC (Watanabe, 1996). Watanabe (1996) has shown that prefrontal cortex (PFC) cells may be involved in computing reward expectancy and that their response is dependent on both the particular reward received for the behavioral response and the way the reward is given. These PFC responses may be related to the DA activations as there is evidence of direct reciprocal projections between DA and PFC cells (Porrino and Goldman-Rakic, 1982; Künzle, 1978; Gaspar et al., 1992), as well as indirect, highly specific, prefrontal projections through the striosomal subsystem to the dopaminergic neurons (Eblen and Graybiel, 1995). In agreement with the anatomical data in the rat, recent studies suggest that medial PFC regulates burst firing and transmitter release in rat mesolimbic dopamine neurons (Murasse et al., 1993). Gariano and Groves (1988) and Tong et al. (1996) have also shown natural burst firing in midbrain dopamine cells by stimulation of the prefrontal cortex in the rat. However, evidence for monosynaptic, as opposed to the indirect pathways through the striatum, projections from prefrontal cortex to the SNc and VTA in the primate is far from clear and accepted. Nevertheless, synaptic innervation of midbrain neurons by extrinsic glutamate-enriched terminals, presumably from prefrontal cortex, in the squirrel monkey have been reported (Smith et al., 1996). The PFC projection to DA cells depicted in Fig. 1 can then be considered an anatomical prediction of the model.

Although the PFC neurons of Watanabe may therefore appear to be responsible for sending the prediction to DA cells, this possibility is weakened by the following observations: PFC delay neurons show changes in activity during the delay period (see, for example, Fig. 2(b) of Watanabe, 1996). If this signal is sent to the DA cells for computing the error in reward prediction, the DA activity would show an increasingly larger depression that would end at the time of delivery of the actual reward (assuming it is delivered) if we follow the DA = reward-prediction rule; unless of course we assume signal changes from PFC to DA neurons via synaptic processing. Furthermore, some PFC cells showed a different pattern of activity changes depending on the task, so any cell population (e.g., DA or striatal cells) that uses this PFC activity must also show task-dependent activity. However, the rather homogeneous DA responses may lack the degree of specialization needed to take into account the context of the task; in fact, DA cells appear to be prompt to generalize (Mirenowicz and Schultz, 1996). The task-dependent and reward-dependent responses seen in ventral striatal cells reflect the changes in PFC activity patterns with task variations (Schultz et al., 1992; Bowman et al., 1996). We instead propose that DA cells receive a predictive timing signal that is generated at the striosomes and that is compared to information about the actual timing of reward delivery to generate an error signal. This is supported by anatomical data suggesting that striosomes, unlike the striatal matrix, project directly to the midbrain DA cells (Eblen and Graybiel, 1995; Gerfen, 1992). Conversely, the structure of the model suggests that sensory-motivational representations of approach behavior may be subserved by cortico-striatal circuits originating in PFC and that project to the extrastriosomal matrix (Eblen and Graybiel, 1995). The model's prediction of dopamine-sensitive timing computations in the basal ganglia is supported by recent human and animal experimental data (Ivry, 1996). Pastor et al. (1992) reported that patients with Parkinson's disease tend to underestimate temporal intervals. Moreover, positron emission tomography (PET) studies show that healthy subjects have a significant increase in regional cerebral blood flow (rCBF), with respect to the control condition, in the striatum and

globus pallidus when estimating time differences by comparing a test interval with a standard interval (Jueptner et al., 1995). Furthermore, after behavioral training, rats injected with DA agonists tend to respond earlier than expected, while rats injected with DA antagonists tend to respond later than expected (Maricq and Church, 1983). The specific predictive timing function assigned here to dopamine cells is important, as current models (see below) do not differentiate between the onset timing and the featural qualities of predicted primary reinforcement.

Dopamine cells appear also to receive mismatch or novelty signals from PFC neurons and use these signals to reset current sensory-motivational expectancies that allow orienting reactions to the new stimuli. Lesions to the DAergic cells or to the frontal cortical areas projecting to DA neurons may be responsible for perseveration deficits of formerly rewarding choices of action (Milner, 1963, 1964). Thus, DA cells seem to be a critical component of reward-related mechanisms that guide learning of approach behavior.

Our model also predicts that anatomical or pharmacological lesions of the DA projections to the matrix modules will disrupt the process of stimulus evaluation without affecting the predictive timing circuit. Therefore, it should be possible to test experimentally whether a pharmacological manipulation of the matrix, as opposed to the striosomal patches, will cause perseveration deficits (as novelty computations will be disrupted), prevent learning of new associations, but still allow the striosomal cells to compute a predictive timing signal.

4.3. Limitations of the Model

The proposed model provides a way to unify data obtained using different methods, preparations and paradigms in terms of hypothetical neural mechanisms underlying the effects of attention, motivation, novelty, and stimulus generalization or discrimination in the response of DA cells during learning of approach behavior. However, it necessarily simplifies the biology due to otherwise computational intractability (such as by using reduced neuron models) or the lack of information relevant to the brain structure under study. Specifically, the model makes the following assumptions or simplifications. First, the double inhibitory feedback loop through the pallidum and the thalamus was simplified as a single adaptive excitatory projection from striatum to cortex, both for computational convenience and because the computational properties of the striopallidal, pallidothalamic, and thalamocortical connections are still poorly understood (Parent and Hazrati, 1995). However, it is known that the basal ganglia are linked to the cerebral cortex via multiple segregated output channels (Hoover and Strick, 1993). Moreover, Sherman and Guillery (1996) have proposed that the functional organization of thalamocortical relays may allow for cortico-thalamic afferents to control response mode of the thalamocortical relay switching between bursting for close monitoring of changes in the relayed signals (such as enhanced vigilance) or tonic mode for accurate representation of relayed events. The latter data may be related to the proposed match or mismatch function of PFC neurons in the present model. Indeed, higher values of the vigilance parameter (ρ) in the ART subsystem would lead to the detection of smaller differences between sensory-motivational expectancies and the current sensory representation.

Second, the assumption of mutual inhibition between neighboring striatal neurons appears to be weakened by recent experimental data suggesting that surround inhibition among striatal neurons is weak or nonexistent in the rat (Jaeger et al., 1994). However, other studies favor the view of the striatum as a lateral inhibitory network (Plenz and Aersten, 1996; Rebec and Curtis, 1988). Moreover, differences in near versus distant striatal neighborhoods or in striatal organization among species may also be important in assessing surround inhibition in the primate striatum. For example, it has been shown that the ratio of projection neurons to striatal interneurons is significantly greater in the rat than in the primate, which suggests that interneuronal systems in the primate striatum play a greater role in its integrative and functional organization (Graveland and DiFiglia, 1985). Also, the absence of DAergic inputs in striatal slices may affect the activity of striatal neurons or their local axon collaterals (and hence inhibitory interactions) through differential effects on the neuropeptides they express (Contreras-Vidal and Stelmach, 1995). Nevertheless, the assumption of mutual inhibition among striatal neurons may be relaxed given that Carpenter (1997) has demonstrated an ART architecture with distributed, rather than local, coding. Therefore, the assumption of surround inhibition in the striatum is not critical for the sensory-motivational network in the model or for the timing model as far as the same distributed coding is selected after learning.

Third, the present model does not include projections from the amygdala to the ventral striatum (Russchen et al., 1985). However, surgical disconnection of the amygdala and the ventral striatum in monkeys does not impair learning on a two-choice visual discrimination task suggesting that the interaction between these structures is not critical for this type of stimulus-reward associative memory (Gaffan et al., 1993).

Finally, the choice of parameter values, including initial values for the long-term memory weights, for the sensory-motivational network were those used by Carpenter and Grossberg (1987) to ensure stability of the self-organization of recognition codes. The ART-2 model dynamics have been shown to be robust under various parameters and alternative ART architectures (Carpenter and Grossberg, 1987), and we do not review them here. A critical aspect of the model is related to the mechanism responsible for generalization among similar stimuli. In the present instantiation of the model, the degree of match between a pattern at PFC and a weight pattern activated by a striosomal representation is a function of the cosine of the angle between these vectors (see Fig. 7 of Carpenter and Grossberg, 1987). Other interpretations of match mechanisms may be used instead. For example, for all-or-none (that is, binary) patterns, a pattern match could be computed by counting matched bits.

On the other hand, the parameter values for the predictive timing network were selected so that a spectrum of striatal activations could be obtained, as the mutual inhibition, positive feedback, passive decay rate, and neurotransmitter depletion must be balanced to ensure a continuous temporal wave across the striatal layer. Of course, other choices of parameters for predictive timing network will likely change the characteristics of the temporal representation (such as number of active populations, temporal resolution, and clock duration).

4.4. Comparison to Other Models

Houk et al. (1995) outlined a reinforcement learning model of the basal ganglia in which unique functions of the striosomes and the matrix are incorporated. Specifically, they propose that the striosomal function may be related to that of an adaptive critic that learns to predict a reward; while matrisomal function may be associated with an actor, thus generating signals that command actions that regularly lead to the achievement of primary reward. Anatomically, both the Houk et al. and our model are built on the well-known reciprocal projections between the DA cells and the striosomal patches (Gerfen et al., 1987; Jimenez-Castellanos and Graybiel, 1987), which allow DA cells to predict earlier predictors of reward. Functionally, however, these models differ in the way the neural substrates are used to learn to predict primary or secondary reinforcement. We postulate that striosomal cells generate a spectrum of timing signals in response to an arbitrary sensory event and that during learning a few of these striosomal signals are selected or reinforced by dopamine signals in response to the delivery of primary reward, so that after learning the net DA response to the actual timed reward delivery and the predicted timing of reward delivery by the GABAergic striosomes cancel out. Our model predicts the high degree of homogeneity observed in the responses of the DA cells recorded by Schultz and colleagues, as the special qualities (such as type and amount) of actual or predicted primary reinforcement are evaluated by the matrix modules. Thus, we propose that the matrisomes form part of an ART architecture supporting processes such as learning, stimulus categorization and memory search, expectation and novelty detection (Carpenter and Grossberg, 1987).

Recently, Montague et al. (1996) proposed a model for DA systems based on predictive Hebbian learning. Specifically, they suggest that variations in DA from the VTA to cortical and subcortical structures globally signal errors in reward prediction. This model predicts both the time and magnitude of future rewarding stimuli. This is shared by our model as far as in the present model the reward prediction error is equal to the difference between the timing of reward and its predicted occurrence. Our model however departs from that of Montague et al. (1996) in two main points. First, although our model also predicts the timing of future rewarding stimuli as in the Montague et al. TD-based model, it is proposed that the prediction of the magnitude of future rewarding stimuli may not be a function of DA neurons per se, but rather a function of PFC neurons. It appears unlikely that DA cells could learn to predict or discriminate the specific reward attributes (besides onset timing of future reward) such as amount or type given that the response of these cells is rather homogenous and uniform, and they appear to be prompt for response generalization (Mirenowicz and Schultz, 1996). In fact, DA activity also appears to be insensitive to changes in internal drive for food or liquid or in the amount of reward consummed. In this regard, we have reanalyzed previous single cell data from five

monkeys, tested during free juice delivery outside a task, during the period 1992 to 1995 by Schultz and colleagues, which were the subject of previous publications on dopamine activity during learning of approach behavior. In these free-juice trials, animals received the same amount of apple juice (0.15 ml) at irregular intervals as DA responses tend to decrease progressively with repetitive presentation of primary reward delivered at regular intervals without performance of any task (Ljungberg et al., 1992).

Changes in dopamine activity were compared from session trials on Mondays to session trials on Fridays. The rationale was that the thrist drive would be higher for the monkeys on Fridays than on Mondays as the monkeys were subjected to limited juice consumption during the weekdays but had free, unlimited consumption during the weekends. Also, DA responses were analized for amount of reward consummed by comparig the first five trials to the twenty-fifth to thirtieth trials of the same session for each cell and for each day that a free reward session was recorded. The rationale was that for the latter trials, the thrist drive would be lower than for the first trials, and therefore the difference could influence the response of DA cells to free delivery of reward. These data were analyzed in terms of peak magnitude and onset timing using a Wilcoxon procedure described in Mirenowicz and Schultz (1996). Specifically, neuronal activations were compared against a 500 ms prestimulus control period by using a Wilcoxon procedure with a constant time window comprising 80 percent of onset and offset times of statistical significant increases ($p < 0.01$, 126 to 270 ms after liquid). Only those DA cells with significant changes in the time window with respect to the baseline activity were assessed (Mirenowicz and Schultz, 1996). Sixty-three neurons recorded on Mondays showed significant increases, and 68 DA cells recorded on Fridays showed significant increases in their response to free-reward delivery outside a task. However, there were not statistically significant differences between the two groups ($p > 0.22$) for magnitude or onset timing of their responses. The results were also nonsignificant when comparing the first five trials against the twenty-fifth to thirtieth trials. These data suggest that phasic changes in DA activation are not influenced by the level of thirst drive or amount of juice consummed. The prediction that DA neurons are insensitive to changes in the magnitude, but not the onset timing, of predicted future reward could be readily tested experimentally by varying unexpectantly the amount of actual reward delivered. If true, this will strongly suggest a purely predictive timing function for DA neurons. Otherwise, it will support the view that "greater than baseline dopamine activity" means the action performed is "better than expected" and "less than baseline" means "worst than expected" (Schultz et al., 1997, p. 1506). The proposed dichotomy of reward prediction could also be tested by differentially impairing the striatal matrix, while sparing the striosomal patches. It is predicted that this lesion would not impair the ability of dopamine cells to predict the timing of reward delivery, but it will cause a prefrontal cortex deficit in the (learned) prediction of reward attributes. In this regard, a similar mechanistic dissociation between the learning-dependent timing and the response (amplitude) characteristics of the conditioned eyelid response in the rabbit by the cerebellum have been postulated recently (Perret et al., 1993). Moreover, an independent, dopamine-dependent timing mechanism in the basal ganglia has been postulated recently (Ivry, 1996; Jueptner et al., 1995; Pastor et al., 1992).

A second novel aspect of the model with respect to the Montague et al.'s model is the mechanism that generates the expectation about the type and significance of future stimuli, the novelty response due to mismatch between sensory representations and sensory expectancies, and the specific differential roles for the striatal matrix-patch system and PFC. It is not clear how the Montague's model could account for the DA activity elicited by novel stimuli, or stimulus generalization, although it can account for learning with many different sensory cues (Schultz et al., 1997). The DA responses to novel stimuli and to reward-predicting stimuli but not to known neutral stimuli (for example, see Fig. 3 of Ljungberg et al., 1992) suggest an evaluation of the sensory stimuli based on its attributes and previous experience that cannot be accounted by current models of dopamine cell activity. Also, models of DA activity during sensory-reward learning should account for (1) the novelty-dependent initial response to sensory stimuli and (2) the time course of the responses to both the trigger stimulus and the primary reward, which resembles a push-pull process (see Fig. 9), rather than a traveling wave that goes backward in time with learning.

Suri and Schultz (1996) have proposed a model for the responses of DA neurons under various behavioral tasks based on temporal difference learning (Sutton and Barto, 1981). In their model, a predictive reinforcement signal associated with DA activity serves for learning

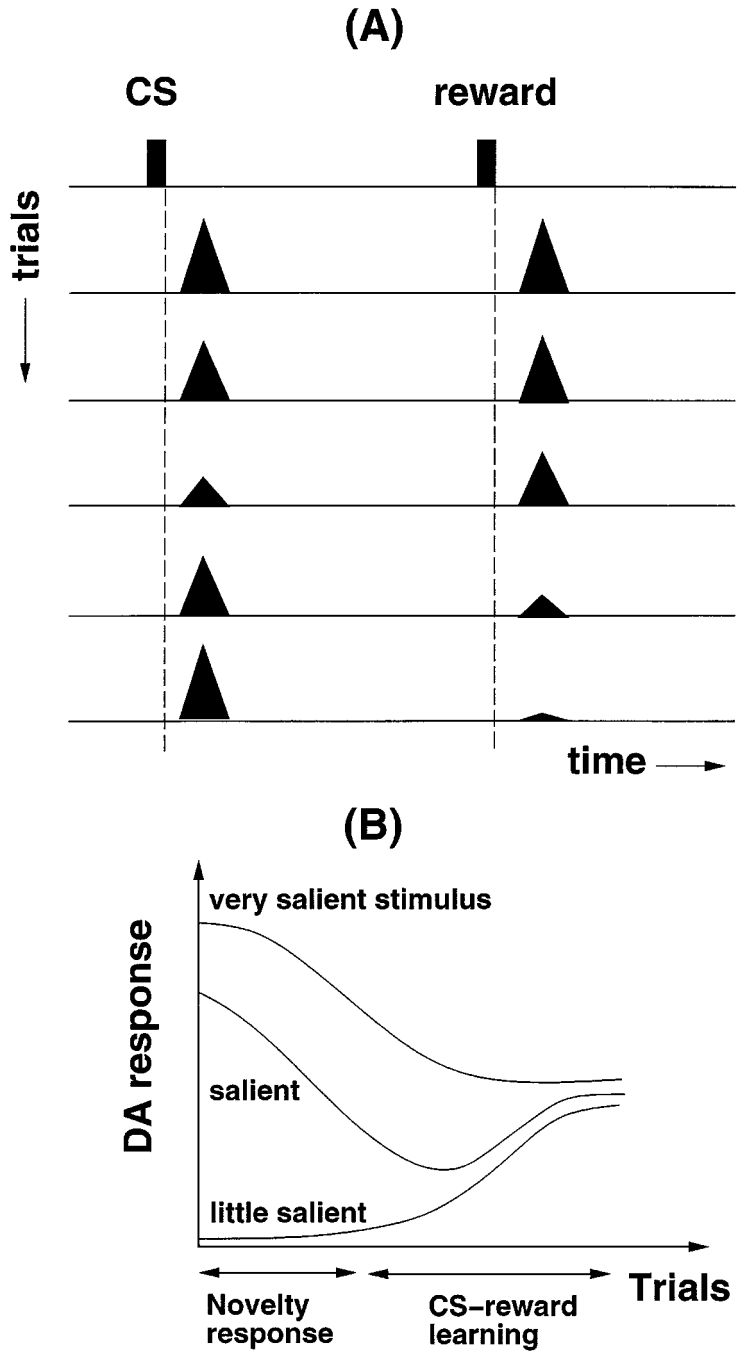

Figure 9. The novelty-dependent time course of dopamine response to the trigger stimulus and the primary reward. A: During the initial trials, DA cells fire due to the novelty of the trigger stimulus (CS) and to the primary reward. As the trigger becomes familiar, the novelty response decreases rapidly; however, as the trials continue, the DA cell increases its response to the trigger as it becomes a predictor of reward but it ceases to respond to the fully predictable primary reward. B: Depending on the degree of saliency or novelty of the sensory stimuli, three possible time courses of dopamine cell activity are possible. Novelty is critical for learning as it indicates the organisms that a new potentially rewarding stimulus has just appeared or that something has changed in the expected environmental conditions, which require a reorienting reaction to further evaluate these stimuli.

stimulus-action associations, stimulus-reward associations, and the motivational salience of stimuli. The Suri and Schultz model is similar to the Montague et al. model in that the former is also based on the view that DA activity codes a deviation between predicted and actually occurring reward. However, the Suri and Schultz model contains an attention mechanism, which allows only the sensory input with the highest motivational and physical salience to be represented in short-term memory, to account for DA responses to salient stimuli. This mechanism is lacking in the Montague et al. model. Nevertheless, it is not clear whether this attention mechanism refers to the selection of a salient stimulus from competing physical stimuli, to a matching process between sensory-motivational expectancies during learning or both. It is also not clear whether the Suri and Schultz or Montague et al. model can explain the PFC activations recorded by Watanabe (1996).

Levine and Prueitt (1989) have proposed an attentional gating mechanism that regulates the formation of stimulus representations within an ART system using reinforcement learning. Levine and Prueitt suggest that frontal cortex may be involved in regulating the effect of reinforcement signals (such as the gain) on "bias" or attentional mechanisms that regulate learning. However, they do not explain how brain subsystems involved in integrating motivational and cognitive information cooperate with other subsystems that link past and future events across time or predict future events, as their system does not explicitally model temporal events.

4.5. Clinical Applications

Basal ganglia have traditionally been implicated in movement initiation and execution. Parkinson's disease (PD), which results from the death of nigrostriatal DAergic neurons, is a good example of the movement impairments caused by basal ganglia dysfunction (Contreras-Vidal and Stelmach, 1995; Wichmann et al., 1995). However, recent experimental and clinical data support the view that basal ganglia are also involved in cognition and other higher brain functions (Brown and Marsden, 1990).

Cognitive dysfunction in PD has been associated with degenerative neostriatal and PFC disorders. Parkinsonian subjects appear to be impaired in conditional associative learning tasks (Canavan et al., 1989; Brown et al., 1993; Gotham et al., 1988; Sahakian et al., 1988). Canavan et al. (1989) asked patients to learn by trial and error which of six different movements of a handle are associated with six different color cues. They found that PD patients were capable of color discrimination, but a large proportion of these patients were impaired in learning the correct stimulus-response associations. Linden et al. (1990) showed that PD patients were also impaired in learning arbitrary pairings between abstract words and selected colors. Thus, these deficits were not restricted to motor conditional learning but were indicative of a more general conditional learning deficit. Sprengelmeyer et al. (1995) tested the possibility that this learning deficit was not due to a deficit in the learning of the associations but in the translation of those associations into motor responses. They found that PD patients had longer decision times compared to the motor component deficits. The data showed that PD patients were able to describe the associations but showed no improvement in their decision times as a result of practice. Knowlton et al. (1996) recently showed that optimally medicated PD patients failed to learn a probabilistic classification task, in which subjects had to learn gradually which of two outcomes would occur on each trial given the particular combination of cues that appeared on that trial. However, the PD patients had intact memory for the training event indicating that declarative memory was intact. The fact that even fully medicated PD patients, who are known to improve their PD motor deficits as a consequence of pharmacotherapy, were impaired in the incremental learning of associations in the probabilistic classification task (Knowlton et al., 1996) and in associative learning tasks (Sprengelmeyer et al., 1995; Linden et al., 1990) suggests that the phasic nature of DA activity or release is a critical component in the learning of these goal-directed tasks. Presumably, systemic levodopa medication cannot restore the phasic DA activity involved in learning but only the tonic DA baseline levels that appear to regulate the motor aspects of behavior (Contreras-Vidal and Stelmach, 1995). Thus, the present model suggests that tasks involving trial-by-error learning are impaired in fully medicated PD patients.

Acknowledgments

We thank Dr. Edward Castañeda and the reviewers for valuable comments on an earlier version of this manuscript. This work was supported in part by the Human Frontier Science Program (SF-412/95), the

McDonnell-Pew Foundation, the Flinn Foundation, and the Swiss National Science Foundation.

References

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Ann. Rev. Neurosci.* 9:357–381.

Apicella P, Ljungberg T, Scarnati E, Schultz W (1991) Responses to reward in monkey dorsal and ventral striatum. *Exp. Brain Res.* 85:491–500.

Apicella P, Scarnati E, Ljungberg T, Schultz W (1992) Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J. Neurophysiol.* 68:945–960.

Bowman EM, Aigner TG, Richmond BJ (1996) Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J. Neurophysiol.* 75:1061–1073.

Brown RG, Marsden CD (1990) Cognitive function in Parkinson's disease: From description to theory. *Trends in Neurosci.* 13:21–29.

Brown VJ, Schwarz U, Bowman EM, Fuhr P, Robinson DL, Hallet M (1993) Dopamine dependent reaction time deficits with Parkinson's disease are task specific. *Neuropsychologia* 31:459–469.

Buonomano DV, Mauk MD (1994) Neural network model of the cerebellum: Temporal discrimination and the timing of motor responses. *Neural Computation* 6:38–55.

Calabresi P, Maj R, Pisani A, Mercuri NB, Bernardi G (1992) Long-term synaptic depression in the striatum: Physiological and pharmacological characterization. *J. Neurosci.* 12:4224–4233.

Canavan AGM, Passingham RE, Marsden CD, Quinn N, Wyke M, Polkey CE (1989) The performance on learning tasks of patients in the early stages of Parkinson's disease. *Neuropsychologia* 27:141–156.

Carpenter GA (1997) Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks* 10:1473–1494.

Carpenter GA, Grossberg S (1987) ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics* 26:4919–4930.

Carpenter GA, Grossberg S (1990) ART-3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks* 3:129–152.

Contreras-Vidal JL, Schultz W (1996) A neural network model of reward-related learning, motivation and orienting behavior (Abstract). *Soc. Neurosci. Abs.* 22:2029.

Contreras-Vidal JL, Stelmach GE (1995) A neural model of basal ganglia-thalamocortical relations in normal and parkinsonian movement. *Biol. Cybern.* 73:467–476.

Eblen F, Graybiel AM (1995) Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J. Neurosci.* 15:5999–6013.

Fiala JC, Grossberg S, Bullock D (1996) Metabotropic glutamate receptor activation in cerebellar purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *J. Neurosci.* 16:3760–3774.

Gaffan D, Murray EA, Fabre-Thorpe M (1993) Interaction of the amygdala with the frontal lobe in reward memory. *Eur. J. Neurosci.* 5:968–975.

Gariano RF, Groves PM (1988) Burst firing induced in midbrain dopamine neurons by stimulation of the medial prefrontal and anterior cingulate cortices. *Brain Res.* 462:194–198.

Gaspar P, Stepniewska I, Kaas J (1992) Topography and collateralization of the dopaminergic projections to motor and lateral prefrontal cortex in Owl monkeys. *J. Comp. Neurol.* 325:1–21.

Gerfen CR (1989) The neostriatal mosaic: Striatal patch-matrix organization is related to cortical lamination. *Science* 246:385–388.

Gerfen CR (1992) The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. *Annu. Rev. Neurosci.* 15:285–320.

Gerfen CR, Herkenham M, Thibault J (1987) The neostriatal mosaic. II. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. *J. Neurosci.* 7:3935–3944.

Goldman-Rakic PS, Porrino LJ (1985) The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *J. Comp. Neurol.* 242:535–560.

Gotham AM, Brown RG, Marsden CD (1988) "Frontal" cognitive function in patients with Parkinson's disease "on" and "off" levodopa. *Brain* 111:299–321.

Grace AA, Bunney BS (1985) Opposing effects of striatonigral feedback pathways on midbrain dopamine cell activity. *Brain Res.* 333:271–284.

Graveland GA, DiFiglia M (1985) The frequency and distribution of medium-sized neurons with indented nuclei in the primate and rodent neostriatum. *Brain Res* 327:307–311.

Graybiel AM (1990) Neurotransmitters and neuromodulators in the basal ganglia. *Trends in Neurosci.* 13:244–254.

Groenewegen HJ (1988) Organization of the afferent connections of the mediodorsal thalamic nucleus in the rat, related to the mediodorsal-prefrontal topography. *Neuroscience* 24:379–431.

Grossberg S, Merrill JWL (1992) A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cogn. Brain Res.* 1:3–38.

Haber SN, Lynd E, Klein C, Groenewegen HJ (1990) Topographic organization of the ventral striatal efferent projections in the rhesus monkey: An anterograde tracing study. *J. Comp. Neurol.* 293:282–298.

Hodgkin AL, Huxley AF (1952) A quatitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117:500–544.

Hollerman J, Schultz W (1996) Activity of dopamine neurons during learning in a familiar task context (Abstract). *Soc. Neurosci. Abs.* 22:1388.

Hoover JE, Strick PL (1993) Multiple output channels in the basal ganglia. *Science* 259:819–821.

Houk JC, Adams JL, Barto AG (1995) A model of how basal ganglia generate and use neural signals that predict reinforcement. In: JC Houk, JL Davis, DG Beiser, eds. Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, MA. pp. 249–270.

Ivry RB (1996) The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology* 6:851–857.

Jaeger D, Kita H, Wilson CJ (1994) Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *J. Neurophysiology* 72:2555–2558.

Jimenez-Castellanos J, Graybiel AM (1987) Subdivisions of the dopamine-containing A8-A9-A10 complex identified by their differential mesostriatal innervation of striosomes and extrastriosomal matrix. *Neurosci.* 23:223–242.

Jimenez-Castellanos J, Graybiel AM (1989) Evidence that histochemically distinct zones of the primate substantia nigra pars compacta are related to patterned distributions of nigrostriatal projection neurons and striatonigral fibers. *Exp. Brain Res.* 74:227–238.

Jueptner M, Rijntjes M, Weiller C, Faiss JH, Timmann D, Mueller SP, Diener HC (1995) Localization of a cerebellar timing process using PET. *Neurology* 45:1540–1545.

Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. *Science* 273:1399–1354.

Kornhuber J, Kim J-S, Kornhuber ME, Kornhuber HH (1984) The cortico-nigral projection: Reduced glutamate content in the substantia nigra following frontal cortex ablation in the rat. *Brain Res.* 322:124–126.

Kötter R, Wickens J (1995) Interactions of glutamate and dopamine in a computational model of the striatum. *J. Computational Neuroscience* 2:195–214.

Künzle H (1978) An autoradiographic analysis of the efferent connections from premotor and adjacent prefrontal regions (Areas 6 and 9) in Macaca fascicularis. *Brain Behav. Evol.* 15:185–234.

Levine DS, Prueitt PS (1989) Modeling some effects of frontal lobe damage: Novelty and perseveration. *Neural Networks* 2:103–116.

Linden A, Bracke-Tolkmitt R, Lutzenberger W, Canavan AGM, Scholz E, Diener H-C, Birbaumer N (1990) Slow cortical potentials in parkinsonian patients during the course of an associative learning task. *J. Psychophysiol.* 4:145–162.

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.* 67:145–163.

Maricq AV, Church RM (1983) The differential effects of haloperidol and methamphetamine on time estimation in the rat. *Psychopharmacology* 79:10–15.

Meck WH (1996) Neuropharmacology of timing and time perception. *Cogn. Brain Res.* 3:227–242.

Milner B (1963) Effects of different brain lesions on card sorting. *Arch. Neurol.* 9:90–100.

Milner B (1964) Some effects of frontal lobectomy in man. In: J Warren, K Akert, eds. The Frontal Granular Cortex and Behavior. McGraw-Hill, New York. pp. 313–334.

Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* 72:1024–1027.

Mirenowicz J, Schultz W (1996) Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379:449–451.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J. Neurosci.* 16:1936–1947.

Murase S, Grenhoff J, Chouvet G, Gonon FG, Svensson TH (1993) Prefrontal cortex regulates burst firing and transmitter release in rat mesolimbic dopamine neurons studied in vivo. *Neurosci. Lett.* 157:53–56.

Nishino H, Ono T, Fukuda M, Sasaki K, Muramoto KI (1981) Single unit activity in monkey caudate nucleus during operant bar pressing feeding behavior. *Neurosci. Lett.* 21:105–110.

Parent A, Hazrati L-N (1995) Functinal anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Rev.* 20:91–127.

Pastor MA, Artieda J, Jahanshahi M, Obeso JA (1992) Time estimation and reproduction is abnormal in Parkinson's disease. *Brain* 115:211–225.

Perret SP, Ruiz BP, Mauk MD (1993) Cerebellar cortez lesions disrupt learning-dependent timing of conditioned eyelid responses. *J. Neurosci.* 13:1708–1718.

Plenz D, Aertsen A (1996) Neural dynamics in cortex-striatum cocultures. II. Spatiotemporal characteristics of neuronal activity. *Neuroscience* 70:893–924.

Porrino LJ, Goldman-Rakic PS (1982) Brainstem innervation of prefrontal and anterior cingulate cortex in the Rhesus monkey revealed by retrograde transport of HPR. *J. Comp. Neurol.* 205:63–76.

Pucak ML, Grace AA (1994) Regulation of substantia nigra dopamine neurons. *Critical Rev. Neurobiol.* 9:67–89.

Raijmakers MEJ, van der Maas HLJ, Molenaar PCM (1996) Numerical bifurcation analysis of distance-dependent on-center off-surround shunting neural networks. *Biol. Cybern.* 75:495–507.

Rebec GV, Curtis SD (1988) Reciprocal zones of excitation and inhibition in the neostriatum. *Synapse* 2:633–635.

Romo R, Schultz W (1990) Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* 63:592–606.

Russchen FT, Bakst I, Amaral DG, Price JL (1985) The amygdolostriatal projections in the monkey: An anterograde tracing study. *Brain Res.* 329:241–257.

Sahakian B, Morris R, Evenden J, Heald A, Levy R, Philpot M, Robins T (1988) A comparative study of visuo-spatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain* 111:695–718.

Schultz W (1986) Activity of pars reticulata neurons of monkey substantia nigra in relation to motor, sensory, and complex events. *J. Neurophysiol.* 55:660–677.

Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* 13:900–913.

Schultz W, Apicella P, Scarnati E, Ljungberg T (1992) Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurophysiol.* 12:4595–4610.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1598.

Schultz W, Romo R (1990) Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioral reactions. *J. Neurophysiol.* 63:607–624.

Sherman SM, Guillery RW (1996) Functional organization of thalamocortical relays. *J. Neurophysiol.* 76:1367–1395.

Smith Y, Charara A, Parent A (1996) Synaptic innervation of midbrain dopaminergic neurons by glutamate-enriched terminals in the squirrel monkey. *J. Comp. Neurol.* 364:231–253.

Sprengelmeyer R, Canavan AGM, Lange HW, Homberg V (1995) Associative learning in degenerative neostriatal disorders: Contrasts in explicit and implicit remembering between Parkinson's and Huntington's diseases. *Mov. Disorders* 10:51–65.

Strick PL, Dum RP, Mushiake H (1995) Basal ganglia "loops" with the cerebral cortex. In: M Kimura, AM Graybiel, eds. Functions of the Cortico-Basal Ganglia Loop. Springer-Verlag, Tokyo. pp. 106–124.

Suri RE, Schultz W (1996) A neural learning model based on the activity of primate dopamine neurons (Abstrac). *Soc. Neurosci. Abs.* 22:1389.

Sutton RS (1988) Learning to predict by the methods of temporal differences. *Machine Learning* 3:9–44.

Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: Expectation and prediction. *Psychol. Rev.* 88:135–171.

Tong ZY, Overton PG, Clark D (1996) Stimulation of the prefrontal cortex in the rat induces patterns of activity in midbrain dopaminergic neurons which resemble natural burst events. *Synapse* 22:195–208.

Watanabe M (1990) Prefrontal unit activity during associative learning in the monkey. *Exp. Brain Res.* 80:296–309.

Watanabe M (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382:629–632.

Wichmann T, Vitek JL, DeLong MR (1995) Parkinson's disease and the basal ganglia: Lessons from the laboratory and from neurosurgery. *Neuroscientist* 1:236–244.

Wickens JR, Alexander ME, Miller R. (1991) Two dynamic modes of striatal function unders dopaminergic-cholinergic control: Simulation and analysis of a model. *Synapse* 8:1–12.

Wickens JR, Begg AJ, Arbuthnott GW (1996) Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* 70:1–5.

Young WS III, Alheid GF, Heimer L (1984) The ventral pallidal projection to the mediodorsal thalamus: A study with fluorescent retrograde tracers and immunohisto-fluorescence. *J. Neurosci*. 4:1626–1638.