

# A Computational Model of How Cholinergic Interneurons Protect Striatal-dependent Learning

F. Gregory Ashby and Matthew J. Crossley

## Abstract

■ An essential component of skill acquisition is learning the environmental conditions in which that skill is relevant. This article proposes and tests a neurobiologically detailed theory of how such learning is mediated. The theory assumes that a key component of this learning is provided by the cholinergic interneurons in the striatum known as tonically active neurons (TANs). The TANs are assumed to exert a tonic inhibitory influence over cortical inputs to the striatum that prevents the execution of any striatal-dependent

actions. The TANs learn to pause in rewarding environments, and this pause releases the striatal output neurons from this inhibitory effect, thereby facilitating the learning and expression of striatal-dependent behaviors. When rewards are no longer available, the TANs cease to pause, which protects striatal learning from decay. A computational version of this theory accounts for a variety of single-cell recording data and some classic behavioral phenomena, including fast reacquisition after extinction. ■

## INTRODUCTION

During skill learning, a response elicited by a specific stimulus might be rewarded, but if this same stimulus is encountered outside the training session, why doesn't the absence of reward extinguish the skill response? This article proposes and tests a computational theory of such context-sensitive learning. Briefly, we propose that a key component of this learning is provided by the tonically active cholinergic interneurons in the striatum (tonically active neurons [TANs]).

The striatum is known to contribute to many aspects of motor, cognitive, and limbic processing, and a huge literature suggests that the striatum is critically important in skill learning (for reviews, see e.g., Ashby & Ennis, 2006; Yin & Knowlton, 2006; Doyon & Ungerleider, 2002; Packard & Knowlton, 2002). In humans, approximately 96% of all striatal neurons are medium spiny neurons (MSNs; Yelnik, Francois, Percheron, & Tande, 1991), which receive cortical input and send axons out of the striatum to BG output structures. The TANs are cholinergic striatal interneurons that have extensive axon fields allowing them to project to large striatal regions (e.g., Calabresi, Centonze, Gubellini, Pisani, & Bernardi, 2000; Kawaguchi, Wilson, Augood, & Emson, 1995).

TANs are tonically active in their resting state, and they have a prominent modulatory effect on MSNs (Pakhotin & Bracci, 2007; Gabel & Nisenbaum, 1999; Akins, Surmeier, & Kitai, 1990; Akaike, Sasa, & Takaori, 1988; Dodt & Misgeld, 1986). These effects are both pre- and postsynaptic.<sup>1</sup> With respect to cortical input, however, the predominant effect of TAN activity on MSN activation is inhibitory. For exam-

ple, Pakhotin and Bracci (2007) reported that a single TAN spike caused a significant reduction in the excitatory postsynaptic current induced by cortical (glutamatergic) input. On the basis of these and other results, they concluded that after a TAN pause, MSNs "will transiently become much more responsive to cortical inputs" (p. 399) and that the resumption of TAN firing "will cause an abrupt reduction of MSN excitation" (p. 399).

Thus, MSNs are especially responsive to cortical input during TAN pauses. To understand the behavioral significance of this phenomenon, it is therefore critical to study the environmental conditions that cause TANs to pause. In fact, it is well established that TANs pause to the delivery of reward and to stimuli that predict the delivery of reward (Apicella, Legallet, & Trouche, 1997; Aosaki, Tsubokawa, et al., 1994; Kimura, 1992; Apicella, Scamati, & Schultz, 1991). They also pause to novel stimuli (Blazquez, Fujii, Kojima, & Graybiel, 2002). Another important result is that whereas most MSNs fire to a restricted set of stimuli from a single sensory modality (e.g., Caan, Perrett, & Rolls, 1984), many TANs respond to stimuli from a number of different modalities (Matsumoto, Minamimoto, Graybiel, & Kimura, 2001). Thus, a TAN might respond to the discriminative cue associated with reward, but it is also likely to respond to other visual, auditory, and olfactory cues (for example) that occur incidentally at the time of reward delivery.

The TANs receive their strongest excitatory glutamatergic input from the caudal intralaminar nuclei of the thalamus (Smith, Raju, Pare, & Sidibe, 2004; Sadikot, Parent, & Francois, 1992; Cornwall & Phillipson, 1988), which includes the center-median (CM) and the parafascicular (Pf) nuclei. The CM/Pf complex receives input from a number of places, including the OFC, the pedunculo-pontine tegmental

University of California, Santa Barbara

nucleus, and the ascending reticular activating system (Van der Werf, Witter, & Groenewegen, 2002)—structures that are well known to participate in reward processing and arousal.

The TANs are also prominent targets of substantia nigra dopamine cells. Two features of this dopaminergic input are relevant to the model proposed here. First, dopamine cell responses and TAN pauses are temporally coincident (Cragg, 2006; Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004). Second, long-term potentiation (LTP) in TANs requires elevated levels of dopamine (Suzuki, Miura, Nishimura, & Aosaki, 2001; Aosaki, Graybiel, & Kimura, 1994). These results suggest that TANs may learn to pause to cues that signal reward via reinforcement learning at CM/Pf–TAN synapses. In support of this idea, simultaneous single-unit recordings from CM/Pf neurons and TANs show that although an intact CM/Pf response is required for the TANs to pause, the CM/Pf response to environmental cues is relatively constant, regardless of the reward contingencies of the task, whereas the TANs pause primarily to reward-predicting cues (Matsumoto et al., 2001). Because TAN pauses are primarily driven by the CM/Pf complex, it therefore seems reasonable that plasticity at CM/Pf–TAN synapses allows the TANs to learn to pause in the presence of cues that predict reward.

To test these ideas formally, we constructed a computational model with the overall architecture shown in Figure 1. The idea is that, in the absence of CM/Pf input, the TAN's high spontaneous firing tonically inhibits the MSN response to cortical input.<sup>2</sup> When cells in the CM/Pf complex fire, reinforcement learning at the CM/Pf–TAN synapse quickly causes the TAN to pause when in a rewarding environment. This releases the MSNs from tonic inhibition,

thereby allowing them to respond to cortical inputs and thus to gate learning at cortical–striatal synapses.

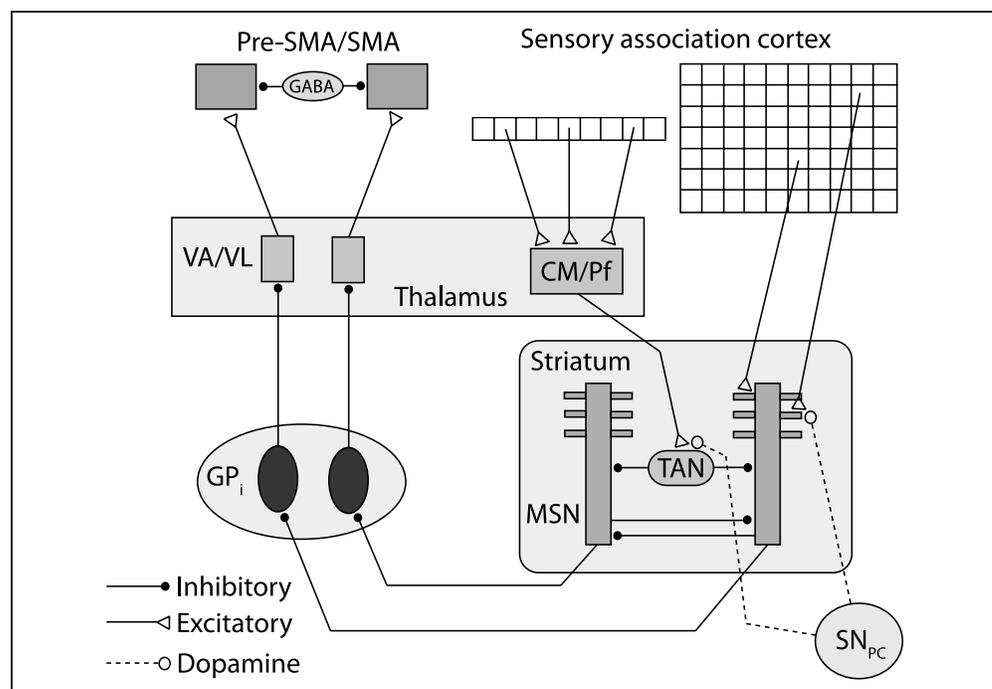
## METHODS

### Activation Equations

The activation of all sensory cortical units and the CM/Pf unit was either off (with activation 0) or on (with activation 1500) during the duration of stimulus presentation (note that we did not specifically model sensory inputs to the CM/Pf). Our model of changes in the membrane potential of striatal MSNs was adapted from a model proposed by Izhikevich (2007). The model includes two coupled differential equations for each medium spiny unit. The first equation models fast changes in membrane potential (measured in mV), and the second equation models slow changes in the activation and inactivation of various intracellular ion channels (e.g.,  $\text{Na}^+$  and  $\text{K}^+$ ). We supplement the Izhikevich model by assuming that the key inputs to the MSNs include (1) excitatory inputs from sensory cortex, (2) inhibitory input from the TAN, and (3) inhibitory input from other MSNs. Specifically, our complete medium spiny unit model assumes that the membrane potential in striatal unit  $J$  at time  $t$ , denoted  $S_J(t)$ , is determined by

$$50 \frac{dS_J(t)}{dt} = \sum_K w_{K,J}(n) I_K(t) - \beta_S T(t) - \gamma_S \sum_{M \neq J} f[S_M(t)] + [S_J(t) + 80][S_J(t) + 25] + E - u_S(t) + \sigma_S \varepsilon(t), \quad (1)$$

**Figure 1.** The neural architecture of the proposed model in a task with two response alternatives.



$$100 \frac{du_s(t)}{dt} = -20[S_J(t) + 80] - u_s(t) \quad (2)$$

where  $\beta_s, \gamma_s, E$ , and  $\sigma_s$  are constants,  $w_{K,J}(n)$  is the strength of the synapse between sensory cortical unit  $K$  and striatal cell  $J$  on trial  $n$ ,  $I_K(t)$  is the input from sensory unit  $K$  at time  $t$ ,  $T(t)$  is the membrane potential of the TAN at time  $t$ , and  $\varepsilon(t)$  is white noise. The third term on the right in Equation 1 models the inhibitory input using a standard model of lateral inhibition (e.g., Usher & McClelland, 2001). Note that this model assumes that the total amount of lateral inhibition on medium spiny unit  $J$  is an increasing function of the total amount of activation in all MSNs. The fourth term is the quadratic integrate-and-fire model (Ermentrout, 1996). To produce spikes, when  $S_J(t) = 40$  mV, then  $S_J(t)$  is reset to  $S_J(t) = -55$  mV. The last term models noise. Equation 2 models the slow changes in various intracellular ion channels. When Equation 1 produces a spike (i.e., when  $S_J(t) = 40$  mV),  $u_s(t)$  is reset to  $u_s(t) + 150$ . All specific numerical values in Equations 1 and 2 and the numerical values used in the resetting procedures are taken from Izhikevich.

The function  $f[S_M(t)]$  in Equation 1 is called the alpha function and is a standard method for modeling the post-synaptic effects of a spike in a presynaptic cell (e.g., Rall, 1967). When a presynaptic cell generates an action potential, synaptic vesicles open and the neurotransmitter is released, diffuses across the synapse, and binds to post-synaptic receptors, which initiates events that eventually effect the membrane potential of the postsynaptic cell. The alpha function models the time course of these effects. The idea is that every time the presynaptic cell spikes, the following input is delivered to the postsynaptic cell:

$$\alpha(t) = \frac{t}{\lambda} \exp\left(-\frac{\lambda - t}{\lambda}\right). \quad (3)$$

This function has a maximum value of 1.0, and it decays to .01 at  $t = 7.64\lambda$ .

The model described by Equations 1 and 2 accurately accounts for patch-clamp data collected from MSNs in the rat (i.e., see Figure 8.37 of Izhikevich, 2007) in the sense that it displays both the up and the down states that characterize MSN firing patterns, and it displays realistic spiking behavior.

The TANs are more challenging to model because of their unusual dynamics. For example, when excitatory input is delivered to the TANs, they fire an initial burst and then pause (Reynolds, Hyland, & Wickens, 2004; Kimura, Rajkowski, & Evarts, 1984). We developed a model of TAN firing that displays these same qualitative properties by modifying the Izhikevich (2003) model of intrinsically bursting cortical neurons. Specifically, we assumed that changes in the TAN membrane potential at time  $t$ ,

denoted  $T(t)$ , are described by the following two coupled equations.

$$100 \frac{dT(t)}{dt} = v(n)Pf(t) + 1.2[T(t) + 75][T(t) + 45] + 950 - u_T(t), \quad (4)$$

$$100 \frac{du_T(t)}{dt} = 5[T(t) + 75] - u_T(t) + 2.7v(n)R(t), \quad (5)$$

where  $v(n)$  is the strength of the synapse between the Pf and the TAN on trial  $n$ , and  $Pf(t)$  is the input from the CM/Pf at time  $t$ . The constant 950 models spontaneous firing, and the function  $R(t) = Pf(t)$  up to the time when CM/Pf activation turns off, then  $R(t)$  decays exponentially back to zero (with rate .0018). To produce spikes, when  $T(t) = 60$  mV, then  $T(t)$  is reset to  $T(t) = -56$  mV and  $u_T(t)$  is reset to  $u_T(t) + 150$ . In a later section, we will consider the dynamical behavior of this model that allows it to mimic the unusual firing properties of TANs.

Note that we modeled the effects of CM/Pf activation as purely excitatory. In fact, the evidence is good that glutamate inputs from the CM/Pf also synapse on GABAergic interneurons in the striatum, which then synapse on TANs. As a result, CM/Pf activation also can induce an inhibitory input to the TANs (Zackheim & Abercrombie, 2005; Suzuki et al., 2001). We chose not to model this indirect inhibitory effect because TANs pause when positive current is injected directly into the cell (e.g., see Figure 4). Thus, whereas the inhibitory input may potentiate the TAN pause, it is apparently not necessary to induce the pause.

For all other units in the model, we excluded the slow regulatory term  $u_i(t)$  and instead modeled membrane potential with the standard quadratic integrate-and-fire model. For example, activation in the globus pallidus at time  $t$ , denoted by  $G_J(t)$ , is described by

$$15 \frac{dG_J(t)}{dt} = -\alpha_G f[S_J(t)] + 71 + 0.7[G_J(t) + 60][G_J(t) + 40], \quad (6)$$

where  $\alpha_G$  is a constant. The first term models the inhibitory input from the striatum, the second term ensures a high tonic firing rate, and the last term is the quadratic integrate-and-fire component that is the same as in Equations 1 and 4. Spikes are produced after  $G_J(t) = 35$  by resetting to  $G_J(t) = -50$ .

Similarly, activation in the thalamus at time  $t$  is given by

$$\frac{dV_J(t)}{dt} = -\beta_T f[G_J(t)] + 71 + .7[V_J(t) + 60][V_J(t) + 40], \quad (7)$$

where  $\beta_T$  is a constant, and  $f[G_J(t)]$  is the alpha function from Equations 1 and 6. Spikes are produced after  $V_J(t) = 35$  by resetting to  $V_J(t) = -50$ . The first term models the inhibitory input from the globus pallidus. The ventral anterior and the ventral lateral thalamic nuclei of the thalamus also receive a variety of excitatory inputs (e.g., cerebellum, pFC). We modeled these via the constant 71. For our purposes, the most important of these excitatory inputs may be from pFC (e.g., Anderson & DeVito, 1987). pFC input is critical because it is thought that striatal firing, by itself, does not trigger a motor response. When the striatum fires, it disinhibits the thalamus (i.e., by reducing pallidal inhibition), but it does not excite the thalamus.<sup>3</sup> For this reason, random sensory stimuli that are encountered as one moves through the world could cause the striatum to fire, but this firing will typically not elicit an unintended motor response. In a skill-learning task, instructions from an experimenter about how to respond could cause the cortical input to thalamus to increase, thereby priming the relevant response goals. Because of the tonic inhibition from the globus pallidus, however, this cortical input is not enough to trigger a response. Instead, the striatum must first inhibit the globus pallidus, an event that would allow the thalamus to trigger one of the primed motor response goals.

Activation in the  $J$ th unit in premotor cortex at time  $t$ , denoted by  $C_L(t)$ ,  $t$  is given by

$$\frac{dC_J(t)}{dt} = \beta_C f[V_J(t)] - \gamma_C \sum_{K \neq J} f[C_K(t)] + 69 + .7[C_J(t) + 60][C_J(t) + 40] + \sigma_C \varepsilon(t), \quad (8)$$

where  $\beta_C$ ,  $\gamma_C$ , and  $\sigma_C$  are constants, and  $\varepsilon(t)$  is white noise. As in other units, spikes are produced after  $C_J(t) = 35$  by resetting to  $C_J(t) = -50$ . The second term on the right models lateral inhibition in the same way as in Equation 1. In tasks with two possible responses, evidence suggests that cortical units in premotor areas are sensitive to the cumulated difference in evidence favoring the two alternatives (e.g., Shadlen & Newsome, 2001). We used a more biologically plausible method that is known to simulate this difference process—that is, we placed a separate threshold on the activation of each unit but included lateral inhibition between the units (Usher & McClelland, 2001).

## Learning Equations

Following standard models, we assumed that synaptic plasticity at all cortical–striatal synapses and at the CM/Pf–TAN synapse is modified according to reinforcement learning that requires three factors: (1) strong presynaptic activation, (2) postsynaptic activation that is strong enough to activate *N*-methyl *D*-aspartate (NMDA) receptors, and (3) dopamine levels above baseline (e.g., Reynolds & Wickens,

2002; Arbuthnott, Ingham, & Wickens, 2000; Calabresi, Pisani, Mercuri, & Bernardi, 1996). If postsynaptic activation is strong but dopamine is below baseline, then the synapse is weakened. The synapse is also weakened, regardless of dopamine level, if postsynaptic activation is below the NMDA threshold but above the  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) threshold (i.e., the AMPA receptor is a low-threshold glutamate receptor).

Let  $w_{K,J}(n)$  denote the strength of the synapse on trial  $n$  between cortical unit  $K$  and striatal unit  $J$ . We model reinforcement learning as follows:

$$\begin{aligned} w_{K,J}(n+1) = & w_{K,J}(n) + \alpha_w \int I_K(t) dt \left[ \int [S_J(t)]^+ dt - \theta_{\text{NMDA}} \right]^+ \\ & \times [D(n) - D_{\text{base}}]^+ [w_{\text{max}} - w_{K,J}(n)] \\ & - \beta_w \int I_K(t) dt \left[ \int [S_J(t)]^+ dt - \theta_{\text{NMDA}} \right]^+ \\ & \times [D_{\text{base}} - D(n)]^+ w_{K,J}(n) \\ & - \gamma_w \int I_K(t) dt \left\{ \theta_{\text{NMDA}} - \int [S_J(t)]^+ dt \right\}^+ \\ & - \theta_{\text{AMPA}} \Big\}^+ w_{K,J}(n). \quad (9) \end{aligned}$$

The function  $[g(t)]^+ = g(t)$  if  $g(t) > 0$ , and otherwise  $[g(t)]^+ = 0$ . The integrals in Equation 10 are all over the time of stimulus presentation. Thus,  $\int [S_J(t)]^+ dt$  is the total positive medium spiny activation during stimulus presentation.  $D_{\text{base}}$  is the baseline dopamine level,  $D(n)$  is the amount of dopamine released after feedback on trial  $n$ , and  $\alpha_w$ ,  $\beta_w$ ,  $\gamma_w$ ,  $\theta_{\text{NMDA}}$ , and  $\theta_{\text{AMPA}}$  are all constants. The first two lines describe the conditions under which LTP occurs (striatal activation above the threshold for NMDA receptor activation and dopamine above baseline) and lines 3–6 describe conditions that produce long-term depression (LTD). The first possibility (Lines 3 and 4) is that postsynaptic activation is above the NMDA threshold but dopamine is below baseline (as on an error trial), and the second possibility (Lines 5 and 6) is that striatal activation is between the AMPA and the NMDA thresholds. Note that synaptic strength does not change if postsynaptic activation is below the AMPA threshold.

Note that these learning equations do not depend on striatal acetylcholine (ACh) levels. The evidence is good, however, that ACh does modulate corticostriatal LTP and LTD (Bonsi et al., 2008; Wang et al., 2006; Centonze, Gubellini, Bernardi, & Calabresi, 1999). In vitro results seem to suggest that (1) steady-state ACh levels are required for normal corticostriatal LTP and (2) reduced ACh levels are required for LTD. An obvious assumption is that that a TAN pause is associated with reduced ACh levels, so these results seem to imply that corticostriatal LTP cannot occur during a TAN pause, only LTD. This creates a paradox, however, because the environmental conditions known to cause TANs to pause (e.g., the appearance of cues that predict reward) are the same as the conditions thought to promote corticostriatal LTP. For example, in conditioning tasks, an animal is rewarded for associating a motor response with a sensory

cue. Many such studies have shown that MSNs learn to fire a burst to the presence of the cue (e.g., Carelli, Wolske, & West, 1997; another example occurs in the Figure 8 data of Barnes, Kubota, Hu, Jin, & Graybiel, 2005), and presumably this increase in MSN activation is mediated by LTP at corticostriatal synapses.

One possibility is that a TAN pause may not cause a simple reduction in striatal ACh levels. The TAN response to sensory cues associated with reward is multiphasic. Frequently, the TAN pause is preceded by an initial burst (as, e.g., in Figure 4) and also followed by a rebound burst (as, e.g., in Figure 6). Thus, ACh levels may fluctuate rapidly during the course of a TAN pause. As a result, an informed model of the role that ACh plays in corticostriatal LTP (and LTD) may require a better understanding of the temporal dynamics of the ACh signal and of corticostriatal LTP and LTD. Lacking such data, we opted for a simpler model that ignores the role of ACh. Even so, as the next section will show, for the applications we considered, this simplified model was sufficient.

We assumed that learning at both cortical–MSN synapses and at Pf/TAN synapses is mediated by this same model. We allowed the learning rates to differ at these two synapse types, but we assumed the same numerical values for  $\theta_{\text{NMDA}}$  and  $\theta_{\text{AMPA}}$ . The numerical values for all parameters are given in the Appendix (i.e., see Table A1).

## Dopamine Model

The Equation 9 model of reinforcement learning requires that we specify the amount of dopamine released on every trial in response to the feedback signal [the  $D(n)$  term]. The more that the dopamine level increases above baseline ( $D_{\text{base}}$ ), the greater the increase in synaptic strength, and the more it falls below baseline, the greater the decrease.

Although there are a number of powerful models of dopamine release, Equation 8 requires only that we specify the amount of dopamine released to the feedback signal on each trial. The key empirical results are as follows (e.g., Tobler, Dickinson, & Schultz, 2003; Schultz, Dayan, & Montague, 1997): (1) the midbrain dopamine cells fire tonically; (2) the dopamine release increases above baseline after unexpected reward, and the more unexpected the reward, the greater the release; and (3) dopamine release decreases below baseline after unexpected absence of reward, and the more unexpected the absence, the greater the decrease. One common interpretation of these results is that over a wide range, dopamine firing increases with the reward prediction error (RPE):

$$\text{RPE} = \text{obtained reward} - \text{predicted reward}. \quad (10)$$

A simple model of dopamine release can be built by specifying how to compute obtained reward, predicted reward, and exactly how the amount of dopamine release is related to the RPE. Our solution to these three problems is as follows.

### Computing Obtained Reward

None of the applications considered in this article vary reward valence. Thus, we can use a simple model to compute obtained reward. Specifically, we defined the obtained reward  $R_n$  on trial  $n$  as +1 if correct or reward feedback is received, 0 in the absence of feedback, and –1 if error feedback is received.

### Computing Predicted Reward

We used a simplified version of the well-known Rescorla–Wagner model (Rescorla & Wagner, 1972) to compute Predicted Reward on trial  $n$ , which we denoted as  $P_n$ . According to this account,

$$P_{n+1} = P_n + .075(R_n - P_n). \quad (11)$$

It is well known that when computed in this fashion,  $P_n$  converges exponentially to the expected reward value and then fluctuates around this value until reward contingencies change.

### Computing Dopamine Release from the RPE

We assumed that the amount of dopamine release is related to the RPE in the manner reported by Bayer and Glimcher (2005). Specifically, we assumed that

$$D(n) = \begin{cases} 1 & \text{if RPE} > 1 \\ .8 \text{ RPE} + .2 & \text{if } -.25 < \text{RPE} \leq 1 \\ 0 & \text{if RPE} < -.25 \end{cases}$$

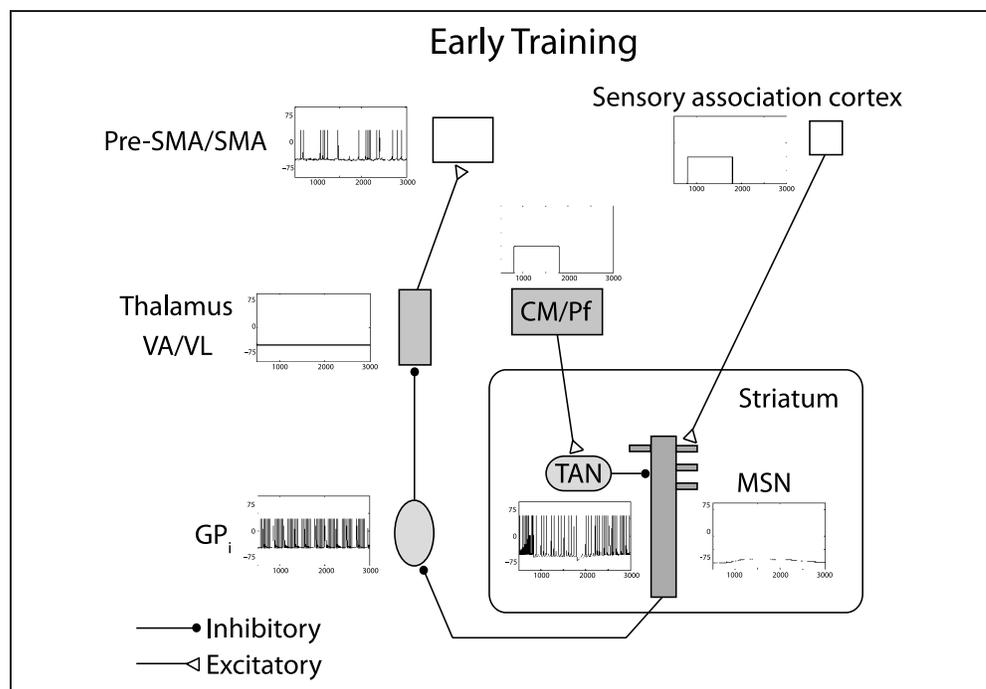
Note that the baseline dopamine level is 0.2 (i.e., when the RPE = 0) and that dopamine levels increase linearly with the RPE. However, note also the asymmetry between dopamine increases and decreases. As is evident in the Bayer and Glimcher (2005) data, a negative RPE quickly causes dopamine levels to fall to zero, whereas there is a considerable range for dopamine levels to increase in response to positive RPEs.<sup>4</sup>

## RESULTS

### Global Dynamics

Figures 2 and 3 illustrate an application of the model to a simple conditioning task in which the participant must execute some specific response (e.g., button press) when a certain sensory cue is presented (e.g., a tone) to receive a reward. Figure 2 shows activation in each brain region in the model during one trial early in training—before the model has learned to reliably respond to the sensory cue. Note that the CM/Pf and the sensory cortex activations are both modeled as simple square waves that are assumed to coincide with the stimulus presentation. Because the TAN has not yet learned that the cue is associated with

**Figure 2.** Model results from a trial early in training before the TAN has learned that the environment is rewarding. Note that the 1-sec presentation of the stimulus (from 800 to 1800 msec) does not cause the TAN to pause, and therefore the MSN does not fire to stimulus presentation. As a result, the firing rate of the premotor unit (pre-SMA/SMA) does not change after stimulus onset.

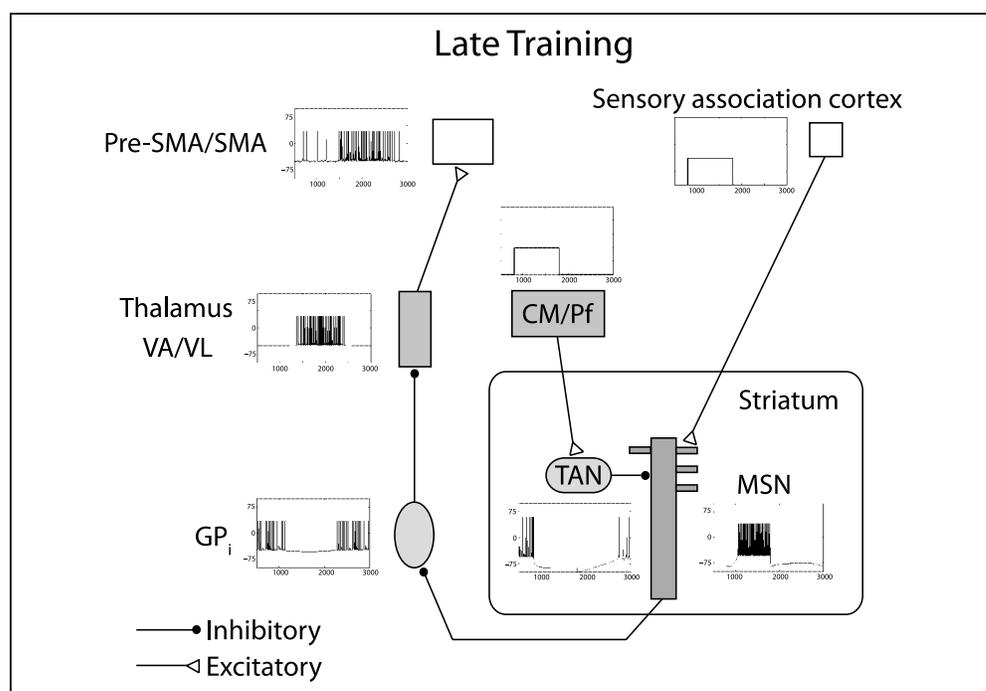


reward, it fails to pause when the stimulus is presented. As a result of the tonic inhibition from the TAN, the MSN does not fire to the stimulus, although stimulus presentation does move it from the down state to the up state. In the absence of any inhibitory input from the striatum, the globus pallidus does not slow its high spontaneous firing rate, and therefore the thalamus is prevented from firing to other excitatory inputs. The premotor unit fires at a slow tonic rate, but note that this rate does not increase during

stimulus presentation. As a result, the model does not respond on this trial.

Figure 3 illustrates a trial in this same experiment, but later in training. Now the TAN pauses when the stimulus is presented, which allows the MSN to fire a vigorous burst, which inhibits the globus pallidus. The pause in pallidal firing allows the thalamus to respond to its other excitatory inputs, and the resulting burst from the thalamus drives the firing rate in the premotor unit above the

**Figure 3.** Model results from a trial late in training. Now the TAN pauses after the 1-sec stimulus presentation (from 800 to 1800 msec), releasing the MSN from its tonic inhibition. This allows the MSN to fire to the stimulus, which causes the firing rate of the premotor unit (pre-SMA/SMA) to increase well above its baseline level.



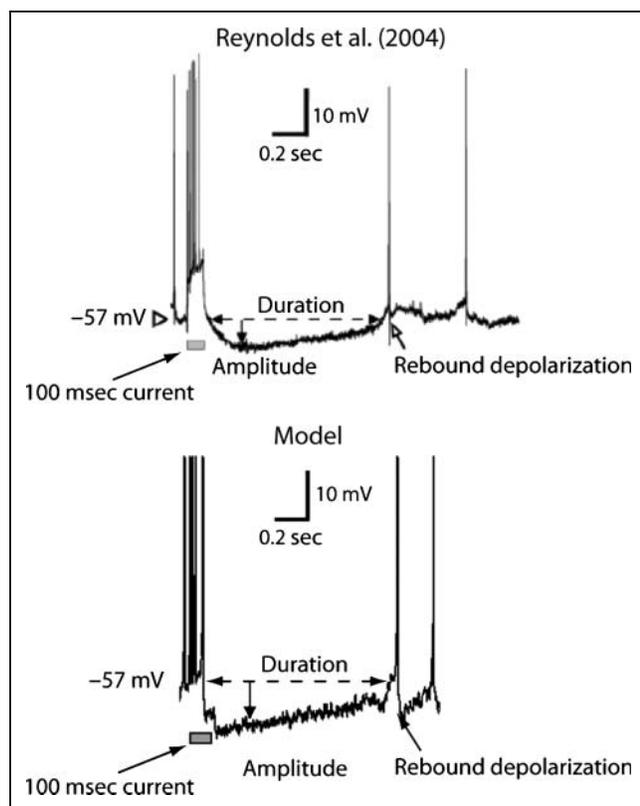
response threshold. The model now responds to the sensory cue.

### Single-unit Recordings from TANs

We begin by testing the model of the TANs against some basic single-unit recording data. The goal is to test whether our model of TAN activity is qualitatively consistent with spiking behavior recorded from real TANs. See the Appendix for technical details of all simulations.

#### *The Patch-Clamp Data of Reynolds et al. (2004)*

Reynolds et al. (2004) collected *in vivo* intracellular recordings from single TANs of anesthetized rats. The results of one such recording are shown in the top panel of Figure 4. In this experiment, a suprathreshold positive current of 100-msec duration was injected into the cell (denoted by the small gray bar in the figure). Figure 4 shows that the TAN responded with an initial burst followed by a prolonged after-hyperpolarization that caused a pause in firing that persisted for approximately 900 msec.



**Figure 4.** Patch-clamp recording from the TAN of a rat (top; from Reynolds et al., 2004; reprinted with permission from the Society for Neuroscience; permission conveyed through Copyright Clearance Center, Inc.) and simulated responses of the TAN model described in Equations 4 and 5 (bottom) during a patch-clamp experiment when positive current is injected into the cell for 100 msec (denoted by the solid gray rectangle).

Note that these data show that excitatory input alone is enough to induce a TAN pause. In other words, although CM/Pf activation may have both excitatory and inhibitory effects on TANs, the Figure 4 data suggest that the excitatory inputs by themselves may be sufficient to induce a pause.

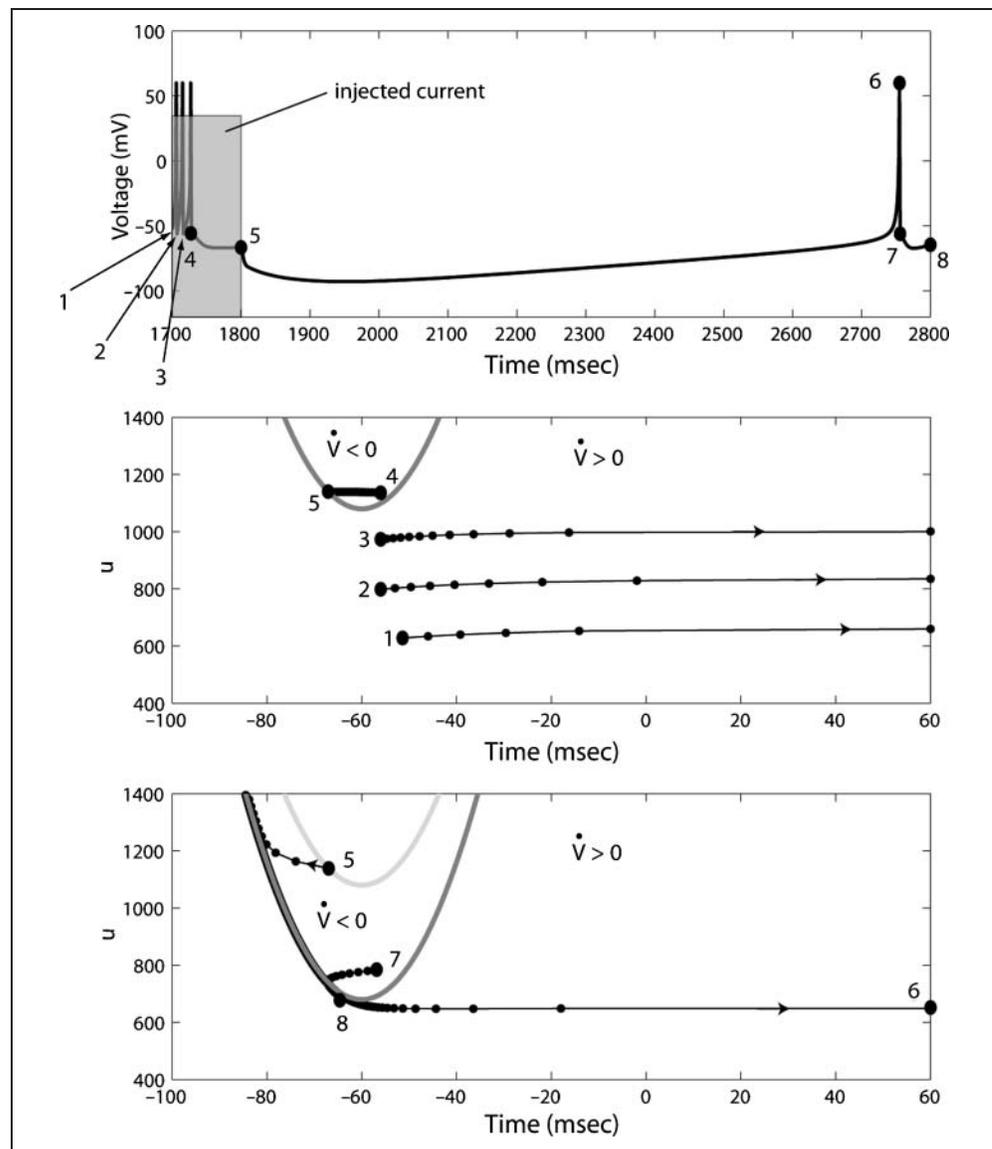
The bottom panel of Figure 4 shows the response of the model's TAN to these same experimental conditions. Note that the model also fires a burst to the injected current and then pauses for roughly 900 msec. Thus, the model displays the same temporal dynamics as real TANs. Figure 5 shows the phase portraits from the Figure 4 application, which explain why the model exhibits its pronounced pause to excitatory input. When the input is turned off, the voltage resetting mechanism in the model moves the model's state to a region where the derivative on voltage is negative (bottom panel). Voltage then decreases until the derivative is zero (i.e., the set of all  $(u,v)$  pairs where the derivative of voltage is zero is called the  $v$ -nullcline). The state then slowly drifts down the  $v$ -nullcline until eventually it breaks free. This prolonged period where voltage does not change produces the pause in firing.

#### *The Learning Data of Aosaki, Tsubokawa, et al. (1994)*

The Figure 4 data of Reynolds et al. (2004) clearly show the characteristic short-term dynamics of TANs, but they fail to show several other well-documented features of the TANs that will be critical to our later modeling. Most obviously, they do not show the high spontaneous firing rate of TANs that inspired their name, and second, they do not show the ability of the TANs to learn to pause to a stimulus that predicts reward. Figure 6 shows data from Aosaki, Tsubokawa, et al. (1994) that illustrate both of these properties. In this experiment, monkeys received a juice reward when a click occurred. This click-reward pairing was repeated many times while extracellular recordings were collected from 858 TANs in two animals. At the beginning of training, the animals ignored the clicks, and only a small percentage of TANs responded to the clicks (i.e., 17%). During training, the number of TANs that paused to the clicks gradually increased, until eventually well over half of the TANs were pausing. Individual TANs learned to pause after as little as 10 min of training, and the pause response was maintained over the course of a 4-week intermission. In addition, when other sensory cues were substituted for the click, TANs that paused to the clicks quickly learned to pause to the new stimulus.

The top panel of Figure 6 shows the spike histogram for a single TAN before and after conditioning. Note the high spontaneous firing rate before the click is presented and that before conditioning the TAN does not respond to the click. After conditioning, however, the TAN pauses about 90 msec after the click for a duration of 189 msec. The bottom panel of Figure 6 shows the responses of the model's TAN under these same conditions. Note that the model's TAN has a high spontaneous firing rate, that it initially does not respond to the click, and that after training, it

**Figure 5.** Phase portrait of the TAN model when applied to the data shown in Figure 4. The top panel is the same as the bottom panel of Figure 4 (without any noise). The middle panel is the phase portrait during the time when positive current is injected into the cell (i.e., from 1700 to 1800 msec). The gray parabola shows the  $v$ -nullcline where the derivative of voltage is zero. Below the  $v$ -nullcline voltage increases with time until it reaches 60 mV, at which point it is reset to  $-56$  mV. Above the  $v$ -nullcline, voltage decreases with time. The bottom panel shows the phase portrait after the current has been turned off. Note that the  $v$ -nullcline has now dropped. The earlier  $v$ -nullcline is shown for reference in light gray. The numbers in all panels denote the same points in time. Note that at Time Point 5 (i.e., 1800 msec), voltage decreases to the  $v$ -nullcline and then slowly follows down the nullcline until it eventually breaks free and increases. During this long period on the  $v$ -nullcline, voltage is constant, and TAN firing is paused.



pauses to the click with roughly the same lag and duration as the monkey's TAN.<sup>5</sup>

These applications suggest that our model of TAN firing mimics the most important properties of real TANs.

## Striatal-dependent Behaviors

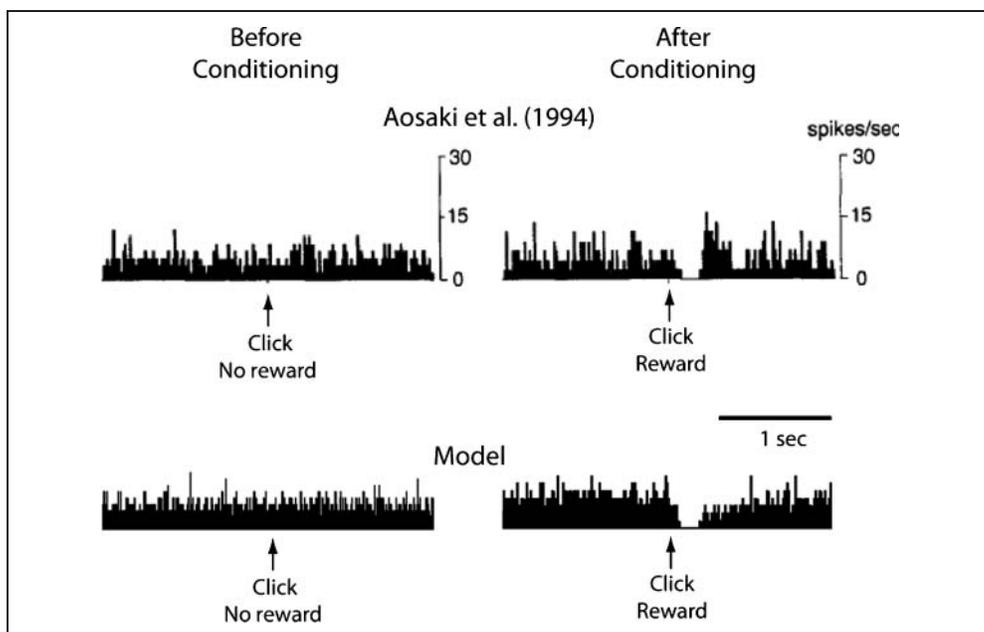
### Instrumental Conditioning

A wide variety of evidence implicates the striatum in instrumental conditioning (e.g., Yin, Ostlund, Knowlton, & Balleine, 2005; O'Doherty et al., 2004). In a typical experiment, a reward-neutral environment is suddenly altered so that rewards become available when certain instrumental behaviors are emitted (acquisition phase). During extinction, the environment is changed again so that any potential to retrieve rewards is eliminated. Finally, during reacquisition, the environment is changed once more so that the instrumental behavior is again rewarded.

The conditioning literature has naturally focused on how the strength of the association between the instrumental behavior and the reward varies during these different conditions, but it is widely recognized that secondary associations are also learned to a variety of environmental cues (e.g., Kamin, 1969). In this section, we examine the role that the TANs might play in these phenomena.

As a model experiment, we considered a task in which an animal must produce a single instrumental response (e.g., a lever press) at the onset of a sensory cue (e.g., a tone) to retrieve a reward. To model initial learning, extinction, and reacquisition in this task, we constructed a version of the model with a single unit in sensory cortex, which was either active or not depending on whether the sensory cue was present. Similarly, because only one response was possible, there was only one unit in all other brain regions. We assumed that a response was emitted when the integrated premotor unit activity crossed a threshold. Because there is only one choice for the model

**Figure 6.** (Top; from Aosaki, Tsubokawa, et al., 1994; reprinted with permission from the Society for Neuroscience; permission conveyed through Copyright Clearance Center, Inc.) Frequency histograms of TAN spikes during extracellular recordings from an experiment in which monkeys were conditioned to expect a juice reward when they heard a click. (Bottom) Frequency histograms of TAN spikes from simulations of the model under these same experimental conditions.

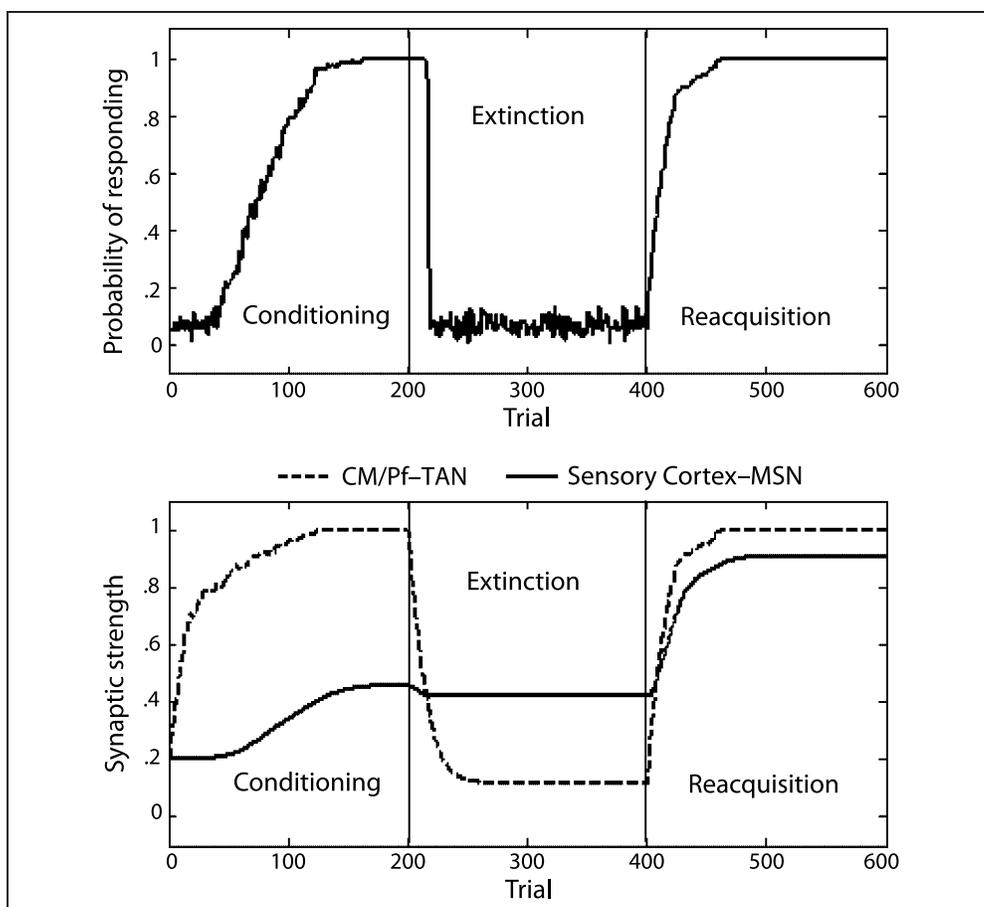


to make, feedback is never negative. Indeed, because the model can either respond and collect a reward or fail to respond and thereby fail to collect a reward, feedback is always either positive or neutral. Figures 2 and 3 show the architecture of this version of the model and predicted

neural activations in each unit during a typical trial early in learning (Figure 2) or much later after the instrumental behavior is well learned (Figure 3).

The behavioral performance of the model in this experiment is shown in the top panel of Figure 7. Note that

**Figure 7.** (Top) Performance of the model in a hypothetical conditioning experiment during conditioning, extinction, and reacquisition phases. (Bottom) Strength of the Pf/TAN synapse (broken line) and the sensory cortex–MSN synapse (solid line) on each trial of this simulated experiment.



the model learns to respond reliably to the cue during initial conditioning, that the cue is gradually ignored during extinction, and that the behavior is quickly reacquired after the reward is reinstated. The most important result in Figure 7, however, is that reacquisition is considerably faster than the initial learning of the behavior. This is one of the most widely known results in the conditioning literature. It is famous because it is a ubiquitous empirical phenomenon that is seen in almost all conditioning–extinction–reacquisition experiments (for an exception, see Ricker & Bouton, 1996) and because it has posed a difficult challenge for learning theories. For example, fast reacquisition has long been known to disconfirm any theory that assumes learning is purely a process of strengthening associations between stimuli and responses (e.g., Redish, Jensen, Johnson, & Kurth-Nelson, 2007). In such models, response rate is completely determined by the strength of these associations. Conditioning increases the strength from some initial value, and extinction decreases it back to its starting point (if the extinction phase is long enough). Thus, at the beginning of the reacquisition phase, the strength of the stimulus–response association is the same as at the beginning of the initial conditioning phase. As a result, relearning must follow the same course as initial learning.

The bottom panel of Figure 7 shows how the model accounts for fast reacquisition. This graph shows the strengths of the CM/Pf–TAN synapse (broken line) and of the sensory cortex–MSN synapse (solid line) for each trial in the experiment. Note that the CM/Pf–TAN synaptic strength increases before the cortex–MSN synaptic strength. Of course, it must rise earlier because the cortical–medium spiny cell synapse cannot be strengthened until the TAN has begun to pause. In addition to increasing sooner, however, note that the CM/Pf–TAN synaptic strength also rises at a greater rate. We hypothesize that this is because TANs are more broadly tuned than MSNs.<sup>6</sup> For example, consider an experiment where an animal must press a lever when a tone is presented to retrieve a food reward. The MSN that is conditioned in this experiment will fire to the tone, but it is unlikely to fire to other cues that are present, especially those from other sensory modalities (e.g., visual and olfactory cues from the testing chamber). As mentioned previously, however, TANs are so broadly tuned that they will respond to stimuli from multiple sensory modalities (Matsumoto et al., 2001). For this reason, the TANs will have many more opportunities to experience synaptic plasticity than the MSNs, and as a result, we hypothesize that they learn more quickly when placed in a rewarding environment.

Note next in Figure 7 that during extinction, the CM/Pf–TAN synaptic strength drops all the way to its preconditioning baseline level, but the sensory cortex–MSN synaptic strength drops only a small amount, where it remains throughout the extinction period. As the CM/Pf–TAN synaptic strength weakens, it becomes less and less likely that CM/Pf activation will induce the TAN to pause. In the ab-

sence of a TAN pause, Figure 2 shows that the MSN will not fire. In Equation 9, this corresponds to a trial in which the postsynaptic activation is below the AMPA receptor activation threshold. As a result, under these conditions, synaptic strength does not change. Thus, the TANs have the desirable property that they protect prior cortical–striatal learning during periods when the environment has changed in such a way that rewards are no longer available.

Figure 7 also shows that reacquisition time is essentially equal to the time it takes the TANs to learn that rewards are again available. At that point, the TANs begin pausing again, and the protected cortical–MSN synaptic strengths allow performance to jump nearly to its preextinction level. Finally, note that during reacquisition, the cortical–MSN synaptic strength grows to an even larger level than it reached during initial acquisition. As a result, the model predicts that after the end of the reacquisition period, the neural representation of the learned behavior is stronger than it was after initial acquisition.

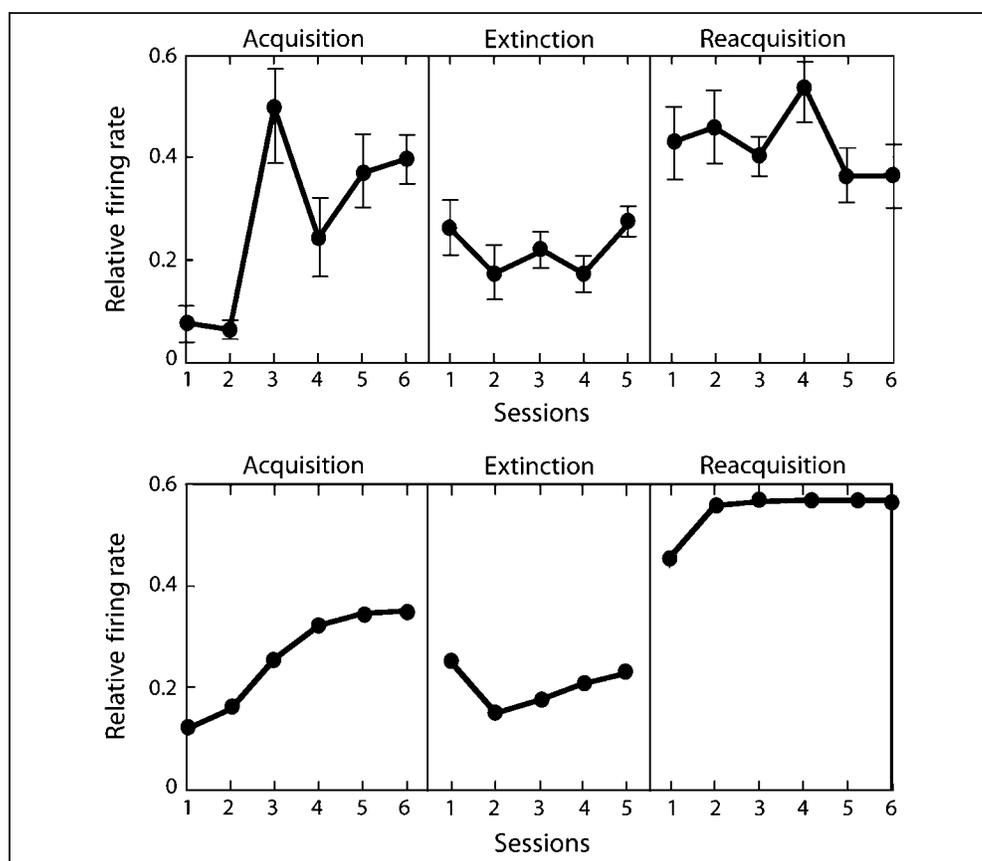
We know of no single-unit recording data from exactly this experiment. Even so, Barnes et al. (2005) reported that single-unit recording results from a similar experiment in which seven rats were trained to run down a T-maze to obtain a food reward. When the animals reached the intersection point, either a high- or a low-pitch tone sounded, which instructed them whether to turn right or left for the reward. Barnes et al. recorded from single (striatal) MSNs during sessions of acquisition, extinction, and reacquisition. The top panel of Figure 8 shows relative firing rates averaged across the MSNs that responded to the auditory tone.<sup>7</sup>

Using the same version of the model that was used to generate Figure 7, we computed this same relative firing rate statistic (averaged across 70 simulated animals) from the spikes elicited from the MSN of the model in response to the stimulus cue (see the Appendix for modeling details). Results are shown in the bottom panel of Figure 8. Note that the model correctly captures many properties of the data. These include (1) an increase in firing rate during acquisition, (2) a reduction in firing rate during extinction, (3) increasing firing rates during reacquisition, (4) lower relative firing rates during extinction than during reacquisition or the end of acquisition but higher than baseline (i.e., Sessions 1 and 2), (5) higher relative firing rates during reacquisition than during acquisition, and (6) numerical firing rates during the entire experiment that are close to the observed relative firing rates. Only one parameter was estimated during this simulation process (baseline firing rate in the absence of any cues), and this parameter value only affected the last of these six properties.

### Category Learning

Next we focus on the ability of the model to account for behavioral and single-unit recording data from the category-learning experiment of Merchant, Zainos, Hernandez, Salinas,

**Figure 8.** (Top) Relative firing rates of a rat MSN during conditioning, extinction, and reacquisition of an instrumental response (adapted from Barnes et al., 2005). (Bottom) Relative firing rates of the MSN of the model under similar experimental conditions.



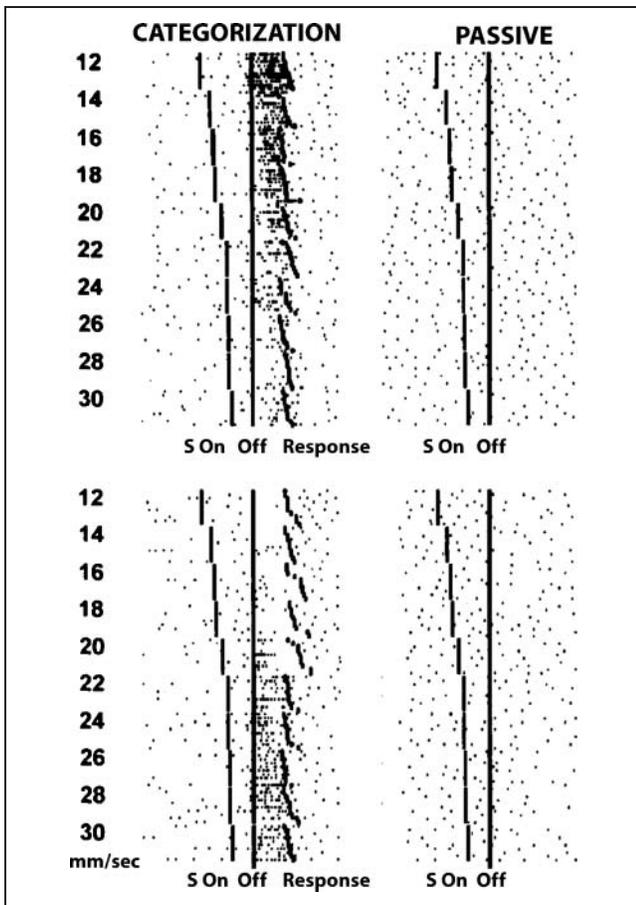
and Romo (1997). In this experiment, a rod was dragged against a monkey's finger at one of 10 different speeds. The animals were trained to push one button if one of the five low speeds occurred and to press a different button if they received one of the five high speeds. After extended feedback training, the animals reliably learned these categories.

After training, the animals completed an additional session during which single-unit recordings were collected from the putamen. Within the putamen, the responses of 695 cells were characterized in detail. Of these, 196 responded to the movement onset of the rod, regardless of category membership, 258 responded to the animal's arm movement, regardless of response, and 165 responded to the category membership of the stimulus. The neurons in this latter category responded to all stimuli in one category but not to any stimuli in the contrasting category. An example from two such cells is shown in the left column of Figure 9.

These same neurons, however, exhibited dramatically different behavior when the monkeys were presented with the same stimuli under passive conditions—that is, when no rewards were available and when their arms were restrained to prevent a response and the device housing the response keys was removed. Under these conditions, as illustrated in the right column of Figure 9, these same category-specific neurons showed no response to stimulus presentation.

According to the theory proposed here, in the passive conditions, the TANs quickly learned that there are no rewards available and therefore failed to pause when the categorization stimuli were presented. In the absence of such a pause, the MSNs were tonically inhibited and therefore unable to respond to any cortical stimulation. To test this prediction rigorously, we constructed a version of the model with 10 sensory cortical units,<sup>8</sup> one tuned to each stimulus, and two pathways through the striatum, globus pallidus, thalamus, and premotor cortex (i.e., as in Figure 1). To model the passive condition, we simply removed feedback delivery from the model (i.e., we set  $R_{z_i} = 0$  in Equation 11).

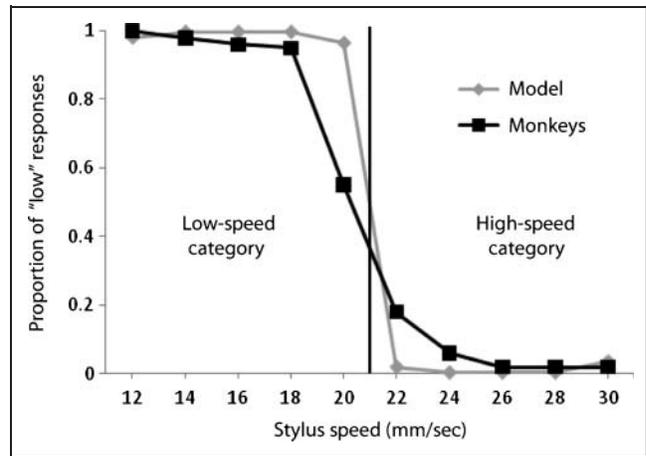
The model easily learned the tactile categories. Figure 10 describes the behavioral performance of the model and the monkeys. The left column of Figure 11 shows the category-specific firing in the model's striatal output units.<sup>9</sup> Comparing Figure 11 with the Merchant et al. (1997) data shown in Figure 9 suggests that the major discrepancy between the model and the data is in the timing of onset and offset of spiny cell firing relative to the stimulus onset and offset. However, it is important to note that we made no attempt to model these temporal dynamics. For example, Merchant et al. varied the duration of each stimulus so that the distance the rod traveled across the monkey's finger was constant. We used the same duration for all stimuli. Also, we made no attempt to model the delay between the time when the rod was first applied to the monkey's finger and the time when cells in somatosensory cortex began



**Figure 9.** Single-unit recording data from two neurons in the putamen of a monkey as a stylus is dragged across its finger at 1 of 10 different speeds (from Merchant et al., 1997; reprinted with permission from the American Physiological Society). The left column shows responses during an active categorization task in which the low speeds required one response and the high speeds required a different response, and the right column shows responses during passive reception of the stimuli when no response was required. The values from 12 to 30 indicate stylus speed (mm/sec), and each row is a different trial. Small dots denote spikes, and large dots denote the animal's response. The top panel shows responses of a cell that responds to the low speed stimuli during categorization, and the bottom panel shows responses of a cell that responds to the high speeds.

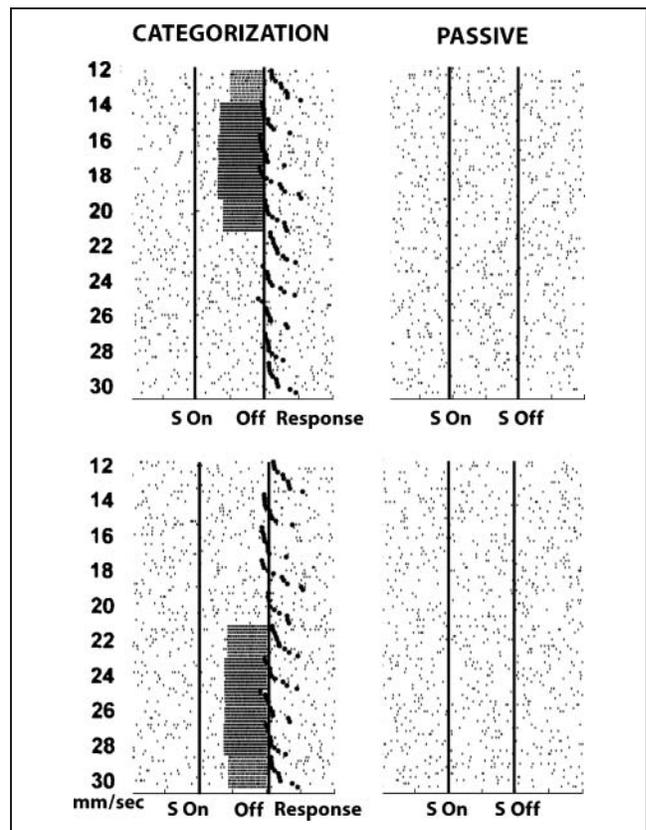
to fire to this stimulation. Making these two changes would improve the correspondence between Figures 9 and 11.

Our main goal in this article is not to construct the most accurate model possible of the active categorization condition but instead to account for the striking difference between the active and the passive conditions. The right column of Figure 11 shows that the model provides a reasonable account of this difference. During active categorization, the TAN learns to pause after stimulus presentation. This allows the medium spiny unit to respond to sensory input and the model to eventually make a response. In the passive condition, the TAN quickly unlearns its pause response. The medium spiny unit is consequently tonically inhibited and cannot respond to sensory input. Thus,



**Figure 10.** Accuracy of the monkeys and the model on each of the ten stimuli in the category-learning experiment of Merchant et al. (1997).

the difference between the two conditions is driven by the model's TAN unit. These results support the hypothesis of Ashby, Ennis, and Spiering (2007) that the TANs might be responsible for mediating the difference in firing



**Figure 11.** Simulated responses of an MSN in the model under the same experimental conditions that were used to collect the data in Figure 8. The values from 12 to 30 indicate stylus speed (mm/sec), and each row is a different trial. Small dots denote spikes and large dots denote the model's response.

properties of category-specific neurons in the Merchant et al. (1997) data.

## DISCUSSION

Many sensory cues are typically present during skill acquisition. It is quite common to encounter some of these in later contexts where the skilled behavior is no longer appropriate. In this article, we showed how cholinergic interneurons in the striatum might protect cortical–striatal synapses during these periods when rewards for the skilled behavior are not available. The idea is that the TANs exert a tonic inhibitory influence over cortical input to striatal MSNs that prevents the execution of striatal-dependent actions. However, the TANs learn to pause in rewarding environments, and this pause releases the striatal output neurons from inhibition, thereby facilitating the learning and expression of striatal-dependent behaviors. When rewards are no longer available, the TANs cease to pause, which protects striatal learning from decay. We showed that the resulting model was consistent with a variety of single-cell recording data and that it also predicted some classic behavioral phenomena, including fast reacquisition after extinction.

### Relations to Other Theories

There have been a number of other proposals that the TANs learn to pause in environments associated with reward (e.g., Apicella, 2007; Shimo & Hikosaka, 2001; Sardo, Ravel, Legallet, & Apicella, 2000). However, to our knowledge, none of these have been developed into predictive theories.

There have also been several computational models that have provided accounts of acquisition and extinction on the basis of a computational model of dopamine release that is similar to the dopamine model used here (Redish et al., 2007; Kakade & Dayan, 2002; O'Reilly & Munakata, 2000). The former two of these models assume extinction is exclusively an unlearning phenomenon and do not account for fast reacquisition. In the O'Reilly and Munakata (2000) model, however, once the strength of what in our model is the cortical–striatal synapse falls low enough for the behavior to disappear, this synaptic strength is no longer weakened by further extinction trials. This allows the model to predict fast reacquisition because the first rewarded trial after extinction brings the synaptic strength above threshold and therefore reinstates the behavior. The TANs endow our model with a similar property.

Tan and Bullock (2008) recently developed a Hodgkin–Huxley-type computational model of TAN firing that is more detailed than the model proposed here. For example, Tan and Bullock specifically modeled changes in several specific ion concentrations, along with the effects on TAN activation of dopamine and of GABAergic interneurons. On the other hand, they did not specifically model activity in striatal MSNs, nor did they model activity in any cells outside the striatum. Thus, their empirical applica-

tions are limited to data collected from single TAN units (e.g., no behavioral data were modeled). The major difference between their model and ours is that they account for intrinsic and learned TAN responses via modulation of TAN activity by GABAergic, and dopaminergic input rather than by synaptic plasticity (as we assume). Because the evidence of LTP at TAN synapses is good (Suzuki et al., 2001; Aosaki, Graybiel, et al., 1994), it seems likely that TAN pauses are modulated by all of these factors. In any case, Tan and Bullock's model may best be seen as a detailed theory of TAN responses within classical conditioning paradigms. Our model, however, is primarily concerned with how TAN responses act to gate learning at corticostriatal synapses and how this function influences behavior in instrumental conditioning paradigms and more general striatal-dependent behaviors.

### Limitations

It is important to acknowledge the rather severe limitations of the theory proposed here. First, with respect to its interactions with cortex, it is important to note that the striatum is organized into a set of functionally separate, parallel loops (Alexander, DeLong, & Strick, 1986). Which loop a particular subregion of striatum is in is determined primarily by the cortical regions that project to it. The theory developed here concerns the learning of stimulus–response associations and therefore applies to striatal regions receiving input from sensory areas of cortex (e.g., see Figure 1). This excludes anterior regions of striatum, which are innervated primarily by areas of frontal cortex. For example, tasks that activate pFC also commonly activate the head of the caudate nucleus because pFC projects strongly to this anterior region of the striatum. pFC and its striatal targets (e.g., dorsal striatum, head of the caudate nucleus) have their own role in context processing, which is beyond the scope of this article. For example, there is considerable evidence that a pFC—head of the caudate circuit—plays a critical role in attentional switching between different contexts (e.g., Robbins, 2007). There is also evidence that the TANs play a critical role in this process (Ragozzino, 2003), so a theory that attempts to account for such switching may postulate a role for the TANs similar to the one proposed here.

It is also important to note that the theory proposed here is meant to apply only to initial skill (or habit) learning. With overtraining, skills eventually come to be executed automatically. Following similar suggestions in the literature (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Miller, 1981, 1988), Ashby et al. (2007) proposed a model in which the development of automaticity is mediated by a transfer of control from the cortical–striatal pathways emphasized here to cortical–cortical pathways from the relevant areas of sensory cortex to the areas of premotor and motor cortices that mediate the selection and execution of the appropriate motor program. For example, several studies have reported evidence that with overtraining, skills of the type modeled here become independent of both dopamine and

the striatum (e.g., Beshpalov, Harich, Jongen-Rêlo, van Gaalen, & Gross, 2007; Choi, Balsam, & Horvitz, 2005; Turner, McCairn, Simmons, & Bar-Gad, 2005; Carelli et al., 1997). It would be straightforward to augment the present model with the cortical–cortical pathways proposed by Ashby et al. (with cortical–cortical plasticity mediated by Hebbian learning). This augmented model should be used to make predictions about the effects of overtraining on the tasks considered in this article.

Third, we have greatly oversimplified the neuroanatomy of the BG, omitting, for example, the GABAergic interneurons, the striosomes (i.e., patch compartments), the ventral striatum, and the indirect and hyperdirect pathways. However, rather than build the most complete model of the BG that was possible, our goal instead was to focus on the effects of the TANs on MSNs. For all regions downstream of the striatum, we simply tried to construct the simplest reasonable model that could account for the limited behavioral phenomena considered in this article. It seems likely that if the model was extended to more complex behaviors, then more biological detail would be needed in downstream areas. This generalization will be a goal of future research.

## APPENDIX: SIMULATION METHODS

### General Methods

Each modeling application was based on the network structure illustrated in Figures 1 or 2, depending on whether there were two or one response alternatives, respectively. Solutions to all differential equations were estimated using Euler’s method. In single-response applications, the model made a response whenever the output of the premotor unit crossed a threshold. In the two-response application, the model responded A or B depending on whether the output from premotor unit A or premotor unit B crossed a threshold first. If the output from neither premotor unit crossed the threshold during the trial, then the model responded A if the maximum output value of the premotor A unit was greater than the maximum output value of the premotor B unit. The behavioral simulations were replicated 100 times, and the results were averaged.

### Parameter Estimation

We began by estimating the intrinsic firing properties of each cell (e.g., the numerical values of 80 and 25 in Equation 2) followed by the synaptic strengths between units (e.g.,  $\beta_S$  in Equation 2). The parameters in the learning equations were estimated last. The parameter estimates from all applications are listed in Table A1.

It is important to point out that although the model includes many numerical parameters, its performance is qualitatively inflexible. There is a range on each parameter that allows the model to make responses and to learn. For numerical values outside this range, the behavior of the

**Table A1.**

	Single-response Model	Two-response Model
<i>Equation 1</i>		
$\beta_S$	125	125
$\gamma_S$	—	1.25
$E$	100	100
$\sigma_S$	5	5
$w_{K,J}$	0.2	Uniform (0.2, 0.225)
<i>Equation 3</i>		
$\lambda$	100	100
<i>Equations 4 and 5</i>		
$v(n), n = 0$	0.2	0.2
<i>Equation 6</i>		
$\alpha_G$	0.4175	0.4175
<i>Equation 7</i>		
$\beta_T$	0.275	0.275
<i>Equation 8</i>		
$\beta_C$	0.35	0.35
$\gamma_C$	—	0.1
$\sigma_C$	10	10
Response threshold	4.5	5.0
<i>Equation 9 (MSN)</i>		
$\alpha_w$	$0.07 \times 10^{-9}$	$1.0 \times 10^{-9}$
$\beta_w$	$0.02 \times 10^{-9}$	$0.9 \times 10^{-9}$
$\gamma_w$	$0.005 \times 10^{-9}$	$0.005 \times 10^{-9}$
<i>Equation 9 (TAN)</i>		
$\alpha_w$	$0.6 \times 10^{-7}$	$0.8 \times 10^{-7}$
$\beta_w$	$0.1 \times 10^{-7}$	$0.2 \times 10^{-7}$
$\gamma_w$	$0.005 \times 10^{-7}$	$0.005 \times 10^{-7}$
<i>Equation 9 (General)</i>		
$\theta_{AMPA}$	10	10
$\theta_{NMDA}$	25	25

network collapses (e.g., a unit never fires, no matter what its input, or it always fires). Within the range of reasonable parameter values, the model tends to always make the same qualitative predictions. For example, the TANs always inhibit the MSNs, so when the TAN activity decreases, the MSNs become more responsive to cortical input. Different numerical values of the parameters within the reasonable range tend to change the predictions of the model only slightly. For example, learning and extinction rates may change, but not whether the model learns or extinguishes. Thus, we believe that all of the predictions derived in this article follow in a necessary fashion from the general architecture of the model and do not depend in any critical way on our ability to find exactly the right set of parameter values.

To verify these observations more formally, we implemented the following sensitivity analysis for the most complex empirical application reported in this article—namely, our demonstration that the model accounts for fast reacquisition after extinction of an instrumental behavior (i.e., top panel of Figure 7). The analysis proceeded as follows. For each of the nine most important parameters in the model (Equation 8 response threshold,  $\theta_{\text{AMPA}}$  and  $\theta_{\text{NMDA}}$  from Equation 9, Equation 9 values of  $\alpha_w$ ,  $\beta_w$ , and  $\gamma_w$  for corticostriatal learning, Equation 9 values of  $\alpha_w$ ,  $\beta_w$ , and  $\gamma_w$  for CM/Pf-TAN learning), we successively changed the parameter estimate from the value used to generate the predictions shown in the top panel of Figure 7 by  $-1\%$ ,  $-10\%$ ,  $+1\%$ , and  $+10\%$ . After each change, we simulated the behavior of the model in the same conditions used to generate Figure 7. Next, after each new simulation, we computed the correlation between the learning curve shown in the top panel of Figure 7 and the learning curve produced by the new version of the model. In all except one case, these correlations exceeded .99, suggesting that the model makes the same qualitative predictions for a wide range of each of its parameters. The only exception occurred for a  $+10\%$  increase in the response threshold parameter (i.e., the threshold for a cortical unit to initiate a motor response). In this case, the correlation was .74. Importantly, however, even in this case, the model predicted that reacquisition was faster than original acquisition.

### Specific Notes on Model Fitting

When updating cortical-striatal synaptic strengths, we summed the total positive medium spiny activation during stimulus presentation to obtain the postsynaptic activity sum. When updating CM/Pf-TAN synapses, we only summed the first 200 msec after stimulus presentation to obtain the postsynaptic activity sum. This was necessary because the TAN pauses for most of the stimulus presentation. By limiting our sum to the first 200 msec after stimulus presentation, we were essentially capturing the short burst of spikes that tends to precede each pause.

The data from the Barnes et al. (2005) experiment were based on an average of 38 trials per block for six blocks

during acquisition, 33 trials per block for five blocks during extinction, and 38 trials per block for six blocks during reacquisition. Acquisition and reacquisition occurred with continuous reinforcement. During extinction, four animals received rewards on 3–9% of trials in each block, and three animals never received a reward. We simulated this experiment with the single-response version of the model (as in Figures 2 and 3) to ensure that the same model was used to generate both Figures 7 and 8. To generate predictions, the model was run through the entire experiment for 70 iterations (i.e., 228 trials of acquisition, 165 trials of extinction, and 228 trials of reacquisition). In all iterations, the model received continuous reinforcement during acquisition and reacquisition. For 40 iterations, the model received a reward on 5% of extinction trials, and for the other 30 iterations, it never received a reward during extinction. Spike frequencies were converted to relative firing rates in the following way. First, note that Barnes et al. defined relative firing rate as the proportion of total spikes recorded while the animal was in the maze that were produced in response to the auditory cue. We assumed that the Session 1 data from the acquisition period could be used to estimate baseline firing levels (i.e., before significant learning has occurred). Figure 8 shows that this value is roughly .10 (actually slightly less), which means that 10% of the total spikes before learning are produced to the tone. Suppose the absolute number of spikes produced to the tone before learning was  $B_0$ , then the total number of spikes produced while the animal is in the maze (before learning) is  $10B_0$ . Let  $N$  equal the number of spikes produced by the model in response to the tone. As mentioned before (see footnote 9), our model of MSNs has a baseline firing level of 0. So before learning,  $n = 0$  (or a very small number). Thus, on each trial, we assumed the number of spikes recorded in response to the tone was  $N + B_0$ , and the total number of spikes recorded for the entire trial was  $N + 10B_0$ . Figure 8 was produced with a value of  $B_0 = 6.5$ .

### Acknowledgments

This research was supported by NIH Grants R01 MH3760-2 and P01 NS044393 and by support from the U.S. Army Research Office through the Institute for Collaborative Biotechnologies under contract DAAD19-03-D-0004. The authors thank John Ennis for his help in developing an earlier version of the model proposed here and Aaron Ettenberg for his helpful comments.

Reprint requests should be sent to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, CA 93106, or via e-mail: ashby@psych.ucsb.edu.

### Notes

1. Evidence suggests that the postsynaptic effect of ACh is to stabilize MSN membrane potential while it is either in the up or down state (Gabel & Nisenbaum, 1999). In contrast, the presynaptic effects seem to be mostly inhibitory (mediated by muscarinic M2 receptors on the axons of cortical pyramidal neurons; e.g., Calabresi et al., 2000).

2. We modeled the inhibitory effect of TANs on activation at cortical–striatal synapses as postsynaptic. As mentioned above, the most significant inhibitory effect may be presynaptic (Pakhotin & Bracci, 2007; Calabresi et al., 2000). Modeling the inhibitory effects as postsynaptic simplifies the model because it allows us to model cortical input as a simple square wave. We are confident that none of the simulations reported in this article would change in any significant way if we changed the model by replacing the square-wave model of cortical input with a more realistic spiking model and making the TAN inhibition presynaptic rather than postsynaptic. Note also that our model ignores postsynaptic excitatory effects of ACh. These are poorly understood, and it is not clear what role they play in cortical–striatal dynamics or how they should be modeled.

3. This is the classical view of the basal ganglia (i.e., as providing a brake on cortex). Thalamic neurons frequently fire a rebound burst when released from inhibition, however (Sherman & Guillery, 2006), so another possibility may be that striatal firing initiates an excitatory input from thalamus to cortex. We believe that the qualitative behavior of our model would not change if we had included rebound spiking in our model of thalamus.

4. Bayer and Glimcher (2007) recently reported evidence that negative RPEs may be coded by the duration of the pause in dopamine cell firing. This suggests that the dynamic range of positive and negative RPEs may be more balanced than we have modeled. However, we also constructed a model with equal dynamic range for positive and negative RPEs and found that the model’s qualitative behavior and ability to account for the data were unaffected.

5. Note that in the data of Aosaki, Tsubokawa, et al. (1994), the TANs fire a burst at the end of the pause and then quickly reduce their firing to baseline levels. We chose not to model this feature of the data of Aosaki, Tsubokawa, et al. (1994) because the data of Reynolds et al. (2004) shown in Figure 4 do not display this property. If we had incorporated a rebound burst into the model, then Figures 4 and 6 would change of course, but none of the other predictions reported in this article would change in any way.

6. We did not explicitly model this broad tuning. Instead, we mimicked the effects of broad tuning by setting the learning rates higher on the CM/Pf–TAN synapse than on the sensory cortex–MSN synapses (i.e., see Table A1 for specific numerical values of all parameters).

7. The relative firing rate plotted in Figure 8 is defined as the number of spikes elicited by the auditory tone divided by the total number of spikes recorded during the entire time the animal was running in the maze.

8. We assumed that the response of each unit decreased as a Gaussian function of the distance in stimulus space between the stimulus preferred by that unit and the presented stimulus. Specifically, if a stimulus with stylus speed  $x_s$  mm/sec was presented, then the response of the unit maximally tuned to speed  $x$  mm/sec was  $\exp[-(x - x_s)^2 / 2.5]$ .

9. The Izhikevich (2007) model of MSN activation used in this article gives a good account of patch–clamp experiments, but the model predicts that MSNs never fire spontaneously. In fact, MSNs do not have a high spontaneous firing rate (e.g., Wilson, 1995). Nevertheless, they do sporadically fire in the absence of significant stimulation. This is easily seen in Figure 9. We chose to model this spontaneous activity by adding a Poisson process to the spike trains that were generated from Equations 1 and 2. In the present application, this noise process added, on average, three spikes per second. We augmented the model in this way only for the two applications where we fit the model to spike trains from MSNs (i.e., this application and the application to the data of Barnes et al., 2005). Note, however, that even without adding these extra random spikes, the model still accounts for the most important qualitative properties of the data

of Merchant et al. (1997)—namely, category-specific responding during the active condition and no response to these same stimuli in the passive condition. It is also important to note that adding or not adding this extra noise source would not affect any other applications in this article.

## REFERENCES

- Akaike, A., Sasa, M., & Takaori, S. (1988). Muscarinic inhibition as a dominant role in cholinergic regulation of transmission in the caudate nucleus. *Journal of Pharmacology and Experimental Therapeutics*, *246*, 1129–1136.
- Akins, P. T., Surmeier, D. J., & Kitai, S. T. (1990). Muscarinic modulation of a transient K<sup>+</sup> conductance in rat neostriatal neurons. *Nature*, *344*, 240–242.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*, 357–381.
- Anderson, M. E., & DeVito, J. L. (1987). An analysis of potentially converging inputs to the rostral ventral thalamic nuclei of the cat. *Experimental Brain Research*, *68*, 260–276.
- Aosaki, T., Graybiel, A. M., & Kimura, M. (1994). Effect of the nigrostriatal dopamine system on acquired responses in the striatum of behaving monkeys. *Science*, *265*, 412–415.
- Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., & Kimura, M. (1994). Responses of tonically active neurons in the primate’s striatum undergo systematic changes during behavioral sensorimotor conditioning. *Journal of Neuroscience*, *14*, 3969–3984.
- Apicella, P. (2007). Leading tonically active neurons of the striatum from reward detection to context recognition. *Trends in Neurosciences*, *30*, 299–306.
- Apicella, P., Legallet, E., & Trouche, E. (1997). Responses of tonically discharging neurons in the monkey striatum to primary rewards delivered during different behavioral states. *Experimental Brain Research*, *116*, 456–466.
- Apicella, P., Scarnati, E., & Schultz, W. (1991). Tonic discharging neurons of monkey striatum respond to preparatory and rewarding stimuli. *Experimental Brain Research*, *84*, 672–675.
- Arbuthnott, G. W., Ingham, C. A., & Wickens, J. R. (2000). Dopamine and synaptic plasticity in the neostriatum. *Journal of Anatomy*, *196*, 587–596.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.
- Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*, 632–656.
- Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*, 1158–1161.
- Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 129–141.
- Bayer, H. M., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, *98*, 1428–1439.
- Bespalov, A. Y., Harich, S., Jongen-Rêlo, A.-L., van Gaalen, M. M., & Gross, G. (2007). AMPA receptor antagonists

- reverse effects of extended habit training on signaled food approach responding in rats. *Psychopharmacology*, *195*, 11–18.
- Blazquez, P. M., Fujii, N., Kojima, J., & Graybiel, A. M. (2002). A network representation of response probability in the striatum. *Neuron*, *33*, 973–982.
- Bonsi, P., Martella, G., Cuomo, D., Platania, P., Sciamanna, G., Bernardi, G., et al. (2008). Loss of muscarinic autoreceptor function impairs long-term depression but not long-term potentiation in the striatum. *Journal of Neuroscience*, *28*, 6258–6263.
- Caan, W., Perrett, D. I., & Rolls, E. T. (1984). Responses of striatal neurons in the behaving monkey. 2. Visual processing in the caudal neostriatum. *Brain Research*, *290*, 53–65.
- Calabresi, P., Centonze, D., Gubellini, P., Pisani, A., & Bernardi, G. (2000). Acetylcholine-mediated modulation of striatal function. *Trends in Neurosciences*, *23*, 120–126.
- Calabresi, P., Pisani, A., Mercuri, N. B., & Bernardi, G. (1996). The corticostriatal projection: From synaptic plasticity to dysfunctions of the basal ganglia. *Trends in Neurosciences*, *19*, 19–24.
- Carelli, R. M., Wolske, M., & West, M. O. (1997). Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *Journal of Neuroscience*, *17*, 1804–1814.
- Centonze, D., Gubellini, P., Bernardi, G., & Calabresi, P. (1999). Permissive role of interneurons in corticostriatal synaptic plasticity. *Brain Research Reviews*, *31*, 1–5.
- Choi, W. Y., Balsam, P. D., & Horvitz, J. C. (2005). Extended habit training reduces dopamine mediation of appetitive response expression. *Journal of Neuroscience*, *25*, 6729–6733.
- Cornwall, J., & Phillipson, O. T. (1988). Afferent projections to the parafascicular thalamic nucleus of the rat, as shown by the retrograde transport of wheat germ agglutinin. *Brain Research Bulletin*, *20*, 139–150.
- Cragg, S. J. (2006). Meaningful silences: How dopamine listens to the ACh pause. *Trends in Neurosciences*, *29*, 125–131.
- Dotz, H. U., & Misgeld, U. (1986). Muscarinic slow excitation and muscarinic inhibition of synaptic transmission in the rat neostriatum. *Journal of Physiology*, *380*, 593–608.
- Doyon, J., & Ungerleider, L. G. (2002). Functional anatomy of motor skill learning. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of memory* (pp. 225–238). New York: Guilford Press.
- Ermentrout, G. B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Computation*, *8*, 979–1001.
- Gabel, L. A., & Nisenbaum, E. S. (1999). Muscarinic receptors differentially modulate the persistent potassium current in striatal spiny projection neurons. *Journal of Neurophysiology*, *81*, 1418–1423.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*, 1569–1572.
- Izhikevich, E. M. (2007). *Dynamical systems in neuroscience: The geometry of excitability and bursting*. Cambridge, MA: MIT Press.
- Kakade, S., & Dayan, P. (2002). Acquisition and extinction in autoshaping. *Psychological Review*, *109*, 544–553.
- Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.
- Kawaguchi, Y., Wilson, C. J., Augood, S. J., & Emson, P. C. (1995). Striatal interneurons: Chemical, physiological and morphological characterization. *Trends in Neurosciences*, *18*, 527–535.
- Kimura, M. (1992). Behavioral modulation of sensory responses of primate putamen neurons. *Brain Research*, *84*, 204–214.
- Kimura, M., Rajkowski, J., & Evarts, E. (1984). Tonicly discharging putamen neurons exhibit set-dependent responses. *Proceedings of the National Academy of Sciences, U.S.A.*, *81*, 4998–5001.
- Matsumoto, N., Minamimoto, T., Graybiel, A. M., & Kimura, M. (2001). Neurons in the thalamic Pf complex supply striatal neurons with information about behaviorally significant sensory events. *Journal Neurophysiology*, *85*, 960–976.
- Merchant, H., Zainos, A., Hernandez, A., Salinas, E., & Romo, R. (1997). Functional properties of primate putamen neurons during the categorization of tactile stimuli. *Journal of Neurophysiology*, *77*, 1132–1154.
- Miller, R. (1981). *Meaning and purpose in the intact brain*. Oxford: Oxford University Press.
- Miller, R. (1988). Cortico-striatal and cortico-limbic circuits: A two-tiered model of learning and memory functions. In H. J. Markowitsch (Ed.), *Information processing by the brain: Views and hypotheses from a cognitive-physiological perspective* (pp. 179–198). Bern: Hans Huber Press.
- Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, *43*, 133–143.
- O’Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*, 452–454.
- O’Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.
- Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*, 563–593.
- Pakhotin, P., & Bracci, E. (2007). Cholinergic interneurons control the excitatory input to the striatum. *Journal of Neuroscience*, *27*, 391–400.
- Ragozzino, M. E. (2003). Acetylcholine actions in the dorsomedial striatum support the flexible shifting of response patterns. *Neurobiology of Learning and Memory*, *80*, 257–267.
- Rall, W. (1967). Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, *30*, 1138–1168.
- Redish, A. D., Jensen, S., Johnson, A., & Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addition, relapse, and problem gambling. *Psychological Review*, *114*, 784–805.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.
- Reynolds, J. N. J., Hyland, B. I., & Wickens, J. R. (2004). Modulation of an afterhyperpolarization by the substantia nigra induces pauses in the tonic firing of striatal cholinergic interneurons. *Journal of Neuroscience*, *24*, 9870–9877.
- Reynolds, J. N. J., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, *15*, 507–521.
- Ricker, S. T., & Bouton, M. E. (1996). Reacquisition following extinction in appetitive conditioning. *Animal Learning & Behavior*, *24*, 423–436.
- Robbins, T. W. (2007). Shifting and stopping: Fronto-striatal substrates, neurochemical modulation and clinical implications. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *362*, 917–932.
- Sadikot, A. F., Parent, A., & Francois, C. (1992). Efferent connections of the centromedian and parafascicular

- thalamus in the squirrel monkey: A PHA-L study of subcortical projections. *Journal of Comparative Neurology*, *315*, 137–159.
- Sardo, P., Ravel, S., Legallet, E., & Apicella, P. (2000). Influence of the predicted time of stimuli eliciting movements on responses of tonically active neurons in the monkey striatum. *European Journal of Neuroscience*, *12*, 1801–1816.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.
- Sherman, S. M., & Guillery, R. W. (2006). *Exploring the thalamus and its role in cortical function*. Cambridge, MA: MIT Press.
- Shimo, Y., & Hikosaka, O. (2001). Role of tonically active neurons in primate caudate in reward-oriented saccadic eye movement. *Journal of Neuroscience*, *21*, 7804–7814.
- Smith, Y., Raju, D. V., Pare, J. F., & Sidibe, M. (2004). The thalamostriatal system: A highly specific network of the basal ganglia circuitry. *Trends in Neurosciences*, *27*, 520–527.
- Suzuki, T., Miura, M., Nishimura, K., & Aosaki, T. (2001). Dopamine-dependent synaptic plasticity in the striatal cholinergic interneurons. *Journal of Neuroscience*, *21*, 6492–6501.
- Tan, C. O., & Bullock, D. (2008). A dopamine-acetylcholine cascade: Simulating learned and lesion-induced behavior of striatal cholinergic interneurons. *Journal of Neurophysiology*, *100*, 2409–2421.
- Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *Journal of Neuroscience*, *23*, 10402–10410.
- Turner, R. S., McCairn, K., Simmons, D., & Bar-Gad, I. (2005). Sequential motor behavior and the basal ganglia: Evidence from a serial reaction time task in monkeys. In J. P. Bolam, C. A. Ingham, & P. J. Magill (Eds.), *The basal ganglia VIII (advances in behavioral biology)* (Vol. 56, pp. 563–574). New York: Springer.
- Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550–592.
- Van der Werf, Y., Witter, M. P., & Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus: Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Research Reviews*, *39*, 107–140.
- Wang, Z., Kai, L., Day, M., Ronesi, J., Yin, H., Ding, J., et al. (2006). Dopaminergic control of corticostriatal long-term synaptic depression in medium spiny neurons is mediated by cholinergic interneurons. *Neuron*, *50*, 443–452.
- Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 29–50). Cambridge, MA: Bradford.
- Yelnik, J., Francois, C., Percheron, G., & Tande, D. (1991). Morphological taxonomy of the neurons of the primate striatum. *Journal of Comparative Neurology*, *313*, 273–294.
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*, 464–476.
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, *22*, 513–523.
- Zackheim, J., & Abercrombie, E. D. (2005). Thalamic regulation of striatal acetylcholine efflux is both direct and indirect and qualitatively altered in the dopamine-depleted striatum. *Neuroscience*, *131*, 423–436.