A Local Circuit Model of Learned Straitial and Dopamine Cell Responses under Probabilistic Schedules of Reward

Tan, Can Ozan

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems

http://hdl.handle.net/2144/1950

Boston University
A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward

Can Ozan Tan and Daniel Bullock

Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215

Corresponding author:
Daniel Bullock
Phone: 617-353-9486
Fax: 617-353-7755
E-mail: danb@cns.bu.edu

Submitted April, 2007
CAS/CNS Technical Report 2007-012

Copyright © 2007
Permission to copy without fee all or part of this material is granted provided that: 1. The copies are not made or distributed for direct commercial advantage; 2. the report title, author, document number, and release date appear, and notice is given that copying is by permission of the BOSTON UNIVERSITY CENTER FOR ADAPTIVE SYSTEMS AND DEPARTMENT OF COGNITIVE AND NEURAL SYSTEMS. To copy otherwise, or to republish, requires a fee and / or special permission.
Abstract

Before choosing, it helps to know both the expected value signaled by a predictive cue and the associated uncertainty that the reward will be forthcoming. Recently, Fiorillo et al. (2003) found that dopamine (DA) neurons of the SNc exhibit sustained responses related to the uncertainty that a cue will be followed by reward, in addition to phasic responses related to reward prediction errors (RPEs). This suggests that cue-dependent anticipations of the timing, magnitude, and uncertainty of rewards are learned and reflected in components of the DA signals broadcast by SNc neurons. What is the minimal local circuit model that can explain such multifaceted reward-related learning? A new computational model shows how learned uncertainty responses emerge robustly on single trials along with phasic RPE responses, such that both types of DA responses exhibit the empirically observed dependence on conditional probability, expected value of reward, and time since onset of the reward-predicting cue. The model includes three major pathways for computing: immediate expected values of cues, timed predictions of reward magnitudes (and RPEs), and the uncertainty associated with these predictions. The first two model pathways refine those previously modeled by Brown et al. (1999). A third, newly modeled, pathway is formed by medium spiny projection neurons (MSPNs) of the matrix compartment of the striatum, whose axons co-release GABA and a neuropeptide, substance P, both at synapses with GABAergic neurons in the SNr and with the dendrites (in SNr) of DA neurons whose somas are in ventral SNc. Co-release enables efficient computation of sustained DA uncertainty responses that are a non-monotonic function of the conditional probability that a reward will follow the cue. The new model’s incorporation of a striatal microcircuit allowed it to reveals that variability in striatal cholinergic transmission can explain observed differences, between monkeys, in the amplitude of the non-monotonic uncertainty function. Involvement of matriceal MSPNs and striatal cholinergic transmission implies a relation between uncertainty in the cue-reward contingency and action-selection functions of the basal ganglia. The model synthesizes anatomical, electrophysiological and behavioral data regarding the midbrain DA system in a novel way, by relating the ability to compute uncertainty, in parallel with other aspects of reward contingencies, to the unique distribution of SP inputs in ventral SN.
1 Introduction

Firing patterns observed in the dopamine cells of the substantia nigra pars compacta (SNC) and ventral tegmental area (VTA) are related to the occurrence and non-occurrence of rewards and cues that predict reward (Schultz et al., 1997; Schultz, 1998). Dopamine cells, which fire tonically at moderate levels, respond immediately to unexpected rewards with a phasic burst. When a reward consistently follows a preceding conditioned stimulus (CS), the phasic burst appears to “transfer” from the time of expected reward delivery to the time of CS onset. That is, the conditioned response to CS onset grows over learning trials whereas the unconditioned response to actual reward delivery declines over learning trials. The amount of the “transfer” depends on the expectation of reward, 
\[ \hat{R} = |R^*| \times p(R^*|CS), \]
that is the conditional probability, \( p(R^*|CS) \), that a reward of magnitude \(|R^*|\) follows the CS (Fiorillo et al., 2003; Schultz et al., 1997; Schultz, 1998, 2004; Tobler et al., 2005). After learning, the omission of an expected reward induces a dopamine cell pause, i.e., a depression in firing rate to a below-baseline level, at the time of expected reward delivery. If a reward occurs at an earlier time than expected, then the conditioned burst to the CS onset will be followed by a second burst at the time of the earlier-than-expected reward. Thus dopamine cells are part of an adaptive system that uses learned expectations to filter reward-related signals. This filtering creates dopamine bursts and pauses that respectively signal positive and negative violations of reward-related predictions. Formal studies of reinforcement learning (Schultz et al., 1997; Schultz, 1998) have shown that such “reward prediction error” (RPE) signals often lead to faster learning than unfiltered reward signals. Thus SNC/VTA dopamine signals, which are broadcast to the dorsal and ventral striatum as well as other brain regions, such as the amygdala and frontal cortex, can be conceptualized as internal teaching signals that foster rapid acquisition of goal-directed behavior (Schultz et al., 1997; Schultz, 1998; Doya, 2002).

Several proposals have been made to explain the adaptive computations that give rise to the RPE responses of dopamine neurons (see review in Wörgötter and Porr, 2005). Some have been implementations of the temporal difference (TD) model (e.g., Nakahara et al., 2004), whereas oth-
ers have been based on local circuit anatomy and physiology (Houk et al., 1995; Brown et al., 1999). Brown et al. (1999) introduced and simulated a local circuit model, based on known basal ganglia (BG) anatomy, that can explain most of the key results outlined above without predicting effects that are known not to occur. In that model, the CS engenders learned fast excitatory and adaptively-timed inhibitory inputs to DA cells. These two learned inputs to DA cells are processed by separate pathways, and the adaptive timing of the inhibitory inputs is mediated by second messenger dynamics in striatal neurons located in striosomes (“patch” compartments of the striatum; Gerfen, 1992). Whereas TD models incorrectly predict that during training the DA burst will gradually slide from the time of reward delivery to the time of CS onset, the Brown et al. (1999) model correctly predicts a gradual transfer of the DA burst from the time of reward delivery to the time of the CS onset: a burst never occurs at times intermediate between the CS onset and the reward delivery.

Recently, however, Fiorillo et al. (2003) discovered a new reward-related component of DA cell discharges, which they called an “uncertainty response”. This component builds up during the interval from CS onset to the expected time of reward if the reward schedule is probabilistic, and contrary to TD model expectations (Niv et al., 2005), Fiorillo et al. (2005) showed that uncertainty responses are robust in single trial data. The size of the buildup depends on two factors, the conditional probability, \( p(R^*|CS) \), and the amount of reward, \(|R^*|\), that is at risk. The maximal “uncertainty” response is observed if \( p(R^*|CS) \) is 0.5, but it disappears if \( p(R^*|CS) \) is 0.0 or 1.0, and a \( p(R^*|CS) \) of 0.75 or 0.25 yields an intermediate-sized response. Thus this component is a non-monotonic, “inverted-U”, function of \( p(R^*|CS) \). Neither the “uncertainty response” nor its functional dependencies were predicted by any BG (basal ganglia) learning model of the time. However, the Brown et al. (1999) model did imply a separate new observation of Fiorillo et al. (2003) and Tobler et al. (2005), namely that the degree of transfer of the phasic DA cell component (from time of reward to time of CS onset) is an increasing function of \(|R^*| \times p(R^*|CS)\). In summary, DA cells exhibit at least two learned reward-related response components. Both are func-
tions of $p(R^*|CS)$ and $|R^*|$, but the sustained uncertainty response is a non-monotonic function of conditional probability $p(R^*|CS)$, whereas the phasic response to CS is a monotonic function of expected reward $\hat{R} = |R^*| \times p(R^*|CS)$. Although some DA neurons display both components, many DA neurons exhibited only one (C. Fiorillo, pers. comm.).

The inability of the Brown et al. (1999) model to account for uncertainty responses probably stem from its omission of many known features of the BG microcircuit that can be expected to play important roles in shaping DA responses. The model SN had only one class of DA cells, and lacked co-release of GABA and neuropeptides. The model striatum lacked tonically active neurons (TANs), notably the large cholinergic interneurons (Aosaki et al., 1994b, 1995), as well as GABA-ergic interneurons. The TANs express DA receptors (Yan et al., 1997; Aosaki et al., 1998) and are recipients of DA-ergic fibers from SNc (Aosaki et al., 1994a; Aosaki, 2003). As with DA neurons, the responses of TANs are both stimulus-dependent and conditionable (Aosaki et al., 1994b; Apicella et al., 1998; Apicella, 2002), although TANs’ dominant responses to a CS are pauses rather than bursts (Aosaki et al., 1995; Apicella et al., 1997; Apicella, 2002). In this paper, we propose an extended model based on such features and corresponding computational hypotheses. The extensions include thalamo-striatal inputs, striatal GABAergic, and cholinergic, interneurons, and neuropeptide co-release with GABA in SN. The new model retains the explanatory successes of the prior model, but also incorporates an efficient basis for the uncertainty responses discovered by Fiorillo et al. (2003).

2 The Model

2.1 Circuit for Learned Phasic Dopamine Responses
The model is schematized in Figure 1, which shows both tonic and phasic DA cell types in the lower right corner. The Figure also labels sites in the circuit with the names of corresponding variables in the formal model. The model for the genesis of phasic DA responses follows the proposal by Brown et al. (1999), with modest modifications. A model PPTN (pedunculopontine
Tegmental nucleus) stage relays signals generated by conditioned and unconditioned rewarding stimuli to DA cells in the substantia nigra pars compacta (SNc) (Nakamura and Ono, 1986; Semba and Fibiger, 1992; Brown et al., 1999). Accordingly, the major excitatory input to phasic DA cells, shown near the lower edge of the Figure, is a signal $P$ from the PPTN. Changes in the level of $P$ are described by the following differential equation, which reflects two afferents to PPTN: a primary reward induced input, $I_R$, from lateral hypothalamus and a conditioned stimulus induced input, $S$, from ventral striatum (via the interposed ventral pallidum):

$$\frac{1}{\tau_P} \frac{d}{dt} P = -(1 + U_P) W_{UP} P + (1 - P) \left[ W_{SP} S + W_{RP} I_R \right]$$

where $W_{RP}$ and $W_{SP}$ are synaptic weights that multiply $I_R$ and $S$, respectively, and $U_P$ is an afterhyperpolarization governed by equation:

$$\frac{1}{\tau_{UP}} \frac{d}{dt} U_P = -U_P + (1 - U_P) P$$

The CS (left side of Figure 1) is assumed to be relayed to the striatum via cortico-striatal afferents (although the relay to ventral striatum may be via amygdala). Within the BG, CS-related inputs excite both ventral striatum via a set of adaptive synaptic weights, and medium spiny projection neurons (MSPNs) in the striosome compartment of the striatum via another set of adaptive synaptic weights. After learning, the former pathway is responsible for eliciting the phasic DA response that immediately follows CS onset, and the learned ventral striatal activation level, $S$, is governed by

$$\frac{1}{\tau_S} \frac{d}{dt} S = -A_S S + (1 - S) \left[ W_{IS} I_i + I_R W_{RS} \right]$$

where $I_R$ again is the primary reward signal, multiplied by fixed weight $W_{RS}$, $I_i$ is a signal coming from the the $i^{th}$ CS representation to the ventral striatal cells, and $W_{IS}$ is the adaptive synaptic weight that multiplies this signal. Weight adaptation is governed by:

$$\frac{1}{\tau_{WS}} \frac{d}{dt} W_{iS} = S \left[ \alpha_{WS} I_i N^+ \left( C_{WS}^{\text{max}} - W_{iS} \right) - \beta_{WS} N W_{iS} \right]$$
where $C_{W_S}^{\text{max}}$ is the upper bound on each weight. Provided that the CS-induced signal $I_i$ is positive, then synaptic weight potentiation and depression are induced, respectively, by phasic dopamine burst or dip signals, $N^+$ and $N^-$ (defined below in equations 16 and 17). These signals are never on simultaneously.

The adaptive excitatory pathway from CSs to DA cells described above is complemented in the model by an adaptive inhibitory pathway from CSs to DA cells via striosomes. An inhibitory projection from striosomes to DA cells is well established (Gerfen, 1992). To explain the data on timing of dopamine dips noted above, the inhibitory signals carried by this projection must be adaptively timed to arrive at DA cells at the expected time of primary reward. Therefore model striosomal MSPN dendritic spines exhibit a spectrum of delayed calcium spikes mediated by second messengers (Fiala et al., 1996; Brown et al., 1999). When a delayed spike coincides with DA burst release engendered by primary reward (via lateral hypothalamus and PPTN, as shown in Figure 1) learning specific to the cortico-striatal synapse on that spine can occur. Such learning potentiates inhibitory outputs at the relevant delay, and such learning is self-terminating because once the inhibition is strong enough, it precisely cancels the excitatory effect of PPTN inputs to DA cells. The formal model approximates this mechanism as follows. A spectrum of striosomal MSPN second messenger activities $x_{ij}$ respond to the $i^{\text{th}}$ input at rates $r_j$:

$$\frac{d}{dt} x_{ij} = r_j [-x_{ij} + (1 - x_{ij})I_i] \quad (5)$$

where the second messenger buildup rates are given by

$$r_j = \frac{\alpha_r}{\beta_r + j}, \quad j = 1, 2, \ldots, n. \quad (6)$$

The activities $x_{ij}$ induce intracellular calcium dynamics within a given spine ($j$) at delays determined by $r_j$. The intracellular calcium spike is represented by the quantity $[G_{ij}Y_{ij}]^+$ (Brown et al., 1999), where

$$\frac{d}{dt} G_{ij} = \alpha_G(B_G - G_{ij})f_G(x_{ij} - \Gamma_G) - \beta_G G_{ij} \quad (7)$$
and

$$\frac{d}{dt} Y_{ij} = \alpha_Y (1 - Y_{ij}) - \beta_Y [G_{ij} Y_{ij} - \Gamma_Y]^+$$  \hspace{1cm} (8)$$

where $f_G(x)$ is a step function: 0 for $x \leq 0$ and 1 for $x > 0$. Parameters $\Gamma_G$ and $\Gamma_Y$ define signal thresholds for calcium accumulation (equation 7) and calcium decay (equation 8), respectively. In the brief interval when the calcium concentration at a particular spine exceeds a threshold activity $\Gamma_S$, CS-striosomal weight $Z_{ij}$ at that particular spine becomes eligible for adaptation that may be induced by dopaminergic bursts ($N^+$) or dips ($N^-$):

$$\frac{d}{dt} Z_{ij} = \alpha_Z [G_{ij} Y_{ij} - \Gamma_S]^+ [(A_Z - Z_{ij})N^+ - Z_{ij}N^-]$$  \hspace{1cm} (9)$$

Thus, phasic DA bursts and dips during learning trials respectively potentiate and depress two sets of adaptive weights, the $Z_{ij}$ and the $W_{iS}$ (treated above). As learning progresses, the CS comes to immediately excite the DA neurons through the ventral striatal pathway and to inhibit the same DA cells after a learned delay. The equation governing activity of phasic DA cells is:

$$\frac{1}{\tau_D} \frac{d}{dt} D_{phasic} = -D_{phasic} + (1 - D_{phasic}) [W_{PD} [P - \Gamma_P]^+ + I_D]$$

$$- (D_{phasic} + h_D) \sum_{i,j} [G_{ij} Y_{ij} - \Gamma_S]^+ Z_{ij}$$  \hspace{1cm} (10)$$

Here the thresholded PPTN signal $[P - \Gamma_P]^+$ excites DA cells, but the summed spectrum of striosomal MSPN signals $\sum_{i,j} [G_{ij} Y_{ij} - \Gamma_S]^+ Z_{ij}$ inhibits DA cells. In equation 10, $I_D$ represents endogenous factors that control the low baseline activation (and associated firing rate) of phasic DA cells.

The ensemble of mechanisms captured in equations 1-10 enables gradual learned transfer of scaled DA responses from the time of rewarding stimulus onset to the (earlier) time of CS onset (Schultz et al., 1997; Schultz, 1998, 2004). The original formulation of this model was shown to successfully explain the reward- and timing-related responses of phasic dopamine bursts (Brown
et al., 1999), and as shown below, all those properties are preserved in the current formulation.

### 2.2 Circuit for Genesis of Learned Tonic Dopamine Responses

Both GABAergic and dopaminergic cells in the substantia nigra exhibit (apparently) unlearned baseline activity levels. However, GABAergic MSPNs of the striatal matrix and striosome compartments, respectively, can inhibit the firing of the GABAergic pars reticulata (SNr) and dopaminergic pars compacta (SNc) cells, respectively (Ragsdale and Graybiel, 1990; Joel and Weiner, 2000). Moreover, the distal dendrites of GABAergic SNr cells are intermingled with dendrites from the dopaminergic cells of the SNc (Conde, 1992; Gerfen and Wilson, 1996). Thus, matriceal MSPNs whose terminals contact SNr dendrites may also contact invading dendrites of DA cells whose somata are in SNc. The model embodies the hypothesis that matriceal MSPNs exert an inhibitory effect on dopaminergic SNc cells as well as on GABAergic output neurons of the SNr, and that the MSPNs that modulate tonic firing rates and uncertainty responses of SNc dopaminergic cells (see below) are located in the matrix compartment of the striatum. This complements the treatment above, in which model MSPNs attributed to striosomes learned to regulate the phasic DA responses that are widely regarded as RPE signals (Schultz et al., 1997; Schultz, 1998; Brown et al., 1999).

Matriceal MSPNs are modeled by:

$$\frac{1}{\tau_M} \frac{d}{dt} M = -A_M M + (B_M - M) \alpha_{iS} W_{iS} I_i - M \left[ F(1 - T) + T \right]$$

where $W_{iS} I_i$ is adaptively weighted cortical input, and the inhibition, $F$ (equation S8 in Supplementary Material), from fast spiking interneurons (FS-INs), is gated by factor $(1 - T)$, which reflects activation of presynaptic receptors on FS-IN axon terminals (Koos and Tepper, 2002) by acetylcholine (ACh) released from TANs (whence “$T$”). The last term, $T$, reflects the direct inhibitory action of ACh on MSPNs (DiChiara et al., 1994). The model used to generate the ACh signal $T$ is treated in Tan and Bullock (2007) and summarized in Section S1 of the Supplementary Material (equation S7). The synaptic strengths of cortical inputs to the matriceal MSPNs are...
assumed to be governed by the same mechanism as ventral striatal cells ($W_{iS}$ in equation 4).

Equation 11 embodies the hypothesis that the primary inputs to the matriceal and striosomal MSPNs in striatum are relayed through glutamatergic cortical afferents via adaptive corticostriatal connections (Joel and Weiner, 2000). It also captures the influence of cholinergic transmission on MSPNs both directly (DiChiara et al., 1994) and indirectly through GABAergic interneurons (Koos and Tepper, 1999, 2002; DeRover et al., 2002; Ravel et al., 2003). This treatment is consistent with reports showing that TANs (the tonically active giant cholinergic interneurons) are preferentially located in, or near the border of, the matrix compartment of the striatum (Kubota and Kawaguchi, 1993; Prensa et al., 1999). Although both phasic and sustained model responses survive without FS-INs and TANs in the striatal microcircuit, their inclusion in the model markedly enhances the sensitivity of sustained responses of model DA neurons to the conditional probability of reward (see Supplementary Material).

In addition to releasing GABA, the MSPNs of the direct pathway release the neuropeptide Substance-P (SP) (Parent, 1986; Conde, 1992; Otsuka and Yoshioka, 1993). Such SP release from the axon terminals of MSPNs projecting to SNr can excite the comingled dendrites of DA cells (Conde, 1992; Otsuka and Yoshioka, 1993). Beaujouan et al. (2004) reported activity-dependent release of SP from striatonigral terminals, and SP released intranigrally increases striatal dopamine release (Reid et al., 1990; Khan et al., 1996). To model this, we assume that the GABA release level, $M_{RG}$, and the substance-P release level, $M_{RS}$, in SN are equal and proportional to the threholded matriceal MSPN activation:

$$M_{RS} = M_{RG} = \alpha_R [M - \Gamma_M]^+$$  

(12)

In accord with data already noted, the SP release level, $M_{RS}$, excites the “tonic” type of model DA cell. However, this excitation by SP level $M_{RS}$ is gated by factor $[1 - M_{RG}]^+$. Because this multiplicative factor declines as co-release of GABA increases, the product $M_{RS} [1 - M_{RG}]^+$ is a non-monotonic, inverted-U, function of matriceal MSPN activation level $M_R$. The activity,
(\(D_{\text{tonic}}\)), of the subset of SNc DA cells that receive this SP/GABA signal is modeled by:

\[
\frac{1}{\tau_D} \frac{d}{dt} D_{\text{tonic}} = -D_{\text{tonic}} + (1 - D_{\text{tonic}}) \left[ I_D + M_{RS} [1 - M_{RG}]^+ \right]
\] (13)

This model emphasizes GABA/SP co-release as a sufficient, and quite efficient, mechanism. However, there may be some redundancy of mechanisms. Non-monotonic sustained DA responses can also be induced by experimental intranigral application of SP, by itself (Reid et al., 1990). However, some other reports (Betarbet and Greenamyre, 2004; Hanson et al., 2002; Martorana et al., 2003) suggest that physiological levels of SP release may cause monotonic excitation, and part of the non-monotonic effect in Reid et al. (1990) may have resulted from SP activation of nigral GABAergic neurons that give off inhibitory collaterals. Indeed, Mendez et al. (1993) reported that GABAergic cells in the SNr that are also recipients of SP terminals make synaptic contacts with the dendrites that DA cells of the SNc extend ventrally into SNr. Moreover, Whitty et al. (1997) reported that NKR1 mRNAs were found in some non-pigmented (i.e., non-DA) cells of the ventral SN. Thus, GABAergic neurons of SNr that become excited by high SP release may help to shunt the excitatory effect of SP on SNc DAergic cell dendrites.

In SNc, there is an additional subset (C. Fiorillo, pers.comm.) of DA neurons that reliably respond to a predictive stimulus with both phasic and sustained activations. To simulate co-existing responses in a single neuron, the model includes a third set of DA neurons affected by all three pathways (PPTN-mediated excitatory projections, and projections from both matriceal and striosomal MSPNs):

\[
\frac{1}{\tau_D} \frac{d}{dt} D_{\text{both}} = -D_{\text{both}} + (1 - D_{\text{both}}) \left[ W_{PD} [P - \Gamma_P]^+ + I_D + M_{RS} [1 - M_{RG}]^+ \right] - (D_{\text{both}} + h_D) \sum_{i,j} [G_{ij}Y_{ij} - \Gamma_{S}]^+ Z_{ij}
\] (14)

### 2.3 Dopamine Release in the Striatum

Dopamine release in the striatum in response to DA cell firing is computed as a time average of momentary dopamine cell firing rate:
\[ \frac{1}{\tau_D} \frac{d}{dt} \bar{D} = D - \bar{D} \]  

(15)

where \( D = D_{\text{phasic}} + D_{\text{tonic}} \) is the overall firing rate of DA cells as a population.

Positive reinforcement signal \( N^+ \) derives from above-baseline fluctuations of the DA signal, whereas complementary negative reinforcement signal \( N^- \) derives from below-baseline fluctuations of the DA signal:

\[ N^+ = [D - \bar{D} - \Gamma_N]^+ \]  

(16)

\[ N^- = [\bar{D} - D - \Gamma_N]^+ \]  

(17)

Because positive and negative reinforcement signals are derived from a running average of DA cell firing rate, only transient, or phasic, changes in DA signal function as a reinforcement signal. Although slow, tonic changes in the DA signal cell firing rate will not be able to induce deviations from the running average, such changes will be reflected in overall DA concentration in the striatum (due to elevated \( \bar{D} \)).

The equations governing model striatal FS-INs and TANs, whose inclusion improves but was not critical for emergence of the basic DA responses (as shown below), are provided in the Supplementary Material (equations S1-S8).

3 Simulations

The model shown in Figure 1 and specified with equations 1-17 and S1-S8 was implemented in Matlab. Numerical integration was performed with an adaptive step size fourth-order Runge-Kutta method. The model was simulated with three different reward magnitudes (\(|R^*|; 0.05, 0.15\) and 0.50; arbitrary units) and five different probabilistic schedules \((p(R^*|CS); 0.00, 0.25, 0.50, 0.75\) and 1.00). Parameter values used for simulations are given in Table 1.
**Scaled DA bursts induced by uncued unconditioned stimuli (US; primary rewards).** In the absence of any predictive stimulus, delivery of reward in the model induces a DA burst, at the onset of reward, that is proportional to the reward magnitude (Figure 2), consistent with neurophysiological observations (Schultz et al., 1997; Schultz, 1998; Tobler et al., 2005). The primary reward input generates phasic firing in the lateral hypothalamus (Nakamura and Ono, 1986), which transiently excites the PPTN (Semba and Fibiger, 1992) and ventral striatum (Schultz et al., 1992). The PPTN signal, in turn, excites SNc dopaminergic cells (Scarnati et al., 1988; Conde, 1992; Futami et al., 1995) and leads to the phasic dopamine burst (Gerfen, 1992).

**Learned DA bursts and pauses that reflect reward prediction errors.** When an initially neutral conditioned stimulus arrives to striatum via cortical afferents, it excites both ventral striatum and striosomal MSPNs through trainable adaptive weights. Ventral striatum disinhibits PPTN through ventral pallidum (Yang and Mogenson, 1987). In the model, phasic DA signals induced by primary reward input act as a teaching signal on these two sets of adaptive weights (Wickens et al., 1996). As learning proceeds, the cortical CS representation learns to excite DA cells by itself through ventral striatum, while it also learns to inhibit DA cells at the expected time of arrival of the primary reward input. The timing of the latter inhibition depends on the internal calcium dynamics of model striosomal MSPNs (Gerfen, 1992; Fiala et al., 1996; Brown et al., 1999). The larger the magnitude of primary reward, the larger the weights of both ventral striatal and striosomal MSPNs become during learning. Hence, after learning, the DA burst induced by a predictive stimulus is proportional to the reward magnitude $|R^*|$. Because the strength of striatal inhibition of model DA cells by MSPNs is matched to the excitation that would be generated by the expected reward, after learning, a larger than expected reward elicits a DA burst while a smaller than expected reward elicits a pause in DA cell firing. These bidirectional DA responses in the model reflect deviations from internal predictions (Figure 3), i.e., are reward prediction error (RPEs). This behavior was also achieved by the precursor dual-pathway model of Brown et al. (1999), and is consistent with many neurophysiological studies.
(Schultz, 1998, 2004; Schultz et al., 1997; Tobler et al., 2005). However, the learning laws used in our model deviate from those in Brown et al. (1999) in several respects. First, due to a misprint in Brown et al. (1999), the published equation governing CS-striosomal synaptic weight, $Z_{ij}$, converged to zero when negative and positive reinforcement signals ($N^+$ and $N^-$) were both equal to zero. Our equation for the CS-striosomal weight (equation 9) corrects this problem while also setting an intrinsic upper bound on weight growth. Second, the increment and decrement scalars, $\alpha_{WS}$ and $\beta_{WS}$ for the non-striosomal striatal weights (equation 4) were adjusted to compensate for the fact that the range of RPEs represented by DA dips is greatly compressed relative to the range of RPEs represented by DA bursts. This change improves the model’s weights’ sensitivities to conditional probability, treated next.

**Learning effects of probabilistic schedules of reward.** During exposure to probabilistic cue-reward contingencies, adaptive weights are incremented by DA bursts induced by delivery of primary reward, but decremented by dips consequent to each omission of the expected reward. Therefore, the net potentiation of the model’s striatal synaptic weights is a function not only of the reward magnitude, as noted above, but also of the conditional probability of reward given the CS, $p(R^*|CS)$. Given this dual dependence of the striatal synaptic strengths (Figure 4A), model neuronal responses to a CS (Figure 4B; see also Figure 7) increase with that CS’s expected reward value, $\hat{R} = |R^*|p(R^*|CS)$. This dual dependence of CS-induced striatal activations is consistent with experimental observations (Tobler et al., 2005).

Weights of cortico-striatal synapses onto matriceal MSPNs, striosomal MSPNs, and ventral striatal cells, all express the same dual dependence. Thus, striatal inhibition of DA cells in the model is also proportional to the CS’s expected reward value, $\hat{R}$, rather than the absolute reward magnitude $|R^*|$. This dependence implies that, for any probabilistic schedule, $\hat{R} \leq |R^*|$ after learning, since $p(R^*|CS) \in [0, 1]$, with $\hat{R} = |R^*|$ only if $p(R^*|CS) = 1.00$. As a consequence, a primary reward elicits a residual DA burst in the model (see below, Figure 7) even after asymptotic learning on a probabilistic schedule, consistent with the observations of Fiorillo et al. (2003) and
Tobler et al. (2005). The latter study also observed that after a probabilistic training schedule with two potential outcomes (no reward, or a cue-specific volume of liquid reward, $|R^*|$), the phasic DA responses discriminate between the two outcomes using near-equal discharges, regardless of the absolute magnitude of the reward, $|R^*|$. A simulation of this paradigm (Figure 5) showed that the model exhibits similar behavior. Although Tobler et al. (2005) noted that “the gain of neural activity with respect to liquid volume appeared to adapt in proportion to the range or standard deviation of the predicted reward outcomes” and went on to speculate that the effect might “be achieved by subtracting the expected value from the absolute reward value and then dividing by the variance (p. 1645)”, it arises in the current model solely from striatal weight learning, and requires no such variance-dependent normalization of the sensitivity of DA neurons to differences.

The adaptively weighted CS inputs to the model matrical MSPNs, and thus these neurons’ CS-induced activities, become proportional to the expected reward during learning. Therefore, in the model as in the data, striatal MSPN activity is modulated by the expected amount of reward (Kawagoe et al., 1998; Watanabe et al., 2003; Samejima et al., 2005). Model simulations (see Supplemental Materials) showed that the depth of this modulation is enhanced because matrical MSPNs also receive inputs related to the reward contingency indirectly, through the DA neuron to TAN to FS-IN to MSPN pathway, in accord with neurophysiological studies (e.g., Samejima et al., 2005).

*Emergence of uncertainty responses on probabilistic schedules of reward.* The distal dendrites of SNc dopaminergic cells are densely intermingled with SNr GABAergic cells (Conde, 1992; Gerfen and Wilson, 1996). Thus, although matrical and striosomal MSPNs project to SNr and SNc, respectively (Flaherty and Graybiel, 1994), matrical MSPNs also inhibit SNc dopaminergic cells. Increasing activity of the MSPNs in the model results in increasing inhibition of the the model SNc. During the time between CS onset and expected time of reward, the matrical MSPN projection of GABA and SP to the SNc is active, and model transmitter release is proportional to MSPN activation, in accord with Khan et al. (1996) and Beaulouan et al. (2004). Since the
model’s SP-induced excitation of SNc is progressively gated off by the co-released GABA, the net effect of this multiplicative gating interaction is an inverted U-shaped excitation, that is lowest for high, \( p(R^*|CS) = 1.00 \), and low, \( p(R^*|CS) = 0.00 \), conditional probabilities, and maximum at \( p(R^*|CS) = 0.5 \). Thus, as shown in Figures 6 and 7, the model is able to explain the genesis of DA uncertainty responses and their non-monotonic dependence on \( p(R^*|CS) \), as reported in Fiorillo et al. (2003).

The latter report did not provide a full parametric characterization of this dependency: probability was varied over its full range, but reward magnitude was not so varied. Hence, extant data do not specify the full relation between reward magnitude (\( |R^*| \)) and the level of sustained DA cell activity. However, the current model provides a novel prediction regarding this dependency. In the model, the probability at which the maximum of the nonmonotonic function occurs depends on the activation threshold of matriceal MSPNs for GABA and substance-P release in SNc (equation 12; \( \Gamma_M \)). Our model predicts that if the relationship between \( p(R^*|CS) \) and sustained DA response is qualitatively invariant across a full range of reward magnitudes, then the GABA and substance-P release threshold should be a function of reward magnitude. For a fixed threshold, our model predicts that the sustained activation is a nonmonotonic function of expected reward, \( \hat{R} = p(R^*|CS)|R^*| \), rather than of the probability alone. It is maximum for moderate values of \( \hat{R} \), and declines for smaller and larger values. See Section S2 of the Supplementary Material.

**Co-existence of sustained and phasic responses in single dopamine neurons.** In the model, DA neurons responding with sustained or phasic activations constitute two of three subpopulations (equations 10 and 13). In a third subpopulation, as shown in Figure 7D, both sustained and phasic responses occur in single cells, and these components exhibit the same functional dependencies noted above. This is in accord with a report that approximately 50% of dopaminergic cells respond to a CS with a phasic activation, 9% with a sustained activation, and 18% with both (C. Fiorillo, pers. comm.).
4 Discussion

A large number of reports have accumulated regarding the sensitivity of striatal (Samejima et al., 2005; Kawagoe et al., 1998; Watanabe et al., 2003) and DA neurons to a number of properties of cue-reward contingencies that are important both for learning to predict outcomes and for intelligent decision making. The present research sought to unify these observations by showing how the sustained uncertainty responses that emerge in DA cells when animals are exposed to probabilistic cue-reward contingencies could be explained by extending a local circuit model Brown et al. (1999) previously used to explain phasic DA responses that emerge during exposure to cue-reward contingencies. To explain the uncertainty responses (Fiorillo et al., 2003) that emerge on probabilistic schedules, the new model incorporates several highly conspicuous and interwoven features of the striato-nigral architecture that were omitted from prior models. In particular, the simulations show that uncertainty responses can be computed efficiently via the co-release of GABA and SP by MSPN fibers that synapse not just on SNr cells but also on the comingled dendrites of DA cells whose somata are in SNC. The simulations further show that this mechanism works even if the adaptive synapses on cue-excited inputs to the matriceal MSPNs giving rise to the GABA/SP fibers obey exactly the same learning law that enables the model to learn to generate phasic DA signals that represent RPEs (Schultz et al., 1997; Schultz, 1998). Finally, the uncertainty responses in the current model are robust on single trials: they do not arise only by averaging across trials. This is important because it means that the uncertainty computation could be useful for single-trial decision making. It also differentiates the current model from TD models. Notably, proponents of TD models (Niv et al., 2005) recently showed how such models could generate uncertainty responses, but only as averaging artifacts. Consistent with the present model, but not TD models, Tobler et al. (2005) recently reported that uncertainty responses are robustly present in DA neuron discharges on single trials. Thus it is important not to confuse the class of RPE models with the class of TD models. Unlike TD models, the present model accommodates both single-trial RPEs and single-trial uncertainty responses.
When Tobler et al. (2005) reported the effect shown in Figure 5, namely that “the gain of neural activity with respect to liquid volume appeared to adapt in proportion to the range or standard deviation of the predicted reward outcomes” they speculated that the effect might “be achieved by subtracting the expected value from the absolute reward value and then dividing by the variance (p. 1645)”. However, they also acknowledged that the apparent adaptation might not arise in DA neurons at all, and might instead arise at upstream sites. From the computational perspective, the choice between these two classes of mechanisms makes an enormous difference. In particular, if the difference between expected and actual inputs were divided by the variance by DA neurons, then those DA neurons could no longer generate RPEs that could guide the learning of synaptic efficacies that reflect absolute reward magnitude. This would be contrary to many published observations. Consistent with those observations, the apparent adaptation arises in the current model solely from “upstream” striatal weight learning, and requires no variance-dependent normalization of the sensitivity of DA neurons to differences. In a brief commentary (Tan, Anderson, Dranias & Bullock, Science, in review), we recently showed that this is true for a generic class of dual-path learning models that assume that DA neurons compute the (un-normalized, absolute) difference between an unlearned excitation scaled by primary-reward magnitude and a learned, cue-specific, delayed inhibition, whose size is adjusted by the history of RPEs engendered by exposure to the cue-reward contingency. The only other mathematical constraint needed to obtain this result was that negative RPEs (DA dips) have a significantly smaller dynamic range than positive RPEs (DA bursts). This mathematical constraint is very well supported by data, and incorporated in the current model.

From the neurobiological perspective, the most novel feature of the new model may be the role proposed for GABA/SP co-release in genesis of sustained DA responses. If correct, this would be of broad interest because it links two conspicuous BG features that play key roles in neurodegenerative disease: SP-MSPN loss is central in Huntington’s disease, whereas nigral DA cell loss is pivotal in Parkinson’s disease. Thus, it will be important to fully test the functional
interaction proposed in the model. Although the best established target of matriceal GABA/SP
MSPNs is GABAergic neurons in SNr (Flaherty and Graybiel, 1994), the distal dendrites of SNC
cells are known to intermingle with SNr GABAergic neurons in the ventral SN (Conde, 1992;
Gerfen and Wilson, 1996). Hence, it is not unreasonable to propose that matriceal MSPNs also
inhibit at least a subset of DA neurons in SNC. Moreover, it is the SP in fibers sent from “direct BG
pathway” MSPNs (Kaneko et al., 2000; Beaujouan et al., 2004) to ventral SN that make it the most
SP-rich part of the brain (Otsuka and Yoshioka, 1993), which undergoes dramatic SP depletion in
Huntington’s disease (Buck et al., 1981; Cicchetti et al., 2000). Data show that activation of direct
pathway does cause MSPNs co-release of SP and GABA at the nigral terminals (Khan et al., 1996;
Beaujouan et al., 2004), and several other studies (Hanson et al., 2002; Martorana et al., 2003;
Betarbet and Greenamyre, 2004) suggested that increased release of SP may cause increased DA
cell firing, as also argued by Conde (1992) and reviewed in Otsuka and Yoshioka (1993). To this
empirical foundation, the model adds the hypothesis that SP-mediated excitation is multiplicatively
gated by the GABAergic inhibition. The model’s resultant inverted-U interaction at the nigral cells
complies with neuropharmacological data based on direct intranigral SP application (Reid et al.,
1990) and with the discovery by Fiorillo et al. (2003) that cue-induced sustained DA responses
scale non-monotonically with the conditional probability of reward, given the cue.

Model simulations revealed the need to further probe neural parameters affecting the peak and
the amplitude of the non-monotonic function relating uncertainty responses to conditional proba-
bility. Although the observed function was demonstrable for a full range of conditional probabili-
ties when cues predict moderate-sized rewards, the model predicts that the peak of the inverted-U
could shift away from \( p = .5 \) with very large or very small rewards, unless the threshold for acti-
vation of co-release is a function of expected reward magnitude. Simulations also revealed that
both sustained and phasic responses of model DA neurons survive without the TANs (presumably
cholinergic interneurons; Aosaki et al., 1994b, 1995) and the FS-INs in the model’s striatal mi-
crocircuit (see Supplementary Material). However, the sensitivity of sustained responses, i.e., the
amplitude of the inverted-U, is enhanced considerably by these constituents of the striatal micro-circuit. Acetylcholine (ACh) release by model striatal TANs modulates the matriceal MSPNs via a dual control of FS-INs by ACh (Koos and Tepper, 2002; DeRover et al., 2002), and via a direct inhibitory influence of ACh on SP-MSPNs (Bernardi et al., 1993; DiChiara et al., 1994). In a prior report (Tan and Bullock, 2007), we showed with simulations that a large set of data are consistent with the hypothesis that the depth of TANs’ pause responses is tightly coupled to, and scales with the magnitude of, the phasic DA signal broadcast to striatum. Notably, information about the cue-reward contingency shapes TAN pause responses via the DA cell to TAN pathway, with no need for additional adaptive weights.

The sustained uncertainty response may be an important component of the total DA signal to target structures including the frontal cortex, amygdala and BG. Its local effects may differ dramatically across these areas due to different complements of DA receptors, different local circuits and DA uptake rates, etc. Predictions vary considerably across current conceptualizations of the role of DA in forebrain function. One proposal consistent with many prior interpretations is that sustained DA release during an interval between CS onset and possible reward delivery might act in frontal cortex to facilitate vigilant attention and working memory storage during the delay, whereas such DA release might act in striatum to facilitate “last second” switches among alternative plans. Both effects would be adaptive because a stimulus that is an uncertain predictor is often the leading edge of an information stream, the later arriving components of which often resolve the uncertainty. Such uncertainty-reducing information will be better detected and learned by a vigilant perceiver who effectively remembers recent events, and better utilized by an actor ready to switch among the several plans suggested by the initial information as further information favors one plan over others.

The model proposed in this study shows how different aspects of CS-Reward contingency, particularly the expected reward value, its timing and magnitude, and its uncertainty, could be learned and reflected in components of the DA signal of SNc. The circuit is efficient, in that it
requires many fewer fibers than would be needed in the absence of SP-GABA co-release. The model goes well beyond prior RPE models in a way that is consistent with known striatal and BG anatomy, as with electrophysiological and neurophysiological features of striatal cell types. The model also makes several novel predictions that can be tested in parametric experiments.

**Grants**

This work was supported by National Science Foundation Grant SBE-354378. C.O.T. was partly supported by Higher Education Council of Turkey and Canakkale Onsekiz Mart University of Turkey.

**References**


Figure Captions

**Figure 1:** Summary diagram of the interactions represented in the model. Arrowheads denote excitatory pathways, circles denote inhibitory pathways, and hemidisks denote synapses at which learning occurs. Thin lines at lower left show the anatomical connections proposed by Brown et al. (1999). These control the phasic dopaminergic responses via one fixed and two adaptive pathways: an excitatory pathway via hypothalamus and the pedunculopontine nucleus (PPTN) that relays primary reward information to dopaminergic cells; a pathway from CS representations via ventral striatum, ventral pallidum and PPTN, that can learn to excite DA cells at the onset of a reward-predicting CS; and an adaptive striosomal MSPN pathway by which a CS can inhibit dopaminergic neurons after a learned time delay. Thick lines show the pathways proposed in the current model. Striatal GABAergic interneurons (both NOS- and FS-INs), TANs, and matriceal SP-D1-MSPNs (M) receive inputs from the CS representation, the latter via synaptic weights that adapt in the same way as at synapses onto ventral striatal cells ($W_{is}$). Both TANs and GABAergic NOS interneurons are recipients of thalamic CM-Pf projections. NOS interneurons inhibit TANs while ACh released by TANs excites matriceal SP-D1-MSPNs and GABAergic interneurons (both NOS- and FS-interneurons). The latter effect is mediated by nicotinic receptors. ACh also inhibits GABAergic transmission between FS-interneurons and matriceal SP-D1-MSPNs presynaptically, via muscarinic receptors. Matriceal SP-D1-MSPNs, in turn, send inhibitory projections to both SNr cells ($S_R$) and SNC dopaminergic cells, whose dendrites reach into SNr. SP-D1-MSPN terminals in the SN co-release GABA and SP, which act simultaneously on distal dendrites of dopaminergic neurons. Dashed arrows indicate sites where axon terminals of DA cells release DA, which acts as a reinforcement signal at corticostriatal synapses onto ventral striatal, matriceal and striosome MSPNs. DA also modulates ACh release from TANs.

**Figure 2:** Neural discrimination of reward magnitude in response to the delivery of an unpredicted reward by model dopaminergic neurons of the phasic type. Bursts in response to reward are a function of the magnitude of reward in the absence of any predictive stimulus. The responses are
normalized to the response after delivery of an unpredicted reward of magnitude 0.5. Inset shows the corresponding neurophysiological data (reprinted with permission from Tobler et al., 2005).

**Figure 3:** Bidirectional responses of model phasic dopamine neurons reflect deviations from the outcome expected after training with reward magnitudes $|R^*| = 0.05$ (▲) and $|R^*| = 0.15$ (●) with a deterministic schedule $p(R^*|CS) = 1.0$ (black lines), and after training with probabilistic schedules $p(R^*|CS) = 0.25$ (□), $p(R^*|CS) = 0.50$ (○) and $p(R^*|CS) = 0.75$ (X) with a reward magnitude $|R^*| = 0.15$ (gray lines). Responses are normalized to the response after the unpredicted delivery of a reward of magnitude 0.15. Inset shows the neurophysiological data (reprinted with permission from Tobler et al., 2005) after training the monkey with 0.15 ml reward with $p(R^*|CS) = 1.0$, equivalent to the model response shown with black circles ●.

**Figure 4:** Model ventral striatal weights ($W_{IS}$) (A) and phasic dopamine cell responses (above baseline) to conditioned stimuli (B) after learning. Activations are normalized by the activation elicited by the conditional stimulus predicting the largest expected value. Responses to conditioned stimuli elicit cell activation as a function of expected reward value, $p(R^*|CS)|R^*|$. Model neurons’ activations discriminate among expected reward values. Black, gray and white squares, respectively, show the outcomes of training with reward magnitudes, $|R^*|$, of 0.05, 0.15 and 0.50. Inset in (B) shows the corresponding neurophysiological data (reprinted with permission from Tobler et al., 2005) where the two curves (squares and circles) represent data from two different monkeys.

**Figure 5:** Adaptation of the gain of phasic DA cell activity in response to primary reward input in the model. Lines connect the phasic activation above baseline in response to the larger of the two possible outcomes, normalized by the phasic response to an unpredicted reward of magnitude 0.15. Slopes of the lines provide a measure of gain, or sensitivity, of neural responses. Insets show the corresponding neurophysiological data from two animals (reprinted with permission from Tobler et al., 2005).
Figure 6: Sustained activity (% above baseline) of model tonic dopamine neurons as a function of reward probability, after training with reward magnitude $|R^*|=0.15$. Inset shows the corresponding neurophysiological data (reprinted with permission from Fiorillo et al., 2003).

Figure 7: Dopaminergic cell responses under different probabilistic schedules (from top to bottom; $p = 0, 0.25, 0.5, 0.75, 1$), with a reward magnitude, if any, of 0.15 ml. (A) Data reprinted with permission from Fiorillo et al. (2003). (B) Model phasic dopamine cell responses and (C) Model tonic dopamine cell responses after 24 trials. (D) Model dopamine cells both with a burst and a sustained activation. Learning in the model reaches its asymptote (i.e., residual phasic activation of DA cells in response to reward delivery diminishes under $p = 1$ schedule) after 94 trials.
Table 1: Parameter values used to simulate the model. Parameter values for the equations given in the Supplementary Material (striatal circuit) are also provided in the table for completeness. $\tau$: time constants, $\Gamma$: thresholds, $\alpha, \beta, \text{ and } \gamma$: gains, $A - D$: firing rate/activation (upper or lower) bounds, $W$: synaptic weights.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_P$</td>
<td>200</td>
<td>$\Gamma_S$</td>
<td>0.20</td>
<td>$A_S$</td>
<td>0.70</td>
</tr>
<tr>
<td>$\tau_{UP}$</td>
<td>4.00</td>
<td>$\Gamma_-$</td>
<td>0.50</td>
<td>$A_Z$</td>
<td>5.00</td>
</tr>
<tr>
<td>$\tau_S$</td>
<td>30.0</td>
<td>$\Gamma_+$</td>
<td>0.50</td>
<td>$A_M$</td>
<td>0.70</td>
</tr>
<tr>
<td>$\tau_{WS}$</td>
<td>20.0</td>
<td>$\Gamma_K$</td>
<td>0.41</td>
<td>$A_T$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\tau_D$</td>
<td>15.0</td>
<td>$\Gamma^{\text{IN}}$</td>
<td>0.73</td>
<td>$B_M$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\tau_M$</td>
<td>600</td>
<td>$\alpha_{WS}$</td>
<td>0.20</td>
<td>$B_K$</td>
<td>0.35</td>
</tr>
<tr>
<td>$\tau_D$</td>
<td>2.00</td>
<td>$\alpha_r$</td>
<td>50.0</td>
<td>$B_T$</td>
<td>0.00</td>
</tr>
<tr>
<td>$\tau_{g-}$</td>
<td>5.00</td>
<td>$\alpha_G$</td>
<td>5.00</td>
<td>$C_{\text{max}}^{\text{WS}}$</td>
<td>3.50</td>
</tr>
<tr>
<td>$\tau_{g+}$</td>
<td>5.00</td>
<td>$\alpha_Y$</td>
<td>1.00</td>
<td>$C_T$</td>
<td>1.50</td>
</tr>
<tr>
<td>$\tau_K$</td>
<td>30.0</td>
<td>$\alpha_Z$</td>
<td>100</td>
<td>$D_T$</td>
<td>0.10</td>
</tr>
<tr>
<td>$\tau_T$</td>
<td>50.0</td>
<td>$\alpha_{iS}$</td>
<td>0.46</td>
<td>$h_D$</td>
<td>0.50</td>
</tr>
<tr>
<td>$\tau_{FS}$</td>
<td>100</td>
<td>$\alpha_R$</td>
<td>10.0</td>
<td>$I_D$</td>
<td>0.15</td>
</tr>
<tr>
<td>$\Gamma_G$</td>
<td>0.37</td>
<td>$\alpha_{Th}$</td>
<td>2.00</td>
<td>$W_{UP}$</td>
<td>140</td>
</tr>
<tr>
<td>$\Gamma_Y$</td>
<td>0.18</td>
<td>$\beta_{WS}$</td>
<td>1.00</td>
<td>$W_{SP}$</td>
<td>2.00</td>
</tr>
<tr>
<td>$\Gamma_P$</td>
<td>0.095</td>
<td>$\beta_r$</td>
<td>1.00</td>
<td>$W_{RP}$</td>
<td>0.80</td>
</tr>
<tr>
<td>$\Gamma_N$</td>
<td>0.00</td>
<td>$\beta_G$</td>
<td>20.0</td>
<td>$W_{RS}$</td>
<td>12.0</td>
</tr>
<tr>
<td>$\Gamma_{D2-\text{dir}}$</td>
<td>0.10</td>
<td>$\beta_Y$</td>
<td>80.0</td>
<td>$W_E$</td>
<td>0.20</td>
</tr>
<tr>
<td>$\Gamma_{D1}$</td>
<td>0.10</td>
<td>$\beta$</td>
<td>15.0</td>
<td>$W_C$</td>
<td>1.67</td>
</tr>
<tr>
<td>$\Gamma_{D2-\text{mod}}$</td>
<td>0.09</td>
<td>$\gamma_g$</td>
<td>3.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{FS}$</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1:
Figure 2:

![Graph showing the relationship between reward magnitude and dopamine cell activation.](image)
Figure 3:
Figure 4:
Figure 5:
Figure 6: 

Change in activity (% Baseline) vs Probability 

Inset: Sustained Activation 

median change in activity (%) 

0.0 0.25 0.5 0.75 1.0 

Probability
Figure 7:
Supplementary Material

S1 Striatal Microcircuit

The striatum receives its main inputs from cortex (Gerfen and Wilson, 1996; Bennett and Wilson, 1998) and intralaminar nuclei of the thalamus (Bennett and Wilson, 1998; Thomas et al., 2000). The former input targets mainly MSPNs (Gerfen and Wilson, 1996) and the latter constitutes a main input to striatal cholinergic interneurons (Thomas et al., 2000). As mentioned in Section 2.2, we have included a striatal microcircuit, notably cholinergic and GABAergic interneurons, in the model for completeness. We have recently proposed a neural circuit model of TANs that is able to account for many response properties of these neurons in terms of their intrinsic properties and synaptic inputs (Tan and Bullock, 2007). For the current study, we embedded that model into the current one without any changes in the parameter values. The CM-Pf input to the striatum and its effects on TANs, both direct and indirect via GABAergic interneurons, are treated in detail in Tan and Bullock (2007).

Posterior intralaminar nuclei of the thalamus, namely the centromedian (CM) and parafascicular (Pf) nuclei, have multimodal response properties, and the responses of most of their neurons show sharp temporal tuning (Matsumoto et al., 2001). Neurons in the CM-Pf complex preferentially process stimuli that have attentional value, thus acting as detectors of behaviorally significant events (Aosaki et al., 1994b; Matsumoto et al., 2001) and these responses often habituate rapidly when the same stimuli are presented repeatedly (Matsumoto et al., 2001). The CM and Pf nuclei receive inputs from similar regions of the brain stem including the reticular formation, superior colliculus, and the pedunculopontine tegmental nucleus (PPTN). Matsumoto et al. (2001) reported that the sensory responses of CM-Pf neurons are largely unaffected by the reward-predictive nature of the stimuli they respond to. Notably, the "response characteristics of CM-Pf neurons sharply contrasted with those of TANs ... which under the same conditions responded preferentially to stimuli associated with reward (p. 960)."
**Striatal Cholinergic Interneurons:** In addition to dopaminergic neurons, striatal tonically active neurons (TANs), which are believed to be large cholinergic interneurons (Aosaki et al., 1994b, 1995) constitute another class of basal ganglia (BG) cells that show conditionable responses. The tonic activity of TANs appears to be a result of intrinsic properties (Bennett and Wilson, 1999). Several studies have indicated that TANs learn to pause in response to stimuli that are predictive of a forthcoming reward (Aosaki et al., 1995, 1998; Aosaki, 2003; Apicella, 2002). After learning, the response of TANs to a conditioned stimulus consists of an initial facilitation, followed by a prolonged pause, then a late rebound facilitation (Apicella et al., 1998; Apicella, 2002). The conditioned pause and rebound, but not the initial unconditioned facilitation, can be largely eliminated by GABA-ergic suppression of activity in the major source of glutamatergic inputs to TANs, namely the CM-Pf complex of the thalamus (Matsumoto et al., 2001; see also Pisani et al., 2001). Because neurons in that complex show neither conditioning effects nor pauses, the pause probably arises from local interactions within the striatum (although cortical afferents to TANs must also be considered), probably with a notable contribution from intrinsic currents (Bennett and Wilson, 1999; Wilson, 2005). Neurons in CM-Pf provide both monosynaptic excitatory and disynaptic inhibitory inputs to TANs, the latter mediated by striatal GABAergic interneurons (Reynolds and Wickens, 2004; Zackheim and Abercrombie, 2005), and the influence of both of these inputs on TANs appear to be dopamine-dependent (Aosaki et al., 1994a; Aosaki, 2003).

Let $T$ be the membrane voltage of a TAN. Then voltage dependent hyperpolarizing SK ($g_-$) and depolarizing HCN ($g_+$) currents are modeled by

$$\frac{1}{\tau_{g_-}} \frac{d}{dt} g_- = -g_- + (1 - g_-) h(T) \left[ 1 - \left[ \tilde{D} - \Gamma_{D1} \right]^+ \right]^+$$

(S1)

and

$$\frac{1}{\tau_{g_+}} \frac{d}{dt} g_+ = -g_+ + (1 - g_+) f(T) \left[ 1 + \left[ D - \Gamma_{D1} \right]^+ - \gamma_g \left[ D - \Gamma_{D2-dir} \right]^+ \right]^+$$

(S2)

where
\[
h(T) = \begin{cases} 
1 & \text{if } T \geq \Gamma_- \\
0 & \text{else}
\end{cases} \quad \text{and} \quad f(T) = \begin{cases} 
1 & \text{if } T \leq \Gamma_+ \\
0 & \text{else}
\end{cases}
\] (S3)

and \(\Gamma_-\) and \(\Gamma_+\) are the voltage thresholds required for activation of hyperpolarizing and depolarizing currents, respectively, and \(\Gamma_{D2-\text{dir}}\) is the threshold for direct post-synaptic DA action via D2 receptors. Hence, in equations S1 and S2, intrinsic currents are assumed to be directly modulated by thresholded DA action on D2 and D1 receptors, with the terms \([D - \Gamma_{D2-\text{dir}}]^+\) and \([D - \Gamma_{D1}]^+\), respectively. Because \(\Gamma_{D2-\text{dir}} < \Gamma_{D1}\), there is a phase during the increase of striatal DA level, \(D\), during which depolarizing current \(g_+\) is suppressed.

The conductance for KIR currents found in TANs is modeled by:

\[
\frac{1}{\tau_K} \frac{d}{dt} g_K = -g_K + (B_K - g_K) u(T) \left[ 1 - [\bar{D} - \Gamma_{D1}]^+ \right]^+ 
\] (S4)

where

\[
u(T) = \begin{cases} 
1 & \text{if } T \leq \Gamma_K \\
0 & \text{else}
\end{cases}
\] (S5)

and \(\Gamma_K\) is the hyperpolarization threshold required for activation of the KIR current.

In addition to intrinsic sources \((g_-, g_+\) and \(g_K\)) there are several external sources modulating the activity of TANs. These include cortical \((I_i)\) and thalamic \((I_{Th})\) inputs, where \(I_{Th} = \alpha_{Th} I_i\) is a phasic, stimulus-locked thalamic input, from CM-Pf nuclei, that reaches NOS interneurons with a 50 ms latency and lasts for the stimulus duration. Other external inputs to TANs include inhibition by NOS interneurons \((V_{IN})\), and DA \((\bar{D}; \text{equation 15})\) which acts on D2 receptors. Formulation of thalamic inputs, as well as NOS interneurons are treated in detail in Tan and Bullock (2007). NOS interneurons are assumed to fire only if their voltage exceeds a threshold \((\Gamma_{IN})\) so a piecewise-linear signal function is used to describe their output:
In addition to the aforementioned direct post-synaptic effects of DA on intrinsic currents of cholinergic interneurons via its action on D1 and D2 receptors, DA has modulatory effects on external inputs to the TANs (Flores-Hernandez et al., 2000; Nicola et al., 2000; Pisani et al., 2000), via D2 receptors. In the model, this modulation is made proportional to DA level by the multiplicative (divisive) terms \(1 \pm \beta \left[ \bar{D} - \Gamma_{D2-mod} \right]^+\), acting on thalamic \(I_{Th}\), cortical \(I_i\) and NOS interneuron \(s(V_{IN})\) inputs. The term \(\beta\) scales this modulation.

Taking all these intrinsic and extrinsic factors into consideration, the activity of TANs is modeled as

\[
\frac{1}{\tau_T} \frac{d}{dt} T = (A_T - T)g_+ - (T + B_T)(g_- + g_K) + (C_T - T) \frac{W_C I_i + W_{Th} I_{Th}}{1 - \beta \left[ \bar{D} - \Gamma_{D2-mod} \right]^+} - (T + D_T) \left[ s(V_{IN}) \left[ 1 + \beta \left[ \bar{D} - \Gamma_{D2-mod} \right]^+ \right] \right]
\]  

(Owing to the coupling to striatal dopamine release, the magnitude of the pause response of TANs increases with increasing dopamine release in the striatum, and thus reflects actual or expected rewards. Indeed, Ravel et al. (2003) showed that the motivational evaluation of the stimulus can shape TAN responses, which differ to an identical stimulus presented before and after appetitive learning. They also showed that TANs adapt their responsiveness to the specific motivational value of stimuli independently of the behavioral reaction itself. Due to the coupling to dopamine release, the amplitude of a CS-induced pause response of model TANs increases with learning, and decreases with extinction (Aosaki et al., 1994b, 1995; Apicella et al., 1998; Apicella, 2002). Notably, the amplitude of TAN pause reflects the relative proportion of rewarded trials, that is the conditional probability of the reward given the CS, \(p(R^*|CS)\). This is a direct consequence of the scaling of phasic dopamine burst magnitudes with reward expectation.
**Fast-Spiking Interneurons:** Striatal GABAergic fast-spiking interneurons (FS-INs) receive monosynaptic inputs from TANs. The acetylcholine (ACh) released by the TANs acts on FS-INs (Koos and Tepper, 2002; Tepper and Bolam, 2004) through a nicotinic acetylcholine receptor (nAChR)-mediated excitation. The nicotinic excitation appears to be a direct postsynaptic effect, independent of glutamatergic afferents (Koos and Tepper, 2002). In addition, ACh presynaptically inhibits GABAergic transmission between FS interneurons and medium spiny projection neurons (MSPN). This is mediated by muscarinic acetylcholine receptors (mAChR) located on FS-IN axon terminals (Koos and Tepper, 2002). A similar mechanism has been shown to operate in the ventral striatum, i.e., the nucleus accumbens (DeRover et al., 2002). The balance between these opposing effect will depend on the magnitude and timing of increments and decrements relative to tonic ACh release. Neostriatal ACh release is, in turn, under a complex regulation involving dopaminergic, GABAergic, glutamatergic and other inputs that induce both rapid and slow changes in ACh levels (see Tan and Bullock, 2007). Model FS-interneurons are governed by equation

\[ \frac{1}{\tau_{FS}} \frac{dF}{dt} = -F + (1 - F)(I_{Th} + W_CI_i + T + [\bar{D} - \Gamma^F_{D}])^+ \]  

(S8)

where \( I_{Th} = \alpha_{Th}I_i \), as defined above, and \( I_i \) is a glutamatergic cortical input, gated by a synaptic weight of \( W_C \). The last two terms in equation S8, \( T \) and \([\bar{D} - \Gamma^F_{D}]\), represent the excitation of FS interneurons by ACh via nicotinic receptors (Koos and Tepper, 2002) and by (thresholded) DA via D1/D5 receptors (Centonze et al., 2003).

As the quantity \(|R^*| \times p(R^*|CS)\) increases, so does the DA burst to CS, as so does the depth of the TAN pause response. Thus, post-CS ACh release in the striatum is inversely related to the expected value signaled by the CS. As a result, the feedforward inhibition exerted on the matriceal MSPNs by FS-INs decreases, leading to a net disinhibition of MSPNs. Therefore, a CS-representing cortical input of a particular amplitude will induce higher activations of MSPNs. Thus, although weights of MSPN themselves reflect expected values, cholinergic modulation of MSPNs amplifies the contrast between matriceal MSPN activations in response to different reward contingencies (Figure S1).
Figure S1: Effect of striatal circuit on sustained responses of model dopamine neurons. Black line shows the percent sustained change in the dopamine activity as % above baseline as simulated by the complete model (same as Figure 6), and gray lines indicate the percent sustained change in the activity as simulated with the striatal circuits (TANs and FS interneurons) excluded. Both simulations are performed with the same reward magnitude ($|R^*| = 0.15$) and with the same parameter set. Although the qualitative behavior of the model is similar in both cases (inverted-U-shaped elevations in sustained responses as a function of conditional probability of reward), inclusion of striatal circuit enhances the contrast between the responses to different probabilistic schedules.

**S2  Effect of Reward Magnitude $|R^*|$ on Sustained Dopamine Response**

As mentioned in Section 3, although Fiorillo et al. (2003) reported an increase in the sustained dopamine response with greater reward magnitude, they did not perform a parametric study to
characterize the nature of the dependency.

In the model, gating of the excitatory substance-P effect by the inhibitory GABAergic effect on substantia nigra dopaminergic cell dendrites,

$$M_{RS}[1 - M_{RG}]^+$$  \hspace{1cm} (S9)

translates the monotonic relation between probability and matriceal MSPN activation at the striatal level into a nonmonotonic relation between probability and sustained dopamine activity. The maximum of the resulting nonmonotonic function is determined by the activation threshold of matriceal MSPNs for GABA and substance-P release in substantia nigra:

$$M_{RS} = M_{RG} = \alpha_R [M - \Gamma_M]^+$$  \hspace{1cm} (S10)

As mentioned in Section 2.2, equation S9 produces an inverted-U shaped response to reward probabilities irrespective of the reward magnitude, if the threshold of GABA and substance-P release is a function of reward magnitude, that is if $\Gamma_M(R) = \gamma |R^*|$. In order to test this, we assumed that after learning, an arbitrary inhibitory signal, proportional to the reward magnitude with $\gamma = 0.05$ arrives to terminals of matriceal MSPNs from an unspecified source at the onset of predictive stimulus. Note that physiologically, such a signal source is required to learn the predictive stimulus in relation to the absolute reward magnitude, irrespective of the conditional probability of reward. If such a signal exists, Figure S2 shows that the sustained response of dopamine cells would be invariant under any absolute reward magnitude.

If such a signal does not exist, and the threshold is constant ($\Gamma_M = 0.035$) irrespective of the reward magnitude, then the model predicts that the sustained response of dopamine neurons would no longer be a function of reward probability alone (Figure S3, left-hand panel). Rather, the sustained response will be a function of expected reward value, $\hat{R} = p(R^*|CS) \times |R^*|$. As shown in the right-hand panel of Figure S3, the response is maximum for moderate values of expected reward, and declines toward smaller and larger values.
Figure S2: Sustained activation of dopamine cells (% above baseline) is an inverted-U shaped function of reward probability, invariant under different reward magnitudes, if and only if MSPN activation threshold for GABA and substance-P release from striatonigral terminals is a function of absolute reward magnitude. Plots are generated by assuming an arbitrary signal at the time of predictive stimulus onset such that $\Gamma_M(R) = \gamma |R^*|$, $\gamma = 0.05$ in equation S10.

Thus, these results show that with a reward-magnitude-dependent MSPN activation threshold for GABA and substance-P release from the striatonigral terminals is required if the sustained response of dopaminergic cells in response to the reward probability is to be invariant under different absolute reward magnitudes.
Figure S3: Sustained activation of dopamine cells is an inverted-U shaped function of expected reward value $\hat{R} = p(R^*|CS)|R^*$ (right-hand panel) but not of conditional probability of reward $p(R^*|CS)$ (left-hand panel), in contrast to Fiorillo et al. (2003) if MSPN activation threshold for GABA and substance-P release from striatonigral terminals is constant, $\Gamma_M = c, c = 0.035$ in equation S10.