Biomimetic Circuits for Autonomously Learning to Selectively Respond to Unexpected Reward Related Events

J. W. Brown[1,2], D. Bullock[1], and S. Grossberg[1]

[1]Department of Cognitive and Neural Systems, Center for Adaptive Systems, Boston
University, 677 Beacon St., Boston, MA 02215
{jwbrown, danb, steve}@cns.bu.edu
http://www.cns.bu.edu
[2]Vanderbilt Vision Research Center, Vanderbilt University, 525 Wilson Hall, 111 21st
Avenue South, Nashville, TN 37240-0009
http://www.vanderbilt.edu/vvrc

Abstract. A truly autonomous mobile robot needs to be sensitive to rewarding and punishing events in a rapidly changing environment. It also needs to assess whether an event is novel or familiar. Putting these two competences together, it should be able to detect both the occurrence of unexpected rewarding events and the omission of expected rewarding events. To form different responses to expected vs. unexpected events - whether reward occurrence or omission - it should be able to adaptively time its expectations of rewarding events as the situation in which they arise becomes familiar. The mammalian brain has circuits, notably in the basal ganglia, that are capable of learning to selectively respond to unexpected rewarding events. A neural model of these circuits that functionally clarifies how this is done is summarized herein. The model quantitatively explains anatomical and neurophysiological data about the cells and connections that form these circuits.

1 Introduction

A variety of models have recently been described that make a quantitative link between brain and behavior. The functional understanding of the problems that the brain has solved is thus rapidly progressing to the point that fine details of brain circuits and of how these circuits control behavior are now becoming clear. These new functional ideas and their circuit realizations may be of value for designing more biomimetic autonomous robots.

In our group, for example, models have recently been described for how the laminar circuits of visual cortex (areas V1 and V2) control perceptual grouping and attention [30], [34] and, more generally, why the mammalian neocortex is organized into laminar circuits in order to generate intelligent behaviors; how the circuits of cortical areas MT and MST control visual navigation and smooth pursuit eye movements [33], [51]; how the cerebellum adaptively times motor actions [21]; how the motor cortex and parietal cortex work with the spinal cord and cerebellum to control planned arm movements [6], [12], [15]; and how multiple areas of the brain, including superior colliculus, reticular formation, cerebellum, and cortical areas cooperatively control reactive and planned saccadic (ballistic) eye movements [26], [27], [35]. The present work develops a neural network model, first described in [4], of how basal ganglia circuits learn how to selectively respond to unexpected rewarding or reward-predicting events as they occur in real time. In order to achieve this competence, such learning needs to be able to adaptively predict the expected times of familiar rewards in a given environment, so that their omission or occurrence can be used to guide behavior and update the internal model of the environment.

Both humans and humanoid robots that are capable of autonomous learning must process multiple cues in parallel when they experience unfamiliar environments. Knowing how to selectively allocate attention to those objects and events that promise to yield valued consequences is essential to successful adaptation by both humans and robots. This is true because value-modulated selective attention helps to guide rapid decision-making and learning in situations for which the robot does not already have well-learned action plans. The present work characterizes a key circuit in this control system. This circuit, which exists in the brain region called the basal ganglia, is able to generate adaptively timed responses to unexpected rewarding events, and can thereby drive selective attention, decision-making, and action to such events. In order to clarify how this circuit may be embedded in a larger control architecture, we are currently developing such an architecture [5] to simulate how the mammalian brain uses these basal ganglia signals to learn the structure of the environment and select among reactive and planned courses of action. The basal ganglia and frontal cortex, among other brain regions, interact together in this architecture to allow animals to learn adaptive responses to acquire rewards when prepotent reflexive responses are insufficient and learned plans are needed for behavioral success. Anatomical studies show a rich pattern of interactions between the basal ganglia and distinct frontal cortical layers. Analysis of the laminar circuitry of the frontal cortex, especially the frontal cortex interaction with the basal ganglia, motor thalamus, superior colliculus, inferotemporal and parietal cortices, and related structures provides new insight into how these brain regions interact to learn and perform complex behaviors. This model architecture simulates these interacting circuits and provides a functional explanation of seventeen major physiologically identified cell types found in these areas. The model predicts how action planning or priming is dissociated from execution, how a cue may serve either as a movement target or as a discriminative cue to move elsewhere, and how the basal ganglia help choose among competing actions. The circuit described herein whereby the brain can learn to respond selectively and at the appropriate times to unexpected rewards can hereby be seen to be a key element in a more complete system for controlling intelligent behavior in real time.

Both humans and animals can learn to predict both the amounts and times of expected rewards. The dopaminergic cells of the substantia nigra pars compacta (SNc) have unique firing patterns related to the predicted and actual times of reward [38], [45], [47], [54], [56], [57]. Figures 1 and 2 summarize some of their main properties. Notably the SNc cells respond immediately to unexpected cues (conditioned stimuli, or CS), show no responses to expected rewards (unconditioned stimuli, or US) if delivered at the expected time, but show an "off" response if an expected reward is omitted. Since these learned firing patterns also act as learning signals in the striatum and elsewhere [65], they have been suggested to play a key role in both addictive behavior [28] and reinforcement learning. In particular, dopaminergic reward signals seem to strengthen the "incentive salience" or "wanting" of a certain reward -- that is, the motivation to work for the reward in a given behavioral context -- as distinct from the affective enjoyment or "liking" of a reward once consumed [2]. The "liking" may be mediated by areas other than the basal ganglia [46]. Recent models [1], [16], [39], [49], [58], [61] of the nigral dopamine cells have noted similarities between dopamine cell properties and well-known learning algorithms, especially Temporal Difference (TD) models [49], [58], [61]. While providing a degree of insight into the information carried by the dopamine signal, the TD approach has not been able to answer the questions of what biological mechanisms actually compute the signal, and how. In particular, how does learning in the circuit that includes these cells enable them to produce a fast excitatory response to conditioned stimuli and a delayed, adaptively timed inhibition of response to subsequent reward-related signals, in all the experimental conditions summarized by Figures 1 and 2? We show here that the known anatomy and cell types in pathways afferent to dopamine cells lead to an explanation with significant advantages over previous models.

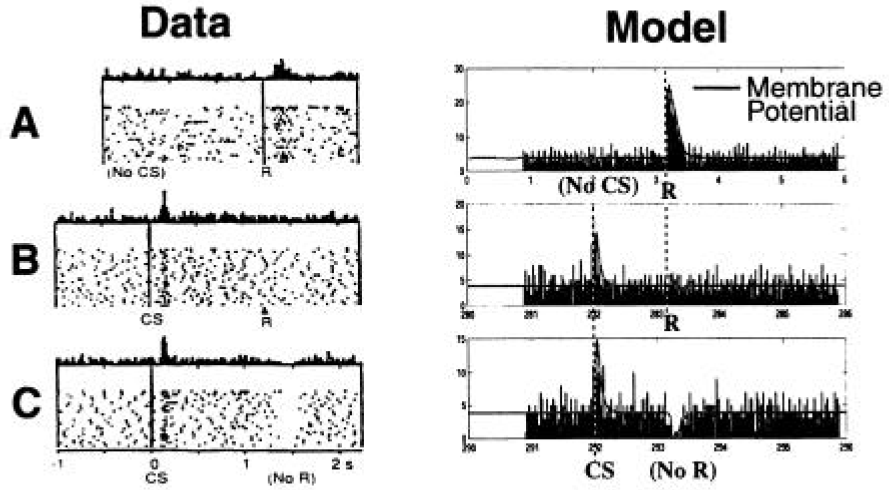

Fig. 1. Dopamine cell firing patterns: Left: Data. Right: Model simulation, showing model spikes and underlying membrane potential. A) In naive monkeys, the dopamine cells fire a phasic burst when unpredicted primary reward R occurs (e.g. if the monkey receives a burst of apple juice unexpectedly). B) As the animal learns to expect the apple juice that reliably follows a sensory cue (conditioned stimulus, CS) that precedes it by a fixed time interval, then the phasic dopamine burst disappears at the expected time of reward, and a new burst appears at the time of the reward-predicting CS. C) After learning, if the animal fails to receive reward at the expected time, a phasic depression in dopamine cell firing occurs. These responses appear to reflect an adaptively-timed expectation of reward that cancels the naïve reward response. [The data in Figure 1 (column 1) are reprinted with permission from Schultz et al. [58]].

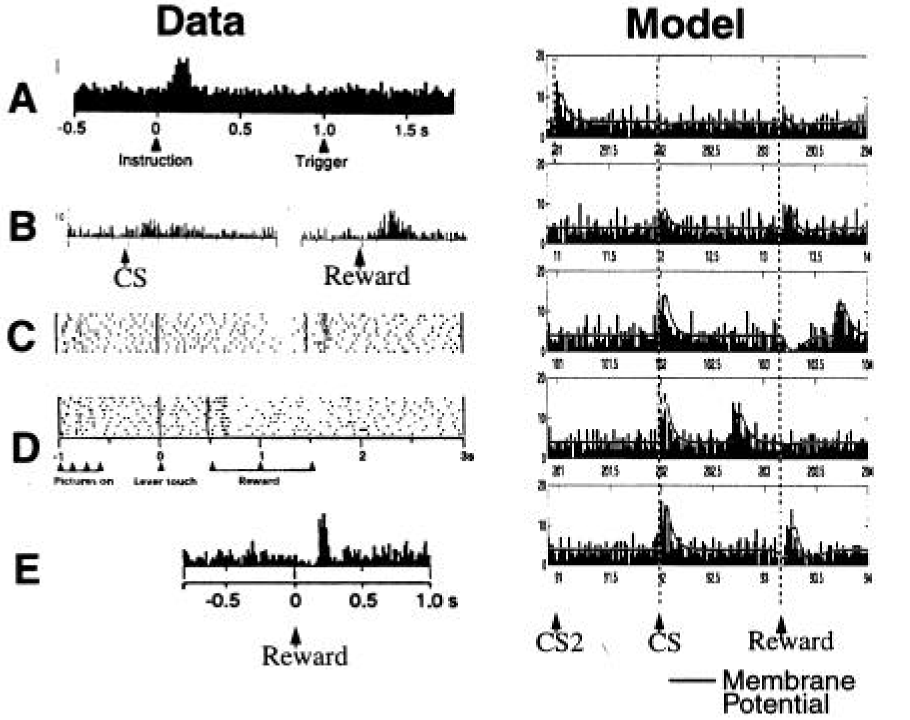

Fig. 2. Dopamine cell firing patterns: Left: Data. Right: Model simulation, showing model spikes and underlying membrane potential. A) The dopamine cells learn to fire in response to the earliest consistent predictor of reward. When CS2 (instruction) consistently precedes the original CS (trigger) by a fixed interval, the dopamine cells learn to fire only in response to CS2. [Data reprinted with permission from [56]]. B) During training, the cell fires weakly in response to both the CS and reward. [Data reprinted with permission from [45]]. C) Temporal variability in reward occurrence: When reward is received later than predicted, a depression occurs at the time of predicted reward, followed by a phasic burst at the time of actual reward. D) Likewise, if reward occurs earlier than predicted, a phasic burst occurs at the time of actual reward. No depression follows since the CS is released from working memory. [Data in C and D reprinted with permission from [38]]. E) When there is random variability in the timing of primary reward across trials (e.g., when the reward depends on an operant response to the CS), the striosomal cells produce a "Mexican hat" depression on either side of the dopamine spike. [Data reprinted with permission from [56]].

We introduce a model in which the learned excitatory and inhibitory responses are subserved by different anatomical pathways, and the adaptively timed inhibitory learning is mediated by metabotropic glutamate receptor (mGluR)-driven Ca2+ spikes in striosomal cells. These Ca2+ spikes occur with a spectrum of temporal delays. When a Ca2+ spike and a dopamine burst occur at the same time, inhibitory learning is enhanced at the corresponding delays. To explicate these excitatory and inhibitory pathways, the model functionally explains and simulates the firing patterns of dopamine cells, striosomal cells of the striatum, pedunculo-pontine tegmental nucleus (PPTN) cells, ventral striatal cells, and lateral hypothalamic cells (Figures 1-3). Its mGluR-based spectral timing mechanism helps to explain more data than the temporal derivative operation that defines the class of TD models previously used to describe dopamine cell behavior. This model is shown schematically in Figure 4. The mathematical details of the model can be found in [4].

2. The Biological Basis of Reinforcement

Dopamine cell responses can be conditioned to phasic cues whose offsets occur long before the reward signals that they predict (e.g., [45]). To bridge the temporal gap, a CS is assumed to activate a sustained working memory input to the model [23]. A subsequent primary reward signal from an unconditioned stimulus, or US, is assumed to trigger a dopamine burst, which augments the weights between the working memory site and the ventral striatum [66]. This allows future CS presentations to elicit an immediate excitatory prediction of reward. The CS also activates a population of lagged inhibitory signals from the striosomes to the SNc. When a dopamine burst occurs at a sufficient lag after CS onset, it strengthens the subset of lagged inhibitory signals that are active at that time. These two types of learning enable a CS to generate an immediate, reward-predictive dopamine signal but also to cancel subsequent SNc excitation that would otherwise be caused by the predicted reward-related signals. When a response is made and reward is received, the working memory input is assumed to shut off [23].

We propose that the PPTN is responsible for the phasic bursts of activity in SNc dopamine cells (see Figures 1 and 2), and thus plays a key role in the learning and maintenance of instrumental tasks. Experiments showing monosynaptic glutamatergic and cholinergic PPTN-to-SNc projections [13], [24], [53] support this hypothesis. Conde [53] has suggested that the PPTN provides the main source of excitation to the SNc, and PPTN cells have been found to fire phasically in response to primary reward or reward-predicting conditioned stimuli, or both, leaving them well situated to provide this kind of SNc input [19] (see Figure 3A). The phasic nature of PPTN signalling is due to habituation, or accommodation, in SNc-projecting PPTN cells [62]. Lesions of the PPTN produced hemiparkinsonian symptoms, as if the SNc itself had been lesioned [44], and reversible PPTN inactivation mimics extinction in an instrumental task, even while rewards, if provided, are readily consumed [14].

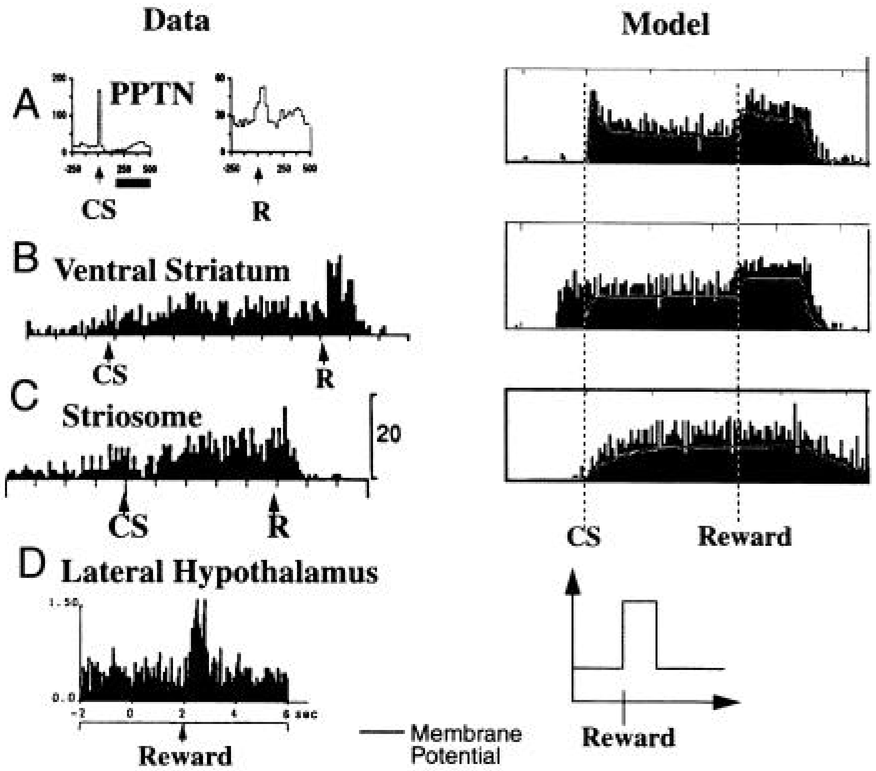

Fig. 3. Trained firing patterns in PPTN, ventral striatum, striosomes, and lateral hypothalamus. Left: Data. Right: Model simulations, showing model spikes and underlying membrane potential. A) PPTN cell (cat), showing phasic responses to both CS and primary reward. [Data reprinted with permission from [19]]. In the model, phasic signalling is due to accommodation or habituation [62], which causes the cell to fire in response to the earliest reward-predicting CS and US reward, but not to subsequent CSs prior to reward. B) Ventral striatal cells show sustained working memory-like response between trigger and a US reward, and a phasic response to the US reward. [Data reprinted with permission from [55]]. C) A ventral striatal cell, predicted here to be a striosomal cell, shows buildup to phasic primary reward response. For the model cell, j = 39. [Data reprinted with permission from [55]]. D) A lateral hypothalamic neuron with a strong, phasic response to glucose reward. [Data reprinted with permission from [50]]. The majority of these neurons fired in response to primary reward but not to a reward-predicting CS. The model lateral hypothalamic input is a rectangular pulse.

PPTN Afferents: From where does the PPTN receive these response motivating reward and reward-predicting signals? We propose that the primary reward signals come from the lateral hypothalamus, while the excitatory reward-prediction signals (which generate a CS-induced dopamine burst) travel via the ventral striatum-ventral pallidum pathway, which receives input mainly from limbic cortex [55] (see Figure 4). Lateral hypothalamic neurons are known to play a role in feeding behavior and to fire phasically in response to primary reward [50], as in Figure 3D. A strong lateral hypothalamus-PPTN projection has been found and confirmed by both anterograde and retrograde labelling [59], and the primary reward signal explains the similar phasic reward response in the PPTN. Thus, the lateral hypothalamus seems to be a principal source of excitation to the PPTN.

Likewise, more than one-quarter of the ventral pallidum projects collaterals to the PPTN [48]. The ventral pallidum receives projections from the matrisomes of the ventral striatum [69], which responds to both predicted and primary reward [55], as in Figure 3B. The double inhibition from ventral striatum to ventral pallidum to PPTN results in net excitation from ventral striatum to PPTN. We predict that the sustained, CS-induced striatal activation that is shown in Fig. 3B is due to receipt of a working memory trace of the CS from limbic cortex, which is enhanced by learning of CS-reward contingencies [18]. The transient component in Fig. 3B results from a phasic primary reward signal from the lateral hypothalamus [3], [50]. We suggest that the ventral striatum is a main pathway of excitatory reward predictions.

Other PPTN afferents are possible candidates for generating phasic PPTN responses. Some other possible sources, found by retrograde labelling from the PPTN, include the central nucleus of the amygdala (CNA) and the subthalamic nucleus (STN) [59]. The amygdala does not appear to provide the main source of excitation, despite its processing of emotional valence information. In particular, it has been shown that rats with amygdala lesions could still learn operant tasks [46]. After CNA damage, rats can learn second order conditioning even though they fail to learn a conditioned orienting response [25]. Similarly, some studies suggest a modulatory rather than an excitatory role of the STN-to-SNc projection [60], and cell recording studies have not yet shown reward-predicting activity in the STN.

Striosomes: What suppresses the dopamine burst response to primary reward after conditioning has occurred, and what causes the transient activity drop when expected reward is not received (see Figure 1)? The striosomal cells provide significant GABA-ergic inhibition to the SNc [29], which could account for both of these phenomena. In turn, striosomal cells receive dopaminergic projections from the SNc [29]. We propose that an intracellular spectral timing mechanism [21], [31], [36] provides the function needed. Specifically, the striosomal cells briefly inhibit SNc dopamine cells, after a learned delay period, to provide an inhibitory expectation of reward. The model incorporates striosomal cells in both the dorsal and ventral aspects of the striatum. Likewise, model dopamine cells correspond to both dorsal and ventral SNc cells which, despite certain differences, have similar inputs and response properties. Gerfen [29] has noted the distinction between the dorsal and ventral tiers of the SNc: dorsal tier SNc cells project to the matrisomes of the striatum (including the model ventral striatal cells), while ventral tier SNc cells project to the striosomes. The model lumps together the ventral and dorsal tiers of the SNc on the basis of their similarities.

It has been suggested that striosomal cells provide adaptively-timed inhibition to the dopamine cells [16], much as cerebellar Purkinje cells provide adaptively timed inhibition of interpositus nucleus cells [21], but this general hypothesis must be coupled to a biologically-supported local mechanism. Given evidence that striatal learning is suppressed by mGluR blockers [8] and Ca2+-chelators [10], we suggest the following striosomal cell model: Conditioned stimuli excite a glutamatergic corticostriatal pathway that activates metabotropic glutamate receptors (mGluR) on striosomal neurons. These in turn cause a delayed transient rise in intracellular Ca2+, at least partly via NMDA channels [9], which are known to be potentiated by mGluR1 receptor activation [52]. This Ca2+ response is proposed to be a basis for both learning and generating an adaptively-timed inhibitory striosomal-SNc signal. The model uses a population of striosomal cells with a range of delayed responses (see Figure 5) which, taken together, constitute the "spectrum" of possible learned delays.

Fiala et al. [21] have proposed a model of adaptively timed conditioning in which cerebellar Purkinje cells generate a spectrum of differently delayed Ca2+ spikes after excitation of mGluR1 receptors. A Ca2+ spike by itself activates a Ca2+-dependent K+ conductance, which is hyperpolarizing. In addition, when a climbing fiber signal is received at the same time as a delayed Ca2+ spike, it causes a long-term increase in the Ca2+-dependent K+ channel conductance (LTD). Thus, in the cerebellar model, the Ca2+ spike is a basis for both immediate hyperpolarization and learned LTD.

We propose that a related but distinct mechanism operates in striosomal cells which, unlike Purkinje cells [17], possess NMDA receptors. In this context, a mGluR1-mediated delayed Ca2+ spike can be amplified and thus serve to transiently increase rather than decrease striosomal cell activity. A class of recently-discovered Ca-inhibited K+ channels [41] may also contribute to a Ca-dependent depolarization. A Ca2+ spike combined with a phasic burst of dopamine acting on striosomal D1 receptors would also allow LTP in striosomal cells. It has been suggested that increased Ca2+ combined with a dopamine burst could result in a potentiation of glutamate receptors (LTP) [39], and dopamine bursts have been shown to reverse corticostriatal LTD and instead cause LTP [66]. Thus, a delayed Ca2+ spike in the striosomal cells could serve as both a signalling gate and as one component of a learning gate.

Recent work on the cerebellum [22], [63] has supported the Fiala et al. [21] cerebellar model, and demonstrated the feasibility of direct calcium imaging in local regions of a dendritic arbor using high-speed confocal microscopy. We suggest that the same technique could be used in neostriatal cells to investigate the predictions regarding striosomal Ca dynamics. Pharmacological inactivation of mGluR1 and second messenger IP3 would test whether they are essential components of the Ca spike cascade, as in the cerebellum.

Functionally, the striosomal cells of the model need to receive a sustained input that is activated when a CS first occurs, as a reference point for the delayed inhibitory signal. Striosomal cells receive excitatory signals from deep layer V of limbic cortex [29]. The sustained working memory signal initiates a steady rise of the intracellular calcium level, e.g., via an mGluR1-IP3-Ca cascade (as in the cerebellum, see [22], [63]), which causes a calcium spike upon reaching a threshold. The sustained input hereby leads to a delayed, phasic response within the striosomal cell. A related property of the model is that, if the sustained input strength is proportional to the CS intensity, then a weaker CS causes an increase in the rise time to threshold, resulting in a slower perceived rate of time passage. This property agrees with behavioral data [67], although due to the complexity of cortical processing, the striosomal inputs may not be directly proportional to external stimulus intensity. The model simulations assume a simple two-state working memory input that is either on or off, and which could be generated by passing a gradually rising input through a sharp sigmoidal signal function. The maximum delay that a single spectrum can adaptively time is still unknown, and needs to be investigated biochemically; cf. [21]. Spectral timing of a single event also needs to be supplemented by inter-event timing mechanisms that involve network interactions, including prefrontal cortex and cerebellum (e.g., [7], [32]).

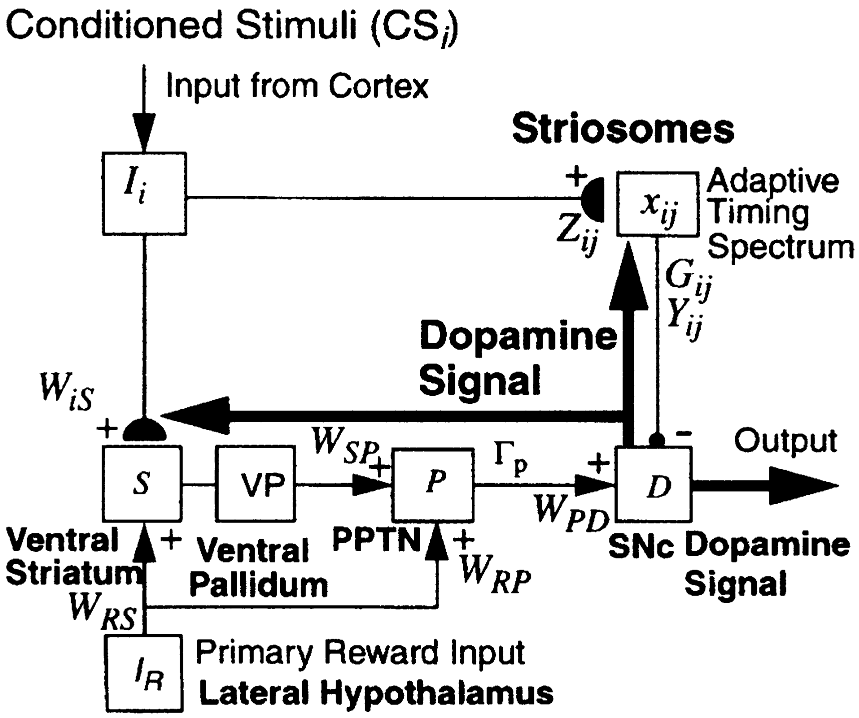

Fig. 4. Model circuit. Cortical inputs (Ii) excited by conditioned stimuli learn to excite the SNc (D) via the ventral striatal (S)-to-ventral pallidal-to-PPTN (P)-to-SNc path. The inputs Ii excite the ventral striatum via adaptive weights W_{iS}, and the ventral striatum excites the PPTN, via double inhibition through the ventral pallidum, with strength W_{SP}. When the PPTN activity exceeds a threshold GP, it excites the dopamine cell with strength W_{PD}. The striosomes, which contain an adaptive spectral timing mechanism (x_{ij}, G_{ij}, Y_{ij}, Z_{ij}), learn to generate lagged, adaptively-timed signals that inhibit reward-related activation of SNc. Primary reward signals (I_R) from the lateral hypothalamus both excite the PPTN directly (with strength W_{RP}) and act as training signals to the ventral striatum S (with strength W_{RS}). Arrowheads denote excitatory pathways, circles denote inhibitory pathways, and hemidisks denote synapses at which learning occurs. Thick pathways denote dopaminergic signals.

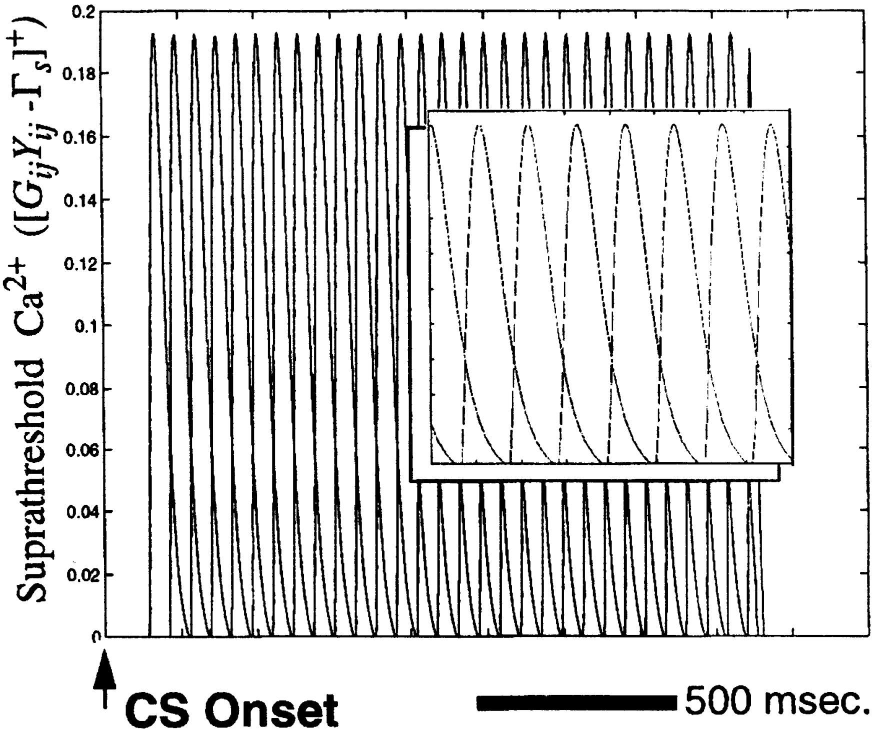

Fig. 5. Striosomal spectral timing model and close-up (inset), showing individual timing pulses. Each curve represents the suprathreshold intracellular Ca2+ concentration $[G_{ij}Y_{ij}$ - $G_s]$+ of one striosomal cell. The peaks are spread out in time so that reward can be predicted at various times after CS onset, by strengthening the inhibitory effect of the striosomal cell with the appropriate delay. The model uses 40 peaks, spanning approximately 2 seconds and beginning 100 msec. after the CSs [36]. Model properties are robust when different numbers of peaks are used. It is important that the peaks be sufficiently narrow and tightly-spaced to permit fine temporal resolution in the reward-cancelling signal. However, a trade-off ensues in that more timed signals must be used as the time between peaks is reduced. The timed signals must not begin too early after the CS, or they will erroneously cancel the CS-induced dopamine burst. The 100 msec. post-CS onset delay prevents this from happening.

3 Results

Given the above background, the model mechanisms can now be summarized as follows (see Figure 4):

First, a primary reward signal is generated in the lateral hypothalamus [50] (Figure 3D). This directly excites the PPTN [59], which fires a brief burst and then accommodates, or habituates [19], [62]. This brief burst directly excites the SNc by cholinergic and/or glutamatergic projections [13] and thereby causes a phasic dopamine burst to the striatum [29] at the time of primary reward.

Suppose that a CS is received and stored in prefrontal working memory at some time t prior to the actual reward. This CS trace generates output signals along adaptive pathways to both the ventral striatum and the striosomes. When primary reward occurs, a dopamine burst facilitates LTP in the limbic cortical-ventral striatal path [3]. Thus, the CS representation in limbic prefrontal cortex learns to excite the dopamine cells via the limbic cortical-ventral striatal-ventral pallidum-PPTN-SNc pathway [69]. In the model, the ventral striatum and ventral pallidum are lumped for simplicity into a single ventral basal ganglia node, which causes net excitation of the PPTN.

The limbic cortical projection to the striosomes [20], [29] activates a spectrum of delayed Ca2+ spikes in the striosomal cells via metabotropic glutamate receptors. When a dopamine burst arrives from the SNc, it strengthens the CS-activated limbic cortical connections to any currently spiking components of the striosomal timing spectrum. The striosomal cells hereby learn to inhibit the dopamine burst at its expected time via the inhibitory striosomal-SNc path [29].

On a later trial in the trained model, when the CS is received at the expected time before an actual reward, its working memory trace tonically activates the ventral striatal model cell, which in turn excites the PPTN, causing an immediate dopamine burst in the SNc. The adaptively-timed inhibition via the striosomal cells then inhibits the SNc so that the subsequent primary reward signal does not elicit a dopamine burst in the SNc. If the primary reward signal is absent on a trial, then the striosomal inhibition causes a phasic dip in the dopamine signal. These properties explain the dopamine cell data of Figure 1.

The model was also used to simulate a variety of other task situations for which dopamine cell responses are known. It successfully reproduced all the key SNc dopamine cell data (Figures 1 and 2) as well as firing patterns of known cell types in the PPTN (Figure 3A) and ventral striatum (Figure 3B), which are afferent to the nigral dopamine cells. In particular, dopamine cell responses were simulated in eight task situations (Figures 1 and 2). First, the model received primary reward R only and showed a strong response to the reward (Figure 1A). We then trained the model with a CS preceding R. During training, the model fired weakly in response to both the CS and R (Figure 2B). As training neared completion, the model SNc responded strongly and only to the CS (Figure 1B). In the trained model, we examined the effect of omitting R and found a transient depression at the predicted time of reward (Figure 1C). To test the effects of higher-order conditioning, we first trained the model with the CS-R association. Then we introduced an additional conditioned stimulus (CS2) which consistently occurred one second prior to the CS. With training, the model dopaminergic cells learned to respond only to CS2 (Figure 2A).

Recent work has examined dopamine cell responses under conditions of variable reward timing [38]. The model successfully simulated these data as well. When the reward R was delayed (Figure 2C), model dopamine cells responded with the characteristic depression at the expected time of R and then showed a burst later when R did occur. Similarly, if R occurred prior to the expected time, model dopamine cells again showed a burst in response to R. They did not, however, show a dip at the expected time of R (Figure 2D), in agreement with the data, since the working memory trace shut off when R was received. In some cases, the timing of primary reward may vary from trial to trial due to its dependence on an operant response. The model dopamine response was simulated when the timing of R varied randomly on an interval spanning 200 msec before and after the expected (mean) time of R, with a uniform random distribution. This caused model striosomal cells to learn to inhibit the dopamine signal during the entire interval in which the dopamine bursts occurred. Since this interval of inhibition is wider than the dopamine burst, model striosomal cells produced tails of depressed firing on either side of the dopamine burst (Figure 2E), generating a kind of temporal Mexican hat function, as in the data [56].

The PPTN model responses also agree with the cell recording data from conditioning tasks [19], which show transient bursts in response to both CS and R (Figure 3A). In addition, when a CS2 preceded the CS, the model PPTN response to the later CS disappeared. This lack of response to subsequent CSs agrees with the data on page 405 of [19], which show a similar disappearance of the CS-induced PPTN response in that delay task.

Model ventral striatal cells also simulated known cell firing patterns (Figure 3B): After the model learned the CS-R association, CS onset produced tonic activity, followed by a phasic burst in response to the R signal from the model hypothalamic cell (Figure 3D).

4 Discussion

The present model explains and predicts significantly more data than previous models through its use of parallel learning pathways. Several models have attempted to describe the dopamine cell behavior by a TD algorithm [49], [58], [61]. These models suggest that the dopaminergic SNc cells compute a temporal derivative of predicted reward. In other words, they fire in response to the sum of the time-derivative of reward prediction and the actual reward received. These models have not been linked with structures in the brain that might compute the required signals. The Suri and Schultz [61] model has simulated much of the known dopamine cell data. However, their model can only learn a single fixed ISI that corresponds to the longest-duration timed signal (xlm(t)) in their model. If the ISI is shorter than this, dopamine bursts will strengthen all of the active stimulus representations predicting reward at the time of the dopamine burst or later. Thus, their model generates inhibitory reward predictions beyond the primary reward time, and predicts a lasting depression of dopamine firing subsequent to primary reward, which is not found in the data.

In contrast to TD models that compute time derivatives immediately prior to dopamine cells, our spectral timing model uses two distinct pathways: the ventral striatum and PPTN for initial excitatory reward prediction, and the striosomal cells for timed, inhibitory reward prediction. The fast excitation and delayed inhibition are hereby computed by separate structures within the brain, rather than by a single temporal differentiator. This separation avoids the problem of the Suri and Schultz [61] model by allowing transient rather than sustained signals to cancel the primary reward signal, thereby enabling precisely-timed reward-canceling signals to be trained, and preventing spurious sustained inhibitory signals to the dopamine cells. This separation also allows the inhibitory system to follow and precisely cancel the real-time dynamics of the primary reward signal, as in Figure 1B, where the striosomal signals cancel the dopamine burst despite its asymmetry. Where temporal uncertainty exists in reward prediction, the tails of inhibition (Figure 2E) in the data are explained by the model's ability to learn temporally distributed net inhibitory signals that track the temporal dispersion of reward.

Like our model, the TD model of [58] uses transient rather than sustained timing signals. However, because this model does not separate the computation of excitation and inhibition, each transient pulse is temporally differentiated to produce an onset burst followed by an offset depression. Over the course of many trials, the onset burst strengthens its preceding timed signal weight, thereby recursively chaining backwards until all timed signal weights between the CS and R have been activated by learning. This predicts that the dopamine burst gradually travels backward in time, and that the reward response extinguishes well before the CS response occurs. The data show instead that dopamine bursts do not occur systematically in the middle of the ISI during training, and moreover, the dopamine burst occurs concurrently at both CS and R during individual training trials [45].

The Contreras-Vidal and Schultz [16] model of the dopamine cell system is based partly on the ART2 model [11]. They first suggested that striosomes may generate a spectrum of adaptively-timed reward predictions, based on the earlier spectral timing models of Grossberg and colleagues [21], [31], [36]. Their striosomal model may face problems because it relies on lateral inhibition among striosomal cells, rather than intracellular timing mechanisms. GABA-ergic lateral inhibition among striosomal cells may be weak [40], [68] and thus not strong enough to mediate the competitive choices required by their model. In addition, their model assumes adaptively-timed inhibitory reward prediction learning at the striosomal-SNc synapses instead of at the cortico-striosomal synapses. This fails to incorporate data on corticostriatal LTP/LTD [65]. In their model, corticostriatal LTP/LTD would cause erroneous timing predictions because the cell with the strongest cortico-striatal input becomes active first and generates its adaptively-timed signal, while suppressing its competing neighbor cells via strong lateral inhibition. After this, the winning cell remains refractory, and the cell with the next-strongest cortico-striosomal weight becomes active, and so on. If learning occurs in the cortico-striosomal path, as much evidence suggests, then the rank ordering of cortico-striosomal weights may change as the synaptic weights change relative to each other. This would cause erroneous reward timing predictions, since the model striosomal cells would become active in the wrong sequential order. Our model avoids these problems by describing an intracellular mGluR-mediated adaptive timing mechanism, rather than an extracellular one.

Another significant difference between the present model and that of Contreras-Vidal and Schultz (1997) is the source of excitation to the dopamine cells. Their model assumes that matrisomal cells provide the excitatory input to SNc cells indirectly, via double inhibition through the SNr. This polysynaptic, matrisomal cell-SNr-SNc pathway cannot be ruled out as a source of net excitation to the dopamine cells, but as noted above, it is not the main pathway of SNc excitation. Although the present model attempts to represent the principal circuitry responsible for dopamine cell responses, additional afferent circuitry exists that may also be capable of eliciting phasic dopamine cell responses; e.g., the SNr-SNc projection, and the STN-PPTN and STN-SNc projections.

Houk et al. [39] model dopamine cell firing using the direct and indirect basal ganglia pathways. They assume that the polysynaptic, net excitatory indirect path through the basal ganglia is faster than the monosynaptic, direct path. The indirect path is proposed to generate the initial excitatory dopamine burst, while the direct path is proposed to mediate the slower inhibition of the dopamine cells.

With regard to the fast excitation of the dopamine cells, Houk et al. [39] cite data showing that striatal stimulation results in a fast EPSP followed by a slower IPSP in the globus pallidus [43]. However, it is unlikely that the EPSPs are polysynaptic, since they could be elicited with as little as 2 msec. latency [43]. Likewise, the fast EPSP that results from cortical excitation [42] might be better explained as from a cortical-STN-pallidal route. Moreover, STN activity may modulate rather than excite the SNc [60]. These data contradict Houk and colleagues' assumption of net striatal-SNc excitation via the model indirect pathway. The data are probably due to STN-SNr excitation and subsequent SNr-SNc inhibition [37], [64].

With regard to the slow inhibition of the dopamine cells, Houk and colleagues [39] propose that the direct path provides a prolonged inhibition of the dopaminergic cells, which persists from the time of the reward-predicting CS through the time at which the reward occurs. This is inconsistent with the data in two distinct but related ways. First, when the reward-predicting CS occurs, it produces a dopamine burst, but the dopamine cell firing then immediately returns to baseline. There is no persistent depression in dopamine cell firing, although the Houk et al. [39] model must predict such a persistent depression. Second, when an expected reward is omitted, there is a brief depression in the DA cell firing, after which it immediately returns to baseline. The [39] model instead predicts a prolonged (though below baseline) response rather than a transient response to the omission of expected reward.

The Berns and Sejnowski [1] model suggests that the primary source of net SNc excitation is the pallidum, via a hypothetical inhibitory neuron. No suggestion is given regarding the location of this neuron, or from which pallidal segment (internal or external) the signal originates. As in our model, the Berns and Sejnowski [1] model assumes that the striosomal cells are the main source of inhibition to the SNc, but their model does not treat dopamine cell temporal dynamics, which would be necessary for it to explain the data of Figures 1 and 2.

Summary: The new spectral timing model of nigral dopamine activity provides functional explanations of known SNc afferents. The model suggests how the ventral basal ganglia stream learns an excitatory prediction of reward via the PPTN, while the striosomal cells learn an adaptively timed inhibitory prediction of reward. This analysis clarifies how the nigral dopamine cells are linked to four other cell types that are directly or indirectly afferent to the SNc: ventral striatal cells, PPTN cells, striosomal cells of the basal ganglia, and cells in the lateral hypothalamus. The model predicts that an adaptive timing mechanism occurs at the striosomal cells. Key explanatory limitations of previous models, including TD and direct/indirect pathway models of nigral dopamine cell responses, are overcome by the present model. Finally, and most important for the present conference, the mathematical equations of the model [4] describe circuits that should be realizable in real-time hardware for compact incorporation into a biomimetic autonomous mobile robot[1].

References

1. Berns, G., Sejnowski, T.: A computational model of how the basal ganglia produce sequences. J. Cog. Neurosci. **10** (1998) 108-121
2. Berridge, K., Robinson, T.: What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res. Rev. **28** (1998) 309-369
3. Brog, J., Salyapongse, A., Deutch, A., Zahm, D.: The patterns of afferent innervation of the core and shell in the "accumbens" part of the rat ventral striatum: immunohistochemical detection of retrogradely transported fluoro-gold. J. Comp. Neurol. **338** (1993) 255-78
4. Brown, J., Bullock, D., Grossberg, S.: How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.* **19** (1999) 10502-10511
5. Brown, J., Bullock, D., Grossberg, S.: How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. In Preparation.
6. Bullock, D., Cisek, P., Grossberg, S.: Cortical networks for control of voluntary arm movements under variable force conditions. Cerebral Cortex **8** (1998) 48-62
7. Buonomano, D.V., Mauk, M.D.: Neural network model of the cerebellum: temporal discrimination and the timing of motor responses. Neural Comput **6** (1994) 38-55
8. Calabresi, P., Maj, R., Pisani, A., Mercuri, N., Bernardi, G.: Long-term synaptic depression in the striatum: Physiological and pharmacological characterization. J. Neurosci. **12** (1992a) 4224-4233
9. Calabresi, P., Pisani, A., Mercuri, N., Bernardi, G.: (1992b) Long-term potentiation in the striatum is unmasked by removing the voltage-dependent magnesium block of NMDA receptor channels. Eur. J. Neurosci. **4** (1992b) 929-935
10. Calabresi, P., Pisani, A., Mercuri, N., Bernardi, G.: Post-receptor mechanisms underlying striatal long-term depression. J. Neurosci. **14** (1994) 4871-4881
11. Carpenter, G., Grossberg, S.: ART 2: Self-organization of stable category recognition codes for analog input patterns. Applied Optics **26** (1987) 4919-4930
12. Cisek, P., Grossberg, S., Bullock, D.: A cortico-spinal model of reaching and proprioception under multiple task constraints. J. Cog. Neurosci. **10** (1998) 425-444
13. Conde, H.: Organization and physiology of the substantia nigra. Exp. Brain Res. **88** (1992) 233-248
14. Conde, H., Dormont, J., Farin, D. The role of the pedunculopontine tegmental nucleus in relation to conditioned motor performance in the cat. II. Effects of reversible inactivation by intracerebral microinjections. Exp. Brain Res. **121** (1998) 411-418
15. Contreras-Vidal, J., Grossberg, S., Bullock, D.: A neural model of cerebellar learning for arm movement control: Cortico-spino-cerebellar dynamics. Learning & Memory **3** (1997) 475-502

[1] Equation (14) of Brown, Bullock, and Grossberg (1999) was misprinted and should read $\frac{d}{dt}Z_{ij} = a_z[G_{ij}Y_{ij} - \Gamma_S]^+(-1000Z_{ij}N^- + g_sN^+)$.

16. Contreras-Vidal, J., Schultz, W.: A predictive reinforcement model of dopamine neurons for learning approach behavior. First International conference on Vision, Recognition, and Action: Neural Models of Mind and Machine. Boston University, Boston, (1997)
17. Crepel, F., Hemart, N., Jaillard, D., Daniel, H.: Cellular mechanisms of long-term depression in the cerebellum. Behav. Brain Sci. **19** (1996) 347-353
18. Dias, R., Robbins, T., Roberts, A.: Dissociation in prefrontal cortex of affective and attentional shifts. Nature **380** (1996) 69-72
19. Dormont, J., Conde, H., Farin, D.: The role of the pedunculopontine tegmental nucleus in relation to conditioned motor performance in the cat I. Context-dependent and reinforcement-related single unit activity. Exp. Br. Res. **121** (1998) 401-410
20. Eblen, F., Graybiel, A.: Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. J. Neurosci. **15** (1995) 5999-6013
21. Fiala, J., Grossberg, S., Bullock, D.: Metabotropic glutamate receptor activation in cerebellar purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. J. Neurosci. **16** (1996) 3760-3774
22. Finch, E.A., Augustine, G.J.: Local calcium signalling by inositol-1,4,5-trisphosphate in Purkinje cell dendrites. Nature **396** (1998) 753-756
23. Funahashi, S., Bruce, C.J., Goldman-Rakic, P.S.: Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. **61** (1989) 331-349
24. Futami, T., Takakusaki, K., Kitai, S.: Glutamatergic and cholinergic inputs from the pedunculopontine tegmental nucleus to dopamine neurons in the substantia nigra pars compacta. Neurosci Res. **21** (1995) 331-342
25. Gallagher, M., Chiba, A.: The amygdala and emotion. Curr. Op. Neurobiol. **6** (1996) 221-227
26. Gancarz, G., Grossberg, S.: A neural model of the saccade generator in the reticular formation. Neural Networks **11** (1998) 1159-1174
27. Gancarz, G., Grossberg, S.: A neural model of the saccadic eye movement control explains task-specific adaptation. Vision Res. **39** (1999) 3123-3143
28. Garris, P.A., Kilpatrick, M., Bunin, M.A., Michael, D., Walker, Q.D., Wightman, R.M.: Dissociation of dopamine release in the nucleus accumbens from intracranial self-stimulation. Nature **398** (1999) 67-69
29. Gerfen, C.: The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. Ann. Rev. Neurosci. **15** (1992) 285-320
30. Grossberg, S.: How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. Spatial Vision **12** (1999) 163-185
31. Grossberg, S., Merrill, J.: A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. Cognitive Brain Res. **1** (1992) 3-38
32. Grossberg, S., Merrill, J.: The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. J. Cogn. Neurosci. **8** (1996) 257-277
33. Grossberg, S., Mingolla, E., Pack, C.: A neural model of motion processing and visual navigation by cortical area MST. Cerebral Cortex **9** (1999) 878-895
34. Grossberg, S., Raizada, R.: Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. Vision Res. **40** (2000) 1413-1432
35. Grossberg, S., Roberts, K., Aguilar, M., Bullock, D.: A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. J. Neurosci. **17** (1997) 9706-9725
36. Grossberg, S., Schmajuk, N.: Neural dynamics of adaptive timing and temporal discrimination during associative learning. Neural Networks **2** (1989) 79-102
37. Hajos, M., Greenfield, S.: Synaptic connections between pars compacta and pars reticulata neurones: electrophysiological evidence for functional modules within the substantia nigra. Brain Res. **660** (1994) 216-224

38. Hollerman, J., Schultz, W.: Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. **1** (1998) 304-309
39. Houk, J., Adams, J., Barto, A.: A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, J., Davis, J., Beiser, D. (eds): Models of Information Processing in the Basal Ganglia. MIT Press Cambridge (1995) 249-270
40. Jaeger, D., Kita, H., Wilson, C.: Surround inhibition among projections neurons is weak or nonexistent in the rat neostriatum. J. Neurosci. **72** (1994) 2555-2558
41. Joiner, W.J., Tang, M.D., Wang, L.Y., Dworetzky, S.I., Boissard, C.G., Gan, L., Gribkoff, V.K., Kaczmarek, L.K.: Formation of intermediate-conductance calcium-activated potassium channels by interaction of Slack and Slo subunits. Nat. Neurosci. **1** (1998) 462-469
42. Kita, H.: Responses of globus pallidus neurons to cortical stimulation: intracellular study in the rat. Brain Res. **589** (1992) 84-90
43. Kita, H., Kitai, S. Intracellular study of rat globus pallidus neurons: Membrane properties and responses to neostriatal, subthalamic and nigral stimulation. Brain Res. **564** (1991) 296-305
44. Kojima, J., Yamaji, Y., Matsumura, M., Nambu, A., Inase, M., Tokuno, H., Takada, M., Imai, H.: Excitotoxic lesions of the pedunculopontine tegmental nucleus produce contralateral hemiparkinsonism in the monkey. Neurosci Lett. **226** (1997) 111-114
45. Ljungberg, T., Apicella, P., Schultz, W.: Responses of monkey dopamine neurons during learning of behavioral reactions. J. Neurophysiol. **67** (1992) 145-163
46. McDonald, R., White, N. A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. Behavioral Neurosci. **107** (1993) 3-22
47. Mirenowicz, J., Schultz, W.: Importance of unpredictability for reward responses in primate dopamine neurons. J. Neurophysiol. **72** (1994) 1024-1027
48. Mogenson, G., Wu, M.: Subpallidal projections to the mesencephalic locomotor region investigated with a combination of behavioral and electrophysiological recording techniques. Brain Res. Bull. **16** (1986) 383-390
49. Montague, P., Dayan, P., Sejnowski, T.: A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. **16** (1996) 1936-1947
50. Nakamura, K., Ono, T.: Lateral hypohtalamus neuron involvement in integration of natural and artificial rewards and cue signals. J. Neurophysiol. **55** (1986) 163-181
51. Pack, C., Grossberg, S., Mingolla, E.: A neural model of smooth pursuit control and motion perception by cortical area MST. In press J. Cog. Neurosci. (2000)
52. Pisani, A., Calabresi, P., Centonze, D., Bernardi, G.: Enhancement of NMDA responses by group I metabotropic glutamate receptor activation in striatal neurones. Br. J. Pharmacol. **120** (1996) 1007-1014
53. Scarnati, E., Hajdu, F., Pacitti, C., Tombol, T.: An EM and Golgi study on the connection between the nucleus tegmenti pedunculopontinus and the pars compacta of the substantia nigra in the rat. J. Hirnforsch **29** (1988) 95-105
54. Schultz, W.: Predictive reward signal of dopamine neurons. J. Neurophysiol. **80** (1998) 1-27
55. Schultz, W., Apicelli, P., Scarnati, E., Ljungberg, T.: Neuronal activity in monkey ventral striatum related to the expectation of reward. J. Neurosci. **12** (1992) 4595-4610
56. Schultz, W., Apicella, P., Ljungberg, T.: Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. **13** (1993) 900-913
57. Schultz, W., Romo, R., Ljungberg, T., Mirenowicz, J., Hollerman, J., Dickinson, A.: Reward-related signals carried by dopamine neurons. In: Houk, J., Davis, J., Beiser, D. (eds): Models of Information Processing in the Basal Ganglia. MIT Press Cambridge (1995) 11-27

58. Schultz, W., Dayan, P., Montague, P.: A neural substrate of prediction and reward. Science **275** (1997) 1593-1598
59. Semba, K., Fibiger, H.: Afferent connections of the laterodorsal and the pedunculopontine tegmental nuclei in the rat: a retro- and antero-grade transport and immunohistochemical study. J Comp Neurol. **323** (1992) 387-410
60. Smith, I., Grace, A.: Role of the subthalamic nucleus in the regulation of nigral dopamine neuron activity. Synapse **12** (1992) 287-303
61. Suri, R., Schultz, W.: Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Exp. Br. Res. **121** (1998) 350-354
62. Takakusaki, K., Shiroyama, T., Kitai, S.: Two types of cholinergic neurons in the rat tegmental pedunculopontine nucleus: electrophysiological and morphological characterization. Neuroscience **79** (1997) 1089-1109
63. Takechi, H., Eilers, J., Konnerth, A.: A new class of synaptic response involving calcium release in dendritic spines. Nature **396** (1998) 757-760
64. Tepper, J., Martin, L., Anderson, D.: GABAA receptor-mediated inhibition of rat substantia nigra dopaminergic neurons by pars reticulata projection neurons. J. Neurosci. **15** (1995) 3092-3103
65. Wickens, J., Kotter, R.: (1995) Cellular models of reinforcement. In: Houk, J., Davis, J., Beiser, D. (eds): Models of Information Processing in the Basal Ganglia. MIT Press Cambridge (1995) 187-214
66. Wickens, J., Begg, A., Arbuthnott, G.: Dopamine reverses the depresdsion of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. Neuroscience **70** (1996) 1-5
67. Wilkie, D.M.: Stimulus intensity affects pigeons' timing behavior: Implications for an internal clock model. Animal Learning and Behavior **15** (1987) 35-39
68. Wilson, C.: The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In: Houk, J., Davis, J., Beiser, D. (eds): Models of Information Processing in the Basal Ganglia. MIT Press Cambridge (1995) 29-50
69. Yang, C., Mogenson, G.: Hippocampal signal transmission to the pedunculopontine nucleus and its regulation by dopamine D2 receptors in the nucleus accumbens: an electrophysiological and behavioural study. Neuroscience **23** (1987) 1041-55