PERGAMON

Neural Networks 15 (2002) 535–547

Neural
Networks

www.elsevier.com/locate/neunet

2002 Special issue

Actor–critic models of the basal ganglia: new anatomical and computational perspectives

Daphna Joel[a,*], Yael Niv[a], Eytan Ruppin[b]

[a]*Department of Psychology, Tel-Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel*
[b]*Schools of Medicine and Mathematical Sciences, Tel-Aviv University, Tel Aviv 69978, Israel*

Received 23 October 2001; revised 27 March 2002; accepted 27 March 2002

Abstract

A large number of computational models of information processing in the basal ganglia have been developed in recent years. Prominent in these are actor–critic models of basal ganglia functioning, which build on the strong resemblance between dopamine neuron activity and the temporal difference prediction error signal in the critic, and between dopamine-dependent long-term synaptic plasticity in the striatum and learning guided by a prediction error signal in the actor. We selectively review several actor–critic models of the basal ganglia with an emphasis on two important aspects: the way in which models of the critic reproduce the temporal dynamics of dopamine firing, and the extent to which models of the actor take into account known basal ganglia anatomy and physiology. To complement the efforts to relate basal ganglia mechanisms to reinforcement learning (RL), we introduce an alternative approach to modeling a critic network, which uses Evolutionary Computation techniques to 'evolve' an optimal RL mechanism, and relate the evolved mechanism to the basic model of the critic. We conclude our discussion of models of the critic by a critical discussion of the anatomical plausibility of implementations of a critic in basal ganglia circuitry, and conclude that such implementations build on assumptions that are inconsistent with the known anatomy of the basal ganglia. We return to the actor component of the actor–critic model, which is usually modeled at the striatal level with very little detail. We describe an alternative model of the basal ganglia which takes into account several important, and previously neglected, anatomical and physiological characteristics of basal ganglia–thalamocortical connectivity and suggests that the basal ganglia performs reinforcement-biased dimensionality reduction of cortical inputs. We further suggest that since such selective encoding may bias the representation at the level of the frontal cortex towards the selection of rewarded plans and actions, the reinforcement-driven dimensionality reduction framework may serve as a basis for basal ganglia actor models. We conclude with a short discussion of the dual role of the dopamine signal in RL and in behavioral switching. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Basal ganglia; Dopamine; Reinforcement learning; Actor–critic; Dimensionality reduction; Evolutionary computation; Behavioral switching; Striosomes/patches

1. Introduction

A large number of computational models of information processing in the basal ganglia have been developed in recent years (Houk, Adams, & Barto, 1995; see Fig. 1 for a general scheme of basal ganglia connections). A recent review groups these models into three main (not mutually exclusive) categories: models of serial processing, models of action selection, and models of reinforcement learning (RL) (Gillies & Arbuthnott, 2000). The first category includes models that assign a central role to the basal ganglia loop structure in generating sequences of activity patterns (Berns & Sejnowski, 1998). The second class focuses on the tonic inhibitory activity that the major basal ganglia output nuclei exert upon their targets, assuming that it provides for action selection via focused disinhibition (Gurney, Prescott, & Redgrave, 2001). In this paper, we focus on the third class of models, which assign a major role for the basal ganglia in RL.

The interest in RL models of the basal ganglia has been initiated by the seminal studies of Wolfram Schultz, which provided experimental evidence suggesting that RL plays an important role in basal ganglia processing (Schultz & Dickinson, 2000; Schultz, Tremblay, & Hollerman, 2000). Recording the activity of dopaminergic (DA) neurons in monkeys during the acquisition and performance of behavioral tasks, Schultz and colleagues found that DA neurons respond phasically to primary rewards, and as the experiment

[*] Corresponding author. Tel.: +972-3-6408996; fax: +972-3-6407391.
E-mail address: djoel@post.tau.ac.il (D. Joel).

0893-6080/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0893-6080(02)00047-3

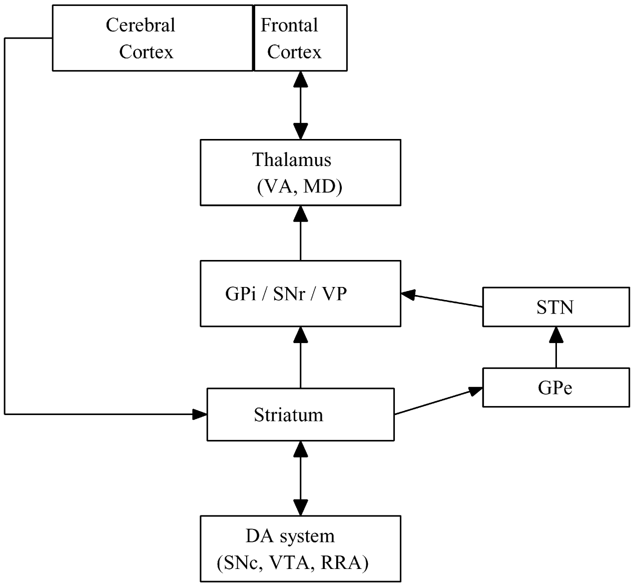

Fig. 1. A general scheme of basal ganglia–thalamocortical connections. The striatum is the main input structure of the basal ganglia. It is divided into dorsal striatum (most of the caudate and putamen) and ventral striatum (nucleus accumbens and the ventromedial parts of the caudate and putamen). The striatum is innervated by the entire cerebral cortex, and projects to the output nuclei of the basal ganglia, the globus pallidus (GPi), the substantia nigra pars reticulata (SNr) and the ventral pallidum (VP). These nuclei project in turn to the ventral anterior (VA) and mediodorsal (MD) thalamic nuclei, which are reciprocally connected with the frontal cortex. Information from the striatum can also reach the output nuclei via the 'indirect pathway', namely, via striatal projections to the external segment of the globus pallidus (GPe), GPe projections to the subthalamic nucleus (STN), and the latter's projections to GPi/SNr/VP. The striatum also projects dopaminergic neurons in the substantia nigra pars compacta (SNC), retrorubral area (RRA) and ventral tegmental area (VTA). Please note that this scheme does not relate to two important principles of organization of the depicted projections. One is the compartmental organization of the dorsal striatum into striosomes (patches, in rats) and matrix. The other is the topographical organization of the projections between the different levels into several 'streams' which form several ganglia–thalamocortical circuits. (For extensive reviews of the organization of basal ganglia–thalamocortical connections, see Alexander & Crutcher, 1990; Gerfen, 1992; Joel & Weiner, 1994, 1997, 2000; Parent, 1990).

progresses, the response of these neurons gradually shifts back in time from the primary reward to reward-predicting stimuli. The firing pattern of DA neurons was also found to reflect information regarding the timing of delayed rewards (relative to the reward-predicting stimulus), as could be seen by the precisely timed depression of DA firing when an expected reward was omitted. This pattern of activity is very similar to that generated by computational algorithms of RL, in particular temporal difference (TD) models (Sutton, 1988), as described in detail in another paper in this issue (Suri, 2002).

In the context of basal ganglia modeling, TD learning is mainly used in the framework of Actor–Critic models (Barto, 1995; Houk et al., 1995). In such models, an actor sub-network learns to perform actions so as to maximize the weighted sum of future rewards, which is computed at every timestep by a critic sub-network (Barto, 1995). The critic is adaptive, in that it learns to predict the weighted sum of future rewards based on the current sensory input and the actor's policy, by means of an iterative process in which it compares its own predictions to the actual rewards obtained by the acting agent. The learning rule used by the adaptive critic is the TD learning rule (Sutton, 1988) in which the error between two adjacent predictions (the TD error) is used to update the critic's weights. Numerous studies have shown that using such an error signal to train the actor results in very efficient RL (Kaelbling, Littman, & Moore, 1996; Tesauro, 1995; Zhang & Dietterich, 1996).

The analogy between the basal ganglia and actor–critic models builds on the strong resemblance between DA neuron activity and the TD prediction error signal, and between DA-dependent long-term synaptic plasticity in the striatum (Calabresi et al., 2000; Wickens, Begg, & Arbuthnott, 1996) and learning guided by a prediction error signal in the actor. Actor–critic models of basal ganglia functioning have gained popularity in recent years, and several models have been proposed. A comparison between these models shows that they mainly differ in two important aspects. Models of the critic differ in the way in which the temporal dynamics of DA firing are reproduced, that is, in the network architecture responsible for producing the short phasic response of DA neurons to unpredicted rewards and reward-predicting stimuli, and the depression induced by reward omission. Models of the actor differ in the extent to which they take into account known basal ganglia anatomy and physiology.

In Section 2 we briefly review several actor–critic models of the basal ganglia with an emphasis on the mechanism responsible for reproducing the temporal dynamics of DA firing and on the architecture of the actor. Section 3 introduces an alternative approach to modeling a critic network, which uses Evolutionary Computation techniques to evolve an optimal RL mechanism. This mechanism is then related to the more classic models of critics presented in Section 2. Section 4 provides a critical discussion of the anatomical plausibility of the implementation of an adaptive critic in basal ganglia circuitry. In Section 5 we return to the actor component of the actor–critic model and describe an alternative model of the basal ganglia which takes into account several important, and previously neglected, anatomical and physiological characteristics of basal ganglia–thalamocortical connectivity. This model sees the main computational role of the basal ganglia as being a key station in a dimension reduction coding–decoding cortico-striato-pallido-thalamo-cortical loop. We conclude with a short discussion of the dual role of the DA signal in RL and behavioral switching.

2. Actor–critic models of reinforcement learning in the basal ganglia

2.1. *Houk, Adams, and Barto (1995)*

One of the first actor–critic models of the basal ganglia was presented by Houk et al. (1995). This model suggests that striosomal modules fulfill the main functions of the adaptive critic, whereas matrix modules function as an actor. Striosomal modules comprise of striatal striosomes, subthalamic nucleus, and dopaminergic neurons in the substantia nigra pars compacta (SNc). According to the model, three sources of input interact in generating the firing patterns of DA neurons. Two of these inputs arise from striatal striosomes and provide information on the occurrence of stimuli that predict reinforcement. One is a direct input to the SNc, which provides prolonged inhibition, and the other is an indirect input, channeled to the DA neurons via the subthalamic nucleus, which provides phasic excitation. The third input to DA neurons, which is assumed to arise from the lateral hypothalamus, is also excitatory and provides information on the occurrence of primary rewards. During acquisition, striatal striosomal neurons learn to fire in bursts when stimuli predicting future primary reinforcement occur, through DA-dependent strengthening of corticostriatal synapses. After learning, the presentation of a reward-predicting stimulus would lead to DA burst firing as a result of indirect excitation from the striosomes. The arrival of an expected primary reward would not lead to a DA response, since the prolonged direct inhibition arising from the striosomes would cancel the excitation arising from the lateral hypothalamus. In terms of the TD equation for the prediction error, the primary reinforcement in the TD equation is equated with the primary reinforcement to DA neurons, the prediction $P(t)$ of future reinforcement is equated with the indirect excitatory input to DA neurons, and the direct inhibitory input is equated with the prediction $P(t-1)$ at the earlier time step.

Houk et al.'s model of the critic does not include an exact timing mechanism, but rather a slow and persistent inhibition of DA neurons. As a result, it does not account for the timed depression of DA activity when an expected reward is omitted. This problem has been tackled in later models by using a different representation of the inputs to the network. The 'complete serial compound stimulus' (Montague, Dayan, & Sejnowski, 1996) is a representation of the stimulus which has a distinct activation component for each timestep during and for a while after the presentation of the stimulus. In general, it is assumed that the presentation of a stimulus initiates an exuberance of temporal representations and the learning rule can select the ones that are appropriate, that is, that correspond to the stimulus–reward interval. The models described later use this computational principle, but describe different neural implementations of this general solution.

In contrast to the detailed discussion of the critic, Houk et al. provide only a general scheme of the implementation of the actor in basal ganglia circuitry. According to their model, matrix modules, comprising of the striatal matrix, subthalamic nucleus, globus pallidus, thalamus, and frontal cortex, generate signals that command various actions or represent plans that organize other systems to generate actual command signals. They note, however, that from a sensory perspective, the signals generated by the matrix modules may signal the occurrence of salient contexts (see also Section 5).

2.2. *Suri and Schultz (1998, 1999)*

Suri and Schultz have extended the basic actor–critic model presented by Barto (1995), both by providing a neural model of the actor and by modifying the TD algorithm with respect to stimulus representation so as to reproduce the timed depression of DA activity at the time of omitted reward. The timing mechanism was implemented by representing each stimulus using a set of neurons, each of which was activated for a different duration (instead of the single prolonged inhibition in Barto's model). The critic learning rule was modified to ensure that only the weight for the stimulus representation component that covers the actual stimulus–reward interval is adapted, whereas the weights for the other neurons remain unchanged. These modifications allowed the model to replicate the firing pattern of DA neurons to reward-predicting stimuli, predicted rewards and omitted rewards (Suri & Schultz, 1998). In an enhancement of their basic model (Suri & Schultz, 1999), the teaching signal was further enriched to better fit the pertaining biological data on the responses of DA neurons to novel stimuli.

The actor in these models was comprised of one layer of neurons, each representing a specific action. It learned stimulus-action pairs based on the prediction error signal provided by the critic. A winner-take-all rule that can be implemented through lateral inhibition between neurons ensured that only one action was selected at a given time.

Using this modified and extended model of the critic, Suri and Schultz (1998, 1999) demonstrated that even a simple actor network was sufficient to solve relatively complex behavioral tasks. However, although these authors acknowledge the general similarity between the actor–critic architecture and basal ganglia structure, and suggest that the components of the temporal stimulus representation may correspond to sustained activity of striatal and cortical neurons, no attempt was made to implement the critic in the known architecture of the basal ganglia. In addition, the extension of the TD algorithm to include novelty responses, generalization responses and some temporal aspects in reward prediction, was achieved by arbitrarily specifying the values of specific parameters of the model (e.g. initializing specific synaptic weights with specific values, using different learning rates for different synapses) rather than by a more biologically plausible implementation in a neural network related to basal ganglia anatomy and physiology. Such an attempt has been made by Contreras-Vidal and Schultz (1999).

2.3. *Contreras-Vidal and Schultz (1999)*

Contreras-Vidal and Schultz (1999) provide a neural network architecture related to basal ganglia anatomy which can account for DA responses to novelty, generalization and discrimination of appetitive and aversive stimuli, by incorporating an additional adaptive resonance neural network originally developed by Carpenter and Grossberg (1987). They further suggest that there are two types of reward prediction errors: a signal representing error in the timing of reward prediction, which may be related to the TD model, and a signal coding for error in the type and amount of reward prediction, which may be related to the adaptive resonance network. Whereas description of this network is beyond the scope of our paper, we will briefly discuss their implementation of the timing mechanism responsible for the depression of DA activity at the time of omitted reward. Similar to Suri and Schultz (1998, 1999), Contreras-Vidal and Schultz postulate that striosomal neurons generate a spectrum of timing signals in response to a sensory input (a ‘complete serial compound’ representation of the stimulus). However, in their model, striosomal neurons are activated successively following stimulus onset and for a restricted period of time, in contrast to the sustained activity of different durations assumed by Suri and Schultz. As in Suri and Schultz’s models, the learning rule ensures that synapses of striosomal neurons active at the time of primary reward delivery (that is, in conjunction with DA activity), are strengthened, but in Contreras-Vidal and Schultz’s model, it is striatonigral rather than corticostriatal synapses that are assumed to be modified by learning. (It should be noted that whereas there is ample evidence for long term plasticity in corticostriatal synapses, there is no such evidence for striatonigral synapses.) After learning, the excitation of DA neurons by predicted primary rewards is canceled by the timed inhibition arising from striosomes. Importantly, in contrast to models based on the general scheme of a critic presented by Barto (1995), in this model the source of excitation to DA neurons is assumed to be different from that of inhibition. Thus, the phasic DA response to reward-predicting stimuli is attributed to excitation arising from the prefrontal cortex (PFC) and channeled to the DA neurons via the striatal matrix and substantia nigra pars reticulata (SNr).

2.4. *Brown, Bullock, and Grossberg (1999)*

Another attempt to answer the question of what biological mechanisms compute the DA response to rewards and reward-predicting stimuli, is provided by Brown et al. (1999). Similar to Contreras-Vidal and Schultz (1999), these authors suggest that the fast excitatory response to conditioned stimuli and the delayed, adaptively timed inhibition of response to rewarding unconditioned stimuli, are subserved by different anatomical pathways. The suppression of DA responses to predicted rewards and the decrease in DA activity when a predicted reward is omitted depend on adaptively timed inhibitory projections from striosomes in the dorsal and ventral striatum to SNc. In contrast to Contreras-Vidal and Schultz (1999), however, the successive bursting of striosomal neurons following stimulus onset depends on an intra-cellular calcium-dependent timing mechanism. As in earlier models, the simultaneous occurrence of striosomal neurons’ spiking and DA burst firing (in response to a primary reward) leads to enhancement of corticostriatal synapses on the active striosomal neurons. A striosomal population that fires at the expected time of reward delivery is thus selected, hence, forward preventing the DA response to predicted rewards. The activation of DA neurons to rewards and reward-predicting stimuli is attributed to excitatory projections from the pedunculopontine tegmental nucleus (PPN) to the SNc. The phasic nature of DA activation is suggested to be due to habituation or accommodation of PPN neurons projecting to the SNc.

2.5. *Suri, Bargas, and Arbib (2001)*

In a recent paper, Suri et al. (2001) extend the actor–critic model employed by Suri and Schultz (1998, 1999) by using an extended TD model, an actor based on the anatomy of basal ganglia–thalamocortical circuitry, and complex interactions between the critic and actor. Similar to the actor in Suri and Schultz (1998, 1999), each model neuron in the striatal layer is thought to correspond to a small population of striatal matrix neurons that is able to elicit an action. However, the mechanism ensuring the selection of only one action at a given time depends on the interaction between the direct and indirect pathways connecting the striatum to the basal ganglia output nuclei and on a winner-take-all rule at the cortical level. In this model DA affects the action of the actor by three types of membrane potential-dependent influences on striatal neurons: long-term adaptation of corticostiatal transmission, and transient effects on striatal neurons’ firing rates and duration of the up- and down-state. The critic receives sensory and reward information, as in earlier models, and in addition receives information regarding the intended and actual action from the thalamic and cortical levels of the actor. As a result, the critic can learn both stimulus–reward and action-stimulus associations.

Suri et al. showed that this extended actor–critic model is capable of sensorimotor learning, as is the original actor–critic model employed by Suri and Schultz (1998, 1999). In addition, this model has planning capabilities, that is, the ability to form novel associative chains and select its action in relation to the outcome predicted by these associative chains. Planning in this model critically depends on the fact that the input to the extended critic includes prediction of future stimuli and information regarding intended actions (provided by the thalamus), which can be used to estimate future prediction signals, and on the fact that the critic is run for two iterations for every action step. Together, these characteristics enable the evaluation of intended actions, based on the formation of new associative chains between an action, the sensory outcome of that action and the reward.

Suri et al. also model the novelty responses of DA neurons, that is, the transient increase in striatal DA upon the encounter of a novel stimulus. This novelty response increases the likelihood of firing in striatal neurons in the up-state, and therefore the likelihood of action, thus generating exploration behavior. The novelty response of DA neurons is achieved through an initial choice of weights effectively equivalent to assigning optimistic initial values to novel places/stimuli. Exploratory behavior also results from the stochastic transitions between up and down states of the striatal neurons in the model. Below we describe another mechanism which may control the tradeoff between exploration and exploitation, which is characteristic of armed bandit situations.

3. Evolution of reinforcement learning—a different approach to modeling the critic

An alternative approach to modeling a RL critic has been taken by us (Niv, Joel, Meilijson, & Ruppin, (2002) in press). We have used Evolutionary Computation techniques to evolve the neuronal learning rules of a simple neural network model of decision-making in bumble-bees foraging for nectar. To this end we formalized a very general framework for evolving learning rules, which encompassed all heterosynaptic Hebbian learning rules and also allowed for neuromodulation of synaptic plasticity. Using a genetic algorithm, bees were evolved based on their nectar-gathering ability in a changing environment. As a result of the uncertainty of the environment, efficient foraging could only result from efficient RL, thus an efficient RL mechanism was evolved.

To avoid the possible confusion of terms, we make a distinction between the notions of heterosynaptic plasticity (Dittman & Regehr, 1997; Schacher, Wu, & Sun, 1997; Vogt & Nicoll, 1999) and neuromodulation of plasticity (Bailey, Giustetto, Huang, Hawkins, & Kandel, 2000; Fellous & Linster, 1998). In contrast to the conventionally used monosynaptic Hebbian learning, heterosynaptic Hebbian learning allows for activity-independent modification of synapses such that a synapse can also be updated when only the pre-synaptic or post-synaptic component has been active, and more generally, even when neither have been active. We term this, 'heterosynaptic' modification as it allows for the firing of a neuron to affect all its synapses, regardless of the activity of the other neurons connected to them. Neuromodulation of synaptic plasticity further enhances the learning rule by allowing a three-factor interaction in the learning process: through neuromodulation the activity of a neuron can gate the plasticity of a synapse between two other neurons. Both heterosynaptic plasticity and neuromodulatory gating of synaptic plasticity have been demonstrated in neural tissues (Bailey et al., 2000; Dittman & Regehr, 1997; Fellous & Linster, 1998; Schacher et al., 1997; Vogt & Nicoll, 1999), and have been recognized to increase the computational complexity of synaptic learning (Bailey et al., 2000; Fellous & Linster, 1998; Wickens and Kotter, 1995). By allowing for heterosynaptic learning and neuromodulation of plasticity, we defined a very large search space in which the genetic algorithm could search for optimal synaptic learning rules.

Within the framework of our model, we showed that only one network architecture could produce effective RL and above-random foraging behavior. The evolved network was similar to an architecture proposed earlier by Montague, Dayan, Person, and Sejnowski (1995) and consisted of a sensory input module which codes changes over time in the sensory input, a reward input module which provides information on nectar intake, and an output unit *P*. The evolved learning rule was indeed heterosynaptic and incorporated neuromodulation of synaptic plasticity (for a detailed description see Niv et al. (2002) in press; http://www.cns.tau.ac.il/∽yaeln/AdaptiveBehavior2002.htm).

The learning mechanism evolved can be closely related to the adaptive critic, with respect to the activity of the output unit and the neuromodulation of synaptic plasticity. Similar to Montague et al. (1995), the output of the model unit *P* quite accurately captures the essence of the activity patterns of midbrain dopaminergic neurons in primates and rodents (Montague et al., 1996; Schultz, Dayan, & Montague, 1997), and the corresponding octopaminergic neurons in bees (Hammer, 1997; Menzel & Muller, 1996). Since in the evolved network the synaptic weights come to represent the expected reward and the inputs represent changes over time in the sensory input, the output of the network represents an ongoing comparison between the expected reward in subsequent timesteps. As in the critic model, this comparison provides the error measure by which the network updates its weights and learns to better predict future rewards.

With regard to neuromodulation, this work has shown that efficient RL critically depends on the evolution of neuromodulation of synaptic plasticity, that is, the gating of synaptic plasticity between two neurons by the activity of a third neuron (a 'three-factor' Hebbian learning rule). This is similar to the DA-dependent plasticity described in corticostriatal synapses (Calabresi et al., 2000; Wickens et al., 1996). The demonstration of the computational optimality of this learning rule to RL contributes to the attempts of computational models to bridge between the complex anatomy and physiology of the basal ganglia–

thalamocortical system and findings from lesion and imaging studies implicating this system in procedural or stimulus-response learning.

In contrast to the monosynaptic learning rules usually employed by actor–critic models, the heterosynaptic learning rules we have evolved enable the modification of a synapse even when its pre- or post-synaptic component (or both) are not activated. This allows for non-trivial interactions between the rewards predicted by different stimuli. For example, the amount of reward predicted by one stimulus can be modified as a result of the disappointment or surprise encountered when facing a different stimulus, and the tendency to perform a certain response can change even when another response was executed. In the model, these micro-level heterosynaptic plasticity dynamics give rise directly to the macro-level tradeoff between exploration and exploitation characteristic of foraging behavior. Evidence from cerebellar (Dittman & Regehr, 1997) and hippocampal (Vogt & Nicoll, 1999) synapses shows that heterosynaptic plasticity indeed occurs in the brain, but this phenomenon has yet to be demonstrated in the striatum. Such a mechanism could provide another intra-striatal mechanism that controls exploration, in addition to those suggested by Suri et al. (2001).

Our model reflects mainly the critic module of the actor–critic framework and consists only of an extremely simplistic actor. Future work focused on elaborating the actor component of the model is needed in order to increase the relevance of the model to learning in the basal ganglia, and to allow for a more detailed account of how this computational model could be implemented in basal ganglia circuitry.

4. Critic networks in the basal ganglia—a discussion

As evident from the earlier description of the models, it is widely accepted that a critic-like function is sub-served by the connections of striatal striosomes with the DA system. Yet, only three studies (Brown et al., 1999; Contreras-Vidal & Schultz, 1999; Houk et al., 1995) have attempted to provide neural network models of the critic based on the known anatomy and physiology of these connections. A comparison between these models in general, and in relation to the implementation of a timing mechanism in particular, can be found in Brown et al. (1999) and Contreras-Vidal and Schultz (1999). Here we would like to focus on two issues: (1) Are there anatomical grounds to support the consensus that striatal striosomes play a critical role in the Critic? (2) Do the excitation of DA neurons when encountering a reward-predicting stimulus and the inhibition of these neurons when a predicted reward is omitted, arise from one origin (as suggested by Houk et al. (1995) and implied in the different models of Suri and colleagues), or do they arise from two different sources with different characteristics (as suggested by Brown et al., 1999; Contreras-Vidal & Schultz, 1999)?

4.1. Striosomes and the adaptive critic

The focus on the connections between the striosomal compartment of the striatum and the DA system stems from the work of Charles R. Gerfen, who showed that in rats there are reciprocal connections between the striosomes of the dorsal striatum and a relatively small group of DA neurons, residing in the ventral part of the SNc and in the SNr (Gerfen, 1984, 1985; Gerfen, Herkenham, & Thibault, 1987). Current data in primates suggest that a group of DA neurons may be reciprocally connected with neurons in the dorsal striatum. There is no evidence, however, regarding the compartmental origin of these striatal neurons (see Joel & Weiner, 2000). Therefore, the implementation of the critic in the connections of striosomal neurons with the DA system is not supported by anatomical evidence in primates. Even when considered only with regard to anatomical evidence in rats, such implementation can account only for the activity of a relatively small group of DA neurons.

Is there another group of striatal neurons which can replace the ‘striosomes’ in the different models? Or, stated differently, is there a group of striatal neurons, which have reciprocal connections with the entire DA system? Two recent meta-analyses of the anatomical data regarding the connections between the striatum and the DA system in primates (Haber, Fudge, & McFarland, 2000; Joel & Weiner, 2000) and rats (Joel & Weiner, 2000) have concluded that an asymmetry rather than reciprocity is an important characteristic of the connections between the striatum and the DA system. That is, the limbic (ventral) striatum projects to most of the DA system but is innervated by a relatively small sub-group of DA neurons, whereas the reverse is true for the motor striatum (mainly putamen), which is innervated by a larger region of the DA system than the one to which it projects. As a result of this organization, the limbic striatum reciprocates its DA input and innervates DA neurons projecting to the associative (mainly caudate nucleus) and motor striatum; the associative striatum reciprocates part of its DA input and innervates DA neurons projecting to the motor striatum, and the motor striatum reciprocates part of its DA input (Haber et al., 2000; Joel & Weiner, 2000). Based on this organization, the authors of both papers suggested that the striato-DA-striatal connections may serve an important role in the transfer of information between basal ganglia–thalamocortical circuits, in addition to the role attributed to these connections in intra-circuit processing.

We conclude that *a critic which builds on reciprocal connections between DA neurons and another group of neurons, cannot be implemented in the connections between the DA system and the striatum*. However, since the ventral striatum (and ventral pallidum (VP), see later) provides a major inhibitory projection to the DA system, and the activity of many ventral striatal neurons is related to rewards and reward-predicting stimuli, it is possible that this structure is part of the mechanism responsible for the activity pattern of DA neurons. Future work will hopefully reveal the role of the topographical organization of the connections between the striatum and the DA system in the computations performed by the basal ganglia.

4.2. Source(s) of excitation and inhibition to DA neurons

All the models we have reviewed, except Contreras-Vidal and Schultz's (1999) model, are based on Barto's (1995) architecture of the critic. In this architecture the computation of the prediction error depends on the activation of a neuron or a group of neurons by the reward-predicting stimulus. This leads both to fast excitation and delayed inhibition of DA neurons (corresponding to $P(t)$ and $-P(t-1)$ in Barto's model, respectively). Since most of these models assume that the source of excitation and inhibition resides in striatal striosomes, the existence of anatomical pathways from the striosomes to the DA system, which carry these signals, is hypothesized, as described in Houk et al.'s model (see earlier). We have already discussed the problem in assuming that striosomes provide direct inhibition to the entire DA system. However, Houk et al.'s model encounters an additional difficulty in assuming the existence of an indirect pathway from the striosomes via the subthalamic nucleus to the DA system, since current anatomical data suggest that striatal projections to the subthalamic nucleus (via the globus pallidus) arise from matrix neurons and not from striosomal neurons (for review see Gerfen, 1992). It is therefore unlikely that striosomes provide the fast excitation to DA neurons.

Is it possible that striatal (not necessarily striosomal) neurons are the source of the early excitatory and late inhibitory input to DA neurons? Electrophysiological data (for review see Bunney, Chiodo, & Grace, 1991; Kalivas, 1993; Pucak & Grace, 1994) and anatomical data (for review see Haber et al., 2000; Joel & Weiner, 2000) indeed suggest that activity of neurons of both the dorsal and ventral striatum can either suppress DA cell activity directly or promote bursting in DA cells indirectly. However, the direct inhibitory effect likely precedes the indirect excitatory effect, which is mediated by at least two inhibitory synapses (e.g. ventral striatal projections to the GABAergic neurons of the VP, which project to most of the DA system). This implies that the signal received by the DA system is $P(t-1) - P(t)$ rather than $P(t) - P(t-1)$. This, of course, predicts an opposite activity pattern of DA neurons to that observed. For example, it will result in inhibition, rather than excitation, of DA activity in response to the encounter of reward-predicting stimuli. In addition to the timing problem, the inhibitory and facilitatory effects likely arise from different subsets of neurons in the dorsal striatum. Regarding the ventral striatum, it remains an open question whether ventral striatal neurons projecting to the VP are distinct from those projecting directly to DA cells (see Joel & Weiner, 2000). Taken together, *it is unlikely that a single group of striatal neurons is the source of both indirect fast excitation and direct delayed inhibition to the DA neurons, as required by most models of the critic*.

An alternative source of such a dual input to the DA system is the limbic PFC. Schultz (1998) suggested that input from this cortical region may be responsible for the excitatory responses of DA neurons to rewards and reward-predicting stimuli. Neurons in the limbic PFC respond to primary rewards and reward-predicting stimuli and show sustained activity during the expectation of reward (for review see Schultz, Tremblay, & Hollerman, 1998; Zald & Kim, 2001), and data in rats suggest that the limbic PFC projects directly to DA neurons (for review see Overton & Clark, 1997). The limbic PFC projects in addition to the limbic (ventral) striatum (Berendse, Galis-de-Graaf, & Groenewegen, 1992; Groenewegen, Berendse, Wolters, & Lohman, 1990; Parent, 1990; Uylings & van Eden, 1990; Yeterian & Pandya, 1991). Via the latter pathway, the limbic PFC can provide the delayed inhibition to DA neurons. This is in line with electrophysiological evidence that neurons in the limbic striatum show reward related activity, including sustained activity during the expectation of rewards and reward-predicting stimuli (Rolls & Johnstone, 1992; Schultz, Apiccela, Scarnati, & Ljungberg, 1992). The finding of neurons with sustained activity in the limbic PFC and limbic striatum is in line with the timing mechanism implemented in the critic models of Suri and Schultz (1998, 1999). As detailed earlier, in their model, sustained activity of the stimulus representation component that covers the actual stimulus–reward interval is responsible for the phasic DA response to reward-predicting stimuli, the lack of DA response during the stimulus–reward interval, and the depression of DA activity when expected rewards are omitted. Assuming that neurons of the limbic PFC provide the timed sustained activity, their direct projections can provide the prediction at time t of future reinforcement ($P(t)$), and their indirect projections, via the limbic striatum, can provide the delayed prediction from the previous timestep ($P(t-1)$), as required by Suri and Schultz's model. We would like to note, however, that although the above suggestion respects known anatomy, it does not incorporate other important projections to the DA cells which may play a role in the production of the DA signal, most notably, the projections from the limbic pallidum.

The assumption that the limbic PFC is the source of the early excitation and late inhibition to DA neurons can also be found in Brown et al.'s (1999) model. However, in their model the translation of sustained activity in cortical neurons to phasic responses of DA neurons (i.e. increase in response to reward-predicting stimuli and decrease in response to the omission of predicted rewards) is attained by the specific properties of the pathways carrying the excitation and inhibition signals. Thus, habituation of the PPN (which is the final station in the pathway providing excitatory input to the DA neurons in their model) ensures that DA neurons receive only a transient excitation following a reward-predicting stimulus, and an intracellular adaptive timing mechanism in the striosomes translates the sustained cortical activity into a transient and timed inhibition of DA cells at the time the reward is expected.

We would like to end this section by concluding that although the connections of the basal ganglia with the DA system are thought to carry a 'critic-like' function, *current implementations of basic critic models in basal ganglia connections build on assumptions that are inconsistent with the known anatomy of these nuclei*. Hopefully, future attempts to implement such models in known neural circuits will both shed light on the functioning of basal ganglia nuclei, and provide additional constraints to the theoretical models.

5. Reinforcement driven dimensionality reduction—reward-biased representation in the basal ganglia

In contrast to the fairly advanced models of basal ganglia processing which relate to the critic component of the actor–critic model, most present models employ very simple actor systems of information processing at the striatal level. The exception is the actor in the model presented by Suri et al. (2001), which is implemented in the basal ganglia–thalamocortical connections. However, even in this model, the output of the striatum is assumed to translate into cortical activity in a relatively straightforward manner. Each striatal neuron corresponds to a specific action. Via a specific neuron at the level of the internal segment of the globus pallidus (GPi) and SNr it disinhibits one thalamic neuron, which projects in turn to a specific cortical neuron, whose persistent activation executes a cortical action.

In this section, we will present a model of basal ganglia processing that may potentially be extended to serve as a basis for a basal ganglia actor model. Based on several important constraints imposed by known basal ganglia anatomy and physiology, this model suggests that the basal ganglia perform an efficient reinforcement driven dimensionality reduction (RDDR) of the cortical representation (Bar-Gad, Havazelet-Heimer, Ruppin, & Bergman, 2000; www.math.tau.ac.il/ ∽ ruppin). We focus on the theoretical part. For a more detailed presentation of the model and for a description of the electrophysiological experiments performed on behaving monkeys to test some of the model predictions, see Bar-Gad et al. (2000).

This model was motivated by two main anatomical and physiological characteristics of basal ganglia–thalamocortical circuitry:

1. *The funneling structure of the basal ganglia*. The number of cortical neurons projecting to the striatum is two orders of magnitude greater than the number of striatal neurons (Kincaid, Zheng, & Wilson, 1998) and an additional reduction of the same magnitude occurs from the striatum to the GPi (Oorschot, 1996; Percheron, Francios, Yelnik, Fenelon, & Talbi, 1994). Although quantitative studies of the neuronal populations at the pallido-thalamic and thalamo-cortical levels are still lacking, most anatomical studies indicate that the cortico-striato-pallido-thalamo-cortical pathway gradually expands after the pallidal level (Arecchi-Bouchhioua, Yelnik, Francios, Percheron, & Tande, 1996; Sidibe, Bevan, Bolam, & Smith, 1997).
2. *The lack of electrophysiological evidence for mutual inhibition between striatal neurons* (Jaeger, Kita, & Wilson, 1994), in spite of the anatomical evidence for extensive lateral connectivity in the striatum (Kita, 1996; Yelnik, Francios, & Tand, 1997).

A possible solution explaining this apparent discrepancy between anatomical and physiological data and the funneling structure along the cortico-basal ganglia–thalamo-cortical loop is the hypothesis that the basal ganglia perform efficient dimensionality reduction of cortical activity. The term 'dimensionality reduction' describes the process of projecting inputs from a high dimensional space to a considerably smaller one. Efficient reduction is achieved when all or most of the information contained within the original space is preserved.

An important assumption of Bar-Gad et al.'s model is that dimensionality reduction in a behaving animal should be affected not only by the statistical properties of the input patterns but also by their behavioral significance. The relative significance of an input is determined by its novelty (Redgrave, Prescott, & Gurney, 1999), incentive salience (Berridge & Robinson, 1998), and ability to predict reward (Robbins & Everitt, 1996). Suri and Schultz's paper in this issue reviews the large amount of evidence gathered in recent years showing that such signals are coded by DA neurons, and can reach the striatum by way of its DA input (as described by Kotter et al. in this issue).

Theoretical studies have already shown that neural networks can perform efficient dimensionality reduction using competitive Hebbian learning rules for inter-layer connectivity (Oja, 1982) and anti-Hebbian rules for the lateral inhibitory intra-layer connectivity (Foldiak, 1989; Kung & Diamantars, 1990). Obviously, these networks typically have a funneled structure. To examine the RDDR hypothesis, Bar-Gad et al. studied a simulated feed-forward neural network, which extracted a principal component subspace using lateral inhibition (Foldiak, 1989; Kung & Diamantars, 1990). This network was comprised of three layers: the first layer represented the cortical input, the intermediate layer represented the striatum, and the output layer represented the GPi. Learning was Hebbian for the feed-forward weights and anti-Hebbian for the lateral weights. A reinforcement signal was combined with the feed-forward input at the intermediate layer to create a three-factor Hebbian learning rule, crudely modeling dopaminergic neuromodulation of corticostriatal synapses. The reinforcement signal was positive for reward-related events and zero for non-reward-related events (baseline DA levels), allocating more encoding resources for rewarded stimuli compared with non-rewarded ones. The network weights were constrained to either positive or negative values to reflect the known neurotransmitter physiology. To measure the information loss of the network due to the RDDR process, the output layer was expanded back to an input-size space, reconstructing the decompressed patterns.

These simulations showed that attributing a larger than baseline reinforcement signal during learning to a selected subset of the 'meaningful' patterns indeed results in discriminative information extraction, providing better reconstructions for the selected, reward-enhanced inputs than for the baseline set of stimuli. Bar-Gad et al. demonstrated that a two-fold increase in the reinforcement signal value versus baseline levels caused an almost five-fold decrease in the compression reconstruction error.

Presenting the network with novel input patterns results in correlated activity of the output neurons. This correlation causes a transient increase in the efficacies of the inhibitory lateral synapses and transient changes in the efficacies of the feed-forward connections. These changes, in turn, lead to decorrelation of neuronal activity within the output layer and to an improvement in information compression. The transient nature of these synaptic alterations explains on the one hand why intra-layer synapses are important for the encoding process, but on the other hand, that at the end of the process they may obtain almost vanishing values. Thus, the learning dynamics of these networks provide a possible explanation to the seeming discrepancy between the anatomical and physiological data pertaining to striatal lateral inhibitory connectivity. These findings suggest that the weak functionality of striatal intra-connectivity and the low correlations of striatal and pallidal neurons' firing can be explained by noting that most of the experiments which obtained these findings were performed in animals which were not actively engaged in learning new behavioral tasks.

To experimentally test this prediction, Bar-Gad et al. have trained a monkey to perform a key pressing task and recorded its pallidal activity during task performance, calculating the correlation coefficients for 151 pairs of pallidal neurons (Bar-Gad et al., 2000). The correlation coefficients were low during performance of a known task leading to an expected reward and during rest periods. A dramatic increase in absolute correlation values was observed following unexpected rewards, following cessation of reward for earlier rewarded actions and following performance of untrained rewarded actions. The periods of enhanced correlation were prolonged and lasted for several tens of seconds. The findings of high correlations during learning, rule out the possibility that the lack of correlated activity found otherwise in striatal and pallidal firing is simply the result of sparse cortico–striatal connectivity. These decreased correlations rather suggest an active decorrelating process.

The RDDR model suggests that the basal ganglia play a role in extraction and pre-processing of information from the whole cortex. Why is it computationally useful? First, it allows for the transmission of large amounts of information within a limited number of axons. Bar-Gad et al. hypothesize that the basal ganglia perform dimensionality reduction of widespread cortical neural activity representing the present state of the animal. The reduced information is projected to the frontal cortex that uses it for planning future actions. The RDDR network thus enables the exposure of neurons in the executive regions of the frontal cortex to maximal incoming cortical information using the anatomically limited number of synapses that each frontal neuron can receive. Second, the RDDR network provides a vehicle by which RL may be carried out in the brain in a central, parsimonious location, by allowing the appetitive value of stimuli to guide their storage and representation. Such selective RDDR storage tends to bias the overall network's response towards rewarded input stimuli. As noted already by Houk et al. (1995), such a biased signaling of complex contexts could be useful in the formulation and implementation of plans and actions. Furthermore, part of the cortical input to the basal ganglia arises from the frontal cortex, and probably represents plans and actions. It is therefore possible that the basal ganglia output acts to bias the representation at the level of the frontal cortex towards the selection of rewarded plans and actions. We thus suggest that the RDDR framework may serve as a basis for basal ganglia actor models.

6. The dual role of the DA signal in reinforcement learning and behavioral switching

Throughout this paper we have related to the DA response to rewards and reward-predicting stimuli as providing a reinforcement signal. This hypothesis is a refinement of the view that DA plays a central role in learning (Le Moal & Simon, 1991; Robbins & Everitt, 1996; White, 1997). An additional central function attributed to the DA system is switching between different behaviors (Le Moal & Simon, 1991; Lyons & Robbins, 1975; Oades, 1985; Robbins & Everitt, 1982; Van den Bos & Cools, 1989; Weiner, 1990). Recently, Redgrave et al. (1999) pointed out that rewarding stimuli serve not only to reinforce the behavior that preceded them, but also to interrupt that behavior and initiate a different behavior (e.g. switching from lever-pressing to approaching the food magazine following reward-delivery). Based on this observation these authors suggested that the short-latency DA response to rewards and reward-predicting stimuli subserves switching rather than learning.

In contrast, based on the dual function of conditioned stimuli in reinforcement and switching, Weiner and Joel (2002) suggested that the phasic response of DA neurons is involved in both learning and switching. They further suggested that these two functions are sub-served, respectively, by the long-term and transient effects of a phasic increase in striatal DA on corticostriatal synaptic transmission. Thus, RL is sub-served by the DA-dependent strengthening of corticostriatal synapses of striatal neurons that were active prior to the increase in DA, and behavioral switching is sub-served by DA-mediated facilitation and attenuation of corticostriatal transmission, which facilitate a change in striatal activity from the set of neurons that had been active to a different set (see Weiner & Joel, 2002, for elaboration of the cellular mechanisms which may underlie these effects). Some support for this hypothesis can be found in the results of Suri et al.'s (2001) simulations, although their work did not directly relate to the issue of behavioral switching. Suri et al.'s (2001) model incorporated both long-term and transient effects of DA on striatal neurons. As may be expected, these authors found that the former is necessary for RL. Suri et al. have also found, that in their model, a phasic increase in DA leads to increased behavioral output, and that this effect is mediated by DA's transient effects on striatal firing.

In the context of the dual role of rewarding events, namely, directing learning and facilitating behavioral switching, we would like to point out that during the course of learning, conditioned stimuli lose the former role, but not the latter. Thus, as learning progresses, each conditioned stimulus becomes predicted by preceding stimuli and actions, and therefore loses its ability to induce a phasic DA response and thus its ability to support learning. However, during the learning process, each conditioned stimulus becomes the elicitor of the next action in the goal-directed behavior, as a result of reinforcement-driven stimulus-response learning. Consequently, during the execution of a learned sequence of actions, each action results in the occurrence of a conditioned stimulus, which in turn elicits the following action in the sequence.

It follows that conditioned stimuli may elicit switching via at least two different mechanisms. One mechanism depends on a phasic increase in striatal DA, and is characteristic of novel situations and of the early stages of learning. This mechanism either increases the likelihood of switching in general, or favors switching to one of the class of behaviors (mostly innate) that are characteristic of novel situation (e.g. orienting). Another mechanism depends on the strengthening of corticostriatal synapses, and is characteristic of well-learned behaviors. This mechanism is responsible for the termination of the current behavior and the initiation of the subsequent behavior, which is specific and learned (Weiner & Joel, 2002). Although this latter type of switching occurs in the absence of a phasic increase in striatal DA, baseline DA levels are thought to sub-serve an important permissive role in movement initiation (Le Moal & Simon, 1991; Robbins & Everitt, 1996; Salamone, 1994). We have recently obtained evidence in rats suggesting that DA also modulates the ability of conditioned stimuli to terminate the preceding behavior (Joel, Avisar, & Doljansky, 2001).

None of the models reviewed above simulates the two types of switching. However, a demonstration of the gradual acquisition and loss of the ability to elicit a DA signal, concomitantly with the acquisition of the ability to elicit 'phasic DA-independent' switching, can be found in the simulations of Suri and Schultz (1998). In their simulations of the acquisition of sequential movements by an actor–critic model, a reward occurred at the end of a correctly performed sequence of stimulus-action pairs. During acquisition of the task, each of the different stimuli gradually acquired the ability to elicit a DA signal and to trigger the correct action. As training progressed, the stimulus became predicted by earlier stimuli, and as a result stopped eliciting the DA signal. However, as a result of learning in the actor, each stimulus continued to trigger the correct action. Thus, following learning, the presentation of a stimulus resulted in the elicitation of the correct action without an increase in DA.

7. Conclusions

Our selective review of actor–critic models of the basal ganglia raises several issues which we believe future models will have to deal with. Models of the critic build on the strong resemblance between DA neuron activity and the TD prediction error signal in the critic. From a computational perspective, these models face two related challenges: One, how to reproduce the specific temporal dynamics of DA firing to rewards, reward-predicting stimuli, and novelty. Two, what are the computational consequences of incorporating DA responses to novelty, generalization and discrimination into a TD RL algorithm.

From an anatomical-physiological perspective it is clear that a critic model which builds on reciprocal connections between DA neurons and another group of neurons, cannot be implemented in the connections between the DA system and the striatum, because these connections are characterized by asymmetry rather than reciprocity. Similarly, a critic which is based on Barto's (1995) architecture, cannot be implemented in these connections, because it is unlikely that a single group of striatal neurons is the source of both indirect fast excitation and direct delayed inhibition to the DA neurons, as required by such models of the critic. One potentially fruitful approach to these quandaries is to harness the power of evolutionary computation techniques to find candidate solution architectures that maximize critic functionality under various anatomical and functional constraints, and then examine these predictions experimentally. The work of Niv et al. (2002, in press) is a first step in this direction. Future models of the critic would have to deal with these problems, and in addition should relate to the question of whether a single projection to the DA system (e.g. from the basal ganglia) is responsible for DA neurons' responses to both rewarding and novel stimuli, or whether these responses are sub-served by different projections (as suggested by Contreras-Vidal & Schultz, 1999).

Models of the actor build on the strong resemblance between DA-dependent long-term synaptic plasticity in the striatum and learning guided by a prediction error signal in the actor. Current models of the actor, however, are very simple and are usually modeled at the striatal level with very little detail. The goal of future studies is to model the known anatomy and physiology of the basal ganglia in a more detailed and faithful manner, and address the question of the computational role of the basal ganglia–thalamocortical connections. There are currently several different neural-network models of these connections that provide different answers to these questions (Berns & Sejnowski, 1998; Gurney et al., 2001). We have described a model of the basal ganglia–thalamocortical connections which suggests that the basal ganglia perform reinforcement-biased dimensionality reduction of cortical inputs (Bar-Gad et al., 2000). This RDDR framework may serve as a basis for future basal ganglia actor models.

In summary, actor–critic models of the basal ganglia have contributed to our thinking on basal ganglia functioning, by integrating some of the central aspects of basal ganglia processing (the DA signal, DA-dependent learning in the striatum) with learning theory. Yet, numerous questions, regarding the function of these nuclei as well as the theoretical aspects of RL, are left unanswered. It is our hope that future models incorporating actor and critic components that are more constrained by the known anatomy and physiology of the basal ganglia will answer some of these questions.

References

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*, 266–271.

Arecchi-Bouchhioua, P., Yelnik, J., Francois, C., Percheron, G., & Tande, D. (1996). 3-D tracing of biocytin-labelled pallido-thalamic axons in the monkey. *Neuroreport*, *7*, 981–984.

Bar-Gad, I., Havazelet-Heimer, G., Ruppin, E., & Bergman, H. (2000). Reinforcement driven dimensionality reductions; a model for information processing in the basal ganglia. *Journal of Basic and Clinical Physiological and Pharmacology*, *11*, 305–320.

Bailey, C. H., Giustetto, M., Huang, Y., Hawkins, R. D., & Kandel, E. R. (2000). Is heterosynaptic modulation essential for stabilizing hebbian plasticity and memory? *Nature Reviews Neuroscience*, *1*, 11–20.

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

Berendse, H. W., Galis-de Graaf, Y., & Groenewegen, H. J. (1992). Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *Journal of Comparative Neurology*, *316*, 314–347.

Berns, G. S., & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience*, *10*, 108–121.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Review*, *28*, 309–369.

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience*, *19*, 10502–10511.

Bunney, B. S., Chiodo, L. A., & Grace, A. A. (1991). Midbrain dopamine system electrophysiological functioning: A review and new hypothesis. *Synapse*, *9*, 79–94.

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., Svenningsson, P., Fienberg, A. A., & Greengard, P. (2000). Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *Journal of Neuroscience*, *20*, 8443–8451.

Carpenter, G. A., & Grossberg, S. (1987). Self organization of stable category recognition codes for analog input patterns. *Applied Optics*, *3*, 4919–4930.

Contreras-Vidal, J. L., & Schultz, W. (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. *Journal of Comparative Neuroscience*, *6*, 191–214.

Dittman, J. S., & Regehr, W. G. (1997). Mechanism and kinetics of heterosynaptic depression at a cerebellar synapse. *Journal of Neuroscience*, *17*, 9048–9059.

Fellous, J.-M., & Linster, C. (1998). Computational models of neuromodulation: A review. *Neural Computation*, *10*, 791–825.

Foldiak, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, *64*, 165–170.

Gerfen, C. R. (1984). The neostriatal mosaic: Compartamentalization of corticostriatal input and striatonigral output systems. *Nature*, *311*, 461–464.

Gerfen, C. R. (1985). The neostriatal mosaic. I. Compartamental organization of projections from the striatum to the substantia nigra in the rat. *Journal of Comparative Neurology*, *236*, 454–476.

Gerfen, C. R. (1992). The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. *Annual Review of Neuroscience*, *15*, 285–320.

Gerfen, C. R., Herkenham, M., & Thibault, J. (1987). The neostriatal mosaic II. Patch- and matrix- directed mesostriatal dopaminergic and non-dopaminergic systems. *Journal of Neuroscience*, *7*, 3915–3934.

Gillies, A., & Arbuthnott, G. (2000). Computational models of the basal ganglia. *Movement Disorders*, *15*, 762–770.

Groenewegen, H. J., Berendse, H. W., Wolters, J. G., & Lohman, A. H. M. (1990). The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. *Progress in Brain Research*, *85*, 95–118.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, *84*, 401–410.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience*, *20*, 2369–2382.

Hammer, M. (1997). The neural basis of associative reward learning in honeybees. *Trends in Neuroscience*, *20*, 245–252.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use reward signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge: MIT Press.

Jaeger, D., Kita, H., & Wilson, C. J. (1994). *Journal of Neurophysiology*, *72*, 2555–2558.

Joel, D., & Weiner, I. (1994). The organization of the basal ganglia–thalamocortical circuits: Open interconnected rather than closed segregated. *Neuroscience*, *63*, 363–379.

Joel, D., & Weiner, I. (1997). The connections of the primate subthalamic nucleus: Indirect pathways and the open-interconnected scheme of basal ganglia–thalamocortical circuitry. *Brain Research Review*, *23*, 62–78.

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: An analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, *96*, 451–474.

Joel, D., Avisar, A., & Doljansky, J. (2001). Enhancement of excessive lever-pressing after post-training signal attenuation in rats by repeated administration of the D1 antagonist SCH 23390 or the D2 agonist quinpirole but not of the D1 agonist SKF 38393 or the D2 antagonist haloperidol. *Behavioural Neuroscience*, *115*, 1291–1300.

Kaelbling, L. P., Littman, M. L., & Moore, A. (1996). Reinforcement learning: A survey. *Journal of AI Research*, *4*, 237–285.

Kalivas, P. W. (1993). Neurotransmitter regulation of dopamine neurons in the ventral tegmental area. *Brain Research Review*, *18*, 75–113.

Kincaid, A. E., Zheng, T., & Wilson, C. J. (1998). Connectivity and convergence of single corticostriatal axons. *Journal of Neuroscience*, *18*, 4722–4731.

Kita, H. (1996). In C. Ohye, M. Kimura, & J. S. McKenzie (Eds.), *The basal ganglia V* (pp. 77–94). New York: Plenum Press.

Kung, S. Y., Diamantars, K. I (1990). IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 2, pp. 861–864).

Le Moal, M., & Simon, H. (1991). Mesocorticolimbic dopaminergic network: Functional and regulatory roles. *Physiological Review*, *71*, 155–234.

Lyon, M., & Robbins, T. W. (1975). The action of central nervous system stimulant drugs: A general theory concerning amphetamine effects (*Vol. 2*) (pp. 80–163). *Current developments in Psychopharmacology*, New York: Spectrum.

Menzel, R., & Muller, U. (1996). Learning and memory in honeybees: From behavior to neural substrates. *Annual Review of Neuroscience*, *19*, 379–404.

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive Hebbbian learning. *Nature*, *377*, 725–728.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framewok for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Niv, Y., Joel, D., Meilijson, I., & Ruppin, E. (2002). Evolution of reinforcement learning in uncertain environments: a simple explanation for complex foraging behaviors. *Adaptive Behavior*, in press.

Oades, R. D. (1985). The role of noradrenaline in tuning and dopamine in switching between signals in the CNS. *Neuroscience Biobehavioural Review*, *9*, 261–282.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, *15*, 267–273.

Oorschot, D. E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: A stereological study using the cavalieri and optical disector methods. *Journal of Comparative Neurology*, *366*, 580–599.

Overton, P. G., & Clark, D. (1997). Burst firing in midbrain dopaminergic neurons. *Brain Research Review*, *25*, 312–334.

Parent, A. (1990). Extrinsic connections of the basal ganglia. *Trends in Neuroscience*, *13*, 254–258.

Percheron, G., Francois, C., Yelnik, J., Fenelon, G., & Talbi, B. (1994). The basal ganglia related systems of primates: definition, description and informational analysis. In G. Percheron, J. S. McKenzie, & J. Feger (Eds.), *The basal ganglia IV: New ideas and data on structure and function* (pp. 3–20). New York: Plenum Press.

Pucak, M. L., & Grace, A. A. (1994). Regulation of substantia nigra dopamine neurons. *Critical Review Neurobiology*, *9*, 67–89.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends in Neuroscience*, *22*, 146–151.

Robbins, T. W., & Everitt, B. J. (1982). Functional studies of the central catecholamines. *International Review of Neurobiology*, *23*, 303–365.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology*, *6*, 228–236.

Rolls, E. T., & Johnstone, S. (1992). Neurophysiological analysis of striatal function. In C. Wallesch, & G. Vallar (Eds.), *Neuropsychological disorders with subcortical lesions* (pp. 61–97). Oxford: University Press.

Salamone, J. D. (1994). The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behavioural Brain Research*, *61*, 117–133.

Schacher, S., Wu, F., & Sun, Z.-Y. (1997). Pathway-specific synaptic plasticity: Activity-dependent enhancement and suppression of long-term heterosynaptic facilitation at converging inputs on a single target. *Journal of Neuroscience*, *17*, 597–606.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1–27.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review Neuroscience*, *23*, 473–500.

Schultz, W., Apiccela, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *Journal of Neuroscience*, *12*, 4595–4610.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Schultz, W., Tremblay, L., & Hollerman, J. R. (1998). Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology*, *37*, 421–429.

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereberal Cortex*, *10*, 272–283.

Sidibe, M., Bevan, M. D., Bolam, J. P., & Smith, Y. (1997). Efferent connections of the internal globus pallidus in the squirrel monkey. 1. Topography and synaptic organization of the pallidothalamic projection. *Journal of Comparative Neurology*, *382*, 323–347.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Networks*, *15*, PII: S0893-6080(02)00046-1.

Suri, R. E., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research*, *121*, 350–354.

Suri, R. E., & Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience*, *91*, 871–890.

Suri, R. E., Bargas, J., & Arbib, M. A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience*, *103*, 65–85.

Sutton, R. (1988). Learning to predict by methods of temporal difference. *Machine Learning*, *3*, 9–44.

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM*, *38*, 58–68.

Uylings, H. B. M., & van Eden, C. G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Progress in Brain Research*, *85*, 31–62.

Van den Bos, R., & Cools, A. R. (1989). The involvement of the nucleus accumbens in the ability of rats to switch to cue-directed behaviors. *Life Science*, *44*, 1697–1704.

Vogt, K. E., & Nicoll, R. E. (1999). Glutamate and gama-aminobutyric acid mediate a heterosynaptic depression at mossy fober synapses in the hippocampus. *Proceedings of the National Academic Science, USA*, *96*, 1118–1122.

Weiner, I. (1990). Neural substrates of latent inhibition: The switching model. *Psychological Bulletin*, *108*, 442–461.

Weiner, I., & Joel, D. (2002). Dopamine in schizophrenia: Dysfunctional information processing in basal ganglia–thalamocortical split circuits. In G. Di Chiara (Ed.), *Handbook of experimental pharmacology: Dopamine in the CNS*, (pp. 417–472). Berlin: Springer.

White, N. M. (1997). Mnemonic functions of the basal ganglia. *Current Opinions in Neurobiology*, *7*, 164–169.

Wickens, J. R., Begg, A. J., & Arbuthnott, G. W. (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience*, *70*, 1–5.

Wickens, J., & Kotter, R. (1995). Cellular models of reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia*, (pp. 187–214). Cambridge, MA: MIT Press.

Yelnik, J., Francois, C., Tand, D (1997). Proceedings if the Third Congress of European Neuroscience Society, Beaurdeax.

Yeterian, E. H., & Pandya, D. N. (1991). Prefrontostriatal connections in relation to cortical architectonic organization in rhesus monkeys. *Journal of Comparative Neurology*, *312*, 43–67.

Zald, D. H., & Kim, S. W. (2001). The orbitofrontal cortex. In S. P. Salloway, P. F. Malloy, & J. D. Duffy (Eds.), *The frontal lobes and neuropsychiatric illness* (pp. 33–69). Washington, DC: American Psychiatric Publishing.

Zhang, W., & Dietterich, T. G. (1996). High performance job shop scheduling with a time delay TD network. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), (*Vol. 8*) (pp. 1024–1030). *Advances in neural information processing systems*, Cambridge: MIT Press.