Exp Brain Res (2010) 200:307–317
DOI 10.1007/s00221-009-2060-6

REVIEW

Striatal action-learning based on dopamine concentration

Genela Morris · Robert Schmidt · Hagai Bergman

Received: 17 July 2009 / Accepted: 8 October 2009 / Published online: 11 November 2009

Abstract The reinforcement learning hypothesis of dopamine function predicts that dopamine acts as a teaching signal by governing synaptic plasticity in the striatum. Induced changes in synaptic strength enable the cortico-striatal network to learn a mapping between situations and actions that lead to a reward. A review of the relevant neurophysiology of dopamine function in the cortico-striatal network and the machine reinforcement learning hypothesis reveals an apparent mismatch with recent electrophysiological studies. It was found that in addition to the well-described reward-related responses, a subpopulation of dopamine neurons also exhibits phasic responses to aversive stimuli or to cues predicting aversive stimuli. Obviously, actions that lead to aversive events should not be reinforced. However, published data suggest that the phasic responses of dopamine neurons to reward-related stimuli have a higher firing rate and have a longer duration than phasic responses of dopamine neurons to aversion-related stimuli. We propose that based on different dopamine concentrations, the target structures are able to decode reward-related dopamine from aversion-related dopamine responses. Thereby, the learning of actions in the basal-ganglia network integrates information about both costs and benefits. This hypothesis predicts that dopamine concentration should be a crucial parameter for plasticity rules at cortico-striatal synapses. Recent in vitro studies on cortico-striatal synaptic plasticity rules support a striatal action-learning scheme where during reward-related dopamine release dopamine-dependent forms of synaptic plasticity occur, while during aversion-related dopamine release the dopamine concentration only allows dopamine-independent forms of synaptic plasticity to occur.

Keywords Basal ganglia · Dopamine · Learning · Action value · Reinforcement learning

G. Morris
Department of Neurobiology and Ethology,
Haifa University, Haifa, Israel

G. Morris · R. Schmidt
Bernstein Center for Computational Neuroscience,
Berlin, Germany

R. Schmidt
Department of Biology, Institute for Theoretical Biology,
Humboldt-Universität zu Berlin, Invalidenstr. 43,
10115 Berlin, Germany

H. Bergman (✉)
Department of Physiology, Hadassah Medical School,
Hebrew University, Jerusalem, Israel
e-mail: hagaibe@ekmd.huji.ac.il

Dopamine in the striatum

The study of the basal-ganglia complex, and of dopamine function in particular, has traditionally been approached from two directions. On one hand, the ventral school, primarily interested in drug addiction and in psychotic disorders, has focused their research on the nucleus accumbens (in the ventral striatum) and its projections, along with its dopamine input structure, the ventral tegmental area (VTA, A10) (Kelley et al. 1982; Bonci and Malenka 1999; Thomas and Malenka 2003; Di Chiara et al. 2004; Cardinal et al. 2002; Arroyo et al. 1998; Ito et al. 2004; Voorn et al. 2004; Kelley 2004; Di Chiara and Bassareo 2007; Dalley et al. 2007; Wheeler and Carelli 2009). On the other hand, the dorsal school, originally occupied with movement disorders, concentrated on the dorsal striatum (caudate and putamen nuclei), with their corresponding dopamine

source—the substanitia nigra pars compacta (SNc, A9) (DeLong and Georgopoulos 1981; Schultz 1982; Schultz et al. 1985; Bergman et al. 1990; Alexander et al. 1990; Schultz 1994). This dissociation was paralleled with the choice of animals to study the structures. Since the motor functions of rodents are not easily quantifiable, the dorsal school quickly converged to primate research, while the ventral school's model animal of choice has been the rat. Although the research areas have now converged, this historical segregation in laboratory animals impedes systematic comparison of experimental results collected from the ventral dopaminergic structure of the VTA and the dorsal structure of the SNc. Still, a growing body of evidence suggests a large similarity between the ventral and dorsal aspects of the basal ganglia, indicating that information processing is similar in both parts. Functional differences probably do not arise from different processing algorithms but instead are due to differences in input and output connectivity (i.e., in the type of information they process) along the dorsal–ventral gradient (Wickens et al. 2007).

The last two decades in striatum and dopamine research have witnessed an abandonment of old controversies in favor of a relative consensus on the view of the role of basal ganglia. In particular, this applies to the role of dopamine in the input structure of the basal ganglia, the striatum. In the 1980s and the first half of the 1990s, the ventral school discussed the hedonic value of dopamine (Berridge 1996; Royall and Klemm 1981; Wise 2008). Anatomical and physiological studies argued whether information processing in the basal ganglia was comprised of parallel or converging circuits (Alexander et al. 1986; Percheron and Filion 1991). The study of motor control and movement disorders focused on the basal ganglia involvement in action initiation versus action selection (Mink 1996). Nowadays, most discourse is comfortable with the notion that the basal ganglia are involved in the mapping of situations (or states) to actions; that dopamine (and other basal-ganglia neuromodulators) plays a major role in learning this mapping, and that partially overlapping circuits, with substantial convergence within each, constitute different facets of the same basic computation. Recent research has shed new light on different aspects of this picture, calling for refining the prevailing theory. In this review, we present the dominant theories of the basal ganglia and dopamine in light of the new perspective offered by the new data.

The striatum serves as an input structure of the basal ganglia (Fig. 1), a group of nuclei which forms a closed loop with the cortex, and which has been implicated with motor, cognitive and limbic roles (Haber et al. 2000). A large majority (90–95%) of neurons in the striatum are medium spiny projection neurons (MSNs). These neurons receive excitatory glutamatergic input from the cortex and the thalamus and project to the globus pallidus (internal and

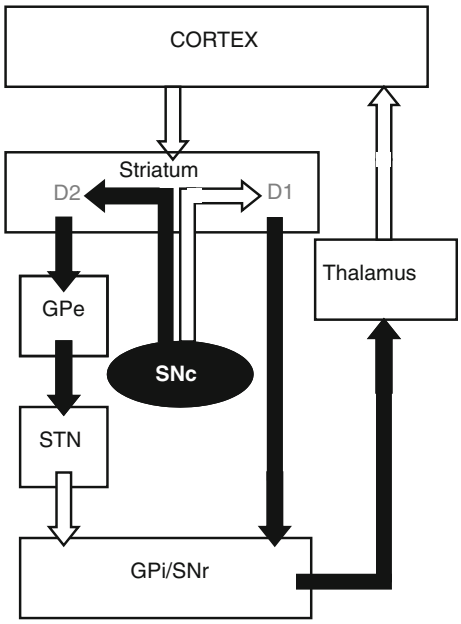

Fig. 1 Schematic view of the connectivity of the two pathway model of the cortex–basal-ganglia network. Direct/Go/D1 pathway is depicted on the *left-hand* side and the indirect/No-Go/D2 pathway on the *right*. *White arrows* indicate excitatory connections, and *black arrows* denote inhibitory connections. *STN* subthalamic nucleus, *GPe,GPi* external, internal segment of the globus pallidus, *SNc SNr* substantia nigra pars compacta and reticulata, respectively

external segments, GPi and GPe, respectively), the substantia nigra (pars reticulata, SNr, and pars compacta, SNc) with inhibitory connections. In fact, with the exception of the projections from the subthalamic nucleus (STN), all the main projections in the basal-ganglia network use the inhibitory neurotransmitter gamma-aminobutyric acid (GABA). The projections from the striatum are classically divided into two pathways (Fig. 1), each of which exerting opposite net effects on the target thalamic (and thus cortical) structures (Albin et al. 1989; Alexander and Crutcher 1990; Gerfen 1992). While activation in the direct pathway results in a net thalamic excitation through dis-inhibition, indirect pathway activation results in net thalamic inhibition through triple inhibition (plus STN excitation). In light of the thalamic and motor control of action, the direct and indirect pathways have been conveniently described as the Go and No-Go pathways, respectively (Frank et al. 2004). Although the two pathways might not be completely segregated (Smith et al. 1998), they do differ in a number of unique biochemical properties (Nicola et al. 2000). Most importantly, the dopaminergic input onto the striatum affects the MSNs of the Go and No-Go pathways differently due to differential expression of dopamine receptors. MSNs of the Go/Direct pathways are equipped with D1 type dopamine receptors, while those of the No-Go/Indirect pathway express D2 type dopamine receptors. Additionally, Go MSNs secrete substance P in addition to GABA and No-Go

MSNs produce enkephalin as well as GABA, and uniquely express A_{2a} type adenosine receptors.

An important feature of basal-ganglia anatomy is the high concentration of neuromodulators in the striatum. Both dorsal (caudate and putamen) and ventral (nucleus accumbens) nuclei of the striatum show the highest density of brain markers for dopamine (Bjorklund and Lindvall 1984; Lavoie et al. 1989) and acetylcholine (Woolf 1991; Holt et al. 1997; Descarries et al. 1997; Zhou et al. 2001, 2003), as well as a high degree of 5-HT immunoreactivity indicating serotoninergic innervation (Lavoie and Parent 1990).

The midbrain dopamine system consists of neurons located in the VTA and the SNc, projecting mainly to the ventral and dorsal striatum, respectively. A third pathway, from the VTA to other frontal targets, is less pronounced, and probably important for other behaviors and pathologies. Furthermore, the synaptic anatomy of the glutamatergic and dopaminergic inputs in the striatum is of vast importance. It has been found that a majority of the glutamatergic cortico-striatal and thalamostriatal synapses are in the functional proximity of dopaminergic innervation (Moss and Bolam 2008).

Functional roles of dopamine

Pioneering physiological self-stimulation studies of the neural correlates of pleasure, motivation and reward centers have identified the brain regions mediating the sensation of pleasure and behavior oriented toward it (Olds and Milner 1954). The structures involved were believed to be the lateral septum, lateral hypothalamus, its connections to the midbrain areas of the tegmentum, as well as the tegmentum itself and its projection to the forebrain via the medial forebrain bundle (MFB). It is now commonly accepted that the optimal region for self-stimulation is the MFB, also known as the meso-limbic pathway, carrying dopamine from the VTA to the ventral striatum or NAc.

Early hypotheses on dopamine function proposed that dopamine signals pleasure or hedonia (Wise 1996). This view is now commonly rejected, giving rise to two lines of thought. The first assigns dopamine with a general function in behavioral arousal, motivation and effort allocation (Salamone et al. 2007). Conversely, the second group of theories argues for a more specific role of dopamine in reward-related behavior (Schultz 2002; Berridge 2007; Redgrave et al. 2008). Among the latter, a further division can be drawn between groups claiming that dopamine plays a causal role in learning (Schultz 2002; Redgrave et al. 2008), and those that assume a reversed causality, according to which the dopamine signal results from learning and is used to guide behavior (Berridge 2007).

There are two major hypotheses for dopamine that propose a causal role for dopamine in learning. The first one can be referred to as 'prediction-error' hypothesis (Schultz 2002; Schultz et al. 1997). It receives support from electrophysiological recordings of dopamine neurons of animals performing reward-related learning tasks. In these experiments it was repeatedly shown that the activity of dopamine neurons exhibits striking resemblance to a teaching signal commonly employed in the machine learning field of reinforcement learning (Sutton and Barto 1998) (see below for further details). These results have led to a model in which the signal emitted by dopamine neurons plays a causal role in reward-related learning, causing reinforcement of actions that lead to the reward. This learning will result in a tendency to repeat rewarded actions. A more recent hypothesis for dopamine in learning (Redgrave et al. 2008) proposes that dopamine is important for learning the association between action-outcome pairs. It is assumed that salient sensory events evoke dopamine responses. This signal is then used to reinforce all actions that preceded the salient event, thereby assisting in identification of the action that caused the event. As a result, the animal learns the consequence of certain behaviors on the environment.

The hypotheses on dopamine playing a causal role in learning have been challenged by Berridge and colleagues (Berridge and Robinson 1998; Berridge 2007). Instead, they propose an alternative theory termed the incentive salience hypothesis of reward dopamine. According to this theory, the similarity of the dopamine signal to a prediction error is relayed from one of its input structures, reflecting learning in upstream neural circuits. Rather, dopamine guides behavior by tagging a particular stimulus as 'wanted' and directing behavior toward it. Thereby dopamine is essential for the expression of learning but not for the learning itself.

Recent findings on dopamine released by aversive stimuli (Joshua et al. 2008; Brischoux et al. 2009; Matsumoto and Hikosaka 2009) challenge current views on dopamine function. In the following we review the reinforcement learning hypothesis of dopamine in more detail and discuss how it can be reconciled to accommodate recent findings on dopamine activity.

Dopamine and reinforcement learning

Despite differences in theories, behavioral psychologists (Thorndike 1911; Pavlov 1927; Skinner 1974) claim that the following basic rule gives a sufficient account for learning: behavior is followed by a consequence, and the nature of the consequence determines the tendency to repeat the same behavior in the future. This rule is best known by its formulation by Thorndike (1898), later coined as *Thorndike's*

law of effect, which reads as follows: "The Law of Effect is that: Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur" (Thorndike 1911). This definition sets the basis for reinforcement learning in the field of psychology, which has subsequently lent its name to the field of machine learning.

Reinforcement learning is situated at an intermediate step between supervised and unsupervised forms of machine learning. In reinforcement learning the learning agent receives limited feedback in the form of rewards and punishments. This feedback is used by the agent to learn to choose the best action in a given situation so that the overall cumulative reward is maximized. Punishments are usually implemented simply as negative rewards. Theorists in the field of artificial intelligence have studied this type of learning intensively. They have developed powerful reinforcement learning algorithms such as temporal-difference (TD) learning (Sutton 1988; Sutton and Barto 1998), which overcomes the major difficulties of learning through unspecific feedback. In this method of learning at each point in time the value (expected reward) at the next point in time is estimated. When external reward is delivered, it is translated into an internal signal indicating whether the value of the current state is better or worse than predicted. This signal is called the *TD error*, and it serves to improve reward predictions and reinforce (or extinguish) particular behaviors.

Physiological and psychological studies have revealed that dopamine plays a crucial role in the control of motivation and learning. Dopaminergic deficits have been shown to disrupt reward-related procedural learning processes (Knowlton et al. 1996; Matsumoto et al. 1999). Insight into the involvement of striatal dopamine release in learning is obtained from the analogy with the TD reinforcement learning algorithm. When presented with an unpredicted reward or with stimuli that predict reward, midbrain dopaminergic neurons display stereotypical responses consisting of a phasic elevation in their firing rate (Schultz et al. 1997; Hollerman and Schultz 1998; Waelti et al. 2001; Kawagoe et al. 2004; Morris et al. 2004; Bayer and Glimcher 2005). Congruent with the TD-learning model we describe next, this response typically shifts to the earliest reward-predicting stimulus (Hollerman and Schultz 1998; Pan et al. 2005).

Temporal-difference learning

The first objective of a reinforcement learning algorithm is to estimate a value function that describes future rewards based on the current state. In the terms of classical conditioning, the relevant information in the states is called the conditioned stimulus (CS), whereas the reward is the unconditioned stimulus (US). The reinforcement learning algorithm must learn to predict upcoming reward based on the state. A very influential approach to this problem was proposed by Rescorla and Wagner (1972). There, learning is induced by the discrepancy between what is predicted and what actually happens. However, this account does not model time within a trial, thereby neglecting several key aspects of natural learning. For example, reward is often delayed, and might also be separated from the action for which it was rewarded by other, irrelevant actions. This poses the problem of 'temporal credit assignment': what action was crucial to obtain the reward? To address this problem, an extension to the Rescorla–Wagner model was put forth by Sutton (1988), which came to be known as TD learning and has been widely used in modeling behavioral and neural aspects of reward-related learning (Montague et al. 1996; Schultz et al. 1997; O'Doherty et al. 2003; Redish 2004; Nakahara et al. 2004; Seymour et al. 2004; Pan et al. 2008; Pan et al. 2005; Ludvig et al. 2008). This learning algorithm utilizes a form of bootstrapping, in which reward predictions are constantly improved by comparing them to actual rewards (see description in Sutton and Barto 1998). A classical conditioning setting is illustrated in Fig. 2, showing the estimated value function and the TD error in two cases: received reward and omitted reward. When the TD error is different from 0, it is linearly related to the expected reward, and thereby also to the learned state values.

Dopamine and synaptic plasticity

As dopamine neurons respond in a manner that is congruent with the TD prediction error signal, it is often suggested that dopamine serves as a teacher in the cortico-striatal system. Since in the neurophysiology literature, 'learning' is generally translated to synaptic plasticity, 'teaching' is attributed to inducing, or at least modulating, synaptic plasticity. Indeed, the cortico-striatal synapses are known to undergo long-term changes in synaptic efficacy in the form of long-term potentiation (LTP) (Calabresi et al. 1998; Reynolds et al. 2001) and long-term depression (LTD) (Centonze et al. 2001; Kreitzer and Malenka 2005). Recently, it has also been shown that similar to cortical (Markram et al. 1997) and hippocampal synapses (Bi and Poo 1999), long-term plasticity of cortico-striatal synapses follows the rules of spike-timing dependent plasticity (STDP) (Shen et al. 2008; Pawlak and Kerr 2008). Furthermore, it appears that dopamine plays a crucial role in cortico-striatal plasticity (Reynolds et al. 2001; Centonze et al. 2001). Induction of LTP in the cortico-striatal pathway appears to be mediated by activation of dopamine D1/D5 receptors (Kerr and Wickens 2001; Reynolds et al. 2001;

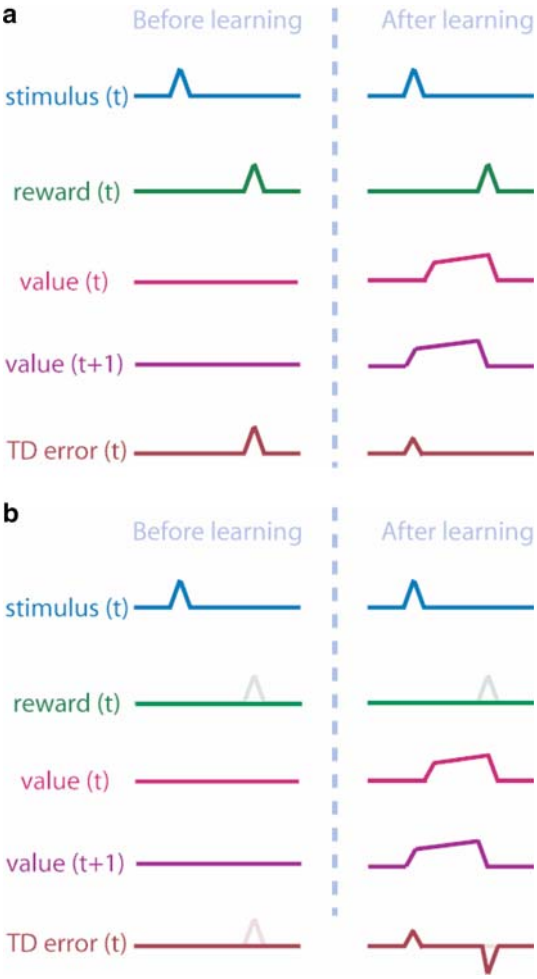

Fig. 2 The TD-learning algorithm. Schematic timeline of TD-learning algorithm in a classical conditioning context. Each *line* represents a different component of the TD error computation. **a** With reward delivery. **b** With omission of predicted reward. *Shaded* expected time of reward

Centonze et al. 2001). LTD is mediated by D2 type receptor activation (Kreitzer and Malenka 2005; Wang et al. 2006; Shen et al. 2008).

In the context of dopamine and synaptic plasticity, there is an interesting connection to drug abuse. Cocaine and amphetamines directly increase the amount of dopamine by inhibiting its reuptake into the synaptic terminals. Opiate narcotics increase dopamine release by disabling tonic inhibition on dopaminergic neurons. Caffeine increases cortical levels of dopamine (Acquas et al. 2002). Nicotine also increases striatal dopamine, probably through the dopamine/ACh interaction (Zhou et al. 2003; Cragg 2006). As addictive drugs increase dopamine levels, the corresponding altered synaptic plasticity might reflect the neural basis for drug addiction.

The reinforcement learning hypothesis of dopamine function relies on a very heavy assumption, namely a dose dependence of the effect of dopamine (and possibly of other neuromodulators) on long-term synaptic plasticity. The notion that dopamine enables reinforcement learning in cortico-striatal synapses through a TD-learning like mechanism received a strong boost from a number of studies showing that the phasic responses of dopamine neurons confirm actual quantitative predictions from the TD model. The TD error signal evoked by unpredicted rewards or reward-predicting stimuli is linearly related to the reward expectancy. This value can be manipulated experimentally, by either systematically changing the size of the reward or its probability of occurrence. Experiments of this nature were performed with primates (Fiorillo et al. 2003; Nakahara et al. 2004; Morris et al. 2004, 2006; Tobler et al. 2005; Bayer and Glimcher 2005), and rats (Roesch et al. 2007). These experiments showed that the phasic positive responses of dopamine neurons exhibit a linear correlation to the state value.

Although an abundance of previous works demonstrated a connection between dopamine and reinforcement of behavior (for review see Wise 2004), their vast majority was oriented toward the topic of drug dependence. Therefore, most studies mainly focused on paradigms such as self-stimulation and self-administration. It was shown that behavior that leads to the increase in meso-limbic dopamine activity is reinforced. This reinforcement is dependent on intact dopaminergic transmission (Cheer et al. 2007; Owesson-White et al. 2008). Other works demonstrated through lesions that dopamine is indeed necessary for learning reward-oriented behaviors (Belin and Everitt 2008; Rassnick et al. 1993; Hand and Franklin 1985). It was also shown that such behavior is paralleled with dopamine-dependent long-term plasticity (Reynolds et al. 2001). In a recent study (Morris et al. 2006), we also showed that the dopamine responses are used to shape the behavior of the monkeys. Finally, a recent ambitious study achieved learning in vivo through optical activation of dopamine neurons in a phasic manner (Tsai et al. 2009). These studies indicate that it is highly likely that these TD error like responses are indeed used in learning.

However, for this to be translated into physiology, the effect on striatal plasticity must also scale with the amount of dopamine released. Therefore, it is essential to perform in vitro experiments in which the dopamine level is dynamically manipulated, on a time-scale which is consistent with phasic activation of dopamine neurons. Recently, this question was addressed by a detailed theoretical study which took into account the dynamics of extracellular dopamine fluctuations (Thivierge et al. 2007). This study predicted that cortico-striatal plasticity depends on the dopamine concentration. Low (non-zero) concentrations caused reverse STDP, while higher concentrations induced regular STDP, the magnitude of which is concentration-dependent. To the best of our knowledge, such an effect has not been

systematically investigated experimentally. Note, however, that concentration dependence must not necessarily occur strictly on single synapse level. Rather, a graded effect on learning could also be achieved through the stochastic nature of binary processes. It may be that increased levels of dopamine enhance the probability of each synapse to undergo a long-term effect, thereby increasing the overall level of potentiation (or depression) in the appropriate circuits.

Striatal decoding of reward and aversive dopamine signaling

Just when the description reward-related activity of dopamine neurons seemed more bullet-proof than ever, several recent studies (Joshua et al. 2008; Brischoux et al. 2009; Matsumoto and Hikosaka 2009) found dopamine neurons that exhibit excitatory responses to aversive stimuli and to cues predicting aversive stimuli. Obviously, similarity in responses of dopamine neurons to aversive and rewarding stimuli poses a serious problem to reinforcement learning accounts of dopamine function (Schmidt et al. 2009). If dopamine acts as a neural reward prediction error signal (Schultz 2002), behavior that leads to punishment should not be reinforced. Alternative theories of dopamine function should encounter similar problems with the new results. Incentive salience accounts for dopamine (Berridge 2007) might have problems explaining why aversive cues are 'wanted'. Similarly, the hypotheses that suggest a role of dopamine for discovering and reinforcing novel actions (Redgrave and Gurney 2006; Redgrave et al. 2008) limit their discussion to rewarding and neutral actions. In fact, only older accounts for dopamine function (Horvitz 2002) that assign dopamine to a general role in motivation and arousal ('activation-sensorimotor hypothesis' (Berridge 2007) appear to be in line with aversive dopamine responses.

Separate aversion- and reward-learning circuits?

Although some of the new studies report a clear dorsal/ventral gradient in the existence of excitatory responses to aversive stimuli, these reports do not seem to be consistent (compare Brischoux et al. 2009 and Matsumoto and Hikosaka 2009). Furthermore, such an anatomical distinction is not likely to be of much impact, as a considerable fraction of projections from each dopaminergic structure diverges to both dorsal and ventral striatal areas (Haber et al. 2000; Matsuda et al. 2009), implying that signals originating at each end of the midbrain dopaminergic area will end up innervating large and spread out regions throughout the striatum. Naively, one might propose a solution to the apparent contradiction between the signals of positive and negative valence involving a hard-wired mapping of dopamine neurons onto subsets of cortico-striatal connections representing different actions. In this case, the dopamine neurons would have to provide differential signals to each target population. However, this does not seem to be in line with the anatomical details of this circuit. Rather, the arborization of dopamine neurons in the striatum supports an information divergence–convergence pattern. Specifically, the broad spatial spread and the enormous number of release sites ($\sim 5 \times 10^5$) of each dopamine axonal tree, additionally imposes extremely high convergence on single striatal projection neurons (Wickens and Arbuthnott 2005; Moss and Bolam 2008; Matsuda et al. 2009). Volume transmission (Cragg et al. 2001) of dopamine in the striatum also enforces population averaging of the dopamine signal on the level of the single target striatal neuron. Finally, the mechanisms removing dopamine from the synapse are highly unreliable, resulting in exceptionally poor spatial and temporal precision of the dopamine signal (Cragg et al. 2000; Venton et al. 2003; Roitman et al. 2004). It is interesting to note in this respect that the low degree of temporal correlations of the spiking activity of dopamine neurons (Morris et al. 2004) provides an optimal substrate for such averaging to yield accurate estimation of the transmitted signal (Zohary et al. 1994).

Decoding aversion- and reward-related dopamine

Another option that may rescue reward-related dopamine hypotheses in light of the new results is that the target structures are able to decode aversive and reward-related dopamine signals. For example, while it is established that dopamine is released for both rewards and punishments, it might be that the amount of released dopamine is different. At least two of the above-mentioned recent studies (Joshua et al. 2008; Matsumoto and Hikosaka 2009) provide some evidence for this idea. In both, the excitatory phasic response to aversive stimuli had a lower firing rate and a shorter duration than the excitatory response to rewarding stimuli. Similarly, responses to cues predicting punishments were weaker and shorter than the response to cues predicting rewards. The difference in duration of the responses is in the range of 50–100 ms. Furthermore, although not discussed in these papers, the activity of dopamine neurons following aversive events seems to decrease below baseline even in those neurons that displayed the initial bursts. Thus, when population averaging is performed (as dictated by the anatomy) the dopamine level after aversive stimuli should be below the level following rewarding stimuli, and perhaps even below baseline. The

latter prediction is in line with recent fast-cyclic-voltammetry studies (Roitman et al. 2008). We propose that the dopamine signal functions on two distinct timescales: while the short initial burst reflects arousal level and initiates immediate action, the long-term plasticity effects (learning) are governed by the average dopamine levels at a more delayed time period. At this later stage, the phasic dopamine responses to reward- and aversion-related stimuli lead to two different dopamine concentrations in the striatum that have opposite effects on synaptic plasticity at the cortico-striatal pathways.

Striatal action-learning based on dopamine concentration

According to the prevailing view of the effect of dopamine on the direct and indirect pathways, a surge of dopamine should increase the excitability of D1 MSNs (Go pathway) and decrease that of D2 (No-Go) MSNs (see Fig. 3a). Thus, the immediate effect of the initial burst would be to execute the default action that is connected to the given set of stimuli, since it would indiscriminately excite all Go circuits, and the strongest circuit will be chosen in a winner-take-all manner. In contrast, the relative strength of the different circuits is established through long-term learning, which should be controlled by the second phase of dopamine signaling.

Learning in the direct and indirect pathways and their control by dopamine has been previously described (Frank et al. 2004). According to this model, in D1 Go, MSNs are the starting points for execution of the actions that will eventually be chosen. Cells representing more likely actions in the current state increase their activity. In the D2 No-Go MSNs, actions, which are unlikely in the current state, show an increase in activity. Reward-related dopamine reinforces current actions in the Go pathway, because these actions seem to be related to obtaining the reward. At the same time, the cells representing the same action cells in the No-Go pathway undergo LTD because these cells were inactive. Finally, projections to all active cells in the No-Go pathway (action alternatives that were not chosen) are potentiated, further decreasing the probability of performing these actions when the animal encounters the same state in the future. All three changes contribute to reinforcement learning: increase the probability of performing a rewarded action in a certain situation. In contrast, aversion-related low levels of dopamine cause LTD in active cells in the Go pathway, decreasing the probability of an action that leads to a punishment. Further, it weakens projections to active cells in the No-Go pathway. Thereby, these action alternatives become more likely the next time the animal is in the troubling situation again.

This simplistic model is somewhat complicated by evidence from cellular neurophysiology. On the neuronal correlate level, aversive learning should translate to inversion of the temporal aspect of normal Hebbian plasticity, or to reversal of the STDP rule. Although used in modeling studies (Bar-Gad et al. 2003; Frank et al. 2007; Thivierge et al. 2007), physiological evidence for this has been lacking. Aside from one report of inverse STDP in striatal MSNs (Fino et al. 2005), the temporal aspects of long-term plasticity induction protocols have not been studied until very recently. It was widely believed, however, that dopamine was essential for both LTP and LTD (Reynolds et al. 2001; Centonze et al. 2001). This view was recently refined by two elegant studies which systematically examined the question of STDP in cortico-striatal synapses and the involvement of dopamine in the process (Shen et al. 2008; Pawlak and Kerr 2008). Both studies revealed that, under normal conditions, both D1 and D2 type MSNs undergo long-term plasticity which follows STDP. Moreover, it appears that adherence to STDP requires activation of dopamine receptors in an asymmetric manner: glutamatergic synapses on D1 (Go circuit) MSNs are potentiated following post-synaptic firing which succeeds pre-synaptic activation, but only if D1 receptors are activated. LTD is displayed after the opposite pairing, but this part is dopamine-independent. Similarly, D2 (No-Go circuit) MSNs are depressed following post-synaptic firing which precedes pre-synaptic activation, but only if D2 receptors are activated. LTP is exhibited following the opposite pairing, but this is again dopamine-independent. Thus, in the absence of dopaminergic activation, plasticity does not merely disappear, but becomes uni-directional: synapses in Go circuits can only undergo LTD, while those in No-Go circuits will only be potentiated.

Figure 3 describes the changes in a hypothetical circuit with four possible actions connected to a single state following reward-related and punishment-related dopamine responses. A careful comparison of the two scenarios depicted in Fig. 3 reveals an interesting feature imposed by the differential dependence of D1 (Go) and D2 (No-Go) MSNs on dopamine. Apparently, the only difference between the ‘Reward’ scenario and the ‘Punishment’ one relates to the connections of the state to the action that was taken. This does not mean that other connections do not change. Rather, these changes are not dopamine-dependent and therefore are indifferent to the delivery of reward/punishment.

The difference in dopamine effect on plasticity in Go and No-Go synapses presents an unexpected answer to another open question in computational modeling of basal-ganglia circuits. So far, TD-learning has been described for classical conditioning situations. However, in settings other than classical conditioning, an agent acts in order to receive

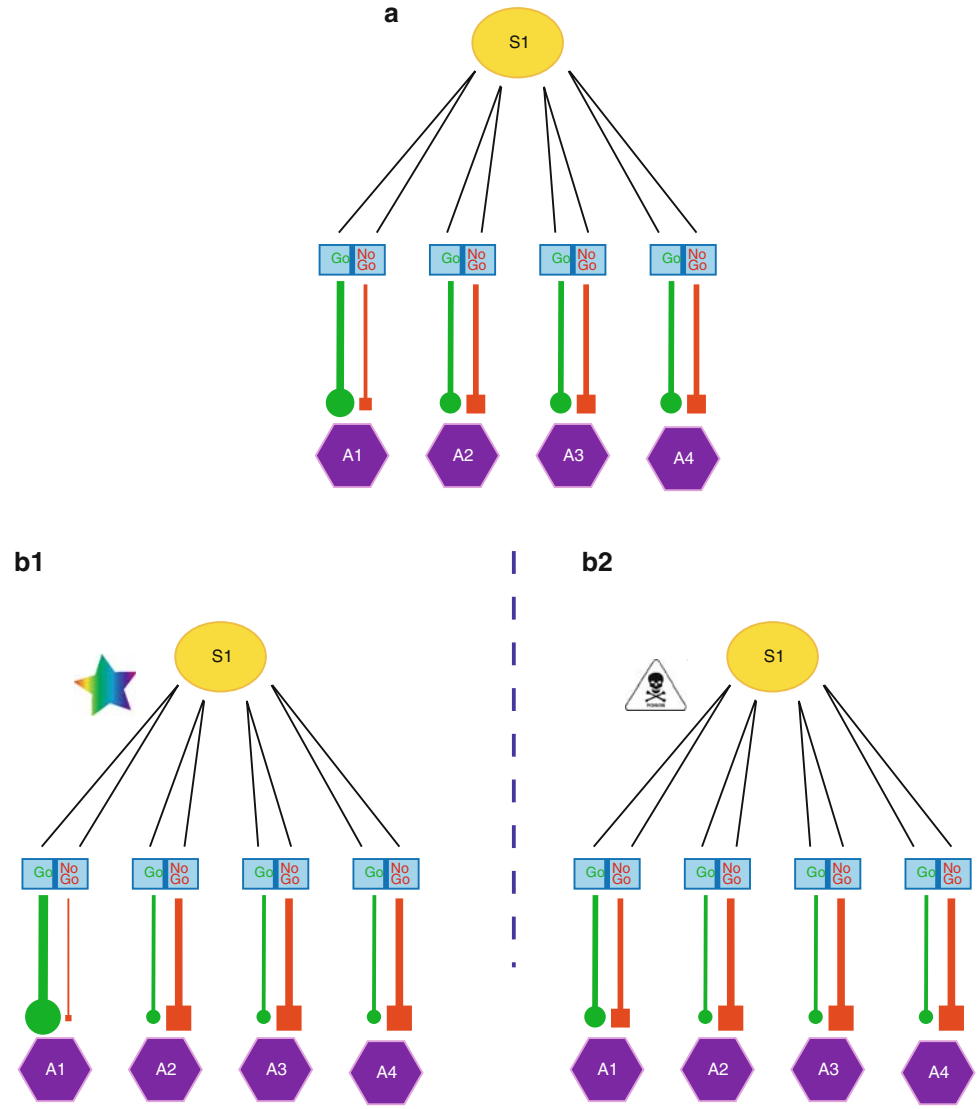

Fig. 3 On-policy learning with dopamine. All connections from the given state to the possible actions that may be taken at that state. *S* state, *A* action, *square* inhibitory connections (Go circuits expressing D1 receptors), *circle* excitatory connections (No-Go circuits expressing D2 receptors). Synaptic strength is represented by line thickness. **a** Occurrence of state S1 yields choice of action A1, as its Go connection is slightly higher and No-Go connection slightly lower than *A2–A4*. **b1** Long-term changes in synaptic strength after receipt of reward for the choice of *A1*. The active A1-Go circuit undergoes dopamine-dependent LTP. The non-active A1-No-Go circuit undergoes dopamine-dependent LTD. We assume that actions *A2–A4* were suppressed, and therefore the corresponding No-Go circuits were active, and thus undergo dopamine-independent LTP; the non-active Go circuits are depressed (dopamine-independent). **b2** Long-term changes in synaptic strength after receipt of punishment for the choice of A1. Dopamine level is too low for dopamine-dependent STDP. Therefore, the active A1-Go circuit undergoes LTD. The non-active A1-No-Go circuit undergoes LTP; as in **b1**, actions *A2–A4* were actively suppressed, and therefore the active corresponding No-Go circuits are potentiated (dopamine-independent), and the non-active No-Go circuits are depressed (dopamine-independent)

rewards. Therefore, an action policy has to be learned which tells the agent how to act in each situation. A number of extensions to the TD-learning scheme of the classical conditioning setting have been proposed. In the so-called Actor/Critic method, the problem at hand is divided between two dedicated components. The *critic* is responsible for value estimation. The action policy is explicitly stored in an *actor* element. Both critic and actor use the same TD error signal for learning. An alternative class of algorithms does not involve an explicit representation of the behavioral policy. Instead, the value function contains action values rather than state values. In this way, the optimal policy emerges from comparing the values of different actions. Algorithms learning action values can be learned *on-policy*, i.e., where only the policy that is currently employed is updated during learning (like SARSA)

or *off-policy* (e.g., Q-learning; Watkins and Dayan 1992), which has the obvious advantage of separation between what is done and what is learnt. Reducing the basal ganglia to an action-selection network, the actor/critic architecture has often been employed to model learning in this framework (Suri and Schultz 1999; Joel et al. 2002). However, two recent studies that examined the activity of dopamine neurons in settings that required explicit action selection in primates (Morris et al. 2006) and rats (Roesch et al. 2007) suggest that dopamine neurons actually code the error in value of state-action pairs, rather than state value as would be expected from an actor/critic learning network. Whereas the primate results (recorded in the SNc) favor the on-policy approach, the rat results (recorded in the VTA) favor the more efficient and complicated Q-learning approach. Since the differential dependence of STDP in Go and No-Go circuits requires that only taken actions are updated according to the dopamine signal, action learning in the basal ganglia must be performed online.

Finally, we would like to note that low levels of dopamine must all be lumped together. As a result, omission of predicted reward and delivery of an aversive stimulus are treated in the same manner. However, behaviorally there seems to be a substantial difference between aversive learning and reward omission. For example, while aversive learning is usually very strong and rapid, often occurs within a single trial and is very difficult to extinguish (Barber et al. 1998). Extinction of rewards (which is learning that a CS no longer reliably predicts reward) is much slower, and the original reward conditioning can easily be reinstated. Therefore, we propose that there is an additional, dopamine-independent neural substrate that is dedicated to aversive learning. Whether this occurs in the same synapses or in a different structure remains an open question.

References

Acquas E, Tanda G, Di Chiara G (2002) Differential effects of caffeine on dopamine and acetylcholine transmission in brain areas of drug-naive and caffeine-pretreated rats. Neuropsychopharmacology 27:182–193

Albin RL, Young AB, Penney JB (1989) The functional anatomy of basal ganglia disorders. Trends Neurosci 12:366–375

Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci 13:266–271

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

Alexander GE, Crutcher MD, DeLong MR (1990) Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Prog Brain Res 85:119–146

Arroyo M, Markou A, Robbins TW, Everitt BJ (1998) Acquisition, maintenance and reinstatement of intravenous cocaine self-administration under a second-order schedule of reinforcement in rats: effects of conditioned cues and continuous access to cocaine. Psychopharmacology Berl 140:331–344

Barber TA, Klunk AM, Howorth PD, Pearlman MF, Patrick KE (1998) A new look at an old task: advantages and uses of sickness-conditioned learning in day-old chicks. Pharmacol Biochem Behav 60:423–430

Bar-Gad I, Morris G, Bergman H (2003) Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. Prog Neurobiol 71:439–473

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141

Belin D, Everitt BJ (2008) Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. Neuron 57:432–441

Bergman H, Wichmann T, DeLong MR (1990) Reversal of experimental parkinsonism by lesions of the subthalamic nucleus. Science 249:1436–1438

Berridge KC (1996) Food reward: brain substrates of wanting and liking. Neurosci Biobehav Rev 20:1–25

Berridge KC (2007) The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl) 191:391–431

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Brain Res Rev 28:309–369

Bi G, Poo M (1999) Distributed synaptic modification in neural networks induced by patterned stimulation. Nature 401:792–796

Bjorklund A, Lindvall O (1984) In: Bjorklund A, Hokfelt T (eds) Classical transmitters in the CNS part I. Elsevier, Amsterdam, pp 55–122

Bonci A, Malenka RC (1999) Properties and plasticity of excitatory synapses on dopaminergic and GABAergic cells in the ventral tegmental area. J Neurosci 19:3723–3730

Brischoux F, Chakraborty S, Brierley DI, Ungless MA (2009) Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. Proc Natl Acad Sci USA 106:4894–4899

Calabresi P, Centonze D, Gubellini P, Pisani A, Bernardi G (1998) Blockade of M2-like muscarinic receptors enhances long-term potentiation at corticostriatal synapses. Eur J Neurosci 10:3020–3023

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352

Centonze D, Picconi B, Gubellini P, Bernardi G, Calabresi P (2001) Dopaminergic control of synaptic plasticity in the dorsal striatum. Eur J Neurosci 13:1071–1077

Cheer JF, Aragona BJ, Heien ML, Seipel AT, Carelli RM, Wightman RM (2007) Coordinated accumbal dopamine release and neural activity drive goal-directed behavior. Neuron 54:237–244

Cragg SJ (2006) Meaningful silences: how dopamine listens to the ACh pause. Trends Neurosci 29:125–131

Cragg SJ, Hille CJ, Greenfield SA (2000) Dopamine release and uptake dynamics within nonhuman primate striatum in vitro. J Neurosci 20:8209–8217

Cragg SJ, Nicholson C, Kume-Kick J, Tao L, Rice ME (2001) Dopamine-mediated volume transmission in midbrain is regulated by distinct extracellular geometry and uptake. J Neurophysiol 85:1761–1771

Dalley JW, Fryer TD, Brichard L, Robinson ES, Theobald DE, Laane K, Pena Y, Murphy ER, Shah Y, Probst K, Abakumova I, Aigbirhio FI, Richards HK, Hong Y, Baron JC, Everitt BJ, Robbins TW (2007) Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement. Science 315:1267–1270

DeLong MR, Georgopoulos AP (1981) Motor functions of the basal ganglia. In: Brookhart JM, Mountcastle VB, Brooks VB, Geiger SR (eds) Handbook of physiology. The nervous system. Motor

control, Sect. 1, Pt. 2, vol II. American Physiological Society, Bethesda, pp 1017–1061
Descarries L, Gisiger V, Steriade M (1997) Diffuse transmission by acetylcholine in the CNS. Prog Neurobiol 53:603–625
Di Chiara G, Bassareo V (2007) Reward system and addiction: what dopamine does and doesn't do. Curr Opin Pharmacol 7:69–76
Di Chiara G, Bassareo V, Fenu S, De Luca MA, Spina L, Cadoni C, Acquas E, Carboni E, Valentini V, Lecca D (2004) Dopamine and drug addiction: the nucleus accumbens shell connection. Neuropharmacology 47(Suppl 1):227–241
Fino E, Glowinski J, Venance L (2005) Bidirectional activity-dependent plasticity at corticostriatal synapses. J Neurosci 25:11279–11287
Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898–1902
Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943
Frank MJ, Samanta J, Moustafa AA, Sherman SJ (2007) Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 318:1309–1312
Gerfen CR (1992) The neostriatal mosaic: multiple levels of compartmental organization. J Neural Transm Suppl 36:43–59
Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J Neurosci 20:2369–2382
Hand TH, Franklin KB (1985) 6-OHDA lesions of the ventral tegmental area block morphine-induced but not amphetamine-induced facilitation of self-stimulation. Brain Res 328:233–241
Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304–309
Holt DJ, Graybiel AM, Saper CB (1997) Neurochemical architecture of the human striatum. J Comp Neurol 384:1–25
Horvitz JC (2002) Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. Behav Brain Res 137:65–74
Ito R, Robbins TW, Everitt BJ (2004) Differential control over cocaine-seeking behavior by nucleus accumbens core and shell. Nat Neurosci 7:389–397
Joel D, Niv Y, Ruppin E (2002) Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw 15:535–547
Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H (2008) Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. J Neurosci 28:11673–11684
Kawagoe R, Takikawa Y, Hikosaka O (2004) Reward-predicting activity of dopamine and caudate neurons—a possible mechanism of motivational control of saccadic eye movement. J Neurophys 91:1013–1024
Kelley AE (2004) Memory and addiction: shared neural circuitry and molecular mechanisms. Neuron 44:161–179
Kelley AE, Domesick VB, Nauta WJH (1982) The amygdalostriatal projection in the rat-an anatomical study by anterograde and retrograde tracing methods. Neuroscience 7:615–630
Kerr JN, Wickens JR (2001) Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. J Neurophysiol 85:117–124
Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399–1402
Kreitzer AC, Malenka RC (2005) Dopamine modulation of state-dependent endocannabinoid release and long-term depression in the striatum. J Neurosci 25:10537–10545
Lavoie B, Parent A (1990) Immunohistochemical study of the serotoninergic innervation of the basal ganglia in the squirrel monkey. J Comp Neurol 299:1–16
Lavoie B, Smith Y, Parent A (1989) Dopaminergic innervation of the basal ganglia in the squirrel monkey as revealed by tyrosine hydroxylase immunohistochemistry. J Comp Neurol 289:36–52
Ludvig EA, Sutton RS, Kehoe EJ (2008) Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. Neural Comput 20:3034–3054
Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275:213–215
Matsuda W, Furuta T, Nakamura KC, Hioki H, Fujiyama F, Arai R, Kaneko T (2009) Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. J Neurosci 29:444–453
Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:837–841
Matsumoto N, Hanakawa T, Maki S, Graybiel AM, Kimura M (1999) Nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. J Neurophysiol 82:978–998
Mink JW (1996) The basal ganglia: focused selection and inhibition of competing motor programs. Prog Neurobiol 50:381–425
Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936–1947
Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H (2004) Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. Neuron 43:133–143
Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057–1063
Moss J, Bolam JP (2008) A dopaminergic axon lattice in the striatum and its relationship with cortical and thalamic terminals. J Neurosci 28:11221–11230
Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. Neuron 41:269–280
Nicola SM, Surmeier J, Malenka RC (2000) Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. Annu Rev Neurosci 23:185–215
O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. Neuron 38:329–337
Olds J, Milner P (1954) Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J Comp Physiol Psychol 47:419–427
Owesson-White CA, Cheer JF, Beyene M, Carelli RM, Wightman RM (2008) Dynamic changes in accumbens dopamine correlate with learning during intracranial self-stimulation. Proc Natl Acad Sci USA 105:11957–11962
Pan WX, Schmidt R, Wickens JR, Hyland BI (2005) Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci 25:6235–6242
Pan WX, Schmidt R, Wickens JR, Hyland BI (2008) Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. J Neurosci 28:9619–9631
Pavlov IP (1927) Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. Oxford university press, London
Pawlak V, Kerr JN (2008) Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. J Neurosci 28:2435–2446

Percheron G, Filion M (1991) Parallel processing in the basal ganglia: up to a point. Trends Neurosci 14:55–56
Rassnick S, Stinus L, Koob GF (1993) The effects of 6-hydroxydopamine lesions of the nucleus accumbens and the mesolimbic dopamine system on oral self-administration of ethanol in the rat. Brain Res 623:16–24
Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci 7:967–975
Redgrave P, Gurney K, Reynolds J (2008) What is reinforced by phasic dopamine signals? Brain Res Rev 58:322–339
Redish AD (2004) Addiction as a computational process gone awry. Science 306:1944–1947
Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Black AJ, Prokasy WF (eds) Classical conditioning II: current research and theory. Appelton-Century Crofts, New York, pp 64–99
Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67–70
Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615–1624
Roitman MF, Stuber GD, Phillips PEM, Wightman RM, Carelli RM (2004) Dopamine operates as a subsecond modulator of food seeking. J Neurosci 24:1265–1271
Roitman MF, Wheeler RA, Wightman RM, Carelli RM (2008) Real-time chemical responses in the nucleus accumbens differentiate rewarding and aversive stimuli. Nat Neurosci 11:1376–1377
Royall DR, Klemm WR (1981) Dopaminergic mediation of reward: evidence gained using a natural reinforcer in a behavioral contrast paradigm. Neurosci Lett 21:223–229
Salamone JD, Correa M, Farrar A, Mingote SM (2007) Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. Psychopharmacology (Berl) 191:461–482
Schmidt R, Morris G, Hagen EH, Sullivan RJ, Hammerstein P, Kempter R (2009) The dopamine puzzle. Proc Natl Acad Sci USA 106:E75
Schultz W (1982) Depletion of dopamine in the striatum as an experimental model of parkinsonism: direct effects and adaptive mechanisms. Prog Neurobiol 18:121–166
Schultz W (1994) Behavior-related activity of primate dopamine neurons. Rev Neurol Paris 150:634–639
Schultz W (2002) Getting formal with dopamine and reward. Neuron 36:241–263
Schultz W, Studer A, Jonsson G, Sundstrom E, Mefford I (1985) Deficits in behavioral initiation and execution processes in monkeys with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-induced parkinsonism. Neurosci Lett 59:225–232
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599
Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS (2004) Temporal difference models describe higher-order learning in humans. Nature 429:664–667
Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. Science 321:848–851
Skinner BF (1974) About behaviorism. Knopf, New York
Smith Y, Bevan MD, Shink E, Bolam JP (1998) Microcircuitry of the direct and indirect pathways of the basal ganglia. Neuroscience 86:353–387
Suri RE, Schultz W (1999) A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 91:871–890
Sutton RS (1988) Learning to predict by the methods of temporal difference. Mach Learn 3:9–44
Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. MIT press, Cambridge, MA
Thivierge JP, Rivest F, Monchi O (2007) Spiking neurons, dopamine, and plasticity: timing is everything, but concentration also matters. Synapse 61:375–390
Thomas MJ, Malenka RC (2003) Synaptic plasticity in the mesolimbic dopamine system. Philos Trans Roy Soc Lond B Biol Sci 358:815–819
Thorndike EL (1898) Animal intelligence: an experimental study of the associative processes in animals. Psychol Rev, Monograph Suppl 8
Thorndike EL (1911) Animal intelligence. Hafner, Darien
Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642–1645
Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K (2009) Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324:1080–1084
Venton BJ, Zhang H, Garris PA, Phillips PE, Sulzer D, Wightman RM (2003) Real-time decoding of dopamine concentration changes in the caudate-putamen during tonic and phasic firing. J Neurochem 87:1284–1295
Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz CM (2004) Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci 27:468–474
Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43–48
Wang Z, Kai L, Day M, Ronesi J, Yin HH, Ding J, Tkatch T, Lovinger DM, Surmeier DJ (2006) Dopaminergic control of corticostriatal long-term synaptic depression in medium spiny neurons is mediated by cholinergic interneurons. Neuron 50:443–452
Watkins CJCH, Dayan P (1992) Q learning. Mach Learn 8:279–292
Wheeler RA, Carelli RM (2009) Dissecting motivational circuitry to understand substance abuse. Neuropharmacology 56(Suppl 1):149–159
Wickens JR, Arbuthnott GW (2005) Structural and functional interactions in the striatum at the receptor level. In: Dunnett SB, Bentivoglio M, Bjorklund A, Hokfelt T (eds) Dopamine. Elsevier, Amsterdam, pp 199–236
Wickens JR, Budd CS, Hyland BI, Arbuthnott GW (2007) Striatal contributions to reward and decision making: making sense of regional variations in a reiterated processing matrix. Ann N Y Acad Sci 1104:192–212
Wise RA (1996) Addictive drugs and brain stimulation reward. Annu Rev Neurosci 19:319–340
Wise RA (2004) Dopamine, learning and motivation. Nat Rev Neurosci 5:483–494
Wise RA (2008) Dopamine and reward: the anhedonia hypothesis 30 years on. Neurotox Res 14:169–183
Woolf NJ (1991) Cholinergic systems in mammalian brain and spinal cord. Prog Neurobiol 37:475–524
Zhou FM, Liang Y, Dani JA (2001) Endogenous nicotinic cholinergic activity regulates dopamine release in the striatum. Nat Neurosci 4:1224–1229
Zhou FM, Wilson C, Dani JA (2003) Muscarinic and nicotinic cholinergic mechanisms in the mesostriatal dopamine systems. Neuroscientist 9:23–36
Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143