

Published in final edited form as:

Behav Brain Res. 2009 April 12; 199(1): 141–156. doi:10.1016/j.bbr.2008.09.029.

Neurocomputational models of basal ganglia function in learning, memory and choice

Michael X Cohen and Michael J. Frank*

Department of Psychology Program in Neuroscience University of Arizona 1503 E University Blvd Tucson AZ 85721

Abstract

The basal ganglia (BG) are critical for the coordination of several motor, cognitive, and emotional functions and become dysfunctional in several pathological states ranging from Parkinson's disease to Schizophrenia. Here we review principles developed within a neurocomputational framework of BG and related circuitry which provide insights into their functional roles in behavior. We focus on two classes of models: those that incorporate aspects of biological realism and constrained by functional principles, and more abstract mathematical models focusing on the higher level computational goals of the BG. While the former are arguably more “realistic”, the latter have a complementary advantage in being able to describe functional principles of how the system works in a relatively simple set of equations, but are less suited to making specific hypotheses about the roles of specific nuclei and neurophysiological processes. We review the basic architecture and assumptions of these models, their relevance to our understanding of the neurobiological and cognitive functions of the BG, and provide an update on the potential roles of biological details not explicitly incorporated in existing models. Empirical studies ranging from those in transgenic mice to dopaminergic manipulation, deep brain stimulation, and genetics in humans largely support model predictions and provide the basis for further refinement. Finally, we discuss possible future directions and possible ways to integrate different types of models.

1 Introduction

The term basal ganglia refers to a collection of subcortical structures that are anatomically, neurochemically, and functionally linked (Mink, 1996). The basal ganglia are critical for several cognitive, motor, and emotional functions, and are integral components of complex functional/anatomical loops (Haber, Fudge, & McFarland, 2000; Haber, 2003). The intricate complexity of the basal ganglia can be seen at several levels, from myriad cortico-basal ganglia-thalamo-cortical loops, to the modulations by neurochemicals such as dopamine, serotonin, and acetylcholine, to differences in action by distinct receptor subtypes (e.g., D1 vs. D2 dopamine receptors) and locations (e.g., presynaptic autoreceptors and heteroreceptors vs. postsynaptic receptors). Investigations into the functional organization of the basal ganglia spans many species, experimental designs, theoretical frameworks, and levels of analysis (e.g., from functional neuroimaging in humans to genetic manipulations in mice to slice preparations). To make matters more complicated, most individual experiments focus only on

© 2008 Elsevier B.V. All rights reserved.

*corresponding author: mfrank@u.arizona.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

one level of analysis in one species, and each method comes with its own interpretive perils, making it a daunting task to integrate findings across studies and methodologies such that the effects of a single manipulation on the cascade of directly and indirectly affected variables can be predicted.

Biologically constrained computational models provide a useful framework within which to (1) interpret results from seemingly disparate empirical studies in the context of larger theoretical approaches, and (2) generate novel, testable, and sometimes counter-intuitive hypotheses, the evaluation of which can be used to refine our understanding of the basal ganglia. Moreover, the mathematical grounding of computational models eliminates semantic ambiguity and vague terminology, allows for more direct comparisons among findings from different experiments, species, and levels of analysis, and allows one to explore the intricate complexity of the basal ganglia circuitry while simultaneously linking functioning of that circuitry to behavior.

Although not without caveats, computational models provide a tool for exploring cognitive and brain processes not possible with classical box-and-arrow diagrams (whether the boxes contain anatomical brain areas, cognitive/functional processes, or both). Box-and-arrow diagrams can be confusing to interpret, provide too much leeway for semantic ambiguous interpretations, and do not allow one to examine the rich temporal dynamic interactions among subsystems, let alone how these dynamics evolve across time with learning. Computational models are dynamic, amenable to quantitative analyses, and can make predictions or inspire novel empirical work that might be difficult to intuit simply by visually inspecting a box-and-arrow diagram.

There have been several instances in which new understandings of basal ganglia functioning arose as a result of computational models operating on multiple levels of analysis. Some models are built to help understand the precise biophysical processes governing neuronal function, such as ion channel gating within the cholinergic interneuron; others are built to help understand the kinds of computations that might lead to cognitive processes such as learning, action selection, and even cognitive control. Each model and class of models has its own strengths and limitations, and each is appropriate for different applications. Given that no model is complete (i.e., no matter how biophysically or functionally/behaviorally constrained, every model necessarily omits several molecular and systems-level effects that are undoubtedly relevant), models should not be judged solely by any of these factors, but instead by their ability to capture interesting phenomena and make novel predictions that may lead to insights regarding their underlying mechanisms.

This review focuses on two classes of models – neural network models and more abstract mathematical models – that have been repeatedly used to understand behavioral functions of the basal ganglia and related circuitry. Neural network models use simplified neuronal units and neural dynamics to help understand how interactions among multiple parts of the circuit, and modulatory actions by dopamine and other neurochemicals, can support cognitive and behavioral phenomena such as action selection, learning and working memory. In contrast, more abstract models comprise mathematical equations, many of which build on research on machine learning and artificial intelligence. These are not necessarily constrained by biological architecture at the implementational level, but nevertheless make contact with these data and are designed to account for a large range of behavioral phenomena using a smaller number of assumptions and parameters. Given the focus on behavior, we do not discuss models that are highly focused on understanding more detailed biophysical processes within individual neurons (e.g. Wilson & Callaway, 2000; Wilson, Weyrick, Terman, Hallworth, & Bevan, 2004; Wolf, Moyer, Lazarewicz, Contreras, Benoit-Marand, O'Donnell, & Finkel, 2005; Zador & Koch, 1994; Lindskog, Kim, Wikström, Blackwell, & Kotaleski, 2006). This omission does

not imply a lack of interest in or excitement about these models – indeed, any abstract or systems-level neural account relying on implementational mechanisms will eventually need to be tested for plausibility using more realistic model neurons, and it is expected that some higher level explanations are likely to be modified by that endeavor. At present, though, it is intractable to use highly detailed biophysical models to develop a model of cognitive and behavioral phenomena which require systems-level analysis. Empirical data reviewed below confirm that despite some simplifications at the neuronal level, models make specific predictions that have been borne out across multiple experiments involving focal lesions, disease, neuroimaging, pharmacology, genetics, and deep brain stimulation on cognitive processes.

In the following sections we present an overview of the first two classes of models, their basic architecture and mathematical groundwork, and novel insights they have provided into the functions of the basal ganglia and related circuitry including empirical experiments testing specific model predictions. Following these overviews, we discuss how these classes of models can be related to each other, both in theoretical and practical aspects. We conclude by discussing the future of computational modeling in understanding the functional organization of the basal ganglia and related circuitry.

2 Neural network models of basal ganglia

By neural network models, we refer to a class of models in which detailed aspects of neuronal function such as geometry of an axon are abstracted, while other processes, such as membrane potential fluctuations over time and dynamic ionic conductances including activity-dependent channels, are simulated by coupled differential equations (Brown, Bullock, & Grossberg, 2004; Frank, Loughry, & O'Reilly, 2001; Frank, 2005, 2006; O'Reilly & Frank, 2006; Humphries, Stewart, & Gurney, 2006; Houk, 2005). Thus these models are far more biologically constrained than simple “connectionist” models but less so than detailed biophysical models. This approach provides a balance between capturing core aspects of underlying neurobiology while allowing the network to scale up to a level that is relevant to global information processing and behavior. Different model neurons are used to simulate neurons with different firing properties, excitatory and inhibitory neurons, as well as some basic neuromodulators such as dopamine and its postsynaptic effects onto different receptor subtypes. Parameters of these processes can be modified to capture different neuronal properties in different regions of the brain (e.g., striatum vs. globus pallidus vs. thalamus; Frank, 2006). Synaptic efficacy typically is simplified to a single modifiable “weight,” which reflects the extent to which a presynaptic neuron will influence the activity of the postsynaptic neuron. Mathematical and implementational details of this modeling approach is outside the scope of the present review; interested readers are referred to dedicated textbooks (O'Reilly & Munakata, 2000; Dayan & Abbott, 1999), and to the specific basal ganglia model references cited above.

2.1 Architecture of basal ganglia models

Broadly, we conceptualize the basal ganglia to be a system that dynamically and adaptively gates information flow in frontal cortex, and from frontal cortex to the motor system (see Figure 1 for a graphical overview of the model). The basal ganglia is richly anatomically connected to the frontal cortex and the thalamocortical motor system, via several distinct but partly overlapping loops (Gerfen & Wilson, 1996; Nakano, Kayahara, Tsutsumi, & Ushiro, 2000; Haber, 2003). This circuitry can facilitate or suppress action representations in the frontal cortex (Mink, 1996; Frank et al., 2001; Frank, 2005; Brown et al., 2004; Aron, Behrens, Smith, Frank, & Poldrack, 2007). These representations can range from simple actions to complex behaviors to cognitive operations such as working memory updating. Representations that are more goal-relevant or have a higher probability of being correct or rewarded are strengthened, whereas representations that are less goal-relevant or have a lower probability of reward are

weakened. Dopamine plays a key role in this process by modulating both excitatory and inhibitory signals in complementary ways, which can have the effect of modulating the signal-to-noise ratio (Winterer & Weinberger, 2004; Nicola, Surmeier, & Malenka, 2000; Frank, 2005).

Our models of this system includes the main architectural structures of the basal ganglia: Striatum; globus pallidus, external and internal segments (GPi and GPe); substantia nigra, pars compacta (SNc); thalamus; and subthalamic nucleus. This covers both the classical “direct pathway”, which sends a Go signal to frontal cortex, and the “indirect” pathway, which sends a NoGo signal to frontal cortex (Albin, Young, & Penney, 1989; Mink, 1996; Gerfen & Wilson, 1996; Frank, 2005). However, as we shall see, our computational models go beyond the classical direct/indirect model to (a) explore dynamics of this system as activity propagates throughout the system and as a function of synaptic plasticity, neither of which are evident in the static model and (b) incorporate more recent anatomical and physiological evidence that is not in the original model but which is essential for its functionality in action selection.

The direct pathway originates in striatonigral neurons, which mainly express D1 receptors and provide direct inhibitory input to the GPi and SNr. We refer to activity in this pathway as “Go signals” because when striatonigral cells are active, they inhibit GPi, which in turn disinhibits the thalamus (Chevalier & Deniau, 1990), and allows frontal cortical representations to be amplified by bottom-up thalamocortical drive. Note that this disinhibition process only *enables* the corresponding column of thalamus to become active *if* that same column also receives top-down cortico-thalamic excitation. This means that the basal ganglia system does not directly select which action to ‘consider’, but instead modulates the activity of already active representations in cortex. This functionality enables cortex to weakly represent multiple potential actions in parallel; the one that first receives a Go signal from striatal output is then provided with sufficient additional excitation to be executed. Lateral inhibition within thalamus and cortex act to suppress competing responses once the winning response has been selected by the BG circuitry.

Complementary to the direct pathway, the indirect pathway originates in striatopallidal cells in the striatum which mainly express D2 receptors and provide direct inhibitory input to the GPe. We refer to activity in this pathway as sending a “NoGo signal” to suppress a specific unwanted response. Because the GPe tonically inhibits the GPi via direct focused projections; striatopallidal NoGo activity removes this tonic inhibition, thereby disinhibiting the GPi, allowing it to further inhibit the thalamus and preventing particular cortical actions from being facilitated. In this way, the model basal ganglia can facilitate (Go) or suppress (NoGo) representations in frontal cortex. Note that a given action can have both Go and NoGo representations, and the probability that it will be selected is a function of the relative Go-NoGo activation difference (Frank, 2005). This is due to the observation that Go and NoGo cells receiving from a given cortical region (and thereby encoding a given action) originate in the same striatal region, and terminate in the same region within GPi (e.g., Féger & Crossman, 1984; Mink, 1996). Neurons in the latter structure can then reflect the relative difference in the two striatal populations, which then influences the likelihood of disinhibiting the thalamus and in turn selecting the action.

Note that the above depiction omits the subthalamic nucleus (STN), classically thought to be a critical relay station within the indirect pathway linking GPe with GPi (Albin et al., 1989). However, more recent evidence indicates that (a) GPe neurons send direct inhibitory projections to GPi rather than having to exert their control indirectly via STN; (b) these GPe-GPi projections are more focused, allowing a *specific* response to be suppressed, whereas those from STN to GPi projections are broad and diffuse (Mink, 1996; Parent & Hazrati, 1995), perhaps providing a more global modulatory function (see below).

This is not to diminish or discount the role of the STN. To the contrary, recent evidence indicates that the STN should be considered part of a third “hyper-direct” pathway (so-named because it bypasses the striatum altogether), rather than just a relay within the indirect pathway. Indeed, the STN receives direct excitatory input from frontal cortex, and sends diffuse excitatory projections to GPi (Nambu, Tokuno, Hamada, Kita, Imanishi, Akazawa, Ikeuchi, & Hasegawa, 2000; Nambu, Tokuno, & Takada, 2002). We refer to activity in the STN as sending a “Global NoGo” signal because its diffuse excitatory effect on many GPi neurons would prevent all responses, rather than just one, from being facilitated (Frank, 2006). Simulations revealed that these signals are dynamic: The Global NoGo signal is observed early during response selection, preventing any response from being selected prematurely, but as STN activity subsides, a response is then more likely to occur. This transient burst in STN activity is consistent with that observed *in vivo* (Wichmann, Bergman, & DeLong, 1994; Magill, Sharott, Bevan, Brown, & Bolam, 2004). Moreover, in the model, the initial Global NoGo signal is adaptively modulated by the degree of cortical response *conflict*: Greater activation of multiple competing cortical motor commands is associated with greater STN excitatory drive and a pronounced Global NoGo signal, enabling the striatum to take more time to “settle” and integrate over noisy intrinsic activity to choose the best response (Frank, 2006; Bogacz & Gurney, 2007). Without this STN functionality, the BG network is more likely to make premature responses, often settling on the suboptimal choice (see Figure 3a), particularly when there is a high degree of response conflict (Frank, 2006). Such premature responding is observed in rats with STN lesions (Baunez & Robbins, 1997; Baunez, Christakou, Chudasama, Forni, & Robbins, 2007).

Dopamine plays a special modulatory role in the basal ganglia. At D1 receptors in the striatum, dopamine is thought to act as a contrast-enhancer, increasing activity on highly active cells while decreasing activity on less active cells (Hernandez-Lopez, Bargas, Surmeier, Reyes, & Galarraga, 1997). This has the effect of amplifying the signal (highly active cells) while simultaneously decreasing the noise (less active cells; Frank, 2005). At D2 receptors, dopamine is inhibitory, regardless of the amount of activity in the cells (Hernandez-Lopez, Tkatch, Perez-Garci, Galarraga, Bargas, Hamm, & Surmeier, 2000). Because D1 receptors are expressed in great abundance on Go cells whereas D2 receptors are expressed in great abundance on NoGo cells (Gerfen, 1992), elevated dopamine has the net effect of facilitating synaptically-driven Go activity while inhibiting NoGo activity. In contrast, low levels of dopamine would decrease the signal-to-noise in Go cells while freeing the NoGo cells from inhibition. This conceptualization explains why reduced dopamine levels as in Parkinson's disease results in over-activation of the NoGo pathway (Surmeier, Ding, Day, Wang, & Shen, 2007; Shen, Tian, Day, Ulrich, Tkatch, Nathanson, & Surmeier, 2007) and slowness of movement, similar to the original proposal (Albin et al., 1989). Moreover, in the context of our dynamic model, these effects of dopamine on Go and NoGo activity are particularly relevant for reinforcement learning (Frank, 2005; Frank, Seeberger, & O'Reilly, 2004; Brown et al., 2004), and can have important implications for how the basal ganglia system can learn which representations to facilitate and which to inhibit, as discussed in the following section.

2.2 Reinforcement learning in basal ganglia models

Synaptic weights between neurons can change dynamically over time and over experience, forming the basis of learning. Weights between units that are strongly and repeatedly co-activated become stronger (as in long-term potentiation; LTP), otherwise weights between units do not change or become weakened (as in long-term depression; LTD). The presence and timing of dopamine release strongly modulates these effects in the striatum (Berke & Hyman, 2000; Reynolds & Wickens, 2002; Kerr & Wickens, 2001; Reynolds, Hyland, & Wickens, 2001; Calabresi, Pisani, Centonze, & Bernardi, 1997; Calabresi, Gubellini, Centonze, Picconi, Bernardi, Chergui, Svenningsson, Fienberg, & Greengard, 2000; Centonze, Picconi, Gubellini,

Bernardi, & Calabresi, 2001). Indeed, the primary mechanism of learning in the basal ganglia model is dependent on dopaminergic modulation of cells already activated by corticostriatal glutamatergic input.

Specifically, dopaminergic neurons in the SNc famously fire in phasic bursts during unexpected rewards, and firing drops below tonic baseline levels when rewards are expected but not received (Schultz, Dayan, & Montague, 1997; Bayer, Lau, & Glimcher, 2007). In the model, SNc dopamine bursts are simulated when the model selects the correct action (depending on the nature of the task). As a result, activated Go units are further potentiated such that the weights to these Go units from sensory and premotor cortex are increased. This means that the next time the same sensory stimulus is presented together with the associated premotor cortical response, these same Go units are likely to become active and facilitate the same rewarding response. In contrast, weakly active Go units are suppressed. These effects are mediated via simulated D1 receptors, consistent with the aforementioned physiological data.

Further, when the model receives a dip in dopamine (i.e., a lack of reward when one is expected; (Schultz et al., 1997; Bayer et al., 2007)), a complementary process occurs. In this case, NoGo units, which are normally inhibited by dopamine via simulated D2 receptors, now become more activated by their cortical glutamatergic inputs. Indeed, striatopallidal neurons receive stronger projections from frontal cortex and show particularly enhanced excitability to cortical stimulation (Berretta, Parthasarathy, & Graybiel, 1997; Berretta, Sachs, & Graybiel, 1999; Kreitzer & Malenka, 2007; Lei, Jiao, Del Mar, & Reiner, 2004). Critically, transiently enhanced NoGo unit activity is associated with long term potentiation (via similar Hebbian learning principles), such that the next time the model is faced with the same sensory stimulus and potential response, that response is more likely to be suppressed. Thus, phasic dips in dopamine induce learning to avoid particular actions in the presence of particular stimuli (Frank, 2005). Recent studies support this basic model prediction, showing that whereas synaptic potentiation in the direct pathway is dependent on D1 receptor stimulation, potentiation in the indirect pathway is dependent on a *lack* of D2 receptor stimulation (Shen, Flajolet, Greengard, & Surmeier, 2008).

It is through this push-pull mechanism that the basal ganglia model can learn to select actions or reinforce frontal cortical representations that are more likely to lead to reward or correct feedback, while simultaneously reducing the probability that incorrect or nonrewarding actions or representations are less likely to occur. The presence of learning in both pathways allows the model to enhance contrast between different stimulus-reinforcement probabilities, making it easier to discriminate between, say, a choice that is 60% vs 40% rewarding. Models learning only to increase and decrease synaptic weights in just the Go pathway were less able to make these subtle discriminations in complex probabilistic environments (Frank, 2005). In dual pathway models, a 60% response is represented in both Go and NoGo pathways, and recall that the BG output (GPi) computes the relative activation differences for each response. Thus the net effect on GPi (ignoring nonlinearities for simplicity) is $60 - 40 = 20\%$. Similarly, a 40% response in GPi would have greater NoGo than Go and therefore would be represented as -20% . The net difference between the two responses, which in reality is 20%, has been contrast-enhanced to 40% at the BG output.

2.2.1 Do Dopamine 'Dips' Contain Sufficient Information for Learning?—Baseline firing rates of dopamine neurons are low – generally around 5 Hz. Thus, while increases in firing rate can scale upward with larger magnitudes of prediction errors, they cannot scale downwards with negative prediction errors (since neurons cannot have negative firing rates). This led to the question of whether separate non-dopaminergic mechanisms in the brain are required to code negative prediction errors (Daw, Kakade, & Dayan, 2002; Bayer & Glimcher, 2005). However, recent empirical work suggests that, rather than change in firing rate, the

duration of the dopamine neuron pause during reward omissions might contain information about the magnitude of the negative prediction error (Bayer et al., 2007). This is interesting in light of the fact that D2 receptors in the striatum are highly sensitive to small changes in dopamine, in part because most D2 receptors are high-affinity (Richfield, Penney, & Young, 1989); thus, differences in the pause duration might have detectable downstream effects (Frank & Claus, 2006; Frank & O'Reilly, 2006).

Recall that the model requires a lack of D2 receptor stimulation to potentiate NoGo units and to promote learning, as supported by recent data (Shen et al., 2008). Thus, longer pause durations provide more time for dopamine transporters to remove DA from the synapse, increasing the likelihood that neurons expressing D2 receptors will become disinhibited. This account is particularly plausible in dorsal striatum, where there are many dopamine transporters and the half-life of dopamine in the synapse is roughly 55–75 ms (Suaud-Chagny, Dugast, Chergui, Msghina, & Gonon, 1995; Gonon, 1997; Venton, Zhang, Garris, Phillips, Sulzer, & Wightman, 2003). This means that longer duration pauses (> 200ms) would give sufficient time for dopamine to be virtually absent, and would allow NoGo units to become disinhibited (in contrast to ventral striatum, and especially prefrontal cortex, in which the time-course of reuptake may be too slow for phasic dips to have any functional effect). Further, depleted striatal dopamine levels, as in Parkinson's disease, would actually enhance this effect. Although tonic dopamine levels are already low, the resulting D2 receptor supersensitivity (Seeman, 2008), together with enhanced excitability of NoGo cells in the DA-depleted state (Surmeier et al., 2007; Shen et al., 2007), would facilitate the postsynaptic detection of DA pauses (such that perhaps they do not have to be as long in duration to be detected). Indeed, recent studies demonstrate enhanced potentiation of NoGo synapses as a result of DA depletion in a mouse model of Parkinson's disease (Shen et al., 2008).

2.2.2 Plasticity in cortical system: From actions to habits—Finally, the model also captures plasticity directly in the cortico-cortical pathway from sensory to premotor cortex. As responses are made to particular stimuli, simple Hebbian learning occurs such that the same pre-motor cortical units are likely to become active in response to this same stimulus in the future, independent of whether that response is rewarded or not (Frank, 2005; Frank & Claus, 2006). This effect allows the cortical units to identify candidate responses based on their prior frequency of choice, providing an initial “best guess” on the suitability of a given action which can then be facilitated or suppressed by the BG based on Go/NoGo reinforcement values. Once these cortical associations are strong enough, they may not need be facilitated by the BG at all, consistent with data suggesting that striatal dopamine is necessary for initial acquisition of learned behaviors, but much less so for their later expression (Smith-Roe & Kelley, 2000; Parkinson, Dalley, Cardinal, Bamford, Fehnert, Lachenal, Rudarakanchana, Halkerston, Robbins, & Everitt, 2002). Similarly, inactivation of the dorsal striatum impairs execution of a learned task, but this effect is minimal once the behavior has been ingrained (Atallah, Lopez-Paniagua, Rudy, & O'Reilly, 2007). According to the model, habit *learning* is dependent on the striatal dopamine system for acquiring responses that lead to rewards, but its expression is mediated by more direct cortico-cortical associations (which, if strong enough, do not require the additional striatal “boost”). Note that this cortical learning implies that eventually premotor cortical areas participate in reward-based action selection themselves – such that responses chosen often in the past immediately take precedence over other options, prior to any facilitation by the BG.

2.3 Limitations and comparison with anatomy of real brains

Our model is far from capturing all the interesting complexity associated with real basal ganglia circuits. Indeed, the basal ganglia are considerably more complex than what is described in the above paragraphs. Although we have simulated various dynamic and anatomical projections

that are not part of the classical model, our model nevertheless continues to be highly simplified, and for any model it is always legitimate to question whether these simplified principles are relevant for the real system. Here we summarize some of the challenges to the framework.

2.3.1 Are Go and NoGo pathways truly segregated?—Despite the success of the classical BG model in providing a predictive framework for interpreting several patterns of data across multiple levels of analysis, there have been several challenges to the basic tenets of the model. First, the model relies on the segregation of D1 and D2 receptors in striatonigral and striatopallidal neurons (Gerfen, 1992; Gerfen & Keefe, 1994; Bloch & LeMoine, 1994; Le Moine & Bloch, 1995; Gerfen, Keefe, & Gauda, 1995; Ince, Ciliax, & Levey, 1997; Aubert, Ghorayeb, Normand, & Bloch, 2000). Earlier challenges suggested that, in fact, D1 and D2 receptors are co-localized on the same neurons, even if this co-localization is small relative to the overall expression of one or the other receptor type (Surmeier, Song, & Yan, 1996; Aizman, Brismar, Uhlen, Zettergren, Levet, Forssberg, Greengrad, & Aperia, 2000). More recent advances, most notably with transgenic mice, have all but put to rest this concern (Surmeier et al., 2007). Nevertheless, a remaining critical challenge is that efferent projections of striatonigral and striatopallidal neurons themselves may not be as clearly segregated as they are in the model. In fact, it appears that although 'striatopallidal' cells exist that project solely to GPe (NoGo cells, in the parlance of our model), many 'striatonigral' (Go cells) also have axon collaterals projecting to GPe (Kawaguchi, Wilson, & Emson, 1990; Lévesque & Parent, 2005; Wu, Richard, & Parent, 2000). On the surface this seems to challenge the idea that 'Go cells' function as such, given that they also project to GPe. However, we argue that this setup is actually useful for ensuring that the activation of the Go pathway remains transient, and implies that the GPI computes the temporal derivative of Go signals rather than raw Go signals (Frank, 2006). That is, because direct projections from striatum to GPI are monosynaptic whereas those to GPe and then GPI are polysynaptic, Go signals will first disinhibit the thalamus, followed by a delayed re-inhibition of the thalamus via the GPe route. This type of system is amenable to rapid facilitation and subsequent inhibition of representations, which would be relevant if a sequence of motor commands or items in working memory had to be activated in succession.

2.3.2 Role of Striatal Interneurons—For simplicity, our model does not explicitly incorporate functions of cholinergic (tonically active) interneurons, and we are only beginning to explore the role of GABA-ergic (fast-spiking) interneurons (Figure 1), which together make up $\sim 5\%$ of striatal neurons (Tepper & Bolam, 2004; Gerfen & Wilson, 1996). The relatively small proportion of these cell types does not necessarily diminish their potential functional significance. For example, cholinergic interneurons are known to be deeply involved in reward-based learning, and they respond dynamically to stimuli as they become predictive of reward (Wilson, Chang, & Kitai, 1990; Aosaki, Tsubokawa, Ishida, Watanabe, Graybiel, & Kimura, 1994). Cholinergic neurons also play a permissive role in striatal long-term plasticity changes (Centonze, Gubellini, Bernardi, & Calabresi, 1999) and appear to indirectly mediate some effects of dopamine-induced plasticity (Wang, Kai, Day, Ronesi, Yin, Ding, Tkatch, Lovinger, & Surmeier, 2006). Recent evidence suggests that acetylcholine and dopamine may play a cooperative role in reward-based learning: during salient events, midbrain dopamine cells and striatal cholinergic cells respond during the same temporal window, but only the dopamine cells fire in proportion to reward probability (Morris, Arkadir, Nevet, Vaadia, & Bergman, 2004). It was argued that the pause in cholinergic firing may serve as a "temporal frame" that determines when to learn based on the magnitude of the dopaminergic signal. Further, Cragg (2006) suggested that the cholinergic pause provides a contrast enhancement effect that discriminates between tonic and phasic dopaminergic states, effectively enhancing learning due to both dopamine bursts and dips. This effect partially arises due to presynaptic effects of acetylcholine on dopamine release via nicotinic receptors (Cragg, 2006). Thus, it is possible

that whereas dopamine facilitates *what* to learn, cholinergic interneurons facilitate *when* to learn.

Although none of these effects is simulated at the biophysical level in our model, we nevertheless implicitly incorporate some of them. That is, the equations that govern learning in our model amount to a form of *contrastive Hebbian learning* in which the effects of phasic dopamine signals on Go/NoGo activity are computed *relative* to those in the immediately preceding states (during which dopaminergic signals are tonic). Thus this mechanism automatically ensures that learning occurs during the correct temporal window and also provides a contrast between tonic and phasic states; both of these functions may be supported by the pause in cholinergic firing, as proposed above (Morris et al., 2004; Cragg, 2006). Nevertheless, it is undoubtedly the case that these interactions are considerably more complex, and may benefit from more explicit simulation.

2.3.3 Thalamic Back-projections—In addition to the recurrent projections between thalamus and frontal cortex, and the feedforward projections from GPi to thalamus, there are also often-neglected back-projections from the parafascicular thalamus to both the striatum and the subthalamic nucleus (e.g., Mouroux & Féger, 1993; Castle, Aymerich, Sanchez-Escobar, Gonzalo, Obeso, & Lanciego, 2005). Given that thalamostriatal projections synapse primarily on cholinergic interneurons and regulate cholinergic efflux (Lapper & Bolam, 1992; Zackheim & Abercrombie, 2005), it is possible the parafascicular thalamus provides an alerting signal during salient events that induces a pause in cholinergic firing and promotes learning. Further, preliminary (unpublished) simulations in our model suggest that back-projections from thalamus to the STN (Castle et al., 2005) might play a role in terminating a motor response once it has been disinhibited.

2.3.4 Ventral vs. Dorsal Striatum—Although the striatum in our and in several other models appears as a unitary structure, it in fact comprises several subregions. These subregions follow a ventromedial to dorsolateral gradient, with afferents from a roughly parallel gradient in the cortex (Haber, 2003; Cohen, Lombardo, & Blumenfeld, 2008). Although precise boundaries between subregions can be difficult to define based on cytoarchitectonic properties (Voorn, Vanderschuren, Groenewegen, Robbiins, & Pennartz, 2004; Liu & Graybiel, 1998), subregions can be delineated by their patterns of input/output fibers (Haber et al., 2000), and, in some cases, by functional dissociations (Cardinal, 2006; Pothuizen, Jongen-Rêlo, Feldon, & Yee, 2005; Atallah et al., 2007; O'Doherty, Dayan, Schultz, Deichmann, Friston, & Dolan, 2004). Dorsal striatal regions are richly interconnected with dorsal prefrontal regions, and therefore are thought to play a central role in modulating cognitive operations such as working memory updating (Frank et al., 2001; Collins, Wilkinson, Everitt, Robbins, & Roberts, 2000; Saint-Cyr, Taylor, & Lang, 1988). In contrast, ventromedial regions, including the nucleus accumbens, are more implicated in reinforcement-guided learning and addiction-related processes (Cardinal, Parkinson, Hall, & Everitt, 2002; Everitt & Robbins, 2005; Koob & Le Moal, 1997). Further distinctions can be made within the nucleus accumbens, between the shell and core regions.

One classic interpretation of the ventral/dorsal functional dissociation in the realm of reinforcement learning has been that between the “critic” and the “actor” (Joel, Niv, & Ruppin, 2002; Houk, Adams, & Barto, 1995). The critic, played by the ventral striatum, evaluates whether the current environmental state is predictive of reward, and learns to do so by experiencing rewards in particular states. Changes in phasic dopamine responses during unexpected rewards (or lack thereof) are thought to drive learning in the critic so that its predictions are more accurate in the future. In contrast, the actor – played by the dorsal striatum – determines which actions to select, and learns to do so via these same phasic dopamine signals following the execution of particular actions, such that it develops action-specific value

representations. (Note that once the critic has learned, it will generate a dopamine burst when encountering an environmental state that is predictive of future reward, which serves to train the actor to produce actions that produced this state – even if they don't immediately precede reward itself). Although evidence exists in favor this viewpoint (Joel et al., 2002; O'Doherty et al., 2004), the story is likely to be more complex (Atallah et al., 2007).

Based on the modeling framework presented above, we would argue that different subregions of the striatum engage in similar computations and interactions with frontal cortex, but that the kind of information that is processed in different regions depends on the subregion of frontal cortex with which the striatal subregion interacts (see also Wickens, Budd, Hyland, and Arbuthnott (2007)). For example, because the dorsal striatum is most densely innervated by dorsal and lateral prefrontal regions, it might gate information flow related to processes engaged by dorsolateral prefrontal cortex, namely working memory, planning, cognitive control, etc (Frank et al., 2001; O'Reilly & Frank, 2006). In contrast, the ventral striatum, with dense connectivity from the orbitofrontal cortex and ventromedial prefrontal cortex, might gate information regarding reward and motivation (Frank & Claus, 2006). Other parts of the accumbens are likely to be involved in learning which environmental states (both external and internal) are associated with reward so that they can drive dopamine signals and train the actor (O'Reilly, Frank, Hazy, & Watz, 2007; Brown, Bullock, & Grossberg, 1999).

More recently, O'Reilly and colleagues have proposed an expanded model of the neurobiological mechanism of dopamine-mediated learning. In the PVLV (primary value-learned value) model (O'Reilly et al., 2007), the single node that corresponded to the SNc is now a network of regions including the ventral striatum, lateral hypothalamus, central nucleus of the amygdala, and SNc. The primary value (PV) system, mediated by patch-like striosomal neurons in the ventral striatum, is responsible for learning when unconditioned rewards will occur, and act to cancel out the dopamine burst when these are expected (due to inhibitory projections from striosomes into SNc and VTA (Joel & Weiner, 2000)). The activity resulting from the PV system matches the initial increase and subsequent decrease of dopamine neuron activity as animals learn to anticipate primary rewards.

The learned value (LV) system of the model learns to assign reward value to arbitrary stimuli that are predictive of later reward (i.e., conditioned stimuli). Learning in this system occurs only if an external reward is present *or* the PV system expects primary reward – that is LV learning is gated by PV activation. In this way, the LV system can express *generalized* reward value at times during which no reward is present in the environment (in contrast, the PV system always learns about rewards or their absence and so does not express reward values in advance of their occurrence). The LV is represented by the central nucleus of the amygdala, which is heavily involved in reward learning and sends excitatory projections to midbrain dopamine neurons. This system is more biologically plausible than previous mathematical estimations of the midbrain dopamine system's functioning using temporal difference learning, and is more robust than that system under certain circumstances (e.g., stimulus-reward timing variability and sensitivity to intervening distracting stimuli (O'Reilly et al., 2007)).

In sum, despite the incompleteness of our computational model, brains are more than the sum of their complex synaptic, neural, and chemical parts: Brains can learn and engage in an impressive array of cognitive and behavioral processes. In this sense, the modeling approach described above is biologically relevant, because, as detailed in the next section, the model can produce outputs that are similar to those of biological organisms, and the model's behavior is modulated from simulations of drugs, disease states, and genetic variation. Thus, the purpose of the neural network approach to modeling is not to capture every known aspect of the neurobiology of the basal ganglia, but instead to relate the key elements of basal ganglia neurobiology to cognitive and behavioral processes.

In the next section, we describe some of the predictions of the model that have been confirmed by empirical results.

2.4 Empirical evidence for predictions from basal ganglia models

This basal ganglia model makes several testable and falsifiable predictions regarding behavioral and neural responses during reinforcement learning, and how those responses should be modulated by drug, disease, or genetic states. The initial model was designed to be constrained by physiological and anatomical data, but also to account for cognitive changes resulting from Parkinson's disease and medication states, including complex probabilistic discrimination between reinforcement values and reversal (Frank, 2005), and the role of the subthalamic nucleus in high-conflict decisions (Frank, 2006).

2.4.1 Dopaminergic modulation of Go and NoGo learning—At the neural level, the model predicted the existence of separate striatal populations that code for positive and negative stimulus-response action values. Such neurons have since been reported in monkeys (Samejima, Ueda, Doya, & Kimura, 2005), although it remains to be determined whether these correspond to the Go and NoGo units (i.e., striatonigral vs striatopallidal), but synaptic plasticity studies support the model's predictions regarding how these separate populations might emerge via differential D1 and D2 receptor mechanisms for potentiating synapses in Go and NoGo synapses (Shen et al., 2008).

At the behavioral level, monkeys' ability to speed reaction times to obtain large rewards (requiring Go learning in our model) is dependent on striatal D1 receptor stimulation, whereas the tendency to slow down for smaller rewards (NoGo learning) is dependent on D2 receptor disinhibition (Nakamura & Hikosaka, 2006). Similarly, our computational model has simulated a constellation of reported findings regarding D2 receptor antagonism effects on expression of catalepsy in rodents, as a form of NoGo learning, including sensitization, context dependency, and extinction (Wiecki, Riedinger, Meyerhofer, Schmidt, & Frank, submitted).

In humans, a direct model prediction is that the ability to learn from positive versus negative feedback should depend on Go and NoGo learning, the balance of which depends on the level of dopamine. Phasic bursts of dopamine promote Go learning from positive feedback, whereas phasic dips promote NoGo learning from negative feedback (Frank, 2005). If these phasic levels of dopamine were modulated or compromised by disease or pharmacology, the way that individuals learn from positive vs. negative feedback should likewise be modulated. Patients with Parkinson's disease provide an opportunity to test these hypotheses: These patients have reduced dopamine signaling when off their medication, but enhanced dopamine levels when on their medication. Previous research has found that Parkinson's patients are impaired at reinforcement learning as a function of feedback (Swainson, Rogers, Sahakian, Summers, Polkey, & Robbins, 2000; Shohamy, Myers, Grossman, Sage, Gluck, & Poldrack, 2004; Cools, 2006; Cools, Barker, Sahakian, & Robbins, 2001a), linked to low levels of dopamine in the striatum and prefrontal cortex. One might therefore expect that dopamine medication would improve performance in these patients. Curiously, however, performance can be improved or impaired depending on which cognitive task is used (Cools, Barker, Sahakian, & Robbins, 2001b; Shohamy, Myers, Gekhman, Sage, & Gluck, 2006; Frank, 2005; Frank et al., 2004).

The computational model might help clarify this apparent inconsistency. Specifically, the model predicts that dopamine levels should differentially affect learning from negative versus positive feedback. When patients are off their medication, they should learn better from negative than from positive feedback, because low levels of dopamine activate the NoGo pathway (e.g., Surmeier et al., 2007), and, together with D2 receptor supersensitivity, may facilitate the detection of DA dips, but prevent the Go pathway from being sufficiently activated during rewards. In contrast, when patients are on their medication, presynaptic dopamine

synthesis increases (Tedroff, Pedersen, Aquilonius, Hartvig, Jacobsson, & Långström, 1996; Pavese, Evans, Tai, Hotton, Brooks, Lees, & Piccini, 2006). Moreover, chronic administration of levodopa (the main DA medications used to treat PD) has been shown to increase phasic (spike-dependent) DA bursts (Harden & Grace, 1995; Wightman, Amatore, Engstrom, Hale, Kristensen, Kuhr, & May, 1988; Keller, Kuhr, Wightman, & Zigmond, 1988), and the expression of *zif-268*, an immediate early gene that has been linked with synaptic plasticity (Knapska & Kaczmarek, 2004), in striatonigral (Go), but not striatopallidal (NoGo) neurons (Carta, Tronci, Pinna, & Morelli, 2005). Thus, the model predicts that medication improves positive feedback learning in the Go pathway. Interestingly, the same model predicts that dopamine medication will impair the ability to learn from negative feedback: because the medication continually stimulates D2 receptors¹, they effectively preclude phasic pauses in DA firing from being detected when rewards are omitted (Frank, 2005).

This pattern of results was recently confirmed in Parkinson's patients who tested on and off their medication in a probabilistic reinforcement learning paradigm in which some choices had greater probabilities of being associated with positive and negative feedback (Frank et al., 2004). Patients off their medication learned better from negative than from positive feedback, whereas patients on medication learned better from positive than from negative feedback. These effects were also produced when DA depletion and medications were simulated in the model (Frank et al., 2004; Frank et al., 2007b), and have been replicated using a different paradigm in a different lab (Cools, Altamirano, & D'Esposito, 2006). Moreover, they are in striking accord with the synaptic plasticity studies described above, in which DA depletion was associated with reduced D1-related potentiation of Go synapses but enhanced D2-related potentiation of NoGo synapses, whereas D2 agonist administration reversed the potentiation of NoGo synapses (Shen et al., 2008). Notably, similar patterns of behavioral results (enhanced Go but reduced NoGo learning) have been reported in mice with genetic knockouts of the dopamine transporter, who have elevated striatal dopamine levels (Costa, Gutierrez, de Araujo, Coelho, Kloth, Gainetdinov, Caron, Nicoletis, & Simon, 2007). All of these findings confirm that dopamine is critically involved in learning not only from positive but also negative prediction errors.

This same modulation of probabilistic Go and NoGo learning has also been observed in young, healthy college students who took small doses of dopamine agonists and antagonists (Frank & O'Reilly, 2006). Further, aged adults (older than 70 years of age), who have striatal DA depletion and damage to DA cell integrity (Bäckman, Ginovart, Dixon, Wahlin, Wahlin, Halldin, & Farde, 2000; Kaasinen & Rinne, 2002; Kravtsov, Kudryavtseva, McKee, Geula, Kowall, & Khrapko, 2006), showed selectively *better* negative feedback learning than their younger counterparts (60–70 years of age), consistent with the Parkinson's findings (Frank & Kong, 2008). The opposite pattern of results was seen in adult ADHD participants, who showed better positive than negative feedback learning while on stimulant medications (Frank, Santamaria, O'Reilly, & Willcutt, 2007c), which block the dopamine transporter and elevate striatal DA (Volkow, Wang, Fowler, Logan, Gerasimov, Maynard, Ding, Gatley, Gifford, & Franceschi, 2001; Madras, Miller, & Fischman, 2005). In sum, across a wide range of populations and manipulations, increases in striatal dopamine are associated with relatively better Go learning and especially, worse NoGo learning, whereas decreases in striatal dopamine is associated with the opposite pattern.

Behaviorally, the model suggests that having independent Go and NoGo pathways improves probabilistic discrimination between different reinforcement probabilities. That is, networks learning only from positive feedback or only from negative feedback do not produce as robust

¹This assumption most clearly applies to D2 agonist medications taken by the majority of patients, but also potentially by levodopa, which may increase tonic DA release in addition to phasic bursts.

learning as those receiving both positive and negative feedback (even if the number of feedback trials is equated). Such a pattern was recently found in a basal ganglia-dependent probabilistic learning task (Ashby & O'Brien, 2007), in which it was concluded that the dual pathway Go/NoGo model is required to capture the basic behavioral findings.

Although we have found in our probabilistic reinforcement paradigm that on average healthy individuals learn equally well from positive and negative feedback, there are nevertheless substantial individual differences in these measures, such that some participants are “positive learners” and some are “negative learners” (Frank, Woroch, & Curran, 2005). We hypothesized that at least some of this variability may be due to genetic factors controlling striatal dopaminergic function. To test this hypothesis, we collected DNA from 69 healthy participants and tested them with the same probabilistic reinforcement learning task (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007a). If individual differences in Go learned are attributed to D1 function and NoGo learning to D2 function, genetic factors controlling striatal D1 and D2 efficacy may be predictive of such learning. Because there is not yet a genetic polymorphism shown to preferentially affect striatal D1 receptors, we analyzed instead a polymorphism that controls the protein DARPP-32, which is heavily concentrated in the striatum, and is required for D1-dependent plasticity and reward learning in animals (Ouimet, Miller, Hemmings, Walaass, & Greengard, 1984; Walaas, Aswad, & Greengard, 1983; Calabresi et al., 2000; Stipanovich, Valjent, Matamales, Nishi, Ahn, Maroteaux, Bertran-Gonzalez, Bami-Cherrier, Enslin, Corbillé, Filhol, Nairn, Greengard, Hervé, & Girault, 2008). Furthermore, in humans, the only brain area that was functionally modulated according to DARPP-32 genotype was the striatum, and its functional connectivity with frontal cortex (Meyer-Lindenberg, Straub, Lipska, Verchinski, Goldberg, Callicott, Egan, Huffaker, Mattay, Kolachana, Kleinman, & Weinberger, 2007). We also analyzed a polymorphism within the DRD2 gene, which codes for postsynaptic striatal D2 receptor density (Hirvonen, Laakso, Rinne, Pohjalainen, & Hietala, 2005). Strikingly, we found that individual differences in DARPP-32 genetic function, as a surrogate measure of striatal D1-dependent plasticity, were predictive of better positive feedback learning, whereas individual differences in DRD2 function, as a measure of striatal D2 receptor density, were predictive of better negative feedback learning (Frank et al., 2007a). This latter effect was also found independently by another group (Klein, Neumann, Reuter, Hennig, von Cramon, & Ullsperger, 2007), who analyzed a different DRD2 polymorphism. Moreover, the Go/NoGo learning effects were specific to striatal genetic function, as a third gene coding primarily for prefrontal dopaminergic function (Tunbridge, Bannerman, Sharp, & Harrison, 2004), was not associated with Go or NoGo incremental probabilistic learning, but instead – and in contrast to the striatal genes – was predictive of participants’ working memory for the most reinforcement outcomes (Frank et al., 2007a). This working memory effect is consistent with other detailed computational models suggesting that prefrontal dopamine is critical for robust maintenance of information in an active state (Durstewitz, Seamans, & Sejnowski, 2000), and that parts of prefrontal cortex support working memory for reward values, guiding trial-to-trial behavioral adaptations and complementing the incrementally learning basal ganglia system (Frank & Claus, 2006).

Such clear genetic findings – where distinct polymorphisms having different functional brain effects are associated with dissociable cognitive functions – are rare in the literature, and without a computational model, it is unlikely that these specific genes would have otherwise been analyzed in the context of these specific types of decisions. Nevertheless, the nature and direction of the prefrontal dopaminergic genetic effects, despite being consistent with the general role of prefrontal cortex in rapid trial to trial adaptations, were inconsistent with our existing model of the role of DA in that system (Frank & Claus, 2006), which will lead us to revisit and refine that model (i.e., such that prefrontal DA plays a role closer to that suggested by Durstewitz et al. (2000).)

2.4.2 Subthalamic nucleus in high-conflict decisions—The basal ganglia model also makes predictions for other non-dopaminergic and non-learning aspects of decision making. As described in a previous section, the subthalamic nucleus (STN) projects diffusely to BG output nuclei (GPi and GPe), and receives direct excitatory input via the hyperdirect pathway from dorsomedial frontal cortex (Nambu et al., 2000; Aron et al., 2007). The model implicates the STN in preventing impulsive decisions, by dynamically (and transiently) adjusting decision thresholds as options are being considered (Frank, 2006). Such a role would be evident when making decisions involving a high degree of response conflict. Neuroimaging studies support this conclusion, whereby increased co-activation between dorsomedial frontal cortex and STN is associated with increasingly slowed response times in high but not low conflict conditions (Aron et al., 2007).

To demonstrate that the STN provides a critical (rather than correlational) role in slowing responses under conflict, it has to be manipulated. Parkinson's patients with deep brain stimulators (DBS) implanted into the STN provide a unique window into the role of the STN in human conflict-related decisions. These stimulators provide electrical current into the STN at abnormally high frequency and voltage, disrupting STN function, effectively acting like a lesion (or like adding noise to the system, preventing it from responding naturally to its cortical inputs) (Benabid, 2003; Benazzouz & Hallett, 2000; Meissner, Leblois, Hansel, Bioulac, Gross, Benazzouz, & Boraud, 2005). However, this virtual lesion is temporary, because the stimulator can be turned on or off by a physician. While the stimulator is switched on, many of the motor-related symptoms of Parkinson's disease are sharply diminished; within minutes to an hour after the stimulator is switched off, symptoms return. In a recent study, Frank and colleagues tested these patients in a reinforcement learning task on and off stimulation, and compared their performance to another group of patients on and off dopaminergic medication (Frank et al., 2007b). High conflict decisions were defined as those choices in which the probability of reinforcement between the two options differed only subtly (e.g., one option had an 80% chance of being rewarded whereas the other had a 70% chance), whereas low conflict decisions were characterized by choices involving disparate reinforcement probabilities (e.g., 80% vs 30%).

Typically, when faced with these high-conflict choices, response times slow down; this pattern was observed in healthy controls, in patients off and on medication, and in patients off DBS. Notably, patients on DBS failed to slow reaction times with increased decision conflict (Figure 3b). Moreover, patients on DBS actually responded *faster* to high than to low conflict choices. These speeded high conflict decision times were even more exaggerated when patients selected the suboptimal choice (that with lower reinforcement probability; (Frank et al., 2007b)), suggesting that the stimulation disrupted the STN's ability to provide a global NoGo signal during high-conflict decisions. Further, when the model was given a STN lesion or when simulated high frequency DBS was applied, it produced the same pattern of results. Together with the medication effects reported above (and replicated in the 2007 study), these findings reveal a double dissociation of treatment type on two aspects of cognitive decision making in PD: Dopaminergic medication influences positive/negative learning biases but not conflict-induced slowing, whereas DBS influences conflict-induced slowing but not positive/negative learning biases.

In sum, although our neural model is simplified relative to the complexity of real basal ganglia circuitry, and abstracts away a host of biophysical and molecular mechanisms, the modeling endeavour has proved to be a valuable tool in developing explicit testable and falsifiable hypotheses, which directly led to empirical experiments providing support for several of these predictions. Nevertheless, we acknowledge that some of the detailed mechanisms by which our model functions, while neurally plausible, are likely over-simplified. We look forward to further refinements and challenging data that will cause us to revisit some of the basic mechanisms.

3 Abstract models of action selection and learning

In contrast to the neural network models described in the previous section, abstract models typically do not capture neurobiological or neuroanatomical processes, but instead focus on the nature of cognitive operations that might lead to specific behavioral outputs, such as learning and decision-making. Although these models have been linked to neurobiological events, and in some cases, incorporate specific neural processes such as the effects of dopamine (Wörgötter & Porr, 2005; Cohen, 2007), these models typically are not constrained by known biological limitations (incorporating neither anatomy nor physiology). Nonetheless, by adapting a 'top-down' functional approach, these models have proven valuable in uncovering the cognitive mechanisms of reward-guided learning and decision-making, and have made several strides in linking these mechanisms to the neurobiology of the basal ganglia, prefrontal cortical, and dopamine systems (Cohen, 2007; Montague, Dayan, & Sejnowski, 1996; Daw, Niv, & Dayan, 2005; O'Doherty et al., 2004).

3.1 The math behind the models

We focus on models that have been used most extensively in understanding basal ganglia functioning. The basic learning mechanism behind these reinforcement learning models can be summarized semantically by Thorndike's Law of Effect (Thorndike, 1911): "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; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond."

In other words, actions associated with positive feedback are more likely to be repeated, whereas actions associated with negative feedback are less likely to be repeated. In models, different actions may be represented with "Q values"; the larger the Q value relative to that of other actions, the more likely the model is to select that action. Nevertheless, the choice function "policy" is typically probabilistic, such that sometimes other choices with lower Q values are selected. This ensures that the model occasionally *explores* alternative actions, thus avoiding situations in which other decision options provide higher rewards but are not selected because the model is stuck continually choosing one decision option (i.e., a local minimum) (Sutton & Barto, 1998). The most common choice function used is termed *softmax* because it assigns a higher probability of choosing the action with the maximum Q value, but the arbitration between Q-values is soft, such that those with only slightly smaller values are almost as likely to be chosen. The slope of the softmax function determines the degree to which maximum Q values are chosen versus the probability of making an exploratory choice (Sutton & Barto, 1998; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Frank et al., 2007a).

To learn which Q values lead to the highest rewards, Q values are adjusted following reinforcements. The most commonly used method for updating Q values is through a reward prediction error, which is the difference between an expected and received reward: $\delta = r - Q$, where δ is the prediction error, r is the reward, and Q is the value of the weight corresponding to the action selected.² This prediction error term might reflect phasic activity of midbrain

²In more sophisticated algorithms the prediction error takes into account not only the current reward but also the predicted reward for future trials, based on prior learning, and where rewards further into the future are discounted (Watkins & Dayan, 1992). This function is important for allowing a reinforcement learning agent to learn not only which actions lead to immediate rewards, but which actions to reinforce when their consequences occur later in time, and to maximize total future rewards (Sutton & Barto, 1998). Nevertheless, we restrict our discussion here to the simple case in which simple actions lead to immediate rewards or lack thereof.

dopamine neurons, described in more detail below (Suri & Schultz, 1998). Thus, when rewards that follow particular actions are greater than the reward expected from that particular action (i.e., the Q value), the prediction error is positive; when rewards are received exactly as expected, the prediction error is zero; and when rewards are smaller than those expected, the prediction error is negative. These prediction errors then adjust the Q value in the subsequent trial: $Q(t+1) = Q(t) + \delta$, where t refers to a trial. Q values that led to rewards (punishments) are strengthened (weakened), thus becoming more (less) likely to be selected in subsequent trials. Thus, the Q value updating equation can be seen as a concise mathematical representation of part of Thorndike's Law of Effect. Note that prediction error terms are multiplied by a learning rate, which scales the impact of the prediction error on the subsequent Q value: $Q(t+1) = Q(t) + \alpha * \delta$.

The learning rate describes the degree to which the prediction error adjusts the Q values, and might correspond to the relative number of AMPA receptors that are mobilized from a single learning experience. These learning rates might also differ among different brain regions. For example, the hippocampal learning system is capable of rapid, single trial learning (high learning rate), whereas the basal ganglia learning system learns by integrating more slowly over time, thus utilizing a lower learning rate. Thus, one important computational issue is determining when to use systems with high versus low learning rates. Some have proposed that the amount of uncertainty plays a role in determining the learning rate (Behrens, Woolrich, Walton, & Rushworth, 2007; Daw et al., 2005; Yu & Dayan, 2005).

Learning Q values might also help explain how we form habits, as is formalized in "Advantage learning" (Dayan & Balleine, 2002). Advantage learning theory states that actions are chosen when the value associated with that action exceeds the average value of the entire set of possible actions at that state (e.g., point in time). Over time, as agents learn optimal response strategies, the advantage of a particular action declines because the overall value of that action state increases. At that point, action selection becomes more automatic, and a stimulus-response habit is formed. Our neural models also show a similar transition from choosing actions according to rewards to choosing actions by habit. In our neural models, this transition occurs gradually over many trials through slow Hebbian learning in cortico-cortical projections, as described above.

Note that the basic principles of reinforcement learning – strengthening representations of rewarded actions while weakening representations of nonrewarded actions – is conserved between the neural network and abstract models. The neural network models are more concerned with putative neural implementation whereas "Q models" abstract the neural implementation in favor of focusing on the essential computational implementation.

Many models that have been used to understand basal ganglia functions are more elaborate and sophisticated than these simple equations. For example, other abstract models address how animals arbitrate between a BG-based habitual system versus a more goal directed system localized in the prefrontal cortex (Daw et al., 2005), when to explore in a dynamic probabilistic environment (Daw et al., 2006; McClure, Gilzenrat, & Cohen, 2006), how much vigor to respond with in variable reward schedules (Niv, Daw, Joel, & Dayan, 2007), and when to supplement basic dopamine mediated reinforcement learning with an explicit rule for detecting when the environment has changed (Hampton, Bossaerts, & O'Doherty, 2006). Nevertheless, the above basic equations are robust in many situations and continue to form the "backbone" of these more sophisticated models. Although many aspects of neurobiology are not incorporated into these models (e.g., membrane potential dynamics, different actions at D1 vs. D2 receptors, role of different BG nuclei), these equations predict activity in specific striatal and prefrontal regions, demonstrating the elegance of these simple but powerful models in

elucidating the computations engaged by the basal ganglia without requiring as many assumptions about the precise implementational form in neural circuitry.

3.2 Neurobiological correlates of abstract models

3.3 Neurobiology of prediction errors

Reward prediction errors have been proposed to be signaled by phasic bursting activity of midbrain dopamine cells in the ventral tegmental area and SNc. This burst induces rapid dopamine release in widespread regions of the striatum and limbic system. Like the prediction error term from reinforcement learning models described above, midbrain dopamine activity phasically increases when unexpected rewards are received, phasically decreases when expected rewards are not received, and does not change from baseline levels when expected rewards are received. Detailed reviews of this evidence can be found elsewhere (Schultz, 2002, 1998; Schultz & Dickinson, 2000).

The link between dopamine cell activity and prediction error terms from computational models has inspired many researchers using noninvasive neuroimaging techniques in humans to investigate neural correlates of reward prediction errors. For example, in functional MRI studies, computational reinforcement learning models, similar to that outlined above, have been used to generate reward prediction errors on each trial. These prediction errors are then used in a regression to identify brain regions areas in which activity correlates with prediction errors derived from the model. These correlations are often observed to be significant in the striatum and frontal cortex, as well as other regions (discussed in more depth below), and are taken to reflect reward prediction error signals from the midbrain to striatal circuitry (Cohen, 2007; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; O'Doherty, 2007; Seymour, O'Doherty, Dayan, Koltzenburg, Jones, Dolan, Friston, & Frackowiak, 2004).

In other work using scalp-recorded EEG in humans, researchers have identified a component called the error-related negativity (ERN) and the feedback-related negativity (FRN) that may reflect a reward prediction error signal (Yasuda, Sato, Miyawaki, Kumano, & Kuboki, 2004; Holroyd & Coles, 2002; Cohen & Ranganath, 2007; Frank et al., 2005; Nieuwenhuis, Holroyd, Mol, & Coles, 2004). These components are located at frontocentral scalp sites from around 200–400 ms following negative compared to positive feedback, or following error compared to correct responses. It has been proposed that the FRN reflects the impact of a negative reward prediction error signal originating in the midbrain dopamine system, which is then used to adapt reward-seeking behavior (Holroyd & Coles, 2002; Brown & Braver, 2005). This is consistent with findings that midbrain dopamine neurons project to, and can modulate activity in, pyramidal cells in the cingulate cortex (O'Connell & Wang, 2005). However, it is unclear whether the cingulate can detect DA dips, given the slow time-course of DA reuptake in frontal cortex (discussed above). Nevertheless, it is possible that these scalp-EEG recordings actually reflect the impact of DA dips in the BG, which activate the NoGo pathway, and then indirectly lead to changes in frontal cortical activity (e.g., via increased post-response “conflict” (Yeung, Botvinick, & Cohen, 2004)).

3.4 Neurobiology of action (Q) values

The other main component of these reinforcement learning models is the Q value, which represents specific actions or decisions. Although the possible neurobiological correlates of Q values has received less attention compared to the neurobiological correlates of prediction errors, evidence suggests that Q values in models might correspond to activity in brain regions responsible for planning and executing those specific actions. For example, activity of neurons in the striatum that represent specific actions (e.g., saccades to the right or left) is modulated by the amount of reward that would be obtained by correct responses (Samejima et al., 2005). In this study, the properties of these neurons were well fit by a Q learning algorithm. Further,

reward-related activity modulations in motor regions can bias decision-making and action selection processes (Gold & Shadlen, 2002; Schall, 2003; Sugrue, Corrado, & Newsome, 2004). Although these findings are not always discussed in terms of Q-values from computational learning models, the observations are consistent with the idea that Q-values or weights in models correspond to activity in sensory-motor systems. Preliminary evidence in humans suggests that activity in cortical motor regions might correspond to Q values. For example, Cohen and Ranganath (2007) reported that EEG activity over lateral frontal electrode sites (sites C3/4, typically taken to index motor cortex activity) resembled Q values obtained from a computational model, while the model played the same strategic game the human subjects played.

One important question is what a “Q” value means in the brain, and where it is stored. As described in the previous paragraph, for simple decisions in which each decision maps onto a particular action or response (e.g., saccade to the left, or pressing the right index finger), the Q value might correspond to the strength of the activation of that motor action in basal ganglia and/or cortical motor regions. But most decisions we face are more complex, and do not have specific, discrete motor actions associated with them (e.g., which college to attend? What to eat for dinner? Should I marry this person?). Relatedly, in some experiments, the same stimuli are associated with different motor responses in different trials. This is useful for counterbalancing motor response requirements, but leaves open the question of whether Q values in such experiments are linked to the stimulus representation, or whether they remain linked to a more abstract response representation that is flexible and changes according to task demands. One possibility is that multiple Q-like representations are maintained by different brain regions, and correspond to reward-modulated weights of different kinds of information. For example, the orbitofrontal cortex or ventral striatum might contain basic value representations of particular world states (divorced from action); the dorsal striatum and supplemental motor area might contain Q-like representations for specific motor actions; and dorsolateral or anterior prefrontal cortex might contain Q-like representations of more abstract goals or plans.

3.5 Individual differences

The equations for reinforcement learning described above are normative, in that they prescribe how all individuals should act and learn from reinforcements. However, human decision-making can be variable; myriad individual differences influence how people make decisions, and different individuals can act and learn quite differently, even when given the same reinforcements following the same actions. One advantage of abstract models is that they can be used to characterize mathematically such individual differences. This is done by fitting the model to each subjects' behavioral data and estimating some model variables through statistical fitting procedures. For example, one could estimate unique learning rates, which scales the impact of prediction errors on adjustments in Q values, for each subject. This approach has been successfully used to link behavioral task performance and brain activity in the basal ganglia and frontal cortex to individual differences in decision-making (Cohen & Ranganath, 2005; Cohen, 2007; Schönberg, Daw, Joel, & O'Doherty, 2007; Behrens et al., 2007). Frank and colleagues recently demonstrated that genetic polymorphisms related to the expression of dopamine receptors in the human striatum and prefrontal cortex are associated with different learning rates (Frank et al., 2007a). Further, separate learning rates for gains (Go) and losses (NoGo) were predicted by the DARPP-32 and DRD2 genes, providing a nice mapping onto the neural network model. Lee and colleagues have shown in monkeys that activity of prefrontal cortical cells is predicted by these estimated model parameters (Lee, Conroy, McGreevy, & Barraclough, 2004; Lee, McGreevy, & Barraclough, 2005).

When subjects vary widely in how they use reinforcements to adjust decision-making (e.g., in a gambling study in which there are no correct answers or policies to learn; Cohen & Ranganath, 2005), fitting model parameters to subjects' data can be critical to elucidating the neurocomputational mechanisms of decision-making. In these cases, ignoring individual differences (i.e., a normative approach) may lead to the misleading interpretation that the models cannot account for the data.

3.6 Uncertainties and inconsistencies in linking abstract models to neurobiology

Although extant literature has shown that activity in fronto-striatal circuits correlates with some aspects of abstract computational models, inconsistencies and uncertainties remain regarding what brain systems are involved to what extent, and how closely brain activity conforms to predictions from the abstract models. Some of this uncertainty is related to the fact that models are far more simplistic than real basal ganglia systems. For example, it is unlikely that the equations detailed above describe all internal mental processes engaged during experimental learning tasks, even in species with simple nervous systems; humans and animals are likely engaging mechanisms akin to these plus other complex and dynamic high level processes, such as hypothesis-testing.

One area of uncertainty concerns positive versus negative prediction errors. As described in the previous section, recent work suggests that the duration of dopamine cell firing pauses may encode negative prediction errors (Bayer et al., 2007). The serotonin systems has also been proposed to play a role in signaling negative prediction errors (Daw et al., 2002). The functional MRI literature is less clear on this issue: Some have found increased/decreased activity in fronto-striatal circuits for positive/negative prediction errors (Cohen, 2007; McClure, Berns, & Montague, 2003; O'Doherty et al., 2003) while others have found that ventral striatal activity correlated with positive prediction errors only (Yacubian, Sommer, Schroeder, Gläscher, Kalisch, Leuenberger, Braus, & Büchel, 2007). Yet others have suggested that different subregions of the striatum are involved in positive versus negative prediction errors (Seymour, Daw, Dayan, Singer, & Dolan, 2007). In studies that have reported activation in the midbrain, some have reported increased activity for rewards compared to punishments (Murray, Corlett, Clark, Pessiglione, Blackwell, Honey, Jones, Bullmore, Robbins, & Fletcher, 2008), others have reported increased midbrain activity for negative compared to positive feedback (Aron, Shohamy, Clark, Myers, Gluck, & Poldrack, 2004), and others have reported increases for positive prediction errors but nonsignificant decreases for negative prediction errors (D'Ardenne, McClure, Nystrom, & Cohen, 2008).

Another area of uncertainty concerns the precise regions in which activity correlates with prediction errors: Although the prediction error-correlated activity in the ventral striatum is commonly reported, different studies have also shown prediction error-like responses in the dorsal striatum, regions of the prefrontal cortex, including orbitofrontal, ventrolateral, and dorsolateral, midbrain, and cerebellum (McClure et al., 2003; Seymour et al., 2004; O'Doherty et al., 2003; Haruno & Kawato, 2006; Ramnani, Elliott, Athwal, & Passingham, 2004). It is possible that prediction errors are utilized by different networks in the brain depending on current task goals, although to our knowledge this has not been investigated.

There is another problem with the interpretation of the BOLD response, particularly with respect to the basal ganglia: The BOLD response is a temporally and spatially sluggish signal that does not distinguish activity of different types of neurons or different functional networks, especially if those networks are spatially overlapping. For example, networks of Go and NoGo cells are spatially overlapping, so one could not distinguish between these systems using functional MRI. Similarly, functional MRI cannot dissociate interneurons from medium spiny neurons, or oscillations of different frequencies, or different subregions within particular basal ganglia structures. This issues may become critically important if one assumes that, for

example, Go and NoGo cells are acting in opposition to each other. In this case, it is possible that there would be no differences in the BOLD response of the striatum between conditions of high Go and low NoGo activity compared to conditions of low Go and high NoGo activity.

Despite these inconsistencies across studies—which may be relatively minor compared to their commonalities—the theory that reward prediction errors are signaled by midbrain dopamine neurons has proven to be a remarkable one in its simplicity, elegance, and ability to tie together vastly different fields of research, from artificial intelligence to cellular electrophysiology to human neuroimaging. It continues to inspire new, creative, and interdisciplinary research, and has shed new light on the role of basal ganglia circuitry on reinforcement learning and decision-making.

4 Integrating neural network and abstract models

These two approaches to understanding the computational functions of the basal ganglia have traditionally been conducted separately, often by separate research groups. As outlined in previous sections, different models have different strengths and weaknesses. To the extent that their strengths and limitations match, combining these two modeling approaches might prove more fruitful than using either in isolation. For example, abstract models, but not neural network models, are amenable to estimating individual differences in a learning rates and other parameters, and relating these individual differences to performance or brain activity; in contrast, neural network models, but not abstract models, make specific predictions regarding how functional computations may arise via interactive dynamics among multiple brain areas, and in turn the effects of focal brain lesions, pharmacological manipulations, and genetics.

One way to combine these modeling approaches is to use abstract mathematical models to estimate learning parameters of neural network models, as if it were a human subject. That is, when estimating individual learning rates, abstract models are typically “fit” to account for a given subject’s actual trial-by-trial choices when faced with their particular sequence of reinforcements. One could instead apply the same procedure and treat the output of the neural network model as “behavioral choices”, and then use the abstract model to estimate learning rates used by the neural network model as an entire system (which may differ substantially from learning rates at a given synapse). This might prove useful in understanding the neurobiology of individual differences in behavioral learning rates. Although several studies have investigated individual differences in learning rates and correlates of those individual differences in behavior and brain activity, it remains unknown what neurobiological factors might lead different individuals to have different learning rates. Is it dopamine system response amplitude, concentration of dopamine receptors, or the efficacy of globus pallidus-thalamus efferents? Empirically, it might be difficult to determine the neuro-biological mechanisms that lead to differences in behavioral learning rates. However, this is where neural network models become useful: Various parameters in a neural network model could be manipulated, and the model could be tested in a virtual experiment. The resulting learning rates from different model versions (e.g., models with intact or impaired dopamine system functioning to simulate Parkinson’s disease) could be compared to different groups of subjects with different learning rates. If the learning rates from different model versions matched the learning rates from different subjects groups (e.g., subjects with different genotypes), one could conclude that the changes made to the model represent one biologically plausible mechanism by which different learning rates are achieved. Of course changes made to the neural network models should be driven by *a priori* hypotheses, constrained by physiological evidence. This would provide an important validation of the abstract models because, although activation in various regions of the brain correlate with model parameters that were derived from individual differences in learning rates (Cohen, 2007; Tanaka, Doya, Okada, Ueda, Okamoto, & Yamawaki, 2004), it remains unknown which biological processes could account for these differences.

5 Conclusions and future directions

Computational models such as those discussed here are theories, and, like all theories, are simplified, limited in scope, and likely to undergo significant revision as new empirical data refines our understanding. Many empirical papers often rely on conceptual models, but these are often static anatomical diagrams that lack the mathematical precision of the models reviewed here, and are often relatively more simplistic. The computational models discussed here are similar in the sense that they are simplistic versions that omit many details. However, computational models have distinct advantages over less mathematically grounded theories: They can go further by considering the computational functions the brain is trying to solve, the implementation of those computations, and the rich dynamics of the basal ganglia circuitry. Ultimately, patterns of data captured by particular models should be replicated by models one level above (for elegance, analytic tractability, and succinctness), and by models one level below (for exploring more biophysically detailed constraints and adjusting models accordingly).

The field of computational modeling, and especially modeling of the basal ganglia system, has grown considerably over the past few decades. We envision several parallel future directions of using computational modeling to understanding basal ganglia and related circuitry. It is likely that more researchers will use more complex and biologically detailed models, due to the emergence of new software that eases the entry into this field, as well as to advancements in computer hardware speed and efficiency. As computers become faster, and parallel processing becomes more commonly used, highly detailed neural models may be scaled up to a level where they can produce behaviorally and cognitively meaningful outputs. We also envision that computational models will be integrated more with empirical research, along the lines discussed in this review about uncovering the putative neural mechanisms of prediction errors and related reinforcement learning variables. Finally, insights from neurobiologically plausible computational models might increasingly find their way into domains outside neuroscience, such as artificial intelligence and robotics (Gurney, Prescott, Wickens, & Redgrave, 2004).

Acknowledgments

This research was supported by NIDA grant DA022630 and NIMH grant MH080066–01 awarded to M.J.F. We thank Christina Figueroa for help with figure preparation.

6 References

- Aizman O, Brismar H, Uhlen P, Zettergren E, Levet AI, Forssberg H, Greengard P, Aperia A. Anatomical and physiological evidence for D₁ and D₂ dopamine receptor colocalization in neostriatal neurons. *Nature Neuroscience* 2000;3:226–230.
- Albin RL, Young AB, Penney JB. The functional anatomy of basal ganglia disorders. *Trends in Neurosciences* 1989;12:366–375. [PubMed: 2479133]
- Aosaki T, Tsubokawa H, Ishida A, Watanabe K, Graybiel AM, Kimura M. Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *Journal of Neuroscience* 1994;14:3969–3984. [PubMed: 8207500]
- Aron AR, Behrens TE, Smith S, Frank MJ, Poldrack RA. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (mri) and functional mri. *J Neurosci* 2007;27(14):3743–3752. [PubMed: 17409238]
- Aron AR, Shohamy D, Clark J, Myers C, Gluck MA, Poldrack RA. Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology* 2004;92(2):1144–1152. [PubMed: 15014103]
- Ashby FG, O'Brien JB. The effects of positive versus negative feedback on information-integration category learning. *Percept Psychophys* 2007;69(6):865–878. [PubMed: 18018967]

- Atallah HE, Lopez-Paniagua D, Rudy JW, O'Reilly RC. Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nature Neuroscience* 2007;10:126–131.
- Aubert I, Ghorayeb I, Normand E, Bloch B. Phenotypical characterization of the neurons expressing the D1 and D2 dopamine receptors in the monkey striatum. *Journal of Comparative Neurology* 2000;418:22–32. [PubMed: 10701753]
- Bäckman L, Ginovart N, Dixon RA, Wahlin TR, Wahlin A, Halldin C, Farde L. Age-related cognitive deficits mediated by changes in the striatal dopamine system. *American Journal of Psychiatry* 2000;157(4):635–637. [PubMed: 10739428]
- Baunez C, Christakou A, Chudasama Y, Forni C, Robbins TW. Bilateral high-frequency stimulation of the subthalamic nucleus on attentional performance: transient deleterious effects and enhanced motivation in both intact and parkinsonian rats. *Eur J Neurosci* 2007;25(4):1187–1194. [PubMed: 17331214]
- Baunez C, Robbins TW. Bilateral lesions of the subthalamic nucleus induce multiple deficits in an attentional task in rats. *European Journal of Neuroscience* 1997;9(10):2086–2099. [PubMed: 9421169]
- Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 2005;47(1):129–141. [PubMed: 15996553]
- Bayer HM, Lau B, Glimcher PW. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J Neurophysiol* 2007;98(3):1428–1439. [PubMed: 17615124]
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. *Nat Neurosci* 2007;10(9):1214–1221. [PubMed: 17676057]
- Benabid AL. Deep brain stimulation for parkinson's disease. *Current Opinion in Neurobiology* 2003;13(6):696–706. [PubMed: 14662371]
- Benazzouz A, Hallett M. Mechanism of action of deep brain stimulation. *Neurology* 2000;55:S13–S16. [PubMed: 11188968]
- Berke JD, Hyman SE. Addiction, dopamine, and the molecular mechanisms of memory. *Neuron* 2000;25(3):515–532. [PubMed: 10774721]
- Berretta S, Parthasarathy HB, Graybiel AM. Local release of GABAergic inhibition in the motor cortex induces immediate-early gene expression in indirect pathway neurons of the striatum. *J Neurosci* 1997;17(12):4752–4763. [PubMed: 9169535]. disinhibition of motor cortex (M1) leads to immediate early gene expression in NoGo neurons!
- Berretta S, Sachs Z, Graybiel AM. Cortically driven Fos induction in the striatum is amplified by local dopamine D2-class receptor blockade. *Eur J Neurosci* 1999;11(12):4309–4319. [PubMed: 10594656]. D2 blockade (systemic haloperidol or intra-striatal sulpiride) enhanced Fos induction in striatum in response to cortical disinhibition, in NoGo neurons! (see also Berretta et al 97).
- Bloch B, LeMoine C. Neostriatal dopamine receptors. *Trends in Neurosciences* 1994;17:3–4. [PubMed: 7511847]
- Bogacz R, Gurney K. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput* 2007;19(2):442–477. [PubMed: 17206871]
- Brown J, Bullock D, Grossberg S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience* 1999;19:10502–10511. [PubMed: 10575046]
- Brown JW, Braver TS. Learned predictions of error likelihood in the anterior cingulate cortex. *Science* 2005;307(5712):1118–1121. [PubMed: 15718473]
- Brown JW, Bullock D, Grossberg S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks* 2004;17:471–510. [PubMed: 15109680]
- Calabresi P, Gubellini P, Centonze D, Picconi B, Bernardi G, Chergui K, Svenningsson P, Fienberg AA, Greengard P. Dopamine and camp-regulated phosphoprotein 32 kda controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *J Neurosci* 2000;20(22):8443–8451. [PubMed: 11069952]
- Calabresi P, Pisani A, Centonze D, Bernardi G. Synaptic plasticity and physiological interactions between dopamine and glutamate in the striatum. *Neuroscience and Biobehavioral Reviews* 1997;21:519–523. [PubMed: 9195611]

- Cardinal RN. Neural systems implicated in delayed and probabilistic reinforcement. *Neural Networks* 2006;19:1277–1301. [PubMed: 16938431]
- Cardinal RN, Parkinson JA, Hall J, Everitt BJ. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews* 2002;26:321–52. [PubMed: 12034134]
- Carta AR, Tronci E, Pinna A, Morelli M. Different responsiveness of striatonigral and striatopallidal neurons to L-DOPA after a subchronic intermittent L-DOPA treatment. *Eur J Neurosci* 2005;21(5):1196–1204. [PubMed: 15813929]
- Castle M, Aymerich MS, Sanchez-Escobar C, Gonzalo N, Obeso JA, Lanciego JL. Thalamic innervation of the direct and indirect basal ganglia pathways in the rat: Ipsi- and contralateral projections. *J Comp Neurol* 2005;483(2):143–153. [PubMed: 15678473]
- Centonze D, Gubellini P, Bernardi G, Calabresi P. Permissive role of interneurons in corticostriatal synaptic plasticity. *Brain Res Brain Res Rev* 1999;31(1):1–5. [PubMed: 10611492]
- Centonze D, Picconi B, Gubellini P, Bernardi G, Calabresi P. Dopaminergic control of synaptic plasticity in the dorsal striatum. *European Journal of Neuroscience* 2001;13:1071–1077. [PubMed: 11285003]
- Chevalier G, Deniau JM. Disinhibition as a basic process in the expression of striatal functions. *Trends in Neurosciences* 1990;13:277–280. [PubMed: 1695403]
- Cohen MX. Individual differences and the neural representations of reward expectation and reward prediction error. *Soc Cogn Affect Neurosci* 2007;2(1):20–30. [PubMed: 17710118]
- Cohen MX, Lombardo MV, Blumenfeld RS. Covariance-based subdivision of the human striatum using t1-weighted mri. *European Journal of Neuroscience* 2008;27:1534–1446. [PubMed: 18364027]
- Cohen MX, Ranganath C. Behavioral and neural predictors of upcoming decisions. *Cogn Affect Behav Neurosci* 2005;5(2):117–126. [PubMed: 16180619]
- Cohen MX, Ranganath C. Reinforcement learning signals predict future decisions. *J Neurosci* 2007;27(2):371–378. [PubMed: 17215398]
- Collins P, Wilkinson LS, Everitt BJ, Robbins TW, Roberts AC. The effect of dopamine depletion from the caudate nucleus of the common marmoset (*callithrix jacchus*) on tests of prefrontal cognitive function. *Behavioral Neuroscience* 2000;114:3–17. [PubMed: 10718258]
- Cools R. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neurosci Biobehav Rev* 2006;30(1):1–23. [PubMed: 15935475]
- Cools R, Altamirano L, D'Esposito M. Reversal learning in parkinson's disease depends on medication status and outcome valence. *Neuropsychologia* 2006;44:1663–1673. [PubMed: 16730032]
- Cools R, Barker RA, Sahakian BJ, Robbins TW. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral Cortex* 2001a;11:1136–1143. [PubMed: 11709484]
- Cools R, Barker RA, Sahakian BJ, Robbins TW. Mechanisms of cognitive set flexibility in parkinson's disease. *Brain* 2001b;124:2503–2512. [PubMed: 11701603]
- Costa RM, Gutierrez R, de Araujo IE, Coelho MRP, Kloth AD, Gainetdinov RR, Caron MG, Nicoletti MAL, Simon SA. Dopamine levels modulate the updating of tastant values. *Genes Brain Behav* 2007;6(4):314–320. [PubMed: 16848782]
- Cragg SJ. Meaningful silences: how dopamine listens to the ach pause. *Trends Neurosci* 2006;29(3):125–131. [PubMed: 16443285]
- D'Ardenne K, McClure SM, Nystrom LE, Cohen JD. Bold responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 2008;319(5867):1264–1267. [PubMed: 18309087]
- Daw ND, Kakade S, Dayan P. Opponent interactions between serotonin and dopamine. *Neural Networks* 2002;15:603–616. [PubMed: 12371515]
- Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 2005;8(12):1704–1711. [PubMed: 16286932]
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature* 2006;441(7095):876–879. [PubMed: 16778890]
- Dayan, P.; Abbott, L. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press; Cambridge, MA: 1999.

- Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. *Neuron* 2002;36:285–298. [PubMed: 12383782]
- Durstewitz D, Seamans JK, Sejnowski TJ. Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology* 2000;83:1733–1750. [PubMed: 10712493]
- Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci* 2005;8(11):1481–1489. [PubMed: 16251991]
- Féger J, Crossman AR. Identification of different subpopulations of neostriatal neurones projecting to globus pallidus or substantia nigra in the monkey: a retrograde fluorescence double-labelling study. *Neurosci Lett* 1984;49(1–2):7–12. [PubMed: 6493600]
- Frank MJ. Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience* 2005;17:51–72. [PubMed: 15701239]
- Frank MJ. Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks* 2006;19:1120–1136. [PubMed: 16945502]
- Frank MJ, Claus ED. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 2006;113(2):300–326. [PubMed: 16637763]
- Frank MJ, Kong L. Learning to avoid in older age. *Psychology and Aging* 2008;23:392–398. [PubMed: 18573012]
- Frank MJ, Loughry B, O'Reilly RC. Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience* 2001;1:137–160.
- Frank MJ, Moustafa AA, Haughey H, Curran T, Hutchison K. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences* 2007a;104:16311–16316.
- Frank MJ, O'Reilly RC. A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience* 2006;120:497–517. [PubMed: 16768602]
- Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: Impulsivity, deep brain stimulation and medication in parkinsonism. *Science* 2007b;318:1309–1312. [PubMed: 17962524]
- Frank MJ, Santamaria A, O'Reilly RC, Willcutt E. Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology* 2007c;32:1583–1599. [PubMed: 17164816]
- Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science* 2004;306:1940–3. [PubMed: 15528409]
- Frank MJ, Wroch BS, Curran T. Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* 2005;47:495–501. [PubMed: 16102533]
- Gerfen CR. The neostriatal mosaic: multiple levels of compartmental organization in the basal ganglia. *Annual Review of Neuroscience* 1992;15:285–320.
- Gerfen CR, Keefe KA. Neostriatal dopamine receptors. *Trends in Neurosciences* 1994;17:2–3. [PubMed: 7511845]
- Gerfen CR, Keefe KA, Gauda EB. D₁ and D₂ dopamine receptor function in the striatum: Coactivation of D₁- and D₂-dopamine receptors on separate populations of neurons results in potentiated immediate early gene response in D₁-containing neurons. *Journal of Neuroscience* 1995;15:8167–8176. [PubMed: 8613751]
- Gerfen, CR.; Wilson, C. The basal ganglia.. In: Swanson, L.; Bjorkland, A.; Hokfelt, T., editors. *Handbook of chemical neuroanatomy*. vol 12: Integrated systems of the CNS. Elsevier; Amsterdam: 1996. p. 371–468.
- Gold JI, Shadlen MN. Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 2002;36:299–308. [PubMed: 12383783]
- Gonon FJ. Prolonged and extrasynaptic excitatory action of dopamine mediated by D1 receptors in the rat striatum in vivo. *Journal of Neuroscience* 1997;17:5972–8. [PubMed: 9221793]
- Gurney K, Prescott TJ, Wickens JR, Redgrave P. Computational models of the basal ganglia: from robots to membranes. *Trends Neurosci* 2004;27(8):453–459. [PubMed: 15271492]

- Haber SN. The primate basal ganglia: parallel and integrative networks. *J Chem Neuroanat* 2003;26(4): 317–330. [PubMed: 14729134]
- Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience* 2000;20:2369–2382. [PubMed: 10704511]
- Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* 2006;26(32):8360–8367. [PubMed: 16899731]
- Harden DG, Grace AA. Activation of dopamine cell firing by repeated L-DOPA administration to dopamine-depleted rats: its potential role in mediating the therapeutic response to L-DOPA treatment. *J Neurosci* 1995;15(9):6157–66. [PubMed: 7666198]
- Haruno M, Kawato M. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Netw* 2006;19(8):1242–1254. [PubMed: 16987637]
- Hernandez-Lopez S, Vargas J, Surmeier DJ, Reyes A, Galarraga E. D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca^{2+} conductance. *Journal of Neuroscience* 1997;17:3334–42. [PubMed: 9096166]
- Hernandez-Lopez S, Tkatch T, Perez-Garci E, Galarraga E, Vargas J, Hamm H, Surmeier DJ. D2 dopamine receptors in striatal medium spiny neurons reduce l-type Ca^{2+} currents and excitability via a novel $plc[\beta]1$ - $ip3$ -calcineurin-signaling cascade. *J Neurosci* 2000;20(24):8987–8995. [PubMed: 11124974]
- Hirvonen M, Laakso A, Abd Nagren K, Rinne J, Pohjalainen T, Hietala J. C957t polymorphism of the dopamine d2 receptor (drd2) gene affects striatal drd2 availability in vivo (corrigendum). *Molecular Psychiatry* 2005;10:889.
- Holroyd CB, Coles MGH. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review* 2002;109:679–709. [PubMed: 12374324]
- Houk JC. Agents of the mind. *Biol Cybern* 2005;92(6):427–437. [PubMed: 15915357]
- Houk, JC.; Adams, JL.; Barto, AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, JC.; Davis, JL.; Beiser, DG., editors. *Models of information processing in the basal ganglia*. MIT Press; Cambridge, MA: 1995. p. 233–248.
- Humphries MD, Stewart RD, Gurney KN. A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *J Neurosci* 2006;26(50):12921–12942. [PubMed: 17167083]
- Ince E, Ciliax BJ, Levey AI. Differential expression of D₁ and D₂ dopamine and m4 muscarinic acetylcholine receptor proteins in identified striatonigral neurons. *Synapse* 1997;27:257–366.
- Joel D, Niv Y, Ruppin E. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks* 2002;15:535–547. [PubMed: 12371510]
- Joel D, Weiner I. 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* 2000;96:451. [PubMed: 10717427]
- Kaasinen V, Rinne JO. Functional imaging studies of dopamine system and cognition in normal aging and Parkinson's disease. *Neurosci Biobehav Rev* 2002;26(7):785–793. [PubMed: 12470690]
- Kawaguchi Y, Wilson CJ, Emson PC. Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. *J Neurosci* 1990;10(10):3421–3438. [PubMed: 1698947]
- Keller RW, Kuhr WG, Wightman RM, Zigmond MJ. The effect of l-dopa on in vivo dopamine release from nigrostriatal bundle neurons. *Brain Res* 1988;447(1):191–194. [PubMed: 3382951]
- Kerr JN, Wickens JR. Dopamine D-1/D-5 receptor activation is required for long-term potentiation in the rat neostriatum in vitro. *Journal of Neurophysiology* 2001;85:117–124. [PubMed: 11152712]
- Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M. Genetically determined differences in learning from errors. *Science* 2007;318(5856):1642–1645. [PubMed: 18063800]
- Knapaska E, Kaczmarek L. A gene for neuronal plasticity in the mammalian brain: *Zif268/egr-1/ngfi-a/krox-24/tis8/zenk?* *Prog Neurobiol* 2004;74(4):183–211. [PubMed: 15556287]
- Koob GF, Le Moal M. Drug abuse: hedonic homeostatic dysregulation. *Science* 1997;278:52–58. [PubMed: 9311926]

- Kraytsberg Y, Kudryavtseva E, McKee AC, Geula C, Kowall NW, Khrapko K. Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat Genet* 2006;38(5):518–520. [PubMed: 16604072]
- Kreitzer AC, Malenka RC. Endocannabinoid-mediated rescue of striatal LTD and motor deficits in parkinson's disease models. *Nature* 2007;445(7128):643–647. [PubMed: 17287809]
- Lapper SR, Bolam JP. Input from the frontal cortex and the parafascicular nucleus to cholinergic interneurons in the dorsal striatum of the rat. *Neuroscience* 1992;51(3):533–545. [PubMed: 1488113]
- Le Moine C, Bloch B. D1 and D2 dopamine receptor gene expression in the rat striatum: sensitive cRNA probes demonstrate prominent segregation of D1 and D2 mRNAs in distinct neuronal populations of the dorsal and ventral striatum. *Journal of Comparative Neurology* 1995;355:418–26. [PubMed: 7636023]
- Lee D, Conroy ML, McGreevy BP, Barraclough DJ. Reinforcement learning and decision making in monkeys during a competitive game. *Brain Res Cogn Brain Res* 2004;22(1):45–58. [PubMed: 15561500]
- Lee D, McGreevy BP, Barraclough DJ. Learning and decision making in monkeys during a rock-paper-scissors game. *Brain Res Cogn Brain Res* 2005;25(2):416–430. [PubMed: 16095886]
- Lei W, Jiao Y, Del Mar N, Reiner A. Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. *J Neurosci* 2004;24(38):8289–8299. [PubMed: 15385612]
- Lévesque M, Parent A. The striatofugal fiber system in primates: a reevaluation of its organization based on single-axon tracing studies. *Proc Natl Acad Sci U S A* 2005;102(33):11888–11893. [PubMed: 16087877]
- Lindskog M, Kim M, Wikström MA, Blackwell KT, Kotaleski JH. Transient calcium and dopamine increase pka activity and darpp-32 phosphorylation. *PLoS Comput Biol* 2006;2(9):e119. [PubMed: 16965177]
- Liu FC, Graybiel AM. Region-dependent dynamics of camp response element-binding protein phosphorylation in the basal ganglia. *Proc Natl Acad Sci U S A* 1998;95:4708–4713. [PubMed: 9539803]
- Madras BK, Miller GM, Fischman AJ. The dopamine transporter and attention-deficit/hyperactivity disorder. *Biol Psychiatry* 2005;57(11):1397–1409. [PubMed: 15950014]
- Magill PJ, Sharott A, Bevan MD, Brown P, Bolam JP. Synchronous unit activity and local field potentials evoked in the subthalamic nucleus by cortical stimulation. *Journal of Neurophysiology* 2004;92(2):700–714. [PubMed: 15044518]
- McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 2003;38:339–346. [PubMed: 12718866]
- McClure, SM.; Gilzenrat, MS.; Cohen, JD. Advances in neural information processing systems. Vol. 18. MIT Press; 2006. p. 867-874. Chap. An exploration-exploitation model based on norepinephrine and dopamine activity
- Meissner W, Leblois A, Hansel D, Bioulac B, Gross CE, Benazzouz A, Boraud T. Subthalamic high frequency stimulation resets subthalamic firing and reduces abnormal oscillations. *Brain* 2005;128 (Pt 10):2372–2382. [PubMed: 16123144]
- Meyer-Lindenberg A, Straub RE, Lipska BK, Verchinski BA, Goldberg T, Callicott JH, Egan MF, Huffaker SS, Mattay VS, Kolachana B, Kleinman JE, Weinberger DR. Genetic evidence implicating darpp-32 in human frontostriatal structure, function, and cognition. *J Clin Invest* 2007;117(3):672–682. [PubMed: 17290303]
- Mink JW. The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology* 1996;50:381–425. [PubMed: 9004351]
- Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience* 1996;16:1936–1947. [PubMed: 8774460]
- Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 2004;43:133–43. [PubMed: 15233923]
- Mouroux M, Féger J. Evidence that the parafascicular projection to the subthalamic nucleus is glutamatergic. *Neuroreport* 1993;4(6):613–615. [PubMed: 8102257]

- Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, Honey G, Jones PB, Bullmore ET, Robbins TW, Fletcher PC. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry* 2008;13(3):239, 267–239, 276. [PubMed: 17684497]
- Nakamura K, Hikosaka O. Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J Neurosci* 2006;26(20):5360–5369. [PubMed: 16707788]
- Nakano K, Kayahara T, Tsutsumi T, Ushiro H. Neural circuits and functional organization of the striatum. *J Neurol* 2000;247(Suppl 5):V1–15. [PubMed: 11081799]
- Nambu A, Tokuno H, Hamada I, Kita H, Imanishi M, Akazawa T, Ikeuchi Y, Hasegawa N. Excitatory cortical inputs to pallidal neurons via the subthalamic nucleus in the monkey. *Journal of Neurophysiology* 2000;84:289–300. [PubMed: 10899204]
- Nambu A, Tokuno H, Takada M. Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway. *Neuroscience Research* 2002;43:111–7. [PubMed: 12067746]
- Nicola SM, Surmeier J, Malenka RC. Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annual Review of Neuroscience* 2000;23:185–215.
- Nieuwenhuis S, Holroyd CB, Mol N, Coles MGH. Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews* 2004;28:441–8. [PubMed: 15289008]
- Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* 2007;191(3):507–520. [PubMed: 17031711]
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 2004;304(5669):452–454. [PubMed: 15087550]
- O'Doherty JP. Lights, camembert, action! the role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Ann N Y Acad Sci.* 2007
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. *Neuron* 2003;38:329–337. [PubMed: 12718865]
- Onn SP, Wang XB. Differential modulation of anterior cingulate cortical activity by afferents from ventral tegmental area and mediodorsal thalamus. *European Journal of Neuroscience* 2005;21:2975–2992. [PubMed: 15978009]
- O'Reilly RC, Frank MJ. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation* 2006;18:283–328. [PubMed: 16378516]
- O'Reilly RC, Frank MJ, Hazy TE, Watz B. PVLV: The primary value and learned value pavlovian learning algorithm. *Behavioral Neuroscience* 2007;121:31–49. [PubMed: 17324049]
- O'Reilly, RC.; Munakata, Y. Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. The MIT Press; Cambridge, MA: 2000.
- Ouimet CC, Miller PE, Hemmings HCJ, Walaass SI, Greengard P. DARPP-32, a dopamine- and adenosine 3':5' monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. III. Immunocytochemical localization. *Journal of Neuroscience* 1984;4:111–24. [PubMed: 6319625]
- Parent A, Hazrati L. Functional anatomy of the basal ganglia. II. the place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Research Reviews* 1995;20:128–54. [PubMed: 7711765]
- Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G, Rudarakanchana N, Halkerston KM, Robbins TW, Everitt BJ. Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behavioral Brain Research* 2002;137:149–63.
- Pavese N, Evans AH, Tai YF, Hotton G, Brooks DJ, Lees AJ, Piccini P. Clinical correlates of levodopa-induced dopamine release in parkinson disease: a pet study. *Neurology* 2006;67(9):1612–1617. [PubMed: 17101892]
- Pothuizen HHJ, Jongen-Rêlo AL, Feldon J, Yee BK. Double dissociation of the effects of selective nucleus accumbens core and shell lesions on impulsive-choice behaviour and salience learning in rats. *European Journal of Neuroscience* 2005;22:2605–2616. [PubMed: 16307603]
- Ramnani N, Elliott R, Athwal BS, Passingham RE. Prediction error for free monetary reward in the human prefrontal cortex. *Neuroimage* 2004;23(3):777–786. [PubMed: 15528079]

- Reynolds JN, Wickens JR. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks* 2002;15:507–521. [PubMed: 12371508]
- Reynolds JNJ, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. *Nature* 2001;412:67–69. [PubMed: 11544526]
- Richfield EK, Penney JB, Young AB. Anatomical and affinity state comparisons between dopamine D1 and D2 receptors in the rat central nervous system. *Neuroscience* 1989;30:767–77. [PubMed: 2528080]
- Saint-Cyr, JA.; Taylor, AE.; Lang, AE. Procedural learning and neostriatal dysfunction in man. 1988.
- Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. *Science* 2005;310(5752):1337–1340. [PubMed: 16311337]
- Schall JD. Neural correlates of decision processes: neural and mental chronometry. *Current Opinion in Neurobiology* 2003;13:182–186. [PubMed: 12744971]
- Schönberg T, Daw ND, Joel D, O'Doherty JP. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci* 2007;27(47):12860–12867. [PubMed: 18032658]
- Schultz W. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* 1998;80:1. [PubMed: 9658025]
- Schultz W. Getting formal with dopamine and reward. *Neuron* 2002;36:241–263. [PubMed: 12383780]
- Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science* 1997;275:1593. [PubMed: 9054347]
- Schultz W, Dickinson A. Neuronal coding of prediction errors. *Annual Review of Neuroscience* 2000;23:473–500.
- Seeman P. Dopamine d2(high) receptors on intact cells. *Synapse* 2008;62(4):314–318. [PubMed: 18241049]
- Seymour B, Daw N, Dayan P, Singer T, Dolan R. Differential encoding of losses and gains in the human striatum. *J Neurosci* 2007;27(18):4826–4831. [PubMed: 17475790]
- Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS. Temporal difference models describe higher-order learning in humans. *Nature* 2004;429(6992):664–667. [PubMed: 15190354]
- Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 2008;321(5890):848–851. [PubMed: 18687967]
- Shen W, Tian X, Day M, Ulrich S, Tkatch T, Nathanson NM, Surmeier DJ. Cholinergic modulation of kir2 channels selectively elevates dendritic excitability in striatopallidal neurons. *Nat Neurosci* 2007;10(11):1458–1466. [PubMed: 17906621]
- Shohamy D, Myers CE, Gekhman KD, Sage J, Gluck MA. L-dopa impairs learning, but spares generalization, in Parkinson's disease. *Neuropsychologia* 2006;44(5):774–784. [PubMed: 16150469]
- Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA. Corticostriatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain* 2004;127:851–9. [PubMed: 15013954]
- Smith-Roe SL, Kelley AE. Coincident activation of NMDA and dopamine D1 receptors within the nucleus accumbens core is required for appetitive instrumental learning. *Journal of Neuroscience* 2000;22:7737–42. [PubMed: 11027236]
- Stipanovich A, Valjent E, Matamales M, Nishi A, Ahn J-H, Maroteaux M, Bertran-Gonzalez J, Bami-Cherrier K, Enslin H, Corbillé A-G, Filhol O, Nairn AC, Greengard P, Hervé D, Girault J-A. A phosphatase cascade by which rewarding stimuli control nucleosomal response. *Nature* 2008;453(7197):879–884. [PubMed: 18496528]
- Suaud-Chagny MF, Dugast C, Chergui K, Msghina M, Gonon F. Uptake of dopamine released by impulse flow in the rat mesolimbic and striatal systems in vivo. *J Neurochem* 1995;65(6):2603–2611. [PubMed: 7595557]
- Sugrue LP, Corrado GS, Newsome WT. Neuroscience: Matching behavior and the representation of value in the parietal cortex. *Science* 2004;304(5678):1782–1786. [PubMed: 15205529]

- Suri RE, Schultz W. Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research* 1998;121:350.
- Surmeier DJ, Ding J, Day M, Wang Z, Shen W. D1 and d2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci* 2007;30(5):228–235. [PubMed: 17408758]
- Surmeier DJ, Song WJ, Yan Z. Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *Journal of Neuroscience* 1996;16:6579–91. [PubMed: 8815934]
- Sutton, RS.; Barto, AG. Reinforcement learning: An introduction. MIT Press; Cambridge, MA: 1998.
- Swainson R, Rogers RD, Sahakian BJ, Summers BA, Polkey CE, Robbins TW. Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia* 2000;38:596–612. [PubMed: 10689037]
- Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience* 2004;7:887–893.
- Tedroff J, Pedersen M, Aquilonius SM, Hartvig P, Jacobsson G, Långström B. Levodopa-induced changes in synaptic dopamine in patients with parkinson's disease as measured by [¹¹C]raclopride displacement and pet. *Neurology* 1996;46(5):1430–1436. [PubMed: 8628494]
- Tepper JM, Bolam JP. Functional diversity and specificity of neostriatal interneurons. *Curr Opin Neurobiol* 2004;14(6):685–692. [PubMed: 15582369]
- Thorndike, EL. Animal intelligence: Experimental studies. MacMillan Press; 1911.
- Tunbridge EM, Bannerman DM, Sharp T, Harrison PJ. Catechol-o-methyltransferase inhibition improves set-shifting performance and elevates stimulated dopamine release in the rat prefrontal cortex. *J Neurosci* 2004;24(23):5331–5335. [PubMed: 15190105]
- Venton BJ, Zhang H, Garris PA, Phillips PEM, Sulzer D, Wightman RM. Real-time decoding of dopamine concentration changes in the caudate-putamen during tonic and phasic firing. *J Neurochem* 2003;87(5):1284–1295. [PubMed: 14622108]
- Volkow ND, Wang GJ, Fowler JS, Logan J, Gerasimov M, Maynard L, Ding Y-S, Gatley SJ, Gifford A, Franceschi D. Therapeutic doses of methylphenidate significantly increase extracellular dopamine in the human brain. *Journal of Neuroscience* 2001;21:RC121. [PubMed: 11160455]
- Voorn, P.; Vanderschuren, LJ.; Groenewegen, HJ.; Robbins, TW.; Pennartz, CM. Putting a spin on the dorsal-ventral divide of the striatum. 2004.
- Walaas SI, Aswad D, Greengard P. A dopamine- and cyclic AMP-regulated phosphoprotein enriched in dopamine-innervated brain. *Nature* 1983;301:69–71. [PubMed: 6296685]
- Wang Z, Kai L, Day M, Ronesi J, Yin HH, Ding J, Tkatch T, Lovinger DM, Surmeier DJ. Dopaminergic control of corticostriatal long-term synaptic depression in medium spiny neurons is mediated by cholinergic interneurons. *Neuron* 2006;50(3):443–452. [PubMed: 16675398]
- Watkins CJCH, Dayan P. Technical note: Q-learning. *Machine Learning* 1992;8:279.
- Wichmann T, Bergman H, DeLong MR. The primate subthalamic nucleus. I. functional properties in intact animals. *Journal of Neurophysiology* 1994;72:494–506. [PubMed: 7983514]
- Wickens JR, Budd CS, Hyland BI, Arbuthnott GW. Striatal contributions to reward and decision making: making sense of regional variations in a reiterated processing matrix. *Ann N Y Acad Sci* 2007;1104:192–212. [PubMed: 17416920]
- Wiecki, TV.; Riedinger, K.; Meyerhofer, A.; Schmidt, WJ.; Frank, MJ. A neurocomputational account of context-dependent catalepsy sensitization induced by haloperidol. submitted
- Wightman RM, Amatore C, Engstrom RC, Hale PD, Kristensen EW, Kuhr WG, May LJ. Real-time characterization of dopamine overflow and uptake in the rat striatum. *Neuroscience* 1988;25(2): 513–523. [PubMed: 3399057]
- Wilson CJ, Callaway JC. Coupled oscillator model of the dopaminergic neuron of the substantia nigra. *Journal of Neurophysiology* 2000;83:3084. [PubMed: 10805703]
- Wilson CJ, Chang HT, Kitai ST. Firing patterns and synaptic potentials of identified giant aspiny interneurons in the rat neostriatum. *J Neurosci* 1990;10(2):508–519. [PubMed: 2303856]

- Wilson CJ, Weyrick A, Terman D, Hallworth NE, Bevan MD. A model of reverse spike frequency adaptation and repetitive firing of subthalamic nucleus neurons. *Journal of Neurophysiology* 2004;91(5):1963–1980. [PubMed: 14702332]
- Winterer G, Weinberger DR. Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci* 2004;27(11):683–690. [PubMed: 15474169]
- Wolf JA, Moyer JT, Lazarewicz MT, Contreras D, Benoit-Marand M, O'Donnell P, Finkel LH. NMDA/AMPA ratio impacts state transitions and entrainment to oscillations in a computational model of the nucleus accumbens medium spiny projection neuron. *J Neurosci* 2005;25(40):9080–9095. [PubMed: 16207867]
- Wörgötter F, Porr B. Temporal sequence learning, prediction, and control: A review of different models and their relation to biological mechanisms. *Neural Computation* 2005;17(2):245–319. [PubMed: 15720770]
- Wu Y, Richard S, Parent A. The organization of the striatal output system: a single-cell juxtacellular labeling study in the rat. *Neurosci Res* 2000;38(1):49–62. [PubMed: 10997578]
- Yacubian J, Sommer T, Schroeder K, Gläscher J, Kalisch R, Leuenberger B, Braus DF, Büchel C. Gene-gene interaction associated with neural reward sensitivity. *Proc Natl Acad Sci U S A* 2007;104(19):8125–8130. [PubMed: 17483451]
- Yasuda A, Sato A, Miyawaki K, Kumano H, Kuboki T. Error-related negativity reflects detection of negative reward prediction error. *Neuroreport* 2004;15:2561–5. [PubMed: 15538196]
- Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review* 2004;111(4):931–959. [PubMed: 15482068]
- Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. *Neuron* 2005;46(4):681–692. [PubMed: 15944135]
- Zackheim J, Abercrombie ED. Thalamic regulation of striatal acetylcholine efflux is both direct and indirect and qualitatively altered in the dopamine-depleted striatum. *Neuroscience* 2005;131(2):423–436. [PubMed: 15708484]
- Zador A, Koch C. Linearized models of calcium dynamics: Formal equivalence to the cable equation. *Journal of Neuroscience* 1994;14:4705–4715. [PubMed: 8046445]

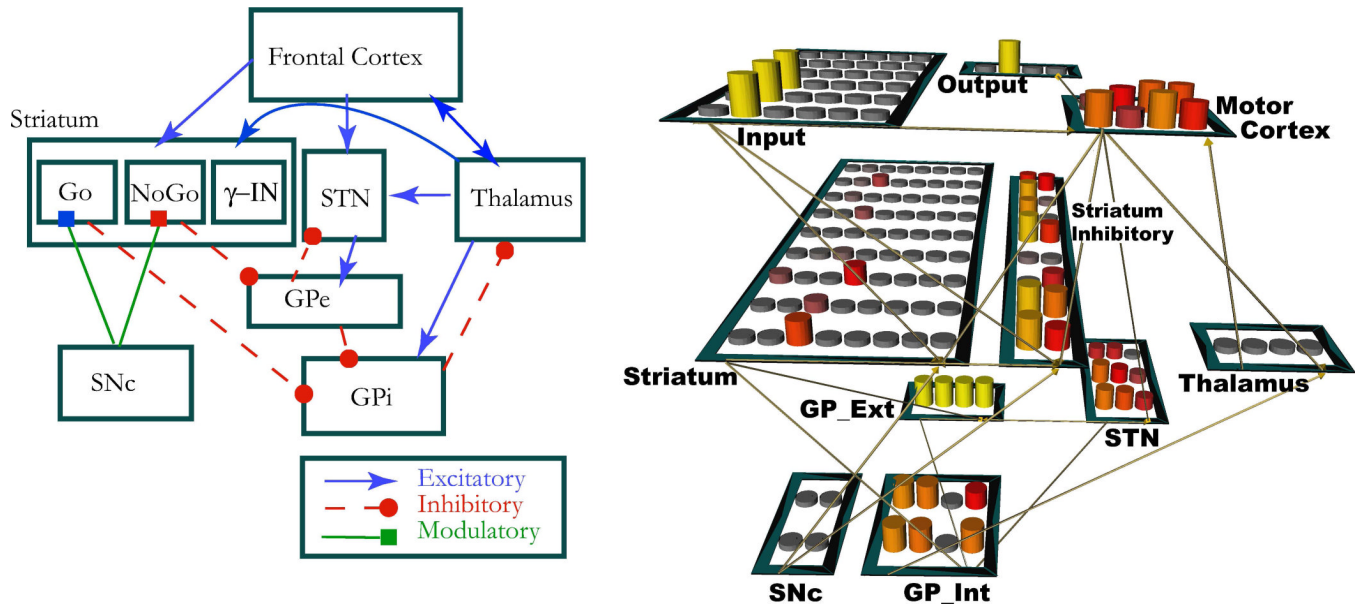


Figure 1.

Left. Functional anatomy of the basal ganglia circuit, showing an updated model of the primary projections. In addition to the classic “direct” and “indirect” pathways from Striatum to BG output nuclei originating in striatonigral (Go) and striatopallidal (NoGo) cells respectively, the revised architecture features focused projections from NoGo units to GPe and strong top-down projections from cortex to thalamus. Further, the STN is incorporated as part of a newly discovered hyperdirect pathway (rather than part of the indirect pathway as originally conceived), receiving inputs from frontal cortex and projecting directly to both GPe and GPi.

Right. Neural network model of this circuit, with four different responses represented by four columns of motor units, four columns each of Go and NoGo units within Striatum, and corresponding columns within GPi, GPe and Thalamus. Fast spiking GABA-ergic interneurons (γ -IN) regulate Striatal activity via inhibitory projections. For implementational details, see Frank (2005, 2006).

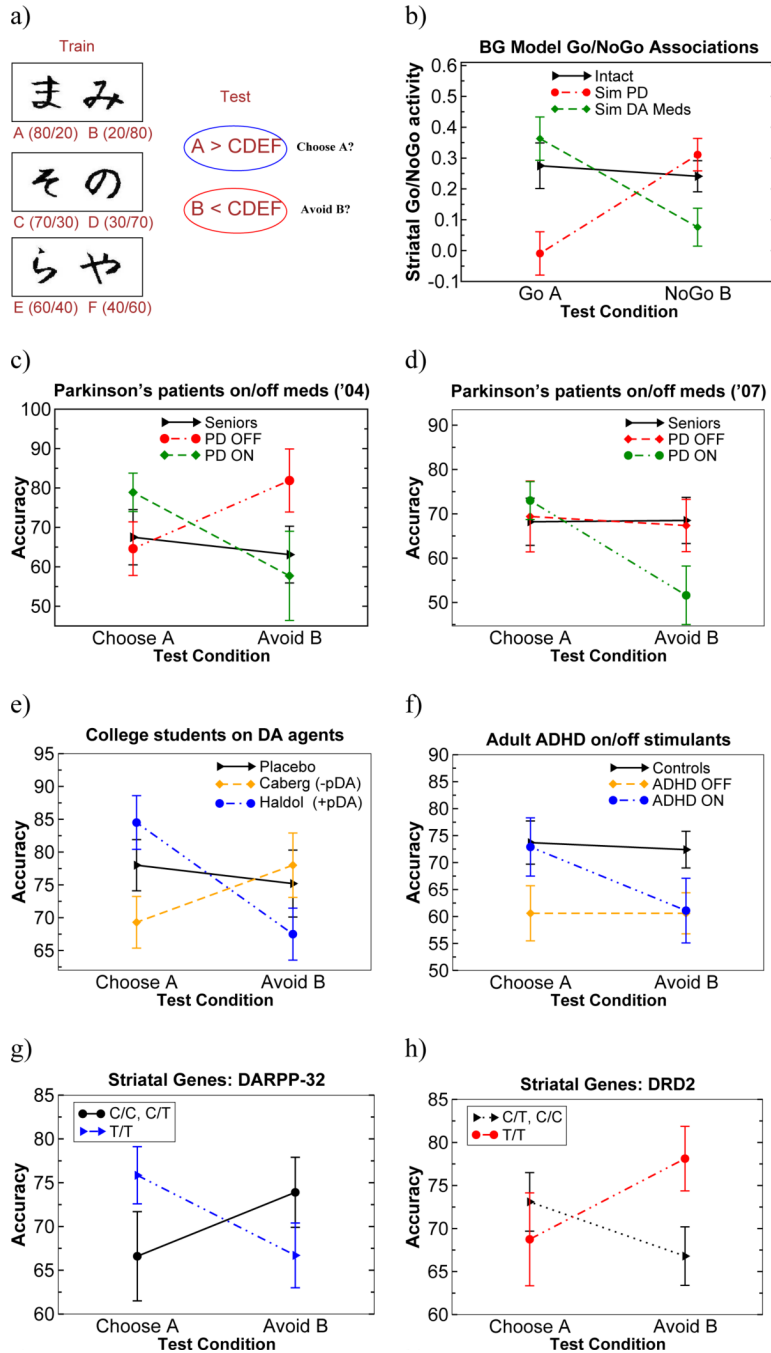


Figure 2.
a) Probabilistic selection reinforcement learning task. During training, participants select among each stimulus pair. Probabilities of receiving positive/negative feedback for each stimulus are indicated in parentheses. In the test phase, all combinations of stimuli are presented without feedback. “Go learning” is indexed by reliable choice of the most positive stimulus A in these novel pairs, whereas “NoGo learning” is indexed by reliable avoidance of the most negative stimulus B. **b)** Striatal Go and NoGo activation states when presented with input stimuli A and B respectively. Simulated Parkinson's (Sim PD) was implemented by reducing striatal DA levels, whereas medication (Sim DA Meds) was simulated by increasing DA levels and partially shunting the effects of DA dips during negative feedback. **c)** Behavioral findings

in PD patients on/off medication supporting model predictions (Frank et al., 2004). **d**) Replication in another group of patients, where here the most prominent effects were observed in the NoGo learning condition (Frank et al., 2007b). **e**) Similar results in healthy participants on dopamine agonists and antagonists modulating presynaptic DA (pDA) and **f**) adult ADHD participants on and off stimulant medications. **g**), **h**) Individual differences in Go/NoGo learning in college students are predicted by genes controlling striatal D1/D2 function.

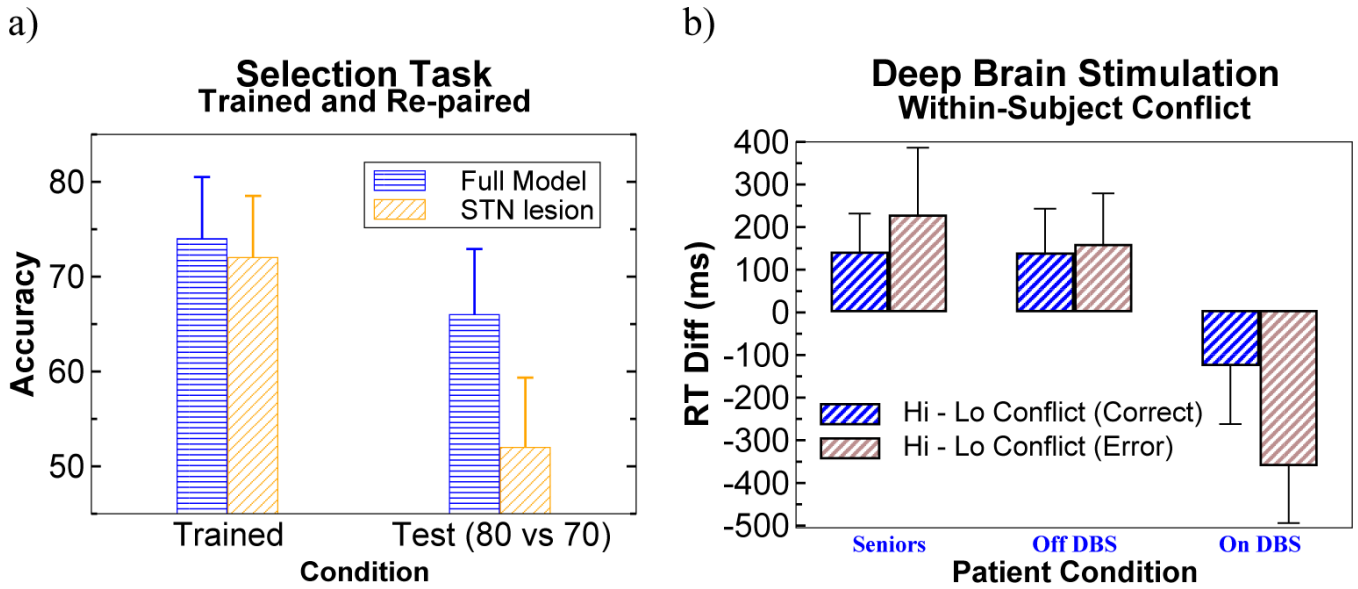


Figure 3.

a) Subthalamic nucleus contributions to model performance in the probabilistic selection task. While not differing from intact networks in selection among trained low-conflict discriminations (80 vs 20 and 70 vs 30), STN lesioned networks were selectively impaired at the high conflict selection of an 80% positively reinforced response when it competed with a 70% response. The model STN Global NoGo signal prevents premature responding when multiple responses are potentially rewarding, increasing the likelihood of accurate choice (Frank, 2006). **b)** Behavioral results in Parkinson's patients on and off DBS, confirming model predictions. Response time differences are shown for high relative to low conflict test trials. Whereas healthy controls, patients on/off medication (not shown) and patients off DBS adaptively slow decision times in high relative to low conflict test trials, patients on DBS respond impulsively *faster* in these trials (adapted from (Frank et al., 2007b)).

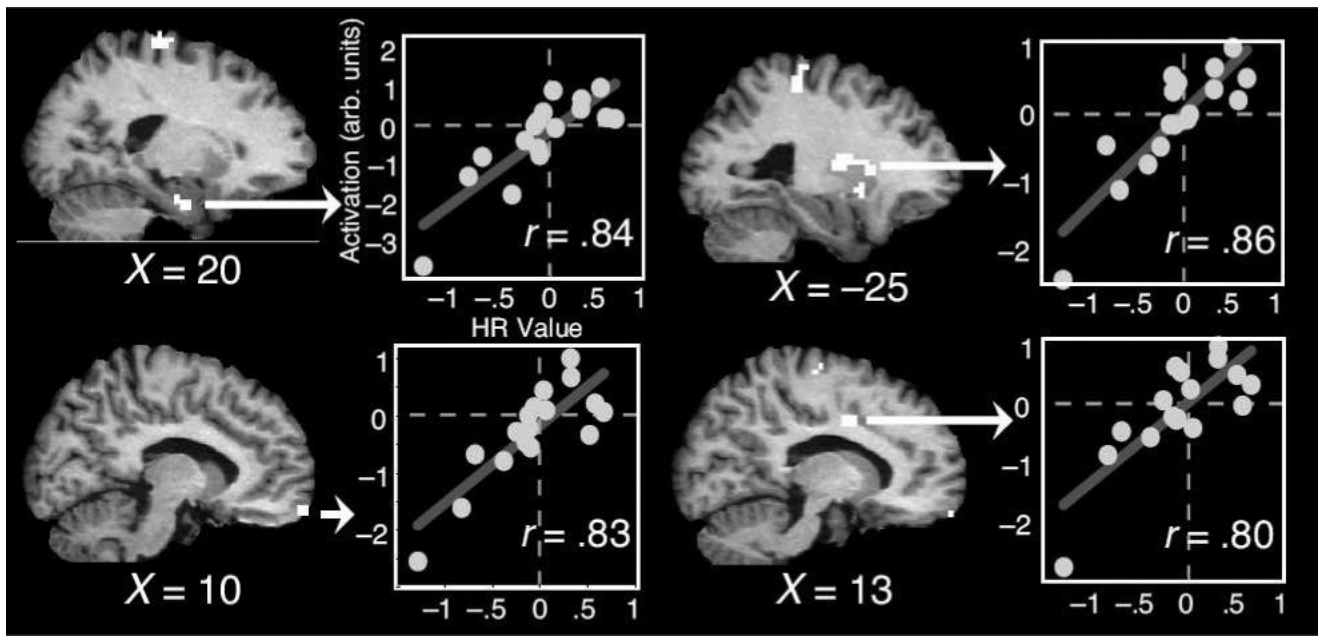


Figure 4.

Abstract reinforcement learning models can be useful for investigating individual differences. Here a model was used to estimate the impact of reinforcement (winning money or not in a gambling task) on the likelihood of making a low- or high-risk gamble in the subsequent trial. The best-fitting parameter for each subject determines the magnitude and sign of the weight change for the high-risk option after obtaining a high-risk reward. Individual differences in this parameter were then correlated with reinforcement-related brain activation. Results indicate that, in a network of regions including the lateral striatum (top right), this weight-update parameter (x-axis) predicts whether brain activations to large rewards are associated with subsequent risky (y-axis positive values) or non-risky (negative values) choices. In this case, individual differences proved critical for understanding how reinforcements guide subsequent decisions: for some subjects reward-related activity predicted increased likelihood of making a subsequent risky choice, whereas for others it predicted decreased likelihood, according to their estimated parameters. See Cohen and Ranganath, 2005, for details.