nature neuroscience

From reinforcement learning models to psychiatric and neurological disorders

Tiago V Maia[1,2] & Michael J Frank[3,4]

Over the last decade and a half, reinforcement learning models have fostered an increasingly sophisticated understanding of the functions of dopamine and cortico-basal ganglia-thalamo-cortical (CBGTC) circuits. More recently, these models, and the insights that they afford, have started to be used to understand important aspects of several psychiatric and neurological disorders that involve disturbances of the dopaminergic system and CBGTC circuits. We review this approach and its existing and potential applications to Parkinson's disease, Tourette's syndrome, attention-deficit/hyperactivity disorder, addiction, schizophrenia and preclinical animal models used to screen new antipsychotic drugs. The approach's proven explanatory and predictive power bodes well for the continued growth of computational psychiatry and computational neurology.

The limitations of the state-of-the-art in nosology in psychiatry have been much debated in the context of the development of the new edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*). There is widespread agreement that the current symptom-based system of classification must eventually be replaced with a system based on pathophysiology[1]. However, the current understanding of the neurobiology and genetics of psychiatric disorders remains too limited to form the backbone of nosology[2]. This limited understanding is also reflected in the state-of-the-art in treatment, with most psychiatric medications having been found by serendipity, rather than through rational design. Neurology typically deals with disorders with better understood etiology (for example, loss of dopaminergic neurons in Parkinson's disease), but even then it is often unclear how these etiological processes produce complex patterns of symptoms and why treatments can alleviate some deficits while exacerbating, or even causing, others[3,4]. Part of the problem is the complexity of the brain and mind and the many levels of analysis that span the two. Computational models are a valuable tool for taming this complexity, as they foster a mechanistic understanding that can span multiple levels of analysis and can explain how changes to one component of the system (for example, increases in striatal D2 receptor density) can produce systems-level changes that translate to changes in behavior.

One area in which substantial progress has been made in integrating computational modeling and empirical research in neuroscience is reinforcement learning[5]. This approach has produced models of the roles of dopamine and cortico-basal ganglia-thalamo-cortical (CBGTC) loops in learning about reinforcers (rewards and punishments) and in guiding behavior so as to acquire rewards and avoid punishments[5]. Existing models address a variety of functions of these circuits, including Pavlovian conditioning, instrumental conditioning and their interactions; habits, goal-directed actions and their interactions; and the inter-related issues of incentive salience, motivation and vigor[5–9].

Organizing behavior in ways that obtain outcomes appropriate for the current motivational state (for example, acquiring food if hungry) and that avoid harmful outcomes is crucial for survival and is therefore a central organizing principle of the nervous system. Not surprisingly, then, disturbances of the dopaminergic system and CBGTC circuits have a key role in several psychiatric and neurological disorders. Reinforcement learning models have recently started to be applied to these disorders and have been shown to have substantial explanatory and predictive power[10–14]. The approach builds on an understanding of the computations that these circuits perform in healthy individuals and investigates how pathophysiological processes alter these computations, producing symptoms. We therefore start by reviewing the computational neurobiology of the normal functioning of these circuits. We then discuss several disorders that have benefited or are ripe to benefit from the use of reinforcement learning models. We close by discussing the future implications of this body of work for nosology and treatment.

In addition to conveying the specifics of how reinforcement learning models provide insights into psychiatric and neurological disorders, we hope that this review will also help to foster the development of the emerging disciplines of computational psychiatry and computational neurology. A powerful set of computational techniques can now be used to investigate pathophysiological processes and their relation to behavior (**Fig. 1**). We hope that the work reviewed here serves as an example that prompts the concerted and widespread use of these techniques across multiple model types and disorders.

Reinforcement learning in the brain

Dopamine and prediction errors. Dopamine neurons code reinforcement prediction errors[5], a key signal in many reinforcement learning models[15]. Prediction errors signal the difference between the observed and expected outcomes: a positive prediction error signals that the outcome was better than expected, and a negative prediction error signals that the outcome was worse than expected. The magnitude of phasic dopamine-neuron bursts quantitatively represents positive

[1]Department of Psychiatry, Columbia University, New York, New York, USA. [2]New York State Psychiatric Institute, New York, New York, USA. [3]Departments of Cognitive, Linguistic and Psychological Sciences, and Psychiatry and Human Behavior, Brown University, Providence, Rhode Island, USA. [4]Brown Institute for Brain Science, Brown University, Providence, Rhode Island, USA. Correspondence should be addressed to T.V.M. (tmaia@columbia.edu) or M.J.F. (michael_frank@brown.edu).

Published online 26 January 2011; doi:10.1038/nn.2723

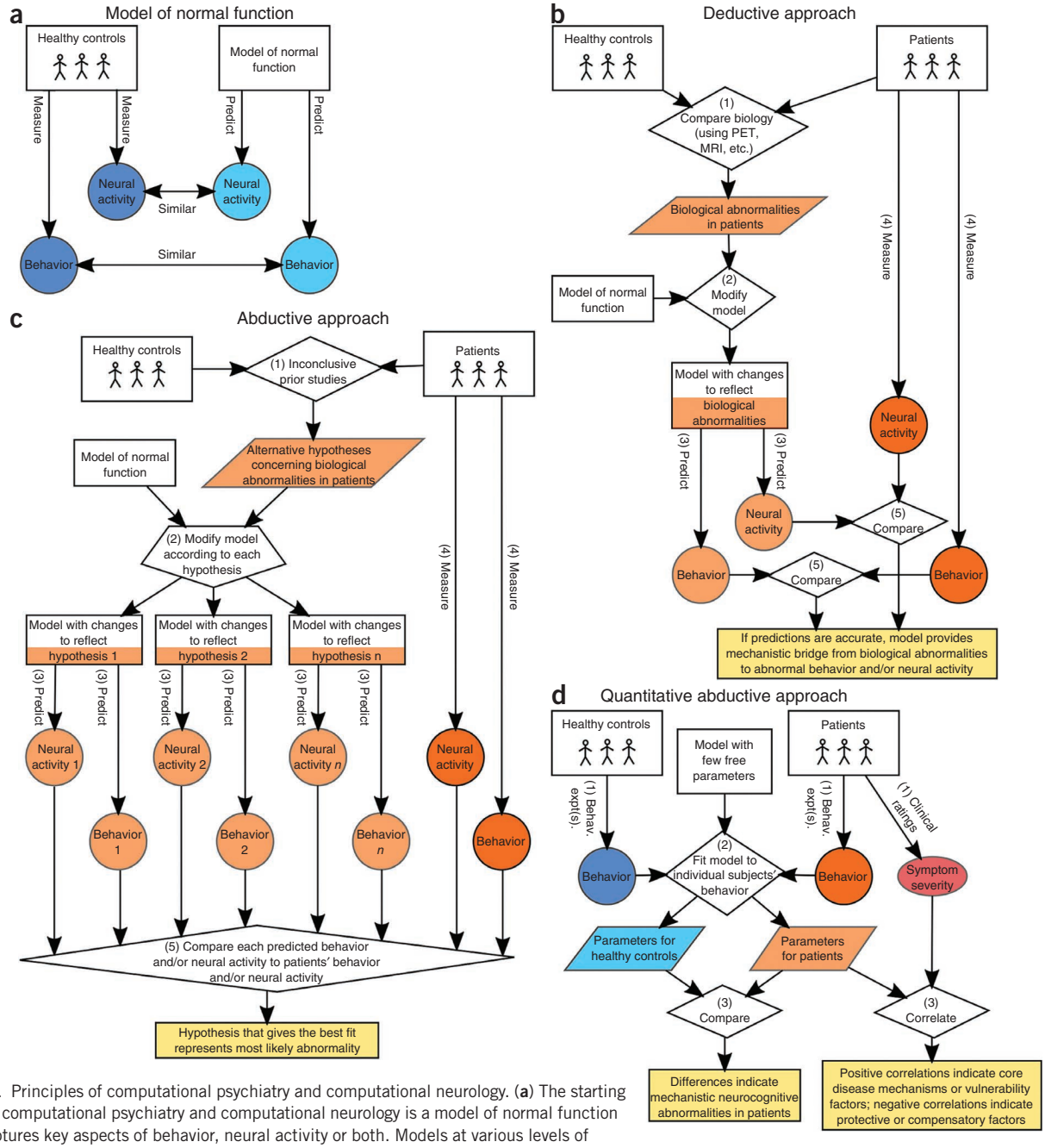

Figure 1 Principles of computational psychiatry and computational neurology. (**a**) The starting point in computational psychiatry and computational neurology is a model of normal function that captures key aspects of behavior, neural activity or both. Models at various levels of abstraction can be useful (for example, algorithmic models from machine learning or neural models from computational cognitive neuroscience). Several approaches can then be pursued. (**b**) With detailed neural models, pathophysiological processes can be simulated by making principled changes to the model that mimic biological alterations in the disorder under consideration (for example, alterations in striatal dopaminergic innervation). The systems-level and behavioral implications of these changes can then be explored, leading to testable predictions. We call this approach 'deductive' because the models are used to recreate the mechanistic link between causes (the biological abnormalities) and their consequences (abnormalities in systems-level neural activity and behavior). (**c**) A second approach involves using a model to try to infer the causes of the observed abnormalities in neural activity or behavior. We call this approach 'abductive' because it involves reasoning from consequences (the behavior or systems-level neural activity) to their possible causes (the underlying biological abnormalities). In this approach, alternative *a priori* hypotheses concerning possible biological abnormalities in a given disorder can be compared to determine which, if any, produce the same abnormalities in behavior and neural activity that are found in the disorder (T.V.M. and B.S. Peterson, unpublished). (**d**) A third approach, used more often with algorithmic than with neural models (largely because the former tend to have fewer parameters), involves fitting the model's parameters to the behavior of individual subjects on a suitable task or set of tasks and then determining if there are parameter differences between diseased and healthy subject groups or correlations between parameters and disease severity. We call this approach 'quantitative abductive' because it also involves reasoning from behavior to its mechanistic causes. A fourth approach (not shown) also involves fitting a model to subjects' behavior, but the goal is to estimate, on a trial-by-trial basis, each subject's putative internal representation of the quantities embedded in the model (for example, state values or prediction errors). These predicted internal representations are then used as regressors in functional imaging (for example, functional magnetic resonance imaging, electroencephalography), to find their neural correlates, which are then compared across the diseased and healthy groups. Each of these four approaches can also be adapted to study the effects of treatments (for example, medication or neurosurgery). Furthermore, additional leverage can sometimes be gained by the synergistic use of different approaches or models at different levels of abstraction. Behav. expt(s)., behavioral experiment(s).

prediction errors[16]. Whether phasic reductions in dopaminergic neuron firing quantitatively represent negative prediction errors is more controversial because the low tonic firing rate of dopamine neurons implies that variations in such reductions are somewhat limited. The duration of the pause in dopaminergic neuron firing, however, seems to represent negative prediction errors quantitatively[17]. This asymmetric coding between positive and negative prediction errors (burst magnitude for positive and pause duration for negative) may be justified biologically, both because it permits lower tonic firing, which is advantageous metabolically, and in terms of the postsynaptic effects of these signals on D1 and D2 receptors, as discussed below. An alternative or complementary hypothesis is that negative prediction errors may be coded by serotonin[18]. This hypothesis, however, has not yet been adequately tested[5].

Although the majority of dopamine neurons burst to positive prediction errors, smaller proportions of these neurons also burst under other conditions[19]. For example, some dopamine neurons burst both to positive prediction errors and to negative events and stimuli that predict negative events[5]. Despite some attempts to reconcile these findings with reinforcement learning[20], additional research is needed in this area. The predominant function of dopamine bursts, however, is to code positive prediction errors[19]. Indeed, phasic optogenetic stimulation of dopamine neurons induces a subsequent preference for the place in which such stimulation occurred[21], just as if reward had been delivered at that place.

The basal ganglia and action selection. Prediction errors are used to learn the values (or 'goodness') of states (stimuli or situations), state-action pairs or both. These values are then used to select optimal actions[15]. A model of this process that has been used to account for many behavioral and neural findings is the actor-critic[5,22,23]. The actor-critic view of action selection in the brain suggests that the cortex represents the current state and the basal ganglia implement two computational modules: the critic, which learns state values and may be implemented in the ventral striatum[5,23] and possibly in the amygdala and orbitofrontal cortex (OFC)[5], and the actor, which learns stimulus-response associations and may be implemented in the dorsal striatum[5,23]. The critic and the actor both use the prediction errors signaled by dopamine to update their estimates (of state values and stimulus-response strengths, respectively). The mapping of the actor-critic to the basal ganglia is consistent with the view that the basal ganglia are crucial for stimulus-response learning, but with different portions of the striatum (which are involved in distinct, parallel CBGTC loops[24]) having distinct roles[25].

A related, but slightly different, view suggests that the basal ganglia are responsible not for generating actions, but for arbitrating between actions that are under consideration in cortex, by facilitating the most appropriate action while suppressing competing actions[26–28]. In other words, whereas according to the actor-critic view, the basal ganglia are fully responsible for selecting the action on the basis of the current state alone, this alternative view suggests that the cortex itself initially generates candidate actions (for example, on the basis of the frequency with which they have previously been executed in the current state[28]) and that the basal ganglia then arbitrate between these actions (likely on the basis of their learned reinforcement probabilities[28,29]) to facilitate (gate) the best one. The commonalities between these views substantially outweigh their differences, however, so we will not delve into this distinction further.

The basal ganglia anatomy consists of direct, indirect and hyperdirect pathways from cortex to basal ganglia output structures[30–32] (**Fig. 2a**). Neurocomputational models[28,33–36] have refined verbal theories[26,30–32] of the role of these pathways in action selection. An influential account, originally proposed to explain the pathophysiology of several neurological disorders[30,31], suggests that the direct pathway provides focused facilitation of the appropriate action(s) for the current state, whereas the indirect pathway suppresses actions that are inappropriate for that state. Although the original version of this

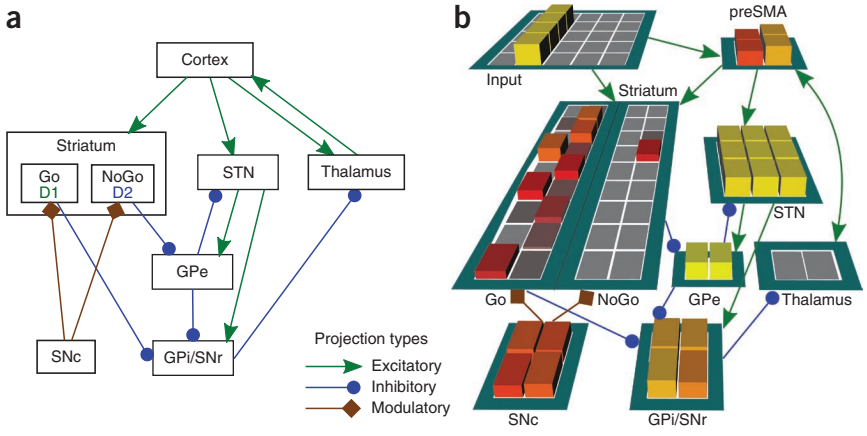

Figure 2 Anatomy and modeling of CBGTC loops. (**a**) Anatomy. Striatal medium spiny neurons in the direct pathway (Go neurons) express mostly D1 receptors[40] and project directly to the globus pallidus internal segment and the substantia nigra pars reticulata (GPi/SNr). Go neurons inhibit the GPi/SNr, which in turn results in disinhibition of the thalamus, thereby facilitating execution of the corresponding action. Striatal medium spiny neurons in the indirect pathway (NoGo neurons) express mostly D2 receptors[40] and project to the globus pallidus external segment (GPe), which in turn projects to the GPi/SNr. NoGo neurons produce a focused removal of the tonic inhibition of the GPe on the GPi/SNr, thereby disinhibiting the GPi/SNr, which in turn results in suppression of the corresponding action in the thalamus. Neurons in the subthalamic nucleus (STN) receive direct projections from the cortex in the hyperdirect pathway and project to both the GPe and GPi/SNr. The projections from the STN to the GPe and GPi/SNr are diffuse[26], so they are believed to modulate all actions rather than a specific action. (**b**) The basal ganglia Go/NoGo model[28,35]. The connections in the model are consistent with the anatomical connections shown in **a**. The model learns to map inputs, representing the current state, to actions in the pre-supplementary motor area (preSMA) (or the SMA). Corticocortical projections from the input layer to preSMA activate in preSMA candidate actions appropriate for the current state. The basal ganglia then facilitate (gate) the best action, that is, the action with the best reinforcement history for the current state, while simultaneously suppressing the other actions (at the level of the thalamus). Distributed populations of Go and NoGo units represent the positive and negative evidence, respectively, for the candidate actions in the current state. Lateral inhibition between the Go and NoGo pathways ensures that the probability of selecting a given action is a function of the difference between the positive and negative evidence for that action. The positive and negative evidence for each action in each state is learned on the basis of past reinforcement history, through the actions of dopamine on D1 and D2 receptors in striatal Go and NoGo units, respectively. The weights from the input layer to preSMA are themselves learned, but through Hebbian mechanisms, thereby allowing these corticocortical projections to activate candidate actions in preSMA in proportion to their prior probability of being executed in the given state. The STN prevents facilitation of suboptimal responses in high-conflict situations[35].

account was based on a somewhat oversimplified view of basal ganglia anatomy and function, recent findings strongly support this distribution of function between the direct (or Go) and indirect (or NoGo) pathways[37–39]. This distribution of function finds formal expression in the basal ganglia Go/NoGo (BG-GNG) model[28] (**Fig. 2b**), in which the probability that a given action is selected is proportional to the difference between the Go and NoGo activity for that action in the current state. Consistent with this scheme, electrophysiological findings demonstrate that the positive and negative values of actions are represented in distinct striatal populations, with greater activity in the neurons that represent the positive value of an action predicting selection of that action and greater activity in the neurons that represent the negative value of an action predicting selection of an alternative action[29]. The BG-GNG model further shows how aspects of basal ganglia anatomy that were not considered in the original account subserve other aspects of basal ganglia function. For example, the model suggests that the hyperdirect (or Global NoGo) pathway provides global inhibition of all actions during the early stages of processing, particularly in high-conflict situations (that is, when multiple actions are strongly activated simultaneously), to prevent premature, suboptimal responding[35].

The BG-GNG model also shows how the direct and indirect pathways can learn which actions to facilitate and suppress in each state, respectively, using the prediction errors conveyed by dopamine[28]. The direct and indirect pathways predominantly express D1 and D2 receptors, respectively[40]. In the model, when an action is followed by a dopamine burst, the corticostriatal synapses in the direct pathway into active Go units (which encode the conjunction between the state and the action that was executed) are strengthened via D1-dependent long-term potentiation (LTP), and the corticostriatal synapses in the indirect pathway into active NoGo units for the executed action are weakened via D2-dependent long-term depression (LTD). When an action is followed by a dopamine dip, the reverse occurs. These dual effects of dopamine on D1- and D2-mediated plasticity have been supported by empirical evidence[41], as has the model prediction that the direct and indirect pathways mediate learning from positive and negative outcomes, respectively[38].

The dynamics of dopamine effects on D1 and D2 receptors may also explain why burst magnitude and pause duration code for positive and negative prediction errors, respectively. D1 and D2 receptors have relatively low and high affinities for dopamine, respectively[42]. D1 stimulation is therefore hypothesized to depend on phasic dopamine bursts, with larger bursts producing greater stimulation. Burst magnitude is therefore crucial for D1-mediated LTP as a result of positive prediction errors. D2 receptors, in contrast, are hypothesized to be stimulated tonically by baseline dopamine levels. The effect of pauses in dopaminergic neuron firing on D2 receptors therefore depends on dopamine reuptake, with longer pauses allowing greater reuptake and therefore producing a larger dip in dopamine concentration. Pause duration is therefore crucial for D2-mediated LTP as a result of negative prediction errors. LTD mechanisms are also consistent with a key role for magnitudes and durations in coding positive and negative prediction errors, respectively. Positive prediction errors stimulate D2 receptors directly, producing LTD in the indirect pathway. Negative prediction errors may not strongly affect D1 receptors, because D1 receptors may not be substantially stimulated by tonic dopamine. In the BG-GNG model, negative prediction errors instead produce LTD in the direct pathway indirectly, via their effects on D2 receptors (dependent on reuptake and therefore pause duration) and subsequent inhibition of the direct pathway by the indirect pathway, leading to activity-dependent LTD in the direct pathway.

In addition to the role of phasic dopamine in learning, tonic dopamine increases excitability in the direct (Go) pathway and decreases excitability in the indirect (NoGo) pathway because D1 receptors are excitatory (at least for neurons receiving strong concomitant glutamatergic input) and D2 receptors are inhibitory[28]. Increases in tonic dopamine therefore produce a Go bias, whereas decreases produce a NoGo bias. Simulations using the BG-GNG model show that these biases exert strong effects on action selection and reaction times[10,28,43], with tonic dopamine promoting the execution and speed of actions (particularly actions with greater positive differences between their previously learned Go and NoGo associations). Dopamine therefore modulates not only learning, but also the expression of prior learning.

Although multiple parallel (albeit interacting[44]) CBGTC loops course through the basal ganglia[24], a common division is into sensorimotor, associative and limbic loops, which connect to sensorimotor cortical areas, dorsolateral prefrontal cortex, and OFC and anterior cingulate cortex, respectively[45]. The BG-GNG model has been applied not only to motor action selection, involving the sensorimotor loop, but also to the selection of cognitive 'actions', particularly working memory updating[46,47], likely involving the associative loop. The idea in the latter case is that Go signals facilitate the gating of a stimulus into working memory, whereas NoGo signals prevent such gating (for example, for stimuli that are irrelevant for the task). Consistent with this idea, basal ganglia damage interferes with the ability to selectively gate only task-relevant stimuli into working memory[48].

The limbic loop, which involves the ventral striatum, may implement the critic, as noted above, learning the values of states. It is tempting to speculate that the direct and indirect pathways learn the positive and negative values of states, respectively. Consistent with this idea, in conditioned place preference, which involves learning the values of places (states) and depends on the nucleus accumbens[49], learning a positive value for a place depends on neurotransmission in the direct, but not the indirect, pathway[38]. Furthermore, cocaine-induced conditioned place preference is increased by optogenetic activation of the direct pathway in the nucleus accumbens during learning and is decreased by activation of the indirect pathway[50].

The OFC has been hypothesized to implement working memory for state reinforcement values, helping to guide action selection when these values change rapidly, among other scenarios[51]. If so, the limbic loop may also implement Go and NoGo for 'actions' that determine when the values maintained in OFC should be updated so as to ensure adaptive, flexible behavior.

Clinical applications

These ideas and models, when applied to both motor and cognitive domains, explain a variety of findings across several disorders. Given dopamine's central role in reinforcement learning, we focus on disorders with strong dopaminergic involvement: Parkinson's disease, Tourette's syndrome, attention-deficit/hyperactivity disorder (ADHD), drug addiction and schizophrenia. Of course, these disorders often also involve nondopaminergic disturbances (for example, Tourette's syndrome involves abnormalities in striatal interneurons[52], ADHD involves noradrenergic abnormalities[53], etc.). Furthermore, in some cases, the dopaminergic disturbances themselves may be caused by upstream abnormalities. For example, schizophrenia involves increased mesolimbic and decreased mesocortical dopamine[54,55], both of which may be caused, at least in part, by dysregulated cortical control of dopamine neurons as a result of NMDA-receptor abnormalities[56]. (In fact, NMDA abnormalities may directly contribute to many of the symptoms of schizophrenia by disrupting the stability of

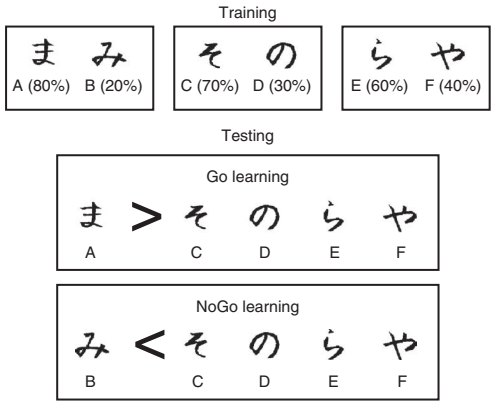

Figure 3 The probabilistic selection task[58]. The probabilistic selection task assesses whether participants learn better from positive or negative outcomes. During training, in each trial, participants are presented with one of the pairs shown on top (AB, CD and EF) and select one of the two stimuli. Feedback then indicates if the choice was correct or incorrect. The probabilities of each stimulus leading to correct feedback are indicated in the figure. Participants may learn to perform accurately during training (that is, learn to select A, C and E) by learning which stimulus in each pair is associated with positive feedback (Go learning), by learning which stimulus in each pair is associated with negative feedback (NoGo learning) or both. The test phase assesses the degree to which participants learned better from positive or from negative feedback. Participants are presented with novel pairs of stimuli consisting of either an A or a B paired with each of the other stimuli (C through F, which on average had a 50% probability of positive feedback during training). No feedback is provided during testing. If participants perform better on the pairs that include A than on those that include B, that indicates that they learned better to select the most positive stimulus (A) than to avoid the most negative stimulus (B), so they learn better from positive feedback (Go learning). If they perform better on the pairs that include B, they learn better from negative feedback (NoGo learning). The test phase can also be used to assess how participants adjust their behavior as a function of conflict (for example, whether they slow down to improve accuracy in high-conflict situations, such as when deciding between stimuli A and C, which have very similar reinforcement histories).

cortical attractors[57].) Comprehensive models of these disorders will ultimately have to integrate these various abnormalities. However, even a more limited focus on the small number of principles articulated above concerning the computational functions of dopamine and CBGTC circuits provides substantial leverage to understand multiple aspects of these disorders.

Parkinson's disease. Dopaminergic cell death in Parkinson's disease results in reduced striatal dopamine, thereby producing an exaggerated tendency for NoGo[10,28]. Simulations using the BG-GNG model have shown that this tendency explains not only Parkinson's disease hypokinetic symptoms, but also a variety of cognitive deficits that accompany this disorder[10,28,58]. For example, as noted above, the model suggests that Go and NoGo signals in the associative loop facilitate and prevent working memory updating, respectively[46,47]. The hyperexcitable NoGo pathway in Parkinson's disease should therefore produce a deficit in working memory updating, while simultaneously producing increased resistance to distractors. Furthermore, L-DOPA and dopamine agonists should reverse these effects. Empirical studies confirmed these predictions[59,60].

Subjects' tendency to learn better from positive or from negative feedback (Go and NoGo learning, respectively) can be assessed using the probabilistic selection task[58] (**Fig. 3**). Healthy controls are equally good at learning to obtain positive outcomes (Go learning) and avoid negative outcomes (NoGo learning) in this task[4,58] (although there are individual differences in Go versus NoGo learning among healthy controls, which are predicted by genetic variations affecting dopamine function in the direct and indirect pathways[37]). Consistent with the predictions of the BG-GNG model, unmedicated individuals with Parkinson's disease are better at NoGo learning than at Go learning[58]. Medication reverses these biases: medicated individuals with Parkinson's disease are better at Go than at NoGo learning and are worse at NoGo learning than unmedicated individuals with Parkinson's disease or controls[4,58]. These medication effects were also predicted by the model, under the assumption that dopaminergic medications reduce dopamine dips during negative prediction errors (because dopaminergic medications result in continued occupation of postsynaptic dopamine receptors during pauses in firing of dopamine neurons)[58]. Such blunting of negative prediction errors reduces learning from negative outcomes, producing the deficit in NoGo learning. Similar findings for both unmedicated and medicated individuals with Parkinson's disease have been obtained using several other tasks that also assess the degree to which subjects learn more from positive or negative outcomes[10]. These asymmetries in learning from positive and negative outcomes may have clinical implications. For example, the medication-induced tendency to learn more from positive than from negative outcomes may explain why medication induces pathological gambling in a subset of individuals with Parkinson's disease.

The BG-GNG model also suggests that the subthalamic nucleus (STN), the key node in the hyperdirect pathway, provides a Global NoGo signal that transiently inhibits all actions during action selection[35]. Activation of the STN, and therefore the Global NoGo signal, is dynamically modulated by the amount of response conflict (see also ref. 33). The Global NoGo signal is therefore particularly strong in situations of high conflict, in which actions with relatively similar reinforcement histories are being considered. In such situations, the Global NoGo inhibition provides time for the best action to win the competition, preventing premature, suboptimal actions from being facilitated. The model therefore predicted that disruption of STN processing—for example, by deep brain stimulation (DBS)—would disrupt subjects' ability to slow down in such high-conflict situations, resulting in faster, but suboptimal, responses[4]. This prediction was confirmed experimentally in individuals with Parkinson's disease undergoing DBS of the STN (and the deficit was resolved when DBS was turned off)[4]. Furthermore, as predicted by the model, dopaminergic medications and DBS had doubly dissociable effects: medications affected the asymmetry in learning from positive and negative outcomes, but not the ability to slow down in high-conflict situations, whereas DBS had the opposite effects[4]. The adverse effects of medications and DBS on real-life behavior in a subset of individuals with Parkinson's disease may therefore be produced by disruptions in distinct neurocognitive processes.

Tourette's syndrome. Tourette's syndrome is characterized by recurrent, stereotyped movements and vocalizations, known as tics. Tics have been hypothesized to reflect abnormal activation of subsets of striatal neurons that provide Go signals for the tic[61]. Evidence from Tourette's clinical pharmacology and from experimental work in animals suggests that tics may result from excessive excitability or plasticity in the direct (Go) relative to the indirect (NoGo) pathway. First, D2 blockers, the standard pharmacological treatment for Tourette's, boost the indirect pathway (because the D2 receptor is inhibitory). Second, administration of dopamine, amphetamine, or a combination of D1 and D2 agonists into the striatum, all of which

simultaneously boost the direct pathway (via D1 receptors) and inhibit the indirect pathway (via D2 receptors), causes stereotypies in animals[62,63]. Several findings suggest that the effects on the two pathways work synergistically to induce stereotypies; for example, stereotypies induced by striatal amphetamine administration are reduced by pretreatment with either D1 or D2 antagonists[62].

Excessive Go relative to NoGo activity in Tourette's syndrome may be a consequence of excessive dopamine or dopaminergic receptor sensitivity in the striatum. Indeed, Tourette's syndrome has been associated with increases in dopamine release, dopaminergic innervation and D2 receptors in the striatum (although the evidence for the latter two alterations is somewhat inconsistent across studies)[55,64]. All of these potential alterations would result in a boosted Go relative to NoGo pathway. Consistent with a possible excess of striatal dopamine, unmedicated individuals with Tourette's syndrome learn better from rewards than from punishments[65]. These biases are the opposite of those shown by unmedicated individuals with Parkinson's disease[4,58,65], mirroring the fact that the symptoms of these two disorders are also, to a limited extent, the opposite: Parkinson's disease is a hypokinetic disease, whereas Tourette's syndrome is a hyperkinetic disease. In fact, medication inverts these learning biases, making individuals with Parkinson's disease medicated with L-DOPA and dopamine agonists perform similarly to unmedicated individuals with Tourette's syndrome (learning better from rewards than from punishments) and making individuals with Tourette's syndrome medicated with D2 blockers perform similarly to unmedicated individuals with Parkinson's disease (learning better from punishments than from rewards)[65]. The mechanistic explanation for the former finding has already been discussed; the latter finding is consistent with the results of simulations showing that D2 blockade increases excitability and plasticity of the indirect pathway, thereby promoting NoGo learning[10]. This enhancement of NoGo learning by D2 blockade suggests that acute administration of D2 antagonists may be an effective adjunct for behavioral therapies that work by assigning negative value to tics (for example, contingency management, in which tics are followed by punishment or the absence of tics is positively reinforced, or massed negative practice, in which tics become aversive owing to fatigue). Whether acute D2 blockade would also be useful as an adjunct to habit reversal training, the best current behavioral treatment for Tourette's syndrome, is unclear because this procedure does not obviously involve aversive learning.

Excessive Go relative to NoGo activity in the motor CBGTC loop may also explain the premonitory urges that are a prominent feature of Tourette's syndrome. These urges are hypothesized to be caused by abnormal activation in the supplementary motor area (SMA) because electrical stimulation of the SMA causes similar urges[66]. Consistent with this idea, the SMA is active before tics[67], and SMA activation is greater with tics than with movements that mimic tics (and that are visually indistinguishable from tics) performed by healthy controls (Z. Wang, T.V.M., R. Marsh, T. Colibazzi, A. Gerber *et al.*, unpublished data). The SMA is the primary target of the motor CBGTC loop[24], so abnormal SMA activation could be a consequence of excessive relative Go activity in that loop. Alternatively, or in addition, the abnormal SMA activation could be driven by the corticocortical projections between the state and the SMA (or preSMA), as in the BG-GNG model (see **Fig. 2b**). In the model, these connections are learned via Hebbian mechanisms, so repeated gating of a tic by the basal ganglia in one or multiple states (initially because of excessive Go relative to NoGo activity) would strengthen the connections between those states and that tic's motor plan in the SMA. That SMA motor plan would then tend to become activated in those states, producing the urge (which could then be gated by the basal ganglia into an actual tic emission, but could also be prevented from doing so). This account, if correct, would explain important clinical features of Tourette's syndrome. For example, it would explain the state dependency of tics (that is, the fact that tics do not occur equally frequently in all contexts). It would also explain why treatments that prevent tic performance, such as habit reversal training, over time result in reduction of the urges: repeated activation of the tic-eliciting states without corresponding tic emission would produce Hebbian unlearning of the association between those states and the tics. More broadly, this account suggests that Tourette's syndrome involves a vicious cycle: performing a tic strengthens the urges to perform that tic (through corticocortical Hebbian learning), which in turn increases the tendency to tic (through corticocortical activation of the tic motor plan in the SMA, which increases the likelihood of basal ganglia gating of that tic).

ADHD. ADHD is characterized by abnormal levels of inattention, hyperactivity and impulsivity. The classical theory of ADHD is that it results from a primary deficit in inhibitory control, which causes several deficits in executive function[68]. Another prominent theory is that ADHD results from excessive discounting of delayed rewards[69]. 'Multiple-pathway' accounts suggest that executive dysfunction and excessive delay discounting are both involved[70]. ADHD seems to involve a hypofunctioning dopaminergic system[11,71].

One reinforcement learning theory suggests that tonic dopamine in the ventral striatum determines the discount factor (the degree to which future reinforcers are discounted relative to immediate ones) in a reinforcement learning system that can look ahead because it includes an internal model of the environment[72]. Reduced tonic dopamine in the ventral striatum in ADHD would produce a smaller discount factor, causing excessive discounting of delayed rewards[72]. This idea seems consistent with some circumstantial evidence: systemic administration of dopamine blockers increases delay discounting, and increasing dopamine via administration of psychostimulants or selective dopamine reuptake inhibitors generally decreases delay discounting[73]. However, dopamine depletion in the nucleus accumbens does not seem to produce excessive delay discounting[74], whereas dopamine depletion in the OFC does[75]. Excessive delay discounting in ADHD may therefore be caused by low dopamine in the OFC.

Biophysically realistic neurocomputational models suggest that dopamine stabilizes representations in PFC[76]. Given the key role of top-down biases from PFC in attention, cognitive control (including inhibitory control) and working memory[77], the hypothesized low PFC dopamine in ADHD could underlie deficits in all of these executive functions. This idea contrasts with the idea that the primary problem in ADHD is with inhibitory control, with problems in the remaining areas being secondary[68]. Cognitive deficits in ADHD need not be caused solely by low PFC dopamine, however: low striatal dopamine in the associative loops may cause, for example, reduced gating of working memory[11]. Conversely, PFC dysfunction need not only cause executive dysfunction: the lateral PFC appears to be involved in the ability to choose delayed rewards[78] (as is the OFC), so a dysfunctional PFC could also contribute to excessive delay discounting. The relationship between neurobiological abnormalities and cognitive and motivational deficits may therefore not be one to one and may vary across individuals with ADHD.

Drug addiction. The importance of fast, phasic-like changes in striatal dopamine in the reinforcing effects of drugs[79] makes addiction a natural candidate for reinforcement learning modeling. An influential

reinforcement learning theory suggests that these fast increases function as positive prediction errors that occur every time the drug is received[14]. This effect of drugs contrasts with the effect of natural rewards, for which the prediction error becomes zero after the reward is expected. The recurring drug-induced positive prediction errors produce a boundless increase in the values of states (or actions) that lead to drug receipt, prompting compulsive drug use. This theory explains important features of addiction[14], but a subsequent study refuted one of its key predictions. The theory predicted that when drugs are used as the unconditioned stimulus, blocking should not occur. Blocking is a procedure in which a stimulus, A, is first paired with an unconditioned stimulus and, subsequently, simultaneous presentations of A and another stimulus, B, are paired with the unconditioned stimulus. Usually, no learning occurs for B, as A already predicts the unconditioned stimulus, so there is no prediction error to support new learning. If, however, drugs always produce a positive prediction error, then subjects should learn to associate B with the unconditioned stimulus—but they do not[80]. This finding prompted a search for alternative reinforcement learning accounts of addiction.

One approach moved from standard temporal-difference learning to average-reward reinforcement learning[81]. In average-reward reinforcement learning, reinforcements are evaluated relative to an average reinforcement value $\bar{R}_t$ calculated using a slowly changing weighted average of past reinforcements. The 'effective' reinforcement at time t is therefore $r_t - \bar{R}_t$, where r_t is the received reinforcement. Part of the motivation for this approach to addiction was to capture the decrease in sensitivity to natural rewards that long-term drug use induces. Intuitively, if drugs are extremely reinforcing, long-term drug use inflates $\bar{R}_t$, making natural rewards less reinforcing. The model, however, added further to this effect by artificially inflating $\bar{R}_t$ even more with each drug use. The model showed decreased sensitivity to natural rewards following long-term drug use and also showed blocking and other relevant effects[81].

The search for simple, single-factor theories of addiction will undoubtedly continue, but multiple aspects of reinforcement learning are likely involved in addiction[82]. For example, chronic drug use induces functional and structural changes in important reinforcement learning brain regions (for example, the OFC[83]), thereby further dysregulating reinforcement learning and potentially contributing to the maintenance or aggravation of addiction. As another example, optogenetic findings in mice demonstrate that direct or indirect pathway stimulation during drug administration increases or decreases the reinforcing effects of the drug, respectively[50], suggesting that reduced indirect relative to direct pathway activity could be a risk factor for addiction. In fact, reduced indirect pathway activity would also explain the reduced sensitivity to negative outcomes that characterizes addiction.

Schizophrenia. Schizophrenia is characterized by positive symptoms (for example, delusions and hallucinations), negative symptoms (for example, anhedonia and avolition) and cognitive symptoms (for example, disturbances in attention and cognitive control). Schizophrenia involves excessive dopamine and D2 receptors in the striatum, but reduced dopamine in PFC[54,55].

One theory, based on the idea that dopamine signals incentive salience[84], suggests that dysregulated dopaminergic firing in schizophrenia imbues percepts, thoughts and memories with abnormal salience and that such abnormal salience experiences underlie delusions and hallucinations[85]. Another theory suggests that psychosis results from abnormal prediction errors that produce inappropriate associations, causal attributions and attentional salience[86]. Individuals with psychosis do exhibit abnormal neural activity during prediction errors[87,88], but a causal relation between these abnormalities and psychosis remains hypothetical.

Negative symptoms might conceivably reflect reduced reward sensitivity, but the evidence for this is mixed. Consistent with this hypothesis, individuals with schizophrenia exhibit reduced neural responses to positive prediction errors, with weaker putamen responses associated with greater avolition[89]. Individuals with schizophrenia also exhibit reduced Go learning[13,43,90], but these deficits do not seem to correlate with negative symptoms[43,90]. In these studies, negative symptoms were instead associated with indicators of PFC dysfunction[13,43]. Anhedonia, in particular, was associated with reduced uncertainty-driven exploration (in which alternative actions are explored in proportion to the uncertainty about their reinforcement statistics relative to the uncertainty about the reinforcement statistics of the currently preferred action)[90]. This may reflect the fact that the anhedonia assessment that was used partly relies on the frequency with which individuals engage in pleasurable activities, which may depend on strategic processes such as exploration.

Reduced striatal responses to positive prediction errors[89] and reduced Go learning[13,43,90] in schizophrenia are suggestive of reduced phasic striatal dopamine. Increased tonic striatal dopamine, as in schizophrenia[54], may reduce phasic dopamine via inhibitory autoreceptors. In fact, individuals with schizophrenia exhibit an overall Go bias (consistent with increased tonic striatal dopamine) coupled with decreased Go learning (consistent with decreased phasic striatal dopamine)[43]. In the associative loop, the tonic Go bias may produce excessive updating of PFC representations with irrelevant information. Reduced PFC dopamine likely adds additional lability to PFC representations[76]. The resulting extreme lability of PFC representations may underlie cognitive symptoms and contribute to positive symptoms.

Preclinical animal models. Reinforcement learning models can also shed light on preclinical animal models used to test new medications. One example is the use of conditioned avoidance to screen antipsychotics[91]. In conditioned avoidance, a warning stimulus is followed by shock unless animals perform a certain avoidance response after the onset of the warning stimulus, but before the shock. The avoidance response produces a transition from a state with negative value (in which shock is expected) to a state with neutral value (in which no shock is expected), so it elicits a positive prediction error[22]. These positive prediction errors are hypothesized to strengthen the stimulus-response association between the warning stimulus and the avoidance response in an actor-critic architecture[22]. Consistent with this idea, lesions of the ventral striatum (expected to damage the critic) and of the nigrostriatal dopaminergic projection (expected to prevent delivery of prediction errors to the actor) disrupt avoidance learning[92,93]. Disrupting dopaminergic signaling in the dorsal striatum only following training (through, for example, nigrostriatal lesions or dorsal striatal D2 blockade) does not, however, disrupt avoidance performance[92,94]. These findings are also consistent with the predictions of the actor-critic account, because dopaminergic signaling of prediction errors in the dorsal striatum is necessary for stimulus-response learning, but not for stimulus-response expression.

Unlike in the dorsal striatum, D2 blockade in the nucleus accumbens following training disrupts avoidance performance[94]. In fact, the standard use of conditioned avoidance to test antipsychotics is to administer them following training. The standard finding is that low antipsychotic doses disrupt avoidance, but not escape from ongoing shock, an effect that is mediated by the nucleus accumbens[91]. Dopamine in the nucleus accumbens can modulate

the activation of instrumental behavior[95,96], possibly via the striato-nigro-striatal spirals that allow the nucleus accumbens to influence the dorsal striatum[44,96]. The disruption of avoidance responding by antipsychotics may therefore reflect a decrease in the activation of instrumental behavior. Escapes would be less affected because the immediacy of pain would activate an innate flight response. An analogous situation is found in the appetitive domain: systemic or intra-accumbens administration of low doses of dopamine antagonists disrupts instrumental lever-pressing for food without affecting food approach or consumption[95].

Conclusions

We focused on a small number of reinforcement learning principles and examined how they can shed light on multiple disorders. Other aspects of reinforcement learning and related computational approaches also seem likely to be relevant for psychiatric and neurological disorders. For example, models of the role of the OFC in reinforcement learning[51] may be relevant for obsessive-compulsive disorder, which involves prominent OFC disturbances[97]; models of the role of serotonin in reinforcement learning[18] may be relevant for disorders that involve serotonergic abnormalities; and models of Pavlovian conditioning and extinction[98,99] may be relevant for some anxiety disorders (and have in fact already been shown to explain complex findings in fear conditioning in humans[100]).

To conclude, reinforcement learning models have been used to explain a wealth of findings across several psychiatric and neurological disorders. Although disorders as seemingly disparate as Parkinson's disease, Tourette's syndrome, ADHD, schizophrenia and addiction might seem to have little in common, they all involve disturbances in dopamine and CBGTC loops. The work reviewed above demonstrates that a mechanistic, computationally grounded understanding of the functions of these circuits sheds important light on all of these disorders. This approach relates to the new Research Domain Criteria initiative from the US National Institute of Mental Health, which calls for research that cuts across diagnostic criteria and focuses instead on neurocognitive domains and how they go awry in a variety of *DSM*-defined conditions. The work reviewed above exemplifies this strategy. This work also demonstrates the new level of theoretical sophistication that computational psychiatry and computational neurology bring to the venerable disciplines of psychiatry and neurology. Such theoretical sophistication and depth is essential if we are to fulfill the promise of a neuroscience-based, mechanistically detailed approach to diagnosis and treatment, which many agree should characterize the psychiatry and neurology of the future.

ACKNOWLEDGMENTS

Preparation of this article was funded by a Research Associate Award from the New York State Psychiatric Institute and the Research Foundation for Mental Hygiene, by National Institute of Mental Health grant R01 MH080066 and by a grant from the Michael J. Fox Foundation for Parkinson's Research.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprintsandpermissions/.

1. Charney, D.S. *et al.* Neuroscience research agenda to guide development of a pathophysiologically based classification system. in *A Research Agenda for DSM-V* (eds. Kupfer, D.J., First, M.B. & Regier, D.A.) 31–83 (American Psychiatric Association, Washington, D.C., 2002).
2. Hyman, S.E. Can neuroscience be integrated into the DSM-V? *Nat. Rev. Neurosci.* **8**, 725–732 (2007).
3. Cools, R. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neurosci. Biobehav. Rev.* **30**, 1–23 (2006).
4. Frank, M.J., Samanta, J., Moustafa, A.A. & Sherman, S.J. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science* **318**, 1309–1312 (2007).
5. Maia, T.V. Reinforcement learning, conditioning, and the brain: successes and challenges. *Cogn. Affect. Behav. Neurosci.* **9**, 343–364 (2009).
6. Dayan, P., Niv, Y., Seymour, B. & Daw, N.D. The misbehavior of value and the discipline of the will. *Neural Netw.* **19**, 1153–1160 (2006).
7. McClure, S.M., Daw, N.D. & Montague, P.R. A computational substrate for incentive salience. *Trends Neurosci.* **26**, 423–428 (2003).
8. Dayan, P. & Balleine, B.W. Reward, motivation, and reinforcement learning. *Neuron* **36**, 285–298 (2002).
9. Niv, Y., Daw, N.D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
10. Wiecki, T.V. & Frank, M.J. Neurocomputational models of motor and cognitive deficits in Parkinson's disease. *Prog. Brain Res.* **183**, 275–297 (2010).
11. Frank, M.J., Santamaria, A., O'Reilly, R.C. & Willcutt, E. Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology* **32**, 1583–1599 (2007).
12. Frank, M.J., Scheres, A. & Sherman, S.J. Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. *Phil. Trans. R. Soc. Lond. B* **362**, 1641–1654 (2007).
13. Waltz, J.A., Frank, M.J., Robinson, B.M. & Gold, J.M. Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biol. Psychiatry* **62**, 756–764 (2007).
14. Redish, A.D. Addiction as a computational process gone awry. *Science* **306**, 1944–1947 (2004).
15. Sutton, R.S. & Barto, A.G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
16. Bayer, H.M. & Glimcher, P.W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
17. Bayer, H.M., Lau, B. & Glimcher, P.W. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J. Neurophysiol.* **98**, 1428–1439 (2007).
18. Daw, N.D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**, 603–616 (2002).
19. Schultz, W. Dopamine signals for reward value and risk: basic and recent data. *Behav. Brain Funct.* **6**, 24 (2010).
20. Morris, G., Schmidt, R. & Bergman, H. Striatal action-learning based on dopamine concentration. *Exp. Brain Res.* **200**, 307–317 (2010).
21. Tsai, H.C. *et al.* Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**, 1080–1084 (2009).
22. Maia, T.V. Two-factor theory, the actor-critic model, and conditioned avoidance. *Learn. Behav.* **38**, 50–67 (2010).
23. O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
24. Alexander, G.E., DeLong, M.R. & Strick, P.L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
25. Yin, H.H. & Knowlton, B.J. The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* **7**, 464–476 (2006).
26. Mink, J.W. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog. Neurobiol.* **50**, 381–425 (1996).
27. Prescott, T.J., Gurney, K. & Redgrave, P. Basal ganglia. in *The Handbook of Brain Theory and Neural Networks* (ed. Arbib, M.A.) 147–151 (MIT Press, Cambridge, Massachusetts, 2003).
28. Frank, M.J. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *J. Cogn. Neurosci.* **17**, 51–72 (2005).
29. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).
30. Albin, R.L., Young, A.B. & Penney, J.B. The functional anatomy of basal ganglia disorders. *Trends Neurosci.* **12**, 366–375 (1989).
31. DeLong, M.R. Primate models of movement disorders of basal ganglia origin. *Trends Neurosci.* **13**, 281–285 (1990).
32. Nambu, A., Tokuno, H. & Takada, M. Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway. *Neurosci. Res.* **43**, 111–117 (2002).
33. Bogacz, R. & Gurney, K. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput.* **19**, 442–477 (2007).
34. Brown, J.W., Bullock, D. & Grossberg, S. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Netw.* **17**, 471–510 (2004).
35. Frank, M.J. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw.* **19**, 1120–1136 (2006).
36. Gurney, K., Prescott, T.J. & Redgrave, P. A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biol. Cybern.* **84**, 411–423 (2001).
37. Frank, M.J. & Fossella, J.A. Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology* **36**, 133–152 (2010).
38. Hikida, T., Kimura, K., Wada, N., Funabiki, K. & Nakanishi, S. Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron* **66**, 896–907 (2010).

39. Kravitz, A.V. *et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).
40. Gerfen, C.R. Molecular effects of dopamine on striatal-projection pathways. *Trends Neurosci.* **23**, S64–S70 (2000).
41. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D.J. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* **321**, 848–851 (2008).
42. Creese, I., Sibley, D.R., Hamblin, M.W. & Leff, S.E. The classification of dopamine receptors: relationship to radioligand binding. *Annu. Rev. Neurosci.* **6**, 43–71 (1983).
43. Waltz, J.A., Frank, M.J., Wiecki, T.V. & Gold, J.M. Altered probabilistic learning and response biases in schizophrenia: behavioral evidence and neurocomputational modeling. *Neuropsychology* published online, doi:10.1037/a0020882 (22 November 2010).
44. Haber, S.N. The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* **26**, 317–330 (2003).
45. Postuma, R.B. & Dagher, A. Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. *Cereb. Cortex* **16**, 1508–1521 (2006).
46. Frank, M.J., Loughry, B. & O'Reilly, R.C. Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect. Behav. Neurosci.* **1**, 137–160 (2001).
47. O'Reilly, R.C. & Frank, M.J. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* **18**, 283–328 (2006).
48. Baier, B. *et al.* Keeping memory clear and stable: the contribution of human basal ganglia and prefrontal cortex to working memory. *J. Neurosci.* **30**, 9788–9792 (2010).
49. Tzschentke, T.M. Measuring reward with the conditioned place preference paradigm: a comprehensive review of drug effects, recent progress and new issues. *Prog. Neurobiol.* **56**, 613–672 (1998).
50. Lobo, M.K. *et al.* Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330**, 385–390 (2010).
51. Frank, M.J. & Claus, E.D. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol. Rev.* **113**, 300–326 (2006).
52. Kalanithi, P.S. *et al.* Altered parvalbumin-positive neuron distribution in basal ganglia of individuals with Tourette syndrome. *Proc. Natl. Acad. Sci. USA* **102**, 13307–13312 (2005).
53. Biederman, J. & Spencer, T. Attention-deficit/hyperactivity disorder (ADHD) as a noradrenergic disorder. *Biol. Psychiatry* **46**, 1234–1242 (1999).
54. Guillin, O., Abi-Dargham, A. & Laruelle, M. Neurobiology of dopamine in schizophrenia. *Int. Rev. Neurobiol.* **78**, 1–39 (2007).
55. Nikolaus, S., Antke, C. & Muller, H.W. *In vivo* imaging of synaptic function in the central nervous system. II. Mental and affective disorders. *Behav. Brain Res.* **204**, 32–66 (2009).
56. Stahl, S.M. Beyond the dopamine hypothesis to the NMDA glutamate receptor hypofunction hypothesis of schizophrenia. *CNS Spectr.* **12**, 265–268 (2007).
57. Rolls, E.T., Loh, M., Deco, G. & Winterer, G. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nat. Rev. Neurosci.* **9**, 696–709 (2008).
58. Frank, M.J., Seeberger, L.C. & O'Reilly, R.C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940–1943 (2004).
59. Moustafa, A.A., Sherman, S.J. & Frank, M.J. A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia* **46**, 3144–3156 (2008).
60. Cools, R., Miyakawa, A., Sheridan, M. & D'Esposito, M. Enhanced frontal function in Parkinson's disease. *Brain* **133**, 225–233 (2010).
61. Mink, J.W. Basal ganglia dysfunction in Tourette's syndrome: a new hypothesis. *Pediatr. Neurol.* **25**, 190–198 (2001).
62. Delfs, J.M. & Kelley, A.E. The role of D1 and D2 dopamine receptors in oral stereotypy induced by dopaminergic stimulation of the ventrolateral striatum. *Neuroscience* **39**, 59–67 (1990).
63. Walters, J.R., Bergstrom, D.A., Carlson, J.H., Chase, T.N. & Braun, A.R. D1 dopamine receptor activation required for postsynaptic expression of D2 agonist effects. *Science* **236**, 719–722 (1987).
64. Steeves, T.D. & Fox, S.H. Neurobiological basis of serotonin-dopamine antagonists in the treatment of Gilles de la Tourette syndrome. *Prog. Brain Res.* **172**, 495–513 (2008).
65. Palminteri, S. *et al.* Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proc. Natl. Acad. Sci. USA* **106**, 19179–19184 (2009).
66. Peterson, B.S. *et al.* Neuroanatomical circuitry. in *Tourette's Syndrome: Tics, Obsessions, Compulsions. Developmental Psychopathology and Clinical Care* (eds. Leckman, J.F. & Cohen, D.J.) 230–259 (John Wiley & Sons, New York, 1999).
67. Bohlhalter, S. *et al.* Neural correlates of tic generation in Tourette syndrome: an event-related functional MRI study. *Brain* **129**, 2029–2037 (2006).
68. Barkley, R.A. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol. Bull.* **121**, 65–94 (1997).
69. Sagvolden, T. & Sergeant, J.A. Attention deficit/hyperactivity disorder—from brain dysfunctions to behaviour. *Behav. Brain Res.* **94**, 1–10 (1998).
70. Sonuga-Barke, E.J. Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biol. Psychiatry* **57**, 1231–1238 (2005).
71. Sagvolden, T., Johansen, E.B., Aase, H. & Russell, V.A. A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behav. Brain Sci.* **28**, 397–419 (2005).
72. Smith, A.J., Becker, S. & Kapur, S. A computational model of the functional role of the ventral-striatal D2 receptor in the expression of previously acquired behaviors. *Neural Comput.* **17**, 361–395 (2005).
73. Pattij, T. & Vanderschuren, L.J. The neuropharmacology of impulsive behaviour. *Trends Pharmacol. Sci.* **29**, 192–199 (2008).
74. Winstanley, C.A., Theobald, D.E., Dalley, J.W. & Robbins, T.W. Interactions between serotonin and dopamine in the control of impulsive choice in rats: therapeutic implications for impulse control disorders. *Neuropsychopharmacology* **30**, 669–682 (2005).
75. Kheramin, S. *et al.* Effects of orbital prefrontal cortex dopamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology (Berl.)* **175**, 206–214 (2004).
76. Durstewitz, D., Seamans, J.K. & Sejnowski, T.J. Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J. Neurophysiol.* **83**, 1733–1750 (2000).
77. Maia, T.V. & Cleeremans, A. Consciousness: converging insights from connectionist modeling and neuroscience. *Trends Cogn. Sci.* **9**, 397–404 (2005).
78. Figner, B. *et al.* Lateral prefrontal cortex and self-control in intertemporal choice. *Nat. Neurosci.* **13**, 538–539 (2010).
79. Volkow, N.D., Fowler, J.S., Wang, G.J., Swanson, J.M. & Telang, F. Dopamine in drug abuse and addiction: results of imaging studies and treatment implications. *Arch. Neurol.* **64**, 1575–1579 (2007).
80. Panlilio, L.V., Thorndike, E.B. & Schindler, C.W. Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward. *Pharmacol. Biochem. Behav.* **86**, 774–777 (2007).
81. Dezfouli, A. *et al.* A neurocomputational model for cocaine addiction. *Neural Comput.* **21**, 2869–2893 (2009).
82. Redish, A.D., Jensen, S. & Johnson, A. A unified framework for addiction: vulnerabilities in the decision process. *Behav. Brain Sci.* **31**, 415–437, discussion 437–487 (2008).
83. Schoenbaum, G., Roesch, M.R. & Stalnaker, T.A. Orbitofrontal cortex, decision-making and drug addiction. *Trends Neurosci.* **29**, 116–124 (2006).
84. Berridge, K.C. & Robinson, T.E. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* **28**, 309–369 (1998).
85. Kapur, S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am. J. Psychiatry* **160**, 13–23 (2003).
86. Corlett, P.R., Honey, G.D. & Fletcher, P.C. From prediction error to psychosis: ketamine as a pharmacological model of delusions. *J. Psychopharmacol.* **21**, 238–252 (2007).
87. Corlett, P.R. *et al.* Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* **130**, 2387–2400 (2007).
88. Murray, G.K. *et al.* Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol. Psychiatry* **13**, 267–276 (2008).
89. Waltz, J.A. *et al.* Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. *Neuropsychopharmacology* **34**, 1567–1577 (2009).
90. Strauss, G.P. *et al.* Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. *Biol. Psychiatry* (in the press).
91. Wadenberg, M.L. & Hicks, P.B. The conditioned avoidance response test re-evaluated: is it a sensitive test for the detection of potentially atypical antipsychotics? *Neurosci. Biobehav. Rev.* **23**, 851–862 (1999).
92. Fibiger, H.C., Phillips, A.G. & Zis, A.P. Deficits in instrumental responding after 6-hydroxydopamine lesions of the nigro-neostriatal dopaminergic projection. *Pharmacol. Biochem. Behav.* **2**, 87–96 (1974).
93. Fantin, G. & Bottecchia, D. Effect of nucleus accumbens destruction in rat. *Experientia* **40**, 573–575 (1984).
94. Wadenberg, M.L., Ericson, E., Magnusson, O. & Ahlenius, S. Suppression of conditioned avoidance behavior by the local application of (–)sulpiride into the ventral, but not the dorsal, striatum of the rat. *Biol. Psychiatry* **28**, 297–307 (1990).
95. Salamone, J.D. & Correa, M. Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behav. Brain Res.* **137**, 3–25 (2002).
96. Yin, H.H., Ostlund, S.B. & Balleine, B.W. Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur. J. Neurosci.* **28**, 1437–1448 (2008).
97. Maia, T.V., Cooney, R.E. & Peterson, B.S. The neural bases of obsessive-compulsive disorder in children and adults. *Dev. Psychopathol.* **20**, 1251–1283 (2008).
98. Dayan, P., Kakade, S. & Montague, P.R. Learning and selective attention. *Nat. Neurosci.* **3** Suppl: 1218–1223 (2000).
99. Gershman, S.J., Blei, D.M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197–209 (2010).
100. Maia, T.V. Fear conditioning and social groups: statistics, not genetics. *Cogn. Sci.* **33**, 1232–1251 (2009).