



# 17 Reinforcement Learning, Conflict Monitoring, and Cognitive Control: An Integrative Model of Cingulate-Striatal Interactions and the ERN

Jeffrey Cockburn and Michael Frank

Fortune favors those who are able to align their plans and goals to accord with the constraints imposed on them by an intricate and dynamic world. However, this presents an exceedingly difficult assignment, since the constraints pressed on an organism are typically complex, uncertain, and even paradoxical. When foodstuffs run low in the fall, should a hungry forager explore new and unfamiliar territory, or should it conserve energy and wait for something to turn up? The situation may appear dire and warrant the hazards of straying from routine, yet knowledge built up over years of experience may suggest that patience will be rewarded.

Flexible goal-directed behavior demands an adaptive system capable of selecting behavior appropriate for a given context. Evidence spanning a range of methodologies suggests that the anterior cingulate cortex (ACC) and the basal ganglia (BG) are two of the core brain structures involved in cognitive control, both contributing to pathways critical for learning and decision making (see also chapters 2, 9, 17, and 18, this volume). At an abstract level of description, the ACC is generally thought to be involved in monitoring performance and instigating rapid behavioral adjustments when required, whereas the BG is thought to facilitate and suppress behavior based on more stable environmental statistics. Both of these functions have been simulated in separate computational models of the ACC<sup>5</sup> and BG.<sup>14</sup> Although considerable debate still surrounds the unique function each system serves, here we take the approach that a better understanding may also emerge from considering the interaction between the ACC and the BG.

We propose a model in which ACC activity is modulated in part by reinforcement learning processes in the BG. In particular, we focus on how this relationship between the ACC and the BG may help clarify our understanding of the error-related negativity (ERN), a component of the event-related potential (ERP) thought to be generated in the ACC. We begin with a brief overview of the two dominant theories explaining the ERN: the reinforcement learning hypothesis advanced by Holroyd, Coles, and colleagues, and the conflict monitoring hypothesis advocated by Botvinick, Yeung, and colleagues. This overview is followed by a sketch of the

T1



core BG model and its role in reinforcement learning and action selection. We then include a novel extension incorporating the ACC into the BG model, using simulated ACC activity to quantify the ERN as response conflict driven by reinforcement learning processes in the BG as a function of feedback processing. We conclude with a discussion of how this model may advance our understanding of both the ACC and the BG, in addition to resolving an ongoing debate between the two dominant models of the ERN.

### **The Error-Related Negativity**

The term error-related negativity (ERN) is most commonly associated with incorrect responses (response ERN or rERN) and incorrect feedback (feedback ERN or fERN). The rERN is typically observed 0 to 150 msec following overt incorrect responses in speeded-response tasks, exhibiting a frontocentral distribution symmetrical about the midline.<sup>24</sup> The fERN is typically observed 200 to 350 msec following unexpected incorrect feedback and has a scalp topography nearly identical to that of the rERN.<sup>33</sup> Dipole modeling has consistently located the neural source of both the rERN and the fERN to be within the medial frontal cortex, most likely in the ACC.<sup>10</sup> However, despite nearly two decades of research on the topic, the cognitive processes that give rise to the ERN are still under debate. In the following, we briefly outline two dominant theories, one relating the ERN to reinforcement learning processes, and the other linking the ERN to performance monitoring (see also chapter 18, this volume).

### **The Reinforcement Learning Theory of the ERN**

An influential theory proposed by Holroyd and Coles has linked the ERN to reinforcement learning processes in the brain.<sup>28</sup> Rooted in a temporal difference reinforcement learning framework, the rERN and fERN are integrated as indices of phasic dopamine signals impacting the ACC. The ACC, in turn, was proposed to use these signals to learn which action should be executed given the current context and goal.

Based on a wealth of behavioral, neurological, and theoretical evidence, the midbrain dopamine system has been shown to encode a powerful reinforcement learning signal referred to as a reward prediction error.<sup>35,45</sup> Within the framework of temporal-difference (TD) reinforcement learning, a prediction error signal provides a mechanism through which actions and events are linked to their outcomes, even when outcomes are temporally distal from their antecedent cause. In short, the goal of a TD learning agent can be reduced to making decisions that maximize its opportunity to encounter rewards in the future. This is achieved by ascribing

values to actions and/or states that reflect the expected value of future rewards. At the heart of this value learning process is the reward prediction error, which quantifies the discrepancy between the expected and actual outcomes of an action or state. This learning signal is used to adjust the agent's expectations such that they accurately predict actual outcomes. By propagating reward values back to the actions and/or states that reliably precede them, the agent learns to accurately predict future outcomes and can therefore make decisions that maximize future rewards.

If the ERN is associated with the arrival of phasic dopamine signals in the ACC, then the characteristics of the ERN should conform to predictions from normative reinforcement learning theory. Critically, TD reinforcement learning theory states that reward prediction error signals propagate back from reward delivery to preceding actions and cues as predictive values are learned. Holroyd and Coles, demonstrated that the ERN exhibits precisely this signature by employing a probabilistic learning task.<sup>28</sup> Early in learning, before reward feedback could be predicted, ERN amplitude was larger following feedback and smaller following responses. Later in learning, once the reward contingencies had been learned, ERN amplitude was smaller following reward feedback and larger following response. These results were extended to show that simply presenting a stimulus predictive of reward also modulated ERN amplitude in correspondence with learning.<sup>1</sup> Subsequent work has also correlated fERN variance with reward expectancy violations. In accordance with normative TD learning, unexpected outcomes elicit larger fERNs than expected outcomes.<sup>29</sup> Together, these results, alongside those from numerous additional studies, suggest that the ERN is indeed related to reinforcement learning processes in the brain, linking the rERN and fERN as indices of the ACC's response to the arrival of phasic dopamine signals. However, a theory has also been proposed that likens the ERN to a response monitoring processes, which we outline here.

### **The Conflict Monitoring Theory of the ERN**

In a model-based investigation of ACC function, Botvinick et al.<sup>5</sup> proposed that the ERN is a product of conflict monitoring processes in the ACC. The authors argued that the ACC is a response conflict monitor, a system that tracks the coactivation of mutually exclusive responses as an indication that additional cognitive resources may be required to help perform the task at hand. As conflict increases, the ACC can engage additional control processes to assist in guiding performance.

Various experimental methodologies have demonstrated that ACC activity increases when a prepotent response must be overridden, when one of several equally permissible responses must be selected, or when errors are made. Botvinick and colleagues hypothesized that response conflict is common to all three of

these scenarios.<sup>5</sup> This hypothesis was tested by applying a connectionist model to simulated tasks known to engage the ACC, most notably a version of the Eriksen flanker task used to investigate the ERN.<sup>5</sup> During the flanker task, subjects are asked to identify the central character in a multiletter array, which can be either congruent (>>>) or incongruent (><>). Modeling results demonstrated that simulated conflict was greater and more sustained on incorrect trials, which were overwhelmingly associated with incongruent trials. On incorrect incongruent trials, flanker information initially drove activation of the incorrect response unit, which was sufficient to generate the incorrect response. Hence, prior to the central character gaining control of response units, both responses exhibited some degree of activation, which is quantified as an increase in response conflict. Simulated response times revealed that incorrect responses were typically associated with faster response times, replicating human performance. An investigation into the model's activation dynamics demonstrated that the probability of eliciting the correct response increased when extra processing time allowed response conflict to resolve before a response was made. Critically, this finding predicted that conflict should be present prior to responses on difficult (incongruent) correct trials.

If the ERN is generated by ACC activation associated with response conflict, then an ERN-like signature should be observed before correct responses on high conflict trials. Yeung and colleagues confirmed this prediction in a detailed investigation of ERPs recorded during a variant of the Eriksen flanker task.<sup>54</sup> As predicted, their results revealed a large rERN following incorrect responses. In addition, their results also revealed modulation of the N2 prior to correct responses. A large negative deflection was observed prior to correct response on high-conflict trials (incongruent, ><>) but not on low-conflict trials (congruent, >>>). Scalp topography and dipole source localization suggested that N2 modulation prior to response shared the same neural generator as the rERN following incorrect responses. Additionally, simulations demonstrated that conflict detection can provide a simple yet reliable mechanism for error detection, thus linking the ERN to conflict monitoring and error detection processes.

In summary, two theories have come to provide the dominant explanations of the ERN: one proposing that the ERN is an index of reinforcement learning processes, and the other positing that the ERN is an index of conflict monitoring. Although these theories are commonly understood to be mutually exclusive, the abundance of data supporting each theory argues against the tenability of rejecting one theory in favor of the other. In the following, we outline preliminary results suggesting that both theories of the ERN may be unified by considering the activation dynamics of the BG in relation to the ACC. However, we first visit evidence for a functional link between the BG and the ACC.

## The Basal Ganglia

The basal ganglia (BG) are composed of an anatomically and functionally linked group of subcortical nuclei located at the base of the forebrain<sup>34</sup> (see also chapter 2, this volume). The BG are thought to facilitate a wide range of faculties, including motor, motivational, and cognitive processes. Such breadth could be interpreted as an indication that multiple disparate roles are subsumed by the BG; however, a more unified and integrative functional role has been suggested by recent model-based investigations of the BG.

In much the same way that discovering the viral source of the common cold unified seemingly unrelated coughing, sneezing, fatigue, and feverish symptoms; broadly speaking, the BG's function can be conceptualized as a system that dynamically and adaptively gates the flow of information among cortical regions via cortico–basal ganglia–thalamocortical loops. Anatomically, the BG is well suited for this role, serving as a central way station where projections from numerous cortical structures converge, including prefrontal cortex, sensory cortex, hippocampus, and amygdala. Hence, the BG is positioned to integrate information from multiple systems, including the candidate motor or cognitive actions represented in frontal cortex, and to gate the most adaptive of these actions to be carried out while suppressing the execution of competing actions.

But how do the BG “know” which action plans should be facilitated and which should be suppressed? Any plausible model of such functionality should avoid appealing to what amounts to a homunculus in the BG, demanding a priori knowledge and pulling the levers as necessary. Following is a high-level overview of a biologically constrained neural-network model of the BG. This model has accounted for a variety of seemingly disparate behaviors, and has led to novel predictions that have been confirmed, as a result of medication, diseases, neuroimaging, and genetics.<sup>14,15</sup> By encapsulating the BG's core structure and activation dynamics, the model explicitly specifies the mechanisms through which the BG learns to integrate and gate information.

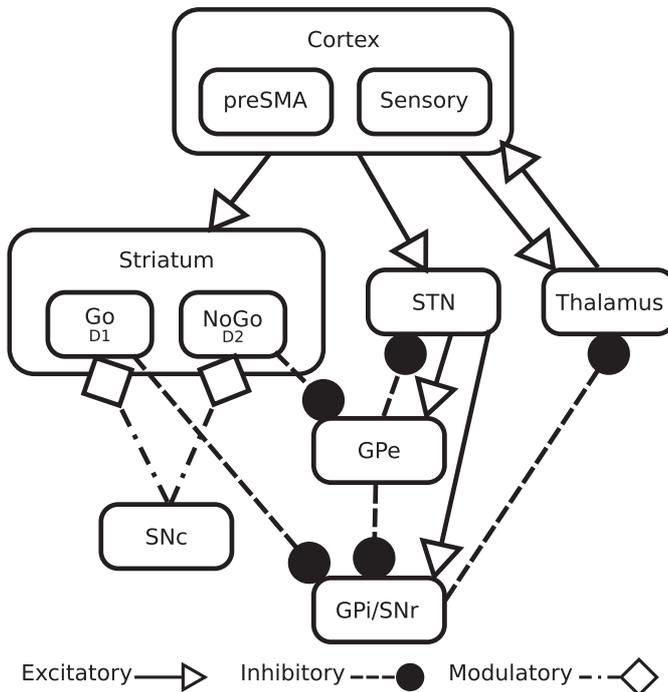
### Neural-Network Model of the Basal Ganglia

The neural-network model described here strikes a balance between biological realism and conceptual clarity. Low-level details such as cell geometry are abstracted away, while other critical biological details such as the temporal dynamics of membrane potentials, the divergent firing properties of neurons in distinct brain regions, and anatomical connectivity are captured. The result is a model with both biologically and functionally oriented constraints, positioning the model between detailed biophysical models and traditional connectionist models. Although the models

described in this chapter focus on the mechanisms of basic action selection, the same core circuitry has been extended to model cognitive functions such as working memory,<sup>36</sup> the role of instructions,<sup>12</sup> and the complementary roles of orbitofrontal representations of reward value and their influence on the BG.<sup>16</sup>

**Model Structure and Dynamics** The BG model consists of several layers that comprise the core structures associated with the BG (see figure 17.1): the striatum, globus pallidus (internal and external segments, GPi, and GPe), substantia nigra pars compacta (SNc), thalamus, and subthalamic nucleus (STN). When a stimulus is represented in sensory cortex, candidate actions are generated in premotor cortex, with both of these cortical regions projecting to the striatum. Columns of striatal units then encode information related to a particular candidate action in the context of the sensory stimulus.

The BG's core gating function emerges as activity from "direct" and "indirect" pathways converges on the thalamus, resulting in either the facilitation or suppression of cortical activations due to recurrent connectivity between thalamus and cortex.<sup>34</sup> The direct pathway originates in striatal D1-receptor-expressing units that



**Figure 17.1**  
Functional anatomy of the BG circuit.

provide focused inhibitory input to the GPi. The GPi is tonically active in the absence of synaptic input and sends inhibitory projections to the thalamus, preventing it from facilitating any cortical response representations. When striatal units in the direct pathway fire, they inhibit the GPi, and remove tonic inhibition of the thalamus. This “disinhibition” process allows premotor representation of the candidate action in question to excite the corresponding column of thalamus, which reciprocally amplifies the cortical activity via recurrent thalamocortical activity. Once a given action is amplified by striatal disinhibition of the thalamus, the competing candidate actions are immediately suppressed due to lateral inhibition in cortex. Thus, we refer to activity in the direct pathway as “Go” signals, which facilitate the selection of particular cortical actions.

Note that, critically, disinhibiting thalamic units only permits those units to become active if those same units also receive top-down excitation from cortex. Hence, the BG is not directly responsible for determining which actions are selected; rather, it modulates the activity of candidate representations already present in cortex. This conceptualization implies that cortex implements its own action-selection process to determine the appropriate candidate actions, which in our models is based on the prior probability of having selected these actions in the context of the current stimulus.

The indirect pathway, so labeled due to its additional synapse passing through the GPe, provides an oppositional force to the direct pathway. Originating in striatal D2 receptor-expressing units, the indirect pathway provides direct and focused inhibitory input to the GPe, which in the absence of input is also tonically active and sends focused inhibitory input to the GPi. We refer to activity in this pathway as “No-Go” signals since these striatal columns inhibit GPe, which transitively disinhibit their respective columns in the GPi further, ultimately preventing the flow of thalamocortical activity. Together, the balance between activity in the direct and in the indirect pathways for each action determines the probability that the action in question is gated.

In addition to the direct and indirect pathways, the model also includes a “hyperdirect” pathway through the STN, so named because this pathway originates in cortex and bypasses the striatum entirely. Like the striatum, the STN receives excitatory input from cortex; however, the STN projects excitatory input to the GPi, which further inhibits thalamocortical activity. Unlike the indirect pathway, STN projections to the GPi are diffuse, and therefore STN activity provides a “global No-Go” signal preventing any response from being gated (see also chapter 11, this volume). Further differentiating the hyperdirect pathway from the striatum, overall STN activity is dynamically regulated over the course of a trial: When a stimulus is presented, an initial STN surge is observed that sends a transient global No-Go signal preventing any response from being selected. When STN activity subsides

(due to feedback inhibition from GPe and neural accommodation), the effective threshold to gate a response declines.<sup>15</sup> Notably, the initial STN surge is amplified when several candidate actions are coactivated in cortex. Thus, the global No-Go signal is adaptively modulated by the degree of response coactivation (response conflict). This self-regulatory mechanism helps prevent premature response selection when multiple actions appear appropriate by allowing the striatum more time to integrate the information provided to it.

**Model Plasticity** As previously discussed, dopamine is central to reinforcement learning processes in the brain. Although our discussion of the reinforcement learning theory of the ERN focused on dopamine signals in the ACC, the relationship between dopamine and learning is perhaps best characterized in the BG. Evidence from patient studies and pharmacological manipulations that primarily affect and target the BG<sup>38,43</sup> suggest that DA processes in the BG are critically involved in reinforcement learning and action selection.<sup>4,7,20,21,37,51</sup> These human studies are complemented by optogenetic manipulations, voltammetry, synaptic plasticity, and single-cell recordings in nonhuman animal studies demonstrating that synaptic plasticity in the striatum depends on phasic dopamine signals.<sup>39,41,42,44,46,48</sup>

Correspondingly, dopamine plays a critical role in the BG model, allowing it to learn which cortical representation should be facilitated and which should be suppressed. Dopamine's impact depends on both the receptor type and the state of the target unit. In Go units, dopamine has a D1 receptor-mediated contrast-enhancing effect by further increasing activity in highly active units while simultaneously decreasing activity in less active units.<sup>14</sup> Thus, only those Go units that have the strongest activity, due to strong weights associated with the stimulus-action conjunction in question, are amplified. In No-Go units, dopamine has a D2 receptor-mediated inhibitory effect, such that greater levels of dopamine inhibit No-Go cells, whereas these cells are more excitable when DA levels are low.<sup>9,53</sup>

The net result is that increased dopamine levels facilitate Go activity while suppressing No-Go activity. Conversely, by reducing inhibition of D2-expressing units, decreased dopamine levels increase the activity of No-Go units while simultaneously reducing activity in Go units. Hence, variation in dopamine levels has differential effects on Go and No-Go signals projected by striatum. In the context of this model, dopaminergic effects play a critical role in both modulating the overall propensity for responding (by modulating Go relative to No-Go activity), and (by modulating activity-dependent plasticity) also allowing the model to learn which representations should be facilitated and which should be suppressed.

The model learns by dynamically changing the connections strength between units over time and experience, according to Hebbian principles. Weights between units that are strongly and repeatedly coactivated are strengthened, simulating



long-term potentiation (LTP), whereas weights between units that do not reliably coactivate are weakened, simulating long-term depression (LTD). Both LTP and LTD processes are strongly modulated by dopamine<sup>46</sup> in a manner that is adaptive in the BG model based on principles of reinforcement learning,<sup>14</sup> as described next.

Dopamine activity in the SNc shows properties consistent with that required by reinforcement learning.<sup>35,45</sup> Specifically, dopamine neurons in the SNc fire in a phasic burst when unexpected rewards are encountered, whereas activity in these same neurons falls below tonic baseline firing rates when expected rewards are not delivered. In the model, phasic DA bursts are simulated by the SNc layer to encode reward feedback following correct responses. Following a phasic SNc burst, activity is increased in Go units associated with the action selected in the current stimulus context, while activity in other Go units and No-Go units declines. Driven by activity-dependent Hebbian learning principles, the weights between representative Go units and active cortical units (sensory and motor) are strengthened, while all other weights are weakened. Thus, Go learning to choose a particular action in the context of a given stimuli is supported by phasic dopamine bursts.

Conversely, dopamine dips are simulated by the SNc layer to encode error feedback following incorrect responses. D2-expressing striatal units in the No-Go pathway, which are typically inhibited by baseline dopamine activity, are disinhibited following a phasic SNc dip, thereby increasing No-Go unit activity driven by excitatory cortical input. This increased activity strengthens the weights between active cortical and No-Go units. Hence, phasic dips of dopamine support learning to avoid certain actions given a particular stimulus.

By coupling simple Hebbian-learning principles with reinforcement learning signals projected to striatal units, the BG comes to reliably facilitate actions that have the highest probability of yielding a positive outcome, and to avoid those actions leading to negative outcomes. In addition, as stimulus-response representations in sensory cortex and pre-SMA coactivate, the connection strengths between these active units increase. As such, the pre-SMA eventually learns to refine its selection of candidate actions based on the probability of having selected these actions in the past for a given context. This slower form of learning ingrains repeated actions as “habits,” enabling action selection to become increasingly independent of BG activity over time.

In summary, the model presented here encapsulates several of the BG’s key structures and the activation dynamics among them. Direct Go and indirect No-Go pathways, differentially modulated by dopamine, allow the model to learn not only what it should do, but what it should not do in particular stimulus contexts. Additionally, the hyperdirect pathway through the STN modulates the within-trial action selection dynamics by preventing premature decisions when a number of potential candidates seem appropriate, thereby allowing striatal units more time to settle on

T1



the most appropriate course of action. Eventually, once reward contingencies are well learned and intercortical connections have been sufficiently strengthened, action selection no longer depends on bottom-up support from the BG. Hence, alongside a wealth of neurological and behavioral evidence, and supported by several empirically confirmed model-based predictions, the model presented here links processes within the BG to action selection, reinforcement learning, and the development of direct intercortical connections. In the following, we investigate an extension of this model aimed at exploring the relationship between the BG and the ACC, with an emphasis on the BG's role in the emergence of the ERN. However, we first consider empirical evidence for a functional relationship between the BG and the ACC.

### **Evidence of a Functional Link between the BG and the ACC**

The ACC is thought to play a critical role in executive control. ACC activation has been noted in tasks involving learning and memory, language, perception, and motor control, which suggests that like the BG, the ACC subserves a relatively general cognitive function. A unifying hypothesis conceptualizes the ACC as monitoring performance with respect to anticipated rewards.<sup>40</sup> Within this framework, the ACC monitors for signals indicating that events are not going as well or as smoothly as expected, and engages additional adaptive control processes when they are required. Thus, broadly speaking, the BG and the ACC appear to share functional concerns as both structures contribute to optimizing performance with respect to anticipated rewards.

A relationship between the BG and the ACC is supported by research spanning multiple methodologies including functional imaging (fMRI), event-related potentials (ERP), genetics, and lesions studies. First and foremost, the BG and the ACC share rich reciprocal connectivity that constitutes a major component of the limbic loop, and are thus thought to play a major role in motivation. Limbic pathways originating in orbitofrontal cortex and ACC pass through the BG and back to the ACC via the thalamus to complete the loop. In addition to thalamic input, the ACC receives input from a number of cortical structures, including prefrontal and motor cortex.<sup>11,13</sup> Pyramidal cells in ACC project to numerous regions involved in motor control, including the BG as well as supplementary and primary motor areas.<sup>50</sup> Hence, in addition to the reciprocal neural connectivity between the BG and the ACC, these two systems also share indirect projections through intermediary structures such as motor cortex.

Empirical support for a functional link between the BG and the ACC is found in a study investigating the ERN in patients with BG lesions.<sup>49</sup> The ERN was significantly reduced in lesion patients compared to age-matched controls, demonstrating that damage to the BG alters activity in the ACC. Further evidence linking ACC

and BG activity has emerged from predictions based on the BG model discussed here. Frank and colleagues found that individuals that learn better from negative feedback exhibited a larger rERN and a larger fERN,<sup>23</sup> with similar effects found in subsequent studies.<sup>6,17,26</sup> Findings across several methodologies, including fMRI,<sup>52</sup> PET,<sup>8</sup> genetics,<sup>19</sup> and targeted genetic engineering manipulations of striatal direct and indirect pathways<sup>27</sup> have all demonstrated that ones' ability to learn from positive and negative feedback is determined by the capacity for reinforcement learning signals to shape activity in the Go and No-Go pathways, respectively.

These studies suggest that either ACC activity that produces the ERN is associated with learning from feedback independent of activity in the BG, or variance in BG activity associated with learning from feedback is correlated with variance in the ERN due to corticostriatal loops driving activity in the ACC. Evidence for the latter hypothesis comes from an experiment combining fMRI and genetics, in which a genetic variation known to affect striatal D2 receptor density predicted ACC activity.<sup>32</sup> Using a variant of the probabilistic learning task applied in Frank et al.,<sup>14</sup> Klein and colleagues grouped subjects according to genetic polymorphisms associated with D2 receptor expression.<sup>32</sup> They found that subjects with the A1 allele (A1+ group), which is associated with a 30% reduction of striatal D2 receptor density, were significantly worse at learning from negative feedback, replicating the findings reported with another D2 polymorphism.<sup>19</sup> Furthermore, they found that the A1+ group exhibited a significantly smaller change in ACC activation following error feedback compared to the A1- group (those without the A1 allele). Since D2 receptors are predominantly expressed by striatal No-Go pathway neurons,<sup>25</sup> and this same genetic variation affects striatal activity during negative outcomes,<sup>31</sup> this result provides further support for a functional link between activity in the BG and the ACC. Similarly, a recent study showed that D2 receptor genetic variation impacts D2 receptor binding in the striatum, which in turn, correlates with prefrontal cortical activity.<sup>3</sup>

### **BG Model of the ERN**

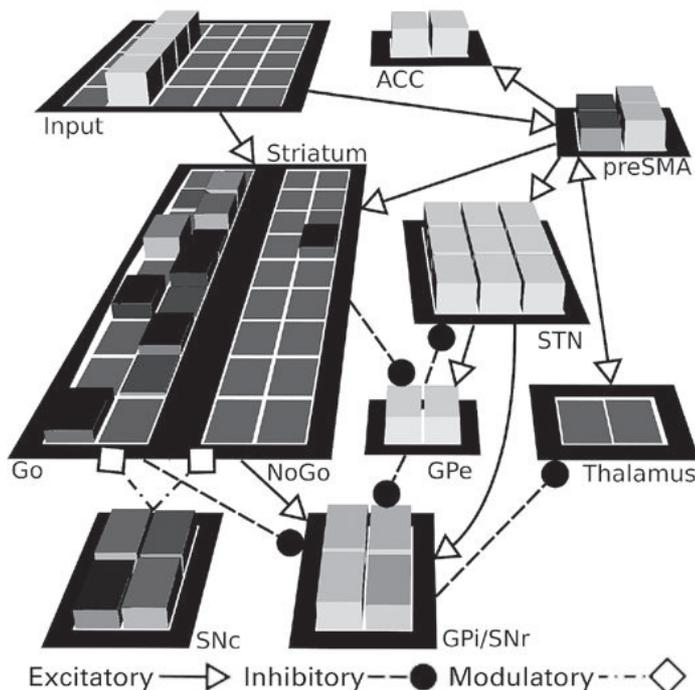
Although the reinforcement learning and conflict monitoring theories of the ERN are both supported by a wealth of empirical data, neither theory explains all ERN phenomenology. The reinforcement learning theory provides an account of the rERN and the fERN but omits the N2, whereas the conflict monitoring theory explains the N2 and rERN but neglects the fERN. Here, we propose an initial step toward a unified theory of the ERN by linking conflict in the ACC with striatal activation dynamics following reinforcement learning signals that encode reward feedback. In doing so, the model provides a means of quantifying the fERN in terms of conflict, allowing for precise predictions regarding variance in the fERN, and also

T1

integrates the well-studied functions of the BG and their implications for ACC function.

Given that the conflict monitoring theory of the ERN outlined by Yeung and colleagues has already demonstrated that variance in the N2 and rERN can be predicted by response conflict,<sup>54</sup> here we focus exclusively on activity following feedback. More specifically, the model presented here demonstrates that conflict in the ACC depends in part on the strength of bottom-up projections from the BG. Since thalamic output to cortical response representations is governed in part by dopaminergic modulation of striatal activity, the model specifies a mechanistic link between reinforcement learning signals and conflict in the ACC.

We augmented the core BG model discussed at the beginning of the chapter with a simple ACC layer (figure 17.2). The ACC layer consists of units representing each response available to the model, with each ACC unit receiving excitatory input from units in the pre-SMA layer encoding the equivalent response. The ACC monitors activity in the pre-SMA layer such that, for example, increased activity in pre-SMA units representing response “A” will drive a corresponding increase in the response



**Figure 17.2**

Neural network model of the BG circuit, with two different responses represented by two columns of units in each of the Go, No-Go, GPe, GPI/SNr, thalamus, pre-SMA, and ACC layers.



“A” unit of the ACC layer. In accordance with previous formalizations of conflict, we quantify conflict as the Hopfield energy<sup>30</sup> in the ACC layer. However, since there are only two possible responses in the current model, and two units in the ACC layer, the energy calculation reduces to

$$\text{Conflict} = -a_1a_2 \quad (17.1)$$

where  $a_1$  and  $a_2$  represent the activity in each unit of the ACC layer.

We applied the model to a simulation of a probabilistic selection task.<sup>15,22</sup> In this task, stimuli are probabilistically associated with reward and punishment, with some stimuli delivering rewards more reliably than others. On each trial, the model is forced to choose one of two stimuli, each response followed immediately by reward feedback. In order to provide a clear description of the mechanisms under investigation, we prevent the model from learning the task’s reward contingencies. Thus, response selection is random and all reward feedback is unpredictable throughout the task.

At the start of each simulated trial, units in the input layer representing the trial stimuli are activated and the network is allowed to freely settle on a response. Since the network has no experience with any of the stimuli or their associated rewards, response selection depends on random activity in the Go and No-Go pathways together with noise in pre-SMA. As outlined earlier, one response is selected via gating mechanisms of the BG and thalamus, and lateral inhibition between mutually exclusive response units in the pre-SMA. This ensures that a response will always be selected: Even a relatively subtle bottom-up bias for one response over the other (due to random initial weights and/or noise in activation dynamics) will be sufficient to select that response once the thalamus is disinhibited. Once a unit in the pre-SMA layer surpasses a threshold of 95% activation, a response is elicited. Feedback is then provided by the SNc layer in the form of a reinforcement learning signal encoded as a phasic burst or dip in activity. Finally, having delivered its phasic learning signal, the SNc returns to baseline activation levels and the network continues to settle until the end of the trial.

As illustrated in figure 17.3b, the activation dynamics of conflict in the ACC following feedback bear a striking resemblance to the temporal signature of the fERN, with conflict rapidly increasing following negative feedback, and returning to baseline levels once the reinforcement learning signal was processed. As discussed previously, input from the SNc differentially modulates activity in the Go and No-Go pathways. Correct feedback, encoded as a burst of output from the SNc, simultaneously strengthens the Go signal and weakens the No-Go signal projected to the thalamus (figure 17.3a). Thus, following correct feedback, bottom-up support for the selected response will increase, and in turn, lateral inhibition of the alternative response will also increase. Hence, should any response coactivation remain in the

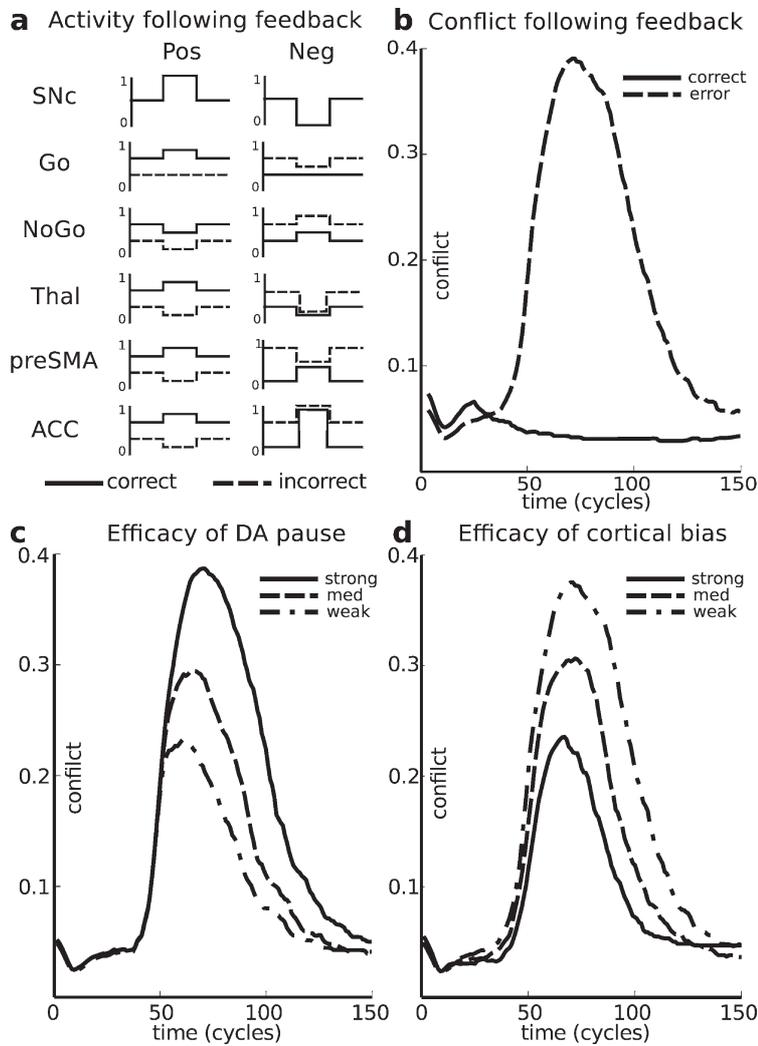


pre-SMA following response selection, a phasic dopamine burst will tend to facilitate the activation of a single response representation. As the ACC layer monitors activity in the pre-SMA layer following feedback, any remaining conflict is correspondingly reduced. When SNc activity returns to baseline levels, Go signal strength is reduced and conflict returns to normal levels.

Conversely, incorrect feedback, encoded as a dip in SNc activity, increases the No-Go signal projecting to the thalamus (figure 17.3a). Hence, following incorrect feedback, increased No-Go activity counteracts the original Go pathway activity, and the thalamus is no longer disinhibited. As a result, bottom-up support for the selected response in the pre-SMA layer is reduced, as is lateral inhibition of the response that was not selected. Consequently, pre-SMA layer activity reflects only the relatively weak influences of sensory cortical activation and noise, which are incapable of inducing activation of a single response in the absence of substantial prior corticocortical learning. Thus, conflict increases as the ACC layer monitors the activity of response representations in the pre-SMA layer. When SNc activity is restored to baseline levels, No-Go signal strength is reduced and Go signals are able to bias response representations once more. As the Go signal gains footing, facilitating the previously selected response in the pre-SMA, activity in the ACC comes to uniquely represent the biased response and conflict returns to baseline levels. Thus, much like the model of the rERN described by Yeung et al.,<sup>54</sup> an increase in ACC conflict provides an account of the fERN, where conflict is driven by negative reinforcement learning signals that induce NoGo signals in the BG.

We detail the emergence of conflict further by manipulating the capacity for information to pass through BG and cortical pathways into the pre-SMA. As was just discussed, conflict in the ACC layer is driven by the relative strength of Go and No-Go signals, which were themselves modulated by SNc activity. Given the current model structure, variance in the amount of conflict was dominated by activity in the No-Go pathway following errors.\* This suggests that altering the potency of reinforcement learning signals in the No-Go pathway (as would be expected due to D2 receptor genetic variation<sup>18</sup>) should impact conflict in the ACC. We investigated this by varying the duration of the phasic pause in SNc activity during negative outcomes. We note that this approach to manipulating dopamine efficacy in the No-Go pathway encompasses variation in the dopamine signal itself (the magnitude of negative prediction errors has been shown to be reflected in the duration of dopamine pauses<sup>2</sup>), and in turn the ability of the No-Go pathway to react to these signals

\* Given the model's response threshold of 95% activation, the Go pathway was unable to significantly reduce conflict following phasic dopamine bursts due to a ceiling effect. However, it is possible for the Go pathway to play a more prominent role in the fERN if responses were elicited at a lower threshold, allowing increased Go pathway activation to further excite the selected response, thereby reducing conflict in the ACC.



**Figure 17.3**

(a) Network activity associated with the correct (solid line) and incorrect (dashed line) response representations in each layer following correct (Pos) and incorrect (Neg) feedback. (b) Measure of conflict in the ACC layer following correct (solid line) and incorrect (dashed line) feedback. (c) Measure of conflict following incorrect feedback as a result of varying the efficacy of negative prediction errors from highly effective (solid line) to weakly effective (dash-dot line). (d) Measure of conflict following incorrect feedback as a result of varying the strength of cortical bias weights projecting to the pre-SMA layer from strong (solid line) to weak (dash-dot line).

(sufficient pause durations are necessary to disinhibit No-Go cells). Figure 17.3c illustrates that as pause duration is progressively reduced and dopamine efficacy is weakened, the duration and magnitude of conflict in the ACC is correspondingly reduced. Thus, our model suggests that variation in the No-Go signal should correlate with conflict in the ACC, and therefore predict the strength of the fERN signal.

We also investigated the relationship between top-down cortical information projected into the pre-SMA and the emergence of conflict in the ACC. Numerous cortical systems influence activity in the pre-SMA, including the ACC, sensory cortex, and prefrontal cortex. Thus, to avoid overcomplicating the model, we simulate top-down input as an excitatory bias weight projecting onto units in the pre-SMA layer, thus simulating aggregate input from all cortical structures signaling the response. Following response selection, the bias weights were applied to the pre-SMA units that encode the selected response for the remainder of the trial. Just as increasing bottom-up response facilitation reduced conflict in the ACC, increasing the top-down bias on response representations in the pre-SMA produced a corresponding reduction in conflict (figure 17.3d). Although a temporary reduction in thalamic facilitation following negative feedback remains as the primary factor contributing to the emergence of conflict, top-down bias weights were able to dampen the amount of activity in the ACC layer when thalamic input was removed. Thus, as the top-down bias weights were strengthened, the modulatory potency of thalamic input into pre-SMA was reduced as was the amount of conflict following negative feedback signals encoded by the SNc.

## Discussion

The model presented here provides preliminary evidence linking the fERN to a conflict-detection mechanism in the ACC. Critically, the model demonstrated that reinforcement learning signals encoded by phasic SNc activity differentially modulated the Go and No-Go signals responsible for facilitating response representations in the pre-SMA. When bottom-up support was removed due to increased No-Go signal strength following negative feedback, response coactivation increased and emerged as conflict in the ACC layer. It was further demonstrated that top-down biasing also played a role in the dynamics of conflict in the ACC. As cortical input to the pre-SMA layer increased, units there were less dependent on thalamic support to maintain activation of a single response representation, thus reducing conflict in the ACC.

As previously discussed, subjects with reduced striatal D2 receptor expression have been found to exhibit a diminished change in ACC activity following negative feedback.<sup>32</sup> Since D2 receptors are predominantly expressed by cells in the No-Go

pathway, this finding suggests that the No-Go pathway plays a critical role in modulating activity in the ACC. In addition, subjects that did not learn well from negative feedback were found to exhibit smaller fERNs.<sup>6,23</sup> Given that learning from negative feedback was also found to depend on activity in the No-Go pathway,<sup>19,27</sup> this further supports a functional link between the No-Go pathway and activity in the ACC responsible for generating the fERN. In line with these findings, our model demonstrated that the manifestation of conflict depends on the efficacy of reinforcement leaning signals in the No-Go pathway (see figure 17.3c). When the No-Go pathway's response to negative outcomes was blunted, as would be the case in populations with impaired D2 receptor function, the No-Go pathway's capacity to withhold thalamic support to the pre-SMA was reduced. This, we suggest, would emerge as a reduction in the fERN (with the critical assumption that the groups in question differ in striatal function but not in cortical measures).

Normative learning theory states that the magnitude of the reward prediction error is correlated with the severity of the prediction violation.<sup>47</sup> Results reported by Holroyd and Krigolson<sup>29</sup> show that the fERN correlates with violation of expectation, suggesting that the fERN is modulated by reward prediction errors encoded by the midbrain dopamine system. Although the reinforcement learning theory of the ERN argued that this correlation is driven by dopamine's arrival in the ACC, the model presented here demonstrates that dopaminergic modulation of striatal activity predicts the same correlation between dopamine signals and conflict in the ACC. Longer phasic pauses in dopamine activity induced greater NoGo signals that withheld bottom-up support more effectively, leading to more conflict in the ACC layer and a larger fERN (see figure 17.3c).

Our understanding of the rERN may benefit from the observation that cortical input to response representations in the pre-SMA also regulated conflict in the ACC layer. This suggests that as cortical projections to the pre-SMA are strengthened, conflict will be progressively less sensitive to influences of the BG. In the model presented here, input representations were externally clamped; however, a more realistic approach would allow these representations to emerge as stimuli are processed and differentially attended. The stability and veridicality of the stimulus representation would likely be determined by biological constraints and attentional processes. As such, activation patterns associated with the stimulus representation would be expected to waver, leading to fluctuations in the response representations with which they are associated. When response selection is more challenging (e.g., incongruent trials in the flanker task), increased levels of conflict would develop as stimulus representations compete for control. A correct response would be more probable on trials when this conflict was allowed to resolve prior to response, whereas an incorrect response would be more probable on trials when a response was elicited prior to conflict resolution. Thus, N2-like modulation of

ACC activity prior to response selection, and a rERN-like increase in conflict following incorrect responses will emerge in the absence of bottom-up conflict-inducing mechanisms.

Although the model presented here specified a mechanism that linked reinforcement learning to conflict monitoring in the ACC, this preliminary model has yet to ascribe a particular function to conflict detection in the ACC: Although conflict detection may provide a simple means of error detection,<sup>54</sup> this cannot be so in the case of conflict associated with the fERN. For the fERN to emerge, the error must first be detected and encoded as a phasic pause in SNc activity. Additionally, the astute reader may have noticed that the hyperdirect pathway through the STN is also monitoring conflict in cortical layers<sup>15</sup>; thus, it may initially appear that including a conflict monitor in the ACC does no more than add redundancy to the model. First, we note that the STN modulates activity in accordance with *instantaneous* conflict in a given trial; however, by itself, the STN is incapable of accumulating information across trials in order to adjust performance to meet more global task demands. Though purely speculative at this point in our research, we suggest that the ACC may provide a means of accumulating task performance over time (e.g., adjusting response caution in proportion to integrated conflict over trials, as proposed by Botvinick and colleagues<sup>5</sup>), allowing the ACC to modify behavior so as to meet more global task demands. By integrating pre-SMA conflict across trials, the ACC would be in a position to regulate behavioral control as a function of global task demands, a process that may be embodied by projecting this integrated signal to the STN as a means of blending global and local task demands. Alternatively (or in addition), the ACC may regulate trial-to-trial behavioral adjustments as a function of negative outcomes (“lose-switch”) as individuals acquire reinforcement contingencies (see Cavanagh et al.<sup>6</sup> for EEG data), in contrast to the BG, which integrates reinforcement probabilities across a longer time horizon.<sup>19</sup>

## Conclusion

We have outlined a biologically inspired and constrained neural network model aimed at investigating the functional relationship between the ACC and the BG. The model defined mechanisms through which dopamine-induced modulation of striatal activation patterns emerged as variance in ACC activity, which we quantified as conflict. By coupling reinforcement learning processes in the BG with activity in the ACC, the model provides a means of quantifying the fERN as an index of conflict-monitoring processes in the ACC. Although much work remains, this model provides preliminary evidence for a unified account of the ERN by including both reinforcement learning and conflict monitoring mechanisms within a single model.



### Outstanding Questions

- If conflict monitoring does indeed provide a means of error detection, as proposed by Yeung et al.,<sup>54</sup> could the phasic reinforcement learning signals encoded by the dopamine system that train the BG be driven in part by conflict processes associated with the ERN?
- How might the learning embodied by the BG and the ACC be differentially specialized, and how might the roles of each system differ with respect to guiding behavior?
- What experiment could be used to shed light on the proposed interactive model (e.g., simultaneous model-based fMRI and EEG to explore the relationship between striatal prediction errors and the ERN on a trial-by-trial basis)?

### Further Reading

Hazy TE, Frank MJ, O'Reilly RC. 2010. Neural mechanisms of acquired phasic dopamine responses in learning. *Neurosci Biobehav Rev* 34: 701–720. An overview and discussion of the biological mechanisms that underlie reward-predictive firing properties of midbrain dopamine neurons. Discusses specifically their relationship to learning.

Houk JC, Davis JL, Beiser DG (eds). 1995. *Models of Information Processing in the Basal Ganglia*. Cambridge, MA: MIT Press. The largest overview of models of the basal ganglia available. Although written quite some time ago, it is still extremely relevant today.

### References

1. Baker TE, Holroyd CB. 2009. Which way do I go? Neural activation in response to feedback and spatial processing in a virtual T-maze. *Cereb Cortex* 19: 1708.
2. Bayer HM, Lau B, Glimcher PW. 2007. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J Neurophysiol* 98: 1428.
3. Bertolino A, Taurisano P, Pisciotto NM, Blasi G, Fazio L, Romano R, Gelao B, et al. 2010. Genetically determined measures of striatal D2 signaling predict prefrontal activity during working memory performance. *PLoS ONE* 5: e9348.
4. Bodi N, Keri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibo G, Takats A, Berezcki D, Gluck MA. 2009. Reward-learning and the novelty-seeking personality: a between-and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. *Brain* 132: 2385.
5. Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. *Psychol Rev* 108: 624–652.
6. Cavanagh JF, Frank MJ, Allen JJB. Social stress reactivity alters reward and punishment learning. *Soc Cogn Affect Neurosci*. In press.
7. Cools R, Altamirano L, D'Esposito M. 2006. Reversal learning in Parkinson's disease depends on medication status and outcome valence. *Neuropsychologia* 44: 1663–1673.
8. Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M. 2009. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *J Neurosci* 29: 1538.





9. Day M, Wokosin D, Plotkin JL, Tian X, Surmeier DJ. 2008. Differential excitability and modulation of striatal medium spiny neuron dendrites. *J Neurosci* 28: 11603.
10. Debener S, Ullsperger M, Siegel M, Fiehler K, Von Cramon D, Engel A. 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *J Neurosci* 25: 11730.
11. Devinsky O, Morrell M, Vogt B. 1995. Contributions of anterior cingulate cortex to behaviour. *Brain* 118: 279–306.
12. Doll BB, Jacobs WJ, Sanfey AG, Frank MJ. 2009. Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain Res* 1299: 74–94.
13. Dum RP, Strick PL. 1993. Cingulate motor areas. In: *Neurobiology of Cingulate Cortex and Limbic Thalamus: A Comprehensive Handbook* (Vogt BA, Gabriel M, eds), pp 415–441. Cambridge, MA: Birkhaeuser.
14. Frank MJ. 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci* 17: 51–72.
15. Frank MJ. 2006. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw* 19: 1120–1136.
16. Frank MJ, Claus ED. 2006. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113: 300–326.
17. Frank MJ, D'Lauro C, Curran T. 2007. Cross-task individual differences in error processing: neural, electrophysiological, and genetic components. *Cogn Affect Behav Neurosci* 7: 297–308.
18. Frank MJ, Hutchison K. 2009. Genetic contributions to avoidance-based decisions: striatal D2 receptor polymorphisms. *Neuroscience* 164: 131–140.
19. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. 2007. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* 104: 16311–16316.
20. Frank MJ, O'Reilly RC. 2006. A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behav Neurosci* 120: 497–517.
21. Frank MJ, Samanta J, Moustafa AA, Sherman SJ. 2007. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science* 318: 1309–1312.
22. Frank MJ, Seeberger LC, O'Reilly RC. 2004. By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* 306: 1940–1943.
23. Frank MJ, Worocho BS, Curran T. 2005. Error-related negativity predicts reinforcement learning and conflict biases. *Neuron* 47: 495–501.
24. Gehring WJ, Goss B, Coles MG, Meyer DE, Donchin E. 1993. A neural system for error detection and compensation. *Psychol Sci* 4: 385–390.
25. Gerfen CR, Engber TM, Mahan LC, Susel Z, Chase TN, Monsma FJ, Jr, Sibley DR. 1990. D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* 250: 1429.
26. Hewig J, Trippe R, Hecht H, Coles MGH, Holroyd CB, Miltner WHR. 2007. Decision-making in blackjack: an electrophysiological analysis. *Cereb Cortex* 17: 865–877.
27. Hikida T, Kimura K, Wada N, Funabiki K, Nakanishi S. 2010. Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron* 66: 896–907.
28. Holroyd CB, Coles MG. 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 109: 679–709.
29. Holroyd CB, Krigolson OE. 2007. Reward prediction error signals associated with a modified time estimation task. *Psychophysiology* 44: 913–917.
30. Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79: 2554.
31. Jocham G, Klein TA, Neumann J, von Cramon DY, Reuter M, Ullsperger M. 2009. Dopamine DRD2 polymorphism alters reversal learning and associated neural activity. *J Neurosci* 29: 3695–3704.





32. Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M. 2007. Genetically determined differences in learning from errors. *Science* 318: 1642–1645.
33. Miltner WHR, Braun CH, Coles MGH. 1997. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *J Cogn Neurosci* 9: 788–798.
34. Mink J. 1996. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog Neurobiol* 50: 381–425.
35. Montague P, Dayan P, Sejnowski T. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936.
36. O’Reilly RC, Frank MJ. 2006. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput* 18: 283–328.
37. Palminteri S, Lebreton M, Worbe Y, Grabli D, Hartmann A, Pessiglione M. 2009. Pharmacological modulation of subliminal learning in Parkinson’s and Tourette’s syndromes. *Proc Natl Acad Sci USA* 106: 19179.
38. Pavese N, Evans AH, Tai YF, Hotton G, Brooks DJ, Lees AJ, Piccini P. 2006. Clinical correlates of levodopa-induced dopamine release in Parkinson disease: a PET study. *Neurology* 67: 1612.
39. Reynolds JN, Hyland BI, Wickens JR. 2001. A cellular mechanism of reward-related learning. *Nature* 413: 67–70.
40. Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. 2004. The role of the medial frontal cortex in cognitive control. *Science* 306: 443–447.
41. Roesch MR, Calu DJ, Schoenbaum G. 2007. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* 10: 1615–1624.
42. Samejima K, Ueda Y, Doya K, Kimura M. 2005. Representation of action-specific reward values in the striatum. *Science* 310: 1337–1340.
43. Sawamoto N, Piccini P, Hotton G, Pavese N, Thielemans K, Brooks DJ. 2008. Cognitive deficits and striato-frontal dopamine release in Parkinson’s disease. *Brain* 131: 1294–1302.
44. Schultz W. 2002. Getting formal with dopamine and reward. *Neuron* 36: 241–263.
45. Schultz W, Dayan P, Montague P. 1997. A neural substrate of prediction and reward. *Science* 275: 1593.
46. Shen W, Flajolet M, Greengard P, Surmeier DJ. 2008. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321: 848.
47. Sutton RS, Barto AG. 1998. Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press.
48. Tsai H, Zhang F, Adamantidis A, Stuber G, Bonci A, de Lecea L, Deisseroth K. 2009. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* 324: 1080.
49. Ullsperger M, von Cramon DY. 2006. The role of intact frontostriatal circuits in error processing. *J Cogn Neurosci* 18: 651–664.
50. Van Hoesen GW, Morecraft RJ, Vogt BA. 1993. Connections of the monkey cingulate cortex. In: *Neurobiology of Cingulate Cortex and Limbic Thalamus* (Vogt BA, Gabriel M, eds), pp 249–284. Cambridge, MA: Birkhäuser.
51. Volkow ND, Wang G-J, Fowler JS, Telang F, Maynard L, Logan J, Gatley SJ, et al. 2004. Evidence that methylphenidate enhances the saliency of a mathematical task by increasing dopamine in the human brain. *Am J Psychiatry* 161: 1173–1180.
52. Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez HH, Dolan RJ, Hallett M. 2010. Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* 65: 135–142.
53. Wiecki TV, Frank MJ. 2010. Neurocomputational models of motor and cognitive deficits in Parkinson’s disease. *Prog Brain Res* 183: 275–297.
54. Yeung N, Botvinick MM, Cohen JD. 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev* 111: 931–959.

